UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Fundació Fisabio
Fundació per al Foment de la Investigació
Sanitària i Biomèdica de la Comunitat Valenciana

# Characterization of genes and functions required by multidrug-resistant enterococci to colonize the intestine

PhD in Biotechnology
January, 2021

Author: Alejandra Flor Duro

Director: Carles Úbeda Morant

Tutor: José Gadea Vacas

## AGRADECIMIENTOS / ACKNOWLEDGMENTS

Este trabajo no hubiese sido posible sin el apoyo de toda la gente que ha estado a mi alrededor durante todos estos años, y a las que quiero mostrar mi agradecimiento en las siguientes líneas.

En primer lugar, quiero dar las gracias a mi director de tesis, Carles Úbeda, ante todo, por haberme dado la oportunidad de iniciarme en el mundo de la investigación ejerciendo un trabajo tan multidisciplinar. Gracias por tener siempre el despacho abierto para cualquier duda, por tu supervisión y consejos constantes para sacar el trabajo adelante, aunque a veces los problemas pareciesen demasiado grandes. También por enseñarnos algo tan fundamental como estructurar las ideas y transmitir los resultados de forma clara.

En segundo lugar, he tenido la suerte de compartir mí día a día con unos compañeros de grupo increíbles que han conseguido hacer, gracias a su apoyo, que esos días malos fueran menos malos. Ana y Sandrine, las veteranas del grupo, dispuestas siempre a echar una mano y resolver dudas, gracias por vuestra paciencia y tiempo invertido en enseñarme. A Bea, mi compañera de correteos por los pasillos en los momentos más desesperantes de la tesis, ahora es imposible no recordarlos entre risas. Gracias por todos estos años trabajando juntas. Majo, mi postdoc favorita, por tu ayuda y por preocuparte siempre tanto por mi salud y de que fueran bien mis experimentos. Anna, llegaste al laboratorio para revolucionarlo con tus carteles, tu orden, espontaneidad y afición por transmitir con tanta gracia lo que hacemos cada día en el laboratorio. Gracias por tantas horas de cabina y animalario, los viajes en tranvía y las conversaciones. Iván, el polifacético del grupo y todo hecho a la perfección, gracias por poder hablar contigo de cualquier cosa, por las risas y el humor cuando algo no sale como se esperaba y por hacer que esta tesis quede mucho más bonita. Gara, tus buenos días y tu acento canario son capaces de alegrarle el día a cualquiera y aunque tu paso por el grupo fue corto, fue suficiente para quedarme con una amiga más allá del laboratorio.

Por supuesto, también quiero agradecer a todos los doctorandos, tanto los que pasaron antes por allí como los que están actualmente, por el buen ambiente, por las comidas llenas de risas y los momentos compartidos tanto fuera como dentro del laboratorio. Dar las gracias también al resto del personal de FISABIO, en especial a Concha, por preparar, aunque te lo pidiéramos a última hora, todo el material necesario para los experimentos y así ayudarnos a que todo salga perfecto.

También, darle las gracias a todo el personal del animalario, especialmente a Inma Noguera por asegurar que todo estuviera bajo control. Gracias a Inma Ballester por su ayuda dentro de P2 y por estar constantemente pendiente de nuestros ratoncitos.

On the other hand, I would like to thank my external supervisors. Firstly, thanks to Willem van Schaik and the people in his lab, especially Fernanda and Vincent, for helping me to find the solution of the first problem that I encountered when I started the thesis. I also thank Willem for always being available after the stay to discuss the results obtained and give us advice on how to improve them. Secondly, thanks to Douwe van

Sinderen for giving me the opportunity to develop part of my work in your lab with your advice. Thanks María and Rocío for welcoming me on my arrival and for helping me in and out of the lab during my stay in Cork.

Fuera del laboratorio también son muchas las personas que me han acompañado desde el principio o que se han unido por el camino y con las que he compartido momentos inolvidables y unos cuántos interrogatorios para tratar de entender qué es lo hacemos en el laboratorio.

Para empezar me gustaría darles las gracias a mis amigos, a los de siempre, a los que están ahí desde hace muchísimos años y que, a pesar de que cada uno estemos en un sitio distinto, cada vez que nos volvemos a juntar es como si esa distancia física, que hace que no nos veamos de forma frecuente, no nos separase en el tiempo.

No puedo olvidarme de aquello que me ha hecho feliz desde pequeña y que me ha acompañado siempre, que han sido la música y la danza. Estos últimos años la salsa ha sido mi compañera de diversión y entretenimiento, gracias a la cual mi estado de ánimo cambiaba por completo con tan solo una hora de clase. Además de eso, me ha servido para hacer grandes amigos. Mis primeros salseros barceloneses, gracias por hacer de mi periodo en Barcelona una época inolvidable, pero sobre todo gracias por seguir ahí después, por las visitas con horchata y paella y porque aunque no bailamos con la misma frecuencia que lo hacíamos, siempre bailamos con las mismas ganas de pasarlo bien. Ya en Valencia, gracias a toda la gente de la academia, tanto profesores como compañeros, por la cantidad de horas bailando y por los consejos para mejorar.

Gracias Mónica por tu paciencia con mis errores en inglés. No pude encontrar mejor profe que me ayudase a mejorar para conseguir mis objetivos, ni mejor persona con la que hablar, largo y tendido, sobre cualquier tema y así hacer esas clases de inglés tan amenas y divertidas. Ha sido toda una suerte haber dado contigo.

Tengo muchísimo que agradecer a los que han sido mis compañeros en el día a día, mi "familia" fuera de casa, porque la palabra compañeros de piso se os queda muy muy pequeña para lo que sois para mí. En primer lugar, a Yael. Gracias a un cambio de mesa el primer día de clase y a unos cuantos viajes en tren formamos la pareja perfecta de *Pili y Mili*, las amigas inseparables para toda la carrera y aunque el contacto no sea diario como antes, tengo claro que para toda la vida. En segundo lugar, gracias a Mar, mi valenciana de pura cepa favorita y mi compañera de batallas en Barcelona. Fue algo más corto pero no menos intenso porque lo nuestro venía ya de la carrera y le añadimos la unión de las comidas familiares de los domingos. Por último, Joan, agradecerte tantísimos años de amistad y es que lo que empezó gracias a compartir un atril y unas cuantas notas musicales siendo apenas unos niños nos llevó a compartir el mismo techo durante la tesis y la residencia, con pandemia y cuarentenas incluidas. El poder llegar a casa y contar las penas o alegrías del día con alguien de toda la vida no tiene precio.

A mi familia, el pilar fundamental al cual tengo tantísimo que agradecer. No hay nada que me haga sonreír más que pensar en lo arropada que estoy por todos vosotros. Primeramente, a mis padres, por haberos dejado la piel para que tanto yo como mi hermano podamos hacer aquello que nos gustara. Esta tesis es el resultado de vuestro esfuerzo incansable por darnos la mejor educación posible. Papá, gracias por recalcarnos tantas veces que estudiar era nuestro trabajo y que la suerte uno la busca con el esfuerzo y la perseverancia. Mamá, gracias por enseñarnos lo que es la valentía demostrándonos en primera persona que se puede conseguir lo que uno se proponga y que, aunque el esfuerzo sea grande, siempre merece la pena. Tete, gracias por haber sido mi mejor regalo, mi persona favorita, por poder contar contigo para todo, por apoyarme y por valorarme siempre tanto. Y es que ni la distancia ni la edad importa cuando la conexión es buena… ¡y es que la nuestra no podría ser mejor! Gracias a mis abuelos por enseñarme lo que es el amor incondicional y la bondad. Soy muy afortunada de poder haber crecido con vosotros y de aprender de vuestra sabiduría sin límite. Gracias tíos y primos, por haberme apoyado siempre y por hacerme disfrutar tanto de esas comidas familiares en las que estamos todos y reímos sin fin.

Por último, este camino habría sido mucho más complicado sin ti Rubén y sin la suerte de que volviésemos a coincidir. Gracias por ser mi soporte moral incondicional, por haberme ayudado tantísimo en tantos aspectos y por estar siempre, incluso esperando a horas intempestivas de un fin de semana a la puerta del laboratorio o del animalario. Es muy cierto eso de que "quién quiere algo encuentra el medio, y quien no, siempre encuentra una excusa", así que gracias por nunca faltarte las ganas de encontrar ese medio.

Para finalizar, nada de esto se podría haber llevado a cabo sin la financiación correspondiente para realizar la tesis por parte del Ministerio de Ciencia e Innovación de España. Este último año la pandemia del coronavirus ha sacado a relucir, aún más si cabe, la carencia y la necesidad de inversión en investigación. La situación precaria de la ciencia es un problema para el país y es que la inversión en investigación científica no tiene que considerarse un gasto, tiene que entenderse como la inversión de futuro tan necesaria que es.

**GLOSSARY**

| | |
|---|---|
| **AMR** | Acquired Antimicrobial Resistance |
| **aa** | Amino acids |
| **Amp$^r$** | Ampicilin resistance |
| **ARPs** | Antibiotic Resistant Pathogens |
| **AUC** | Area Under the Curve |
| **bp** | Base pair |
| **BEA** | Bile Esculin Azide Agar |
| **BEA+A** | Bile Esculin Azide Agar with Ampicillin |
| **BEA+AV** | Bile Esculin Azide Agar with Ampicillin and Vancomycin |
| **BEA+AG** | Bile Esculin Azide Agar with Ampicillin and Gentamicin |
| **BHI** | Brain Heart Infusion Broth |
| **CaCl$_2$** | Calcium chloride |
| **CRE** | Carbapenem-resistant *Enterobacteriaceae* |
| **CDM** | Chemically Defined Medium |
| **Cm$^r$** | Chloramphenicol resistance |
| **CDI** | *Clostridium difficile* infection |
| **CDSs** | Coding Sequences |
| **CR** | Colonization Resistance |
| **CI** | Competitive Index |
| **Δ** | Deletion |
| **DNA** | Deoxyribonucleic acid |
| **DEPC** | Diethylpyrocarbonate |
| **EDTA** | Ethylenediaminetetraacetic acid |
| **FMT** | Faecal Microbiota Transplantation |
| ***gen*** | Gentamicin |
| **Gm$^r$** | Gentamicin resistance |
| **HAIs** | Healthcare-Associated Infections |
| **Tn-seq** | High-throughput transposon mutant library sequencing |
| **::** | Insertion |
| **iPCR** | Inverse PCR |
| **Kn$^r$** | Kanamycin |
| **LOD** | Limit of detection |
| **LB** | Luria-Bertani  Broth |
| **min** | Minutes |
| **SGM17** | M17 medium + Sucrose (0.5 M) +  glycine (2%) + glucose (0.5%) |
| **GM17** | M17 medium + glucose (0.5 %) |
| **MTC** | Maximal Tolerable Concentration |
| **ICU** | Medical Intensive Care Unit |
| **Tm** | Melting temperature |
| **mRNA** | Messenger RNA |

| | |
|---|---|
| **M-tram** | Microarray-based transposon mapping |
| **MSCRAMMs** | Microbial surface components recognizing adhesive matrix molecules |
| **MICs** | Minimum Inhibitory Concentrations |
| **MLST** | Multilocus sequence typing |
| **mut** | Mutant |
| **NIs** | Nosocomial Infections |
| **OD$_{600}$** | Optical Density measured at 600 nm |
| **OTUs** | Operational Taxonomical Units |
| **ON** | Overnight |
| **PAI** | Pathogenicity island |
| **PBS** | Phosphate Buffer Solution |
| **ProBac** | Protective Bacteria |
| **rpm** | Revolutions per minute |
| **RNA** | Ribonucleic acid |
| **RT** | Room Temperature |
| **SCFA** | Short Chain Fatfy Acids |
| **NaC$_2$H$_3$O$_2$** | Sodium acetate |
| **NaCl** | Sodium chloride |
| **s** | Seconds |
| **SDS** | Sodium dodecyl sulphate |
| **SOC** | Super Optimal broth with Catabolite repression (Glucose) |
| **TAE** | Tris-acetate-EDTA |
| **TBE** | Tris-borate-EDTA |
| **UC** | Unclassified |
| **VRE** | Vancomycin Resistant *Enterococcus* |
| **WT** | Wild-type |
| **WHO** | World Health Organization |

# Contents

**ABSTRACT**

Multidrug-resistant bacteria, such as vancomycin-resistant-*Enterococcus* (VRE), are an increasing problem in hospitalized patients. Some VRE strains can be resistant to most available antibiotics, thus, alternative strategies to antibiotics are urgently needed to combat these challenging pathogens. Infections caused by VRE frequently start by colonization of the intestinal tract, a crucial step that is impaired by the presence of the intestinal microbiota. Administration of antibiotics disrupts the microbiota, which promotes VRE intestinal colonization. Once VRE has colonized the gut, it reaches very high levels, which promotes its dissemination to other organs and its transfer to other patients. Despite the relevance of VRE gut colonization, very little is known about the genes encoded by this pathogen to colonize the gut and about the mechanisms by which the microbiota suppresses VRE gut colonization.

In this thesis, we have utilized a previously described methodology (Zhang et al., 2017, BMC Genomics), based on the generation of a transposon mutant library coupled with high-throughput sequencing, in order to identify VRE encoded genes required for colonization of the mouse intestinal tract. In addition, we have performed metatranscriptomic analysis in mice to identify VRE genes specifically expressed in the gut. Our analysis has identified genes whose disruption significantly reduces VRE gut colonization in the large intestine. The genes that most affected VRE gut colonization encoded for proteins related to the uptake or transport of diverse nutrients such as carbohydrates (PTS mannose transporter subunit EIIAB, LacI family DNA-binding transcriptional regulator, N-acetylmuramic acid 6-phosphate etherase) or ions (phosphate ABC transporter ATP-binding protein and proteins from [Fe-S] cluster). The role of these genes in gut colonization has been confirmed through targeted mutagenesis and competition experiments against a wild type strain. Moreover, these genes affect gut colonization under different antibiotic treatments (clindamycin and vancomycin). To elucidate the mechanism by which each gene influences gut colonization, we have performed *in vitro* and *ex vivo* experiments besides transcriptomic analysis. *In vitro* experiments confirm that proteins from [Fe-S] cluster are involved in the transport of different forms of iron ions, mostly $Fe^{3+}$. On the other hand, the PTS mannose transporter subunit EIIAB and N-acetylmuramic acid 6-phosphate etherase genes are required for the utilization of mannose and N-acetyl-muramic acid, respectively, sugars that are usually present in the intestinal environment. We have also confirmed that LacI family DNA-binding transcriptional regulator is a repressor that affects the expression of genes encoding for an ABC transporter probably involved in the uptake of carbohydrates. Furthermore, we have confirmed that some of these genes are encoded mainly by *E. faecium* clinical strains but not or to a lower extent by commensal strains.

Secondly, we studied the mechanisms of protection of a consortium of five commensals bacteria, previously shown to restrict VRE gut colonization in mice. Functional transcriptomics in combination with targeted metabolomics and *in vivo* assays performed in this thesis indicated that the bacterial consortium inhibits VRE growth

through nutrient depletion, specifically by reducing the levels of fructose. Finally, *in vivo* RNA-Seq analysis of each bacterial isolate of the consortium in combination with *ex vivo* and *in vivo* assays demonstrated that a single bacterium (*Olsenella* sp.) could recapitulate the protective effect.

Altogether, the results obtained have identified the function of specific genes required by VRE to colonize the gut. In addition, we have identified a specific mechanism by which the microbiota confers protection against VRE colonization. These results could lead to novel therapeutic approaches to prevent infections caused by this multidrug-resistant pathogen.

## RESUMEN

Las bacterias resistentes a múltiples antibióticos, como el Enterococo resistente a vancomicina (ERV), son un problema creciente en los pacientes hospitalizados, por lo que se necesita estrategias alternativas para combatir estos patógenos. Las infecciones causadas por ERV suelen comenzar con la colonización del tracto intestinal, un paso crucial que se afectado por la presencia de la microbiota. Sin embargo, los antibióticos alteran la microbiota y esto promueve la colonización de ERV. Una vez que el patógeno ha colonizado el intestino, alcanza niveles muy altos pudiendo diseminar a otros órganos y pacientes. A pesar de su importancia, se sabe muy poco sobre los genes que codifica para colonizar el intestino y sobre el mecanismo por el cual la microbiota suprime su colonización intestinal, siendo los dos objetivos principales.

En primer lugar hemos utilizado una metodología previamente descrita (Zhang et al., 2017, BMC Genomics), basada en la generación de una librería de mutantes por transposición junto a secuenciación masiva, con el fin de identificar los genes codificados por ERV necesarios para la colonización del intestino en ratones. Además, hemos realizado análisis metatranscriptómicos para identificar aquellos genes más expresados. El análisis ha identificado genes cuya interrupción reduce significativamente la colonización intestinal en el intestino grueso. Los genes que más afectaron a la colonización codifican proteínas relacionadas con la absorción o el transporte de diversos nutrientes como los carbohidratos (subunidad EIIAB del transportador PTS de manosa, el regulador transcripcional de la familia LacI, ácido N-acetilmurámico 6-fosfato eterasa) o iones (proteína transportadora dependiente de ATP (ABC) y proteínas del grupo [Fe-S]). El papel de estos genes en la colonización se ha confirmado mediante experimentos de mutagénesis directa y de competición con la cepa salvaje. Además, estos genes afectan a la colonización intestinal con diferentes antibióticos (clindamicina y vancomicina). Para identificar el mecanismo molecular por el cual cada gen afecta a la colonización, hemos realizado experimentos *in vitro* y *ex vivo* además del análisis transcriptómico. Los experimentos *in vitro* confirman que las proteínas del grupo [Fe-S] están involucradas en el transporte iones de hierro, principalmente $Fe^{3+}$. Por otra parte, los genes de la subunidad EIIAB del transportador de manosa y del ácido N-acetilmurámico 6-fosfato eterasa son necesarios para la utilización de la manosa y el ácido N-acetilmurámico, respectivamente, azúcares que suelen estar presentes en el intestino. También confirmamos que el regulador transcripcional de la familia LacI es un represor que afecta a proteínas transportadoras ABC, probablemente implicadas en la absorción de carbohidratos. Además, algunos de estos genes están codificados principalmente por cepas clínicas de *E. faecium* y en menor medida por cepas comensales.

En segundo lugar, estudiamos los mecanismos de protección de un consorcio de cinco bacterias comensales, que anteriormente se había demostrado que disminuían la colonización intestinal por ERV en ratones. Mediante transcriptómica, metabolómica y los ensayos *in vivo* observamos que el consorcio bacteriano inhibe el crecimiento de ERV mediante la reducción de nutrientes, concretamente fructosa. Por último, el análisis

ARN-Seq *in vivo* de cada aislado en combinación con los ensayos *ex vivo* e *in vivo* demostraron que una sola bacteria (*Olsenella* sp.) proporciona protección.

En conjunto, los resultados obtenidos han identificado la función de genes específicos requeridos por ERV para colonizar el intestino. Además, hemos identificado un mecanismo mediante el cual la microbiota confiere protección. Estos resultados podrían conducir a nuevos enfoques terapéuticos para prevenir las infecciones causadas por este patógeno multiresistente a los antibióticos.

**RESUM**

Els bacteris resistents a múltiples antibiòtics, com el Enterococo resistent a vancomicina (ERV), són un problema creixent en els pacients hospitalitzats, que són resistents a la majoria d'antibiòtics disponibles per la qual cosa es necessita estratègies alternatives per a combatre aquests patògens. Les infeccions causades per ERV solen començar amb la colonització del tracte intestinal, un pas crucial que es veu afectat per la presència de la microbiota. No obstant això, els antibiòtics alteren la microbiota i això promou la colonització de ERV. Una vegada que el patogen ha colonitzat l'intestí, aconsegueix nivells molt alts podent disseminar a altres òrgans i pacients. Malgrat la seua importància, se sap molt poc sobre els gens que codifica ERV per a colonitzar l'intestí i sobre el mecanisme pel qual la microbiota suprimeix la seua colonització intestinal.

En primer lloc hem utilitzat una metodologia prèviament descrita (Zhang et al., 2017, BMC Genomics), basada en la generació d'una llibreria de mutants per transposició junt amb seqüenciació massiva, amb la finalitat d'identificar els gens codificats per ERV necessaris per a la colonització de l'intestí en ratolins. A més a més, hem realitzat anàlisi metatranscriptòmics per a identificar aquells gens més expressats. L'anàlisi ha identificat gens quina interrupció redueix significativament la colonització intestinal en l'intestí gros. Els gens que més van afectar la colonització codifiquen proteïnes relacionades amb l'absorció o el transport de diversos nutrients com els carbohidrats (subunitat EIIAB del transportador PTS de manosa, el regulador transcripcional de la família LacI, àcid N-acetilmuràmic 6-fosfat eterasa) o ions (proteïna transportadora dependent d'ATP (ABC) i proteïnes del grup [Fe-S]). El paper d'aquests gens en la colonització s'ha confirmat mitjançant experiments de mutagènesis directa i de competició amb el cep salvatge. A més, aquests gens afecten la colonització intestinal amb diferents antibiòtics (clindamicina i vancomicina). Per a identificar el mecanisme molecular pel qual cada gen afecta a la colonització, hem realitzat experiments *in vitro* i *ex viu* a més de l'anàlisi transcriptòmic. Els experiments *in vitro* confirmen que les proteïnes del grup [Fe-S] estan involucrades en el transport d'ions de ferro, principalment $Fe^{3+}$. D'altra banda, els gens de la subunitat EIIAB del transportador PTS de manosa i de l'àcid N-acetilmuràmic 6-fosfat eterasa són necessaris per a la utilització de la manosa i l'àcid N-acetilmuràmic, respectivament, sucres que solen estar presents en l'intestí. També confirmem que el regulador transcripcional de la família LacI és un repressor que afecta proteïnes transportadores ABC, probablement implicades en l'absorció de carbohidrats. A més a més, alguns d'aquests gens estan codificats principalment per ceps clínics de *E. faecium* i en menor mesura per ceps comensals.

En segon lloc, estudiem els mecanismes de protecció d'un consorci de cinc bacteris comensals, que adès s'havia demostrat que disminuïen la colonització intestinal per ERV en ratolins. Amb l'ús de transcriptòmica, metabolòmica i els assajos *in vivo* observem que el consorci bacterià inhibeix el creixement de ERV mitjançant la reducció de nutrients, concretament fructosa. Finalment, l'anàlisi ARN-Seq *in vivo* de cada aïllat en combinació amb els assajos *ex viu* i *in vivo* van demostrar que un sol bacteri (*Olsenella* sp.) proporciona protecció.

En conjunt, els resultats obtinguts han identificat la funció de gens específics requerits per ERV per a colonitzar l'intestí. A més, hem identificat un mecanisme mitjançant el qual la microbiota confereix protecció. Aquests resultats podrien conduir a nous enfocaments terapèutics per a previndre les infeccions causades per aquest patogen multiresistent als antibiòtics.

# I.    INTRODUCTION

# 1. *ENTEROCOCCUS*

## 1.1 General characteristics

The enterococci are a group of non-spore Gram-positive cocci, disposed individually, in pairs, chains of various lengths or in groups[1,2]. They are facultative anaerobes and chemo-organotrophic organisms[2,3] . Enterococci have an optimal growth between 35 ºC to 37 ºC and a wide range of growth temperature, from 10 to > 45 ºC[2,4].

They have the ability to survive under very harsh conditions, including high salt concentration (6.5 % sodium chloride), in the presence of 40 % bile salts and sodium azide (up to 40 %) and they are tolerant to alkaline pH[2–4]. Most enterococci are catalase-negative organisms with the ability to hydrolyse esculin in the presence of bile salts[2,3].

They have been considered as part of lactic acid bacteria being classified as homofermentative. They produce lactic acid as the main and predominant product of carbohydrate fermentation, without gas production[2,3,5].

Enterococci have low G+C content, approximately from 34 % to 45 %, and the size of their genomes have a wide range, from 2.3 Mb to 5.3 Mb with around 2100 to 5100 predicted genes[6,7].

## 1.2 Enterococci taxonomy

The taxonomy has change significantly in the recent years. The enterococci were first classified in the *Streptococcus* genus due to some morphological and biochemical similarities. They were classified in *Streptococcus* group D until 1984, when the *Enterococcus* genus was proposed, recognised and separated from the previous one[2,8,9]. In 1984, Shleifer and Kilpper-Balz performed 16S rRNA sequencing and studies of DNA-DNA and DNA-rRNA hybridisation that demonstrated that *Streptococcus faecalis* and *Streptococcus faecium* were not related to streptococci, being recognised as a separated genus and renamed as *Enterococcus faecalis* and *Enterococcus faecium*[8]. Afterwards, other bacteria previously classified as streptococci were reclassified into the *Enterococcus* genus[10].

To date, the *Enterococcus* genus has 59 species with validly published and correct name according to the LPNS (List of Prokaryotic Names with Standing in Nomenclature).

In 2009, enterococci were classified in the *Enterococcaceae* family based on 16S rRNA, inside Lactobacillales order, Bacilli class in the phylum Firmicutes[11].

## 1.3 Habitat

*Enterococcus* genus is present in diverse environments and habitats that comprise the gastro-intestinal (GI) tracts of humans and varied animals including a wide range of mammals, reptiles and avians[12]. They are natural colonizers of the normal microbiota[1] known as harmless commensals. They are ancient and evolved members of GI from

diverse animals studied[13]. Indeed, the GI represents the highest reservoir for enterococci[3]. Enterococci can also be found in plants[14], insects[15], water, soil and man-made products like fermented food items[3].

In humans, enterococci represent a minor population within the adult microbiota, less than 1 %, after analysing faeces by 16S rRNA-targeted oligonucleotide probes[16]. *E. faecalis* is the most prevalent *Enterococcus* species found in human faeces followed by *E. faecium,* whereas all the other species account low percentage of representation[17].

Despite the differences in the microbiota composition in the different compartments of the intestine, enterococci are localized in both small and large intestine. They inhabit the jejunal and ileal microbiota, consisting of simple microbial communities formed by facultative anaerobes and aerobes. They are also present in caecum and recto-sigmoidal parts, where the microbiota contains higher variety of anaerobe communities[18] and a high bacterial density ($10^9$ to $10^{12}$ colony forming units - CFUs per gram of content)[3]. They are rarely found in the stomach[19] which has $10^2$ to $10^4$ CFUs[20] and whose acidic pH limits the presence of enterococci[3].

On the other hand, they are commonly present in the oral cavity, although enterococci are considered transient bacteria of the oral microbiome[21]. *E. faecalis* is the predominant *Enterococcus* found and it has been isolated from plaque[22], saliva and root canals[23].

## 1.4 Nosocomial pathogen

The presence of enterococci as part of the gastrointestinal tract along with its tolerance and capability to adapt and survive in extreme environments has allowed these organisms to easily and widely spread in the hospital setting causing diverse and severe diseases to hospitalized patients[1].

The antibiotic treatments administered to a hospitalized patient decrease the protective layer of the mucus and in addition induce drastic and long-lasting changes in the intestinal microbiota composition that in normal conditions confers protection and inhibits the expansion of resistant *Enterococcus*[24,25]. Thus, the antibiotics, by altering the intestinal microbiota, increase the density of resistant *Enterococcus* in the gastro-intestinal tract, facilitating the overgrowth in the gut and their spreading[26]. During antibiotic treatment, these opportunistic pathogens increases their density in the intestine, which promotes their dissemination to the bloodstream, causing diseases that can put in danger the patient's life[27,28]. Since hospitalized patients are often treated with antibiotics, *Enterococcus* has become one of the most frequent pathogens causing nosocomial infections (NIs). Indeed, it is considered the second leading cause of NIs in the United States[1,29] and the third most prevalent cause in Europe[30]. NIs produced by enterococci can put in serious danger the life of the patient. Indeed the mortality of enterococci infections varies from 20 to 60 % among hospitalized patients, which

depend on the type of infection caused, the risk factor of the host and the exposure to antibiotics[26,31].

Besides having an impact on patients health, NIs are economically relevant since they increase considerably the cost of the hospital stay due to the more expensive treatments required to treat these type of opportunistic pathogens in addition to the longer hospital stays of the patients[32].

The spread of enterococci among hospitalized patients can occur through different routes: (i) proximity between colonized and susceptible patients, especially with long-term hospitalized patients; (ii) faecal contamination of the hospital environment[1], and (iii) close contact to contaminated health care personnel or objects[1,33]. Interestingly, it has been studied the relevance of the *Enterococcus* density in the intestine for the contamination of the environment with *Enterococcus* and it subsequent dissemination among patients[34]. Importantly, the authors found that the contamination of the hospital environment with *Enterococcus* occurs only when this bacterium was present in the large intestine of a hospitalized patient with a density higher than $10^4$ per gram of content[34]. Indicating that the expansion of *Enterococcus* in the large intestine could contribute to its dissemination through contamination of the environment. Moreover, enterococci can survive from 5 days to 4 months on surfaces[35], which further increase their capacity of dissemination among hospitalized patients. Note that in this thesis, we will study which genes are encoded by clinical strains of *E. faecium* to expand in the intestinal tract and how the microbiota inhibits *E. faecium* intestinal colonization. Thus, the obtained results could have a direct impact on developing novel strategies to diminish the spread of enterococci among patients by decreasing *E. faecium* gut levels.

Besides contamination of the hospital environment, the hands of the health workers are a vector of transmission because enterococci are able to survive up to 60 minutes on hands or gloves[36]. It is necessary therefore to implement hygiene practices and wash hands frequently with soap for at least 30 seconds in order to eradicate this bacterium and decrease its dissemination[1,36].

*Enterococcus faecalis* and *Enterococcus faecium* are not only the most prevalent species within the *Enterococcus* genus found in human faeces but also the most pathogenic for humans[26]. Within these two, *E. faecalis* was more prevalent and pathogenic, causing 85-95 % of *Enterococcus* infections, whereas 5-10 % were caused by *E. faecium*[26]. However, since 2002 the prevalence of vancomycin-resistant *E. faecium* has increased, being more common than vancomycin-resistant *E. faecalis*. *E. faecium* has acdquired resistance to many different classes of antibiotics and has higher mortality rates than *E. faecalis*[37,38].

Indeed, in the last years, there has been an important increase of *E. faecium* infections in Europe. More specifically, it has been an increased in the proportion of *E. faecium* resistance to vancomycin, from 8.1 % in 2012 to 19 % in 2018 of the *E. faecium* infections (data from the European Antimicrobial Resistance Surveillance Network (EARS-Net) database (TESSy))[39]. This increase is mainly associated with the resistance rise to

antibiotics, especially resistance to vancomycin, which has led the infections produced by this pathogen to become more difficult to treat[1,2].

### 1.4.1 Type of infections caused by *Enterococcus*

Nosocomial enterococci infections frequently start by colonization of the gastro-intestinal tract. Normally *Enterococcus* is present in low abundance in the GI tract but the density increases due to the exposure to antibiotic treatments and the weakened immune system of the host[1,40]. The increase in the intestinal density of enterococci facilitates its dissemination to the bloodstream through the intestinal lining and also translocating through the liver[1]. The presence in the bloodstream enable its spread to other body organs[1] (Figure 1). The microbiota has a relevant role in inhibiting *Enterococcus* expansion in the intestine and the antibiotics, by altering the microbiota composition, facilitates *Enterococcus* overgrowth in the intestine[28]. The role of the microbiota related to enterococcal infections will be explained in detail in section 2.

On the other hand, the bacterium can get into the urinary tract causing urinary tract infections. This is usually associated to previous exposure to antibiotics, catheters or instrumentation such as urinary drainage and ventilators. These type of infections are the most common ones caused by enterococci and are frequently detected in older men[41]. Indeed, 15 % of the catheter-associated urinary tract infections are caused by enterococci[33,41].

The second most common infections caused by enterococci are endocarditis (5-20 %)[33,41], a life threating infection of the endothelial lining where *Enterococcus* attach to endocardial epithelium and colonize the heart[40,42].

Bacteremia without endocarditis is a recurrent disease present in weakened patients who receive antibiotics. It is frequently acquired from urinary and intravascular catheters[33]. The risks for patient's life increase when bacteremia is present with endocarditis[41]. Enterococcal bacteremia is usually associated to older people that are hospitalized in the Medical Intensive Care Units (ICU) for a prolonged period[43].

Other type of infections caused by enterococci are intra-abdominal, pelvic and skin wounds or soft tissue infections[41,44].
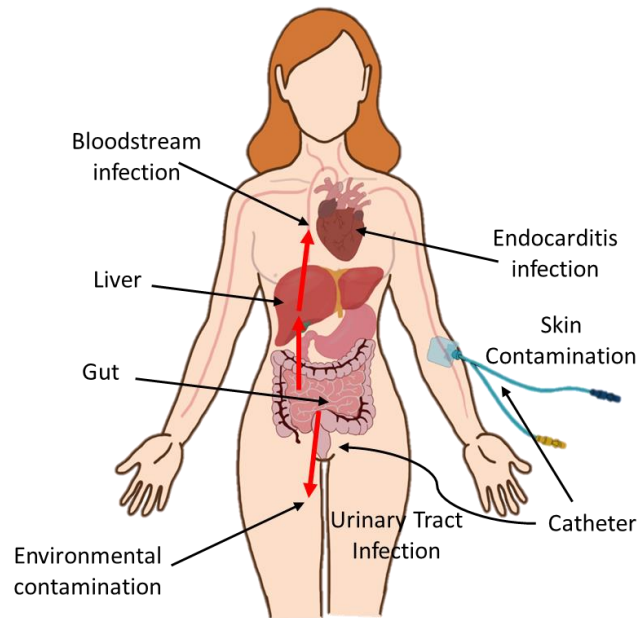
**Figure 1. Enterococci infections in the human body.** High density enterococci access to bloodstream through the intestinal lining reaching first the liver. From the bloodstream, enterococci can reach different organs causing severe infections such as endocarditis. Enterococcal infections could also occur by skin contamination and the use of catheters. Enterococci in the gut also promote the dissemination to other patients through environmental contamination. Image modified and adapted from Murray and Arias[1], Macmillan Publishers Limited, Nature Reviews Microbiology©, 2012.

## 1.5 Antibiotic resistance

The emergence of enterococci as a nosocomial pathogen has been due to the selective advantage of intrinsic resistance to antibiotics and the ability to acquire new resistant traits[1,2].

Enterococci are resistant to a high number of antibiotics of different spectrum, but this inherent resistance has increased in the last decades, especially in *E. faecium* species[33,45] (Figure 2). There has been a rise of virulence enterococcal infections associated with the acquired antimicrobial resistance (AMR), being the most important the acquired resistance to vancomycin and β-lactam antibiotics[2,45].
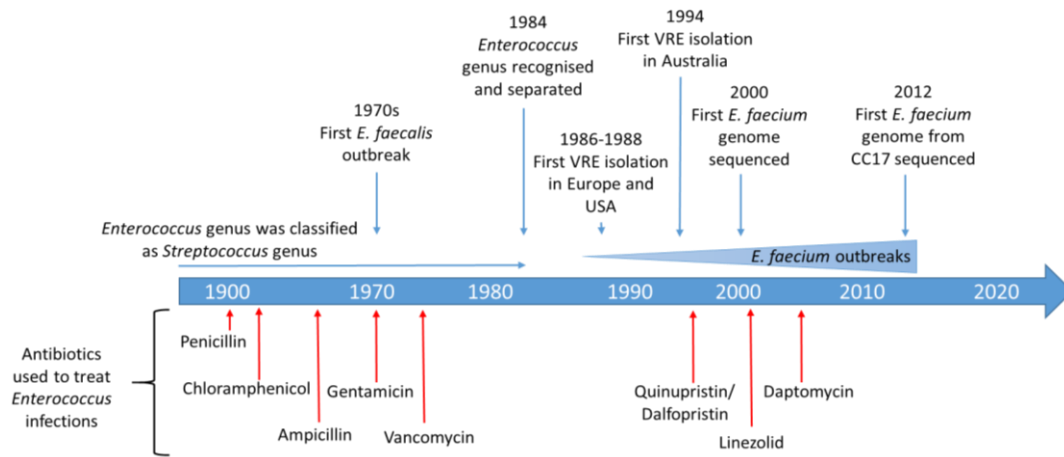
**Figure 2. Timeline of the development of the enterococcal infections and the antibiotics treatments used.** Subsequently to the use of vancomycin as usual treatment for enterococcal hospital infections, the incidence of *E. faecium* outbreaks resistant to vancomycin (VRE) increased. A recent group of antibiotics (including daptomycin and linezolid) have been used to treat VRE infections but some strains have become resistant to them. Figure adapted from Rice and García-Solache[2], 2019, Clinical Microbiology Reviews.

Vancomycin-resistant Enterococcus (VRE) has been classified by the World Health Organization (WHO) within the high priority group of the list of antibiotic-resistant bacteria for research and development of new antibiotics. In this study carried out by WHO, VRE is classified in the 4th position of the final ranking of the most important antibiotic-resistant bacteria, being the first Gram-positive bacteria in the ranking[46]. Since the 80s, VRE have been a global problem for public health, especially in USA, followed by Australia, where up to 50 % of infectious *E. faecium* isolates are resistant to vancomycin (Figure 3)[45]. Resistance to vancomycin and other glycopeptides will be explained in more detail in the section: "Resistance to glycopeptides".
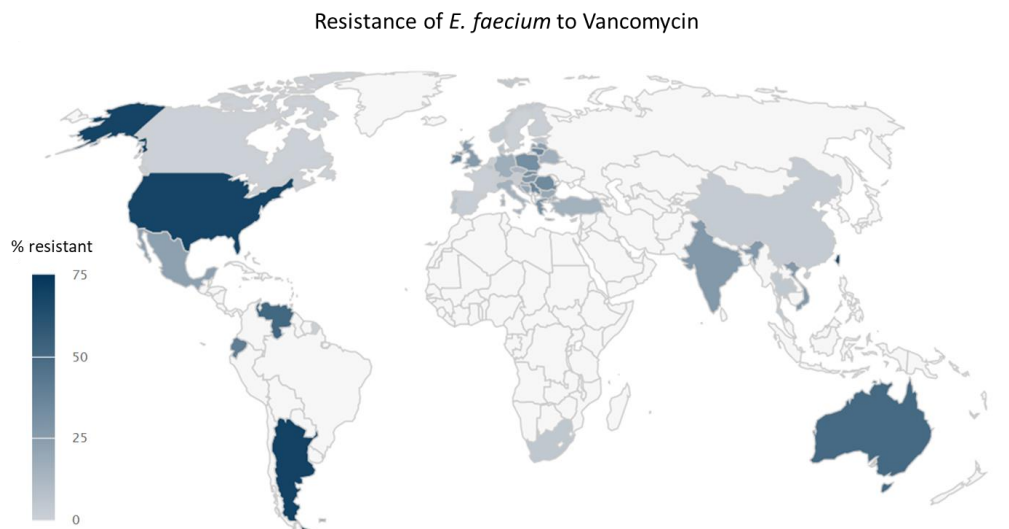


**Figure 3. Worldwide distribution of *E. faecium* strains resistant to vancomycin (VRE).** The Centre for Disease Dynamics, Economics and Policy. Resistance Map: Vancomycin resistance 2020. https://resistancemap.cddep.org/AntibioticResistance.php. Date accessed: April 8, 2020. The intensity of the blue colour indicates the % of resistant strains to vancomycin (including intermediate resistance) obtained from blood and cerebrospinal fluid (i.e., invasive) from inpatients (hospitalized) of all ages.

**Resistance to β-Lactams**

β-Lactams antibiotics have bactericidal activity, inhibiting the peptidoglycan synthesis of the bacteria cell wall after binding covalently[47,48]. The most frequent β-Lactams used to treat enterococci has been ampicillin and penicillin[1].

Resistance to these antibiotics has been developed through two different mechanism: production of β-lactamases and production of low-affinity penicillin binding protein 5 (PBP5)[2,33]. The first one is less common, only observed rarely in *E. faecalis* outbreak strains that expressed β-lactamases but at low level and that do not confer high levels of resistance[1,2]. However, *E. faecium* strains present high-level resistance to penicillin due to the overproduction of PBP5, proteins that have low affinity to penicillins. Specific amino acid changes in the *pbp*5 gene are responsible for the resistance by interfering with the binding site of the β-lactam[1,2,33]. This high-resistance level is much rare in *E. faecalis* even though some strains produce a similar protein, PBP4, but with a lower level of expression[2].

**Resistance to aminoglycosides**

Enterococci are intrinsically resistant to low-moderate concentrations of aminoglycosides, but this resistance can be overcome by the addition of cell-wall synthesis inhibitors to increase cellular permeability into the cell, reaching bactericidal synergism[33,48]. This combination has been extensively used in clinical practice, which has produced an emergence of high levels of resistance, precluding the potential use of synergetic treatments.

High-levels of resistance appears due to the expression aminoglycoside-modifying enzymes. The most common enzyme that confers resistance to gentamicin is acc-(6')-le-aph-(2"), conferring resistance to other aminoglycosides, except to streptomycin. Resistance to streptomycin is conferred by ribosomal mutations or the encoded ant-6 gene. *E. faecium* is intrinsically resistance to kanamycin and tobramycin by the expression of AAC(6')[1,2].

**Resistance to glycopeptides**

Glycopeptides antibiotics, such as vancomycin and teicoplanin, block the synthesis of the bacterial wall preventing the cross-linking of peptidoglycans by binding to the peptidoglycan precursors D-alanine-D-alanine (D-ala-D-ala)[47], having bactericidal activity. Vancomycin was commonly used as an antibiotic to treat infections cause by ampicillin resistant strains until early 80s, when *E. faecium* vancomycin resistance emerged[1,2].

The mechanism of glycopeptides resistance involves the modification of the peptidoglycan synthesis pathway, resulting in the modification of the precursor D-ala-D-ala in the Uridine diphosphate (UDP)-N-Acetylmuramic acid (MurNac) pentapeptide[2] (Figure 4). The peptidoglycan precursor is substituted by either D-Alanine-D-lactate (D-Ala-D-Lac) or D-Alanine-D-Serine (D-Ala-D-Ser)[47,49]. The resistance level is given by the type of amino acid change[47]: D-Lac produce high level-resistance resulting in 1000-fold

efficiency decrease of the binding affinity whereas D-Ser produces low-level resistance producing a decrease about sevenfold[1,47].
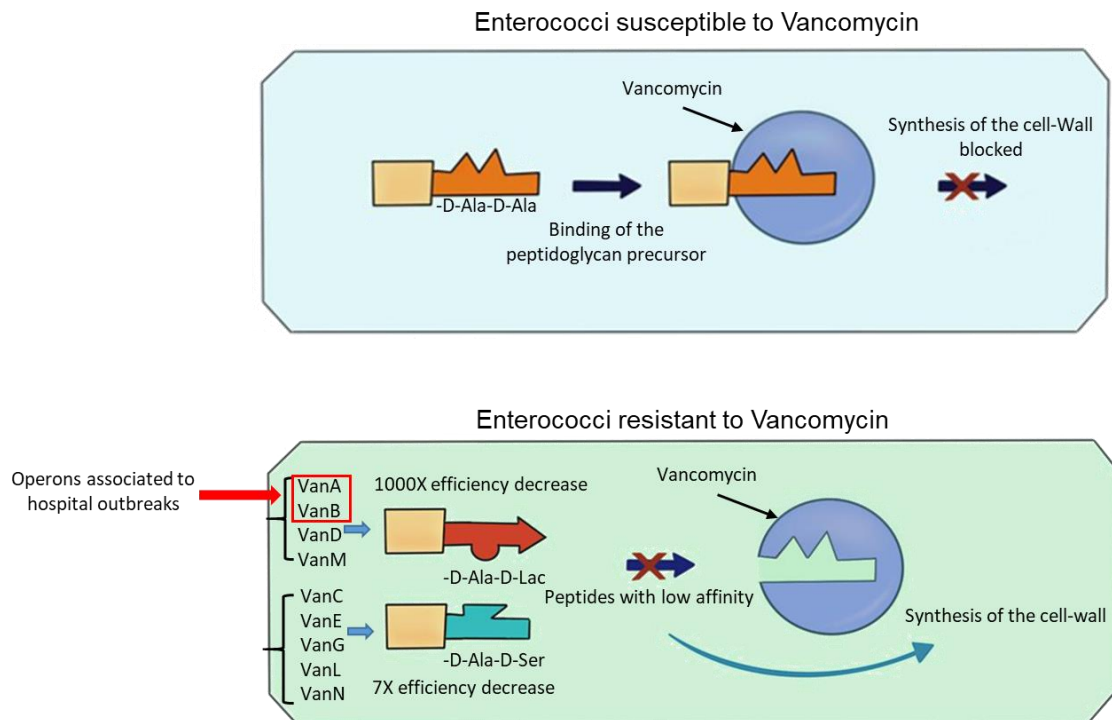


**Figure 4. Enterococci mechanism of resistance to vancomycin.** The peptidoglycan precursor of vancomycin-susceptible enterococci ends in D-Ala-D-Ala which can bind to vancomycin. In vancomycin-resistant enterococci, the amino acid of the peptidoglycan is substituted by Lac or Ser, conferring low affinity to vancomycin and, therefore resistance. Figure adapted from Murray[38], 2000, The New England Journal of Medicine.

Up to now, nine operons related to vancomycin resistance have been described in enterococci, some of them conferring also resistance to teicoplainin[47,49] (Table 1). These operons are differentiated by the terminal peptidoglycan, the degree of vancomycin resistance, the teicoplainin resistance and the organization of the genes in the different operons[47]. The main classification is the terminal peptidoglycan, dividing the nine operons in two groups: the terminal D-Ala-D-Lac with *van*A, *van*B, *van*D and *van*M operons; and the terminal D-Ala-D-Ser, that includes *van*C, *van*E, *van*G, *van*L and *van*N operons[2].

By far, the most common operons found in vancomycin resistant species related to hospital outbreaks have been VanA and VanB operons, carried by *E.faecium* strains[33,45]. In general, *van*A is more common whereas *van*B has been recently a major problem in certain regions. The VanA operon is carried on Tn1546 transposon that belongs to Tn3 family and it can be located in the genome or harboured in transferrable plasmids. The VanB operon can be present on the chromosome or in a plasmid, being carried by Tn5382 transposon (or known as well as Tn1549), belonging to Tn916 family[2,45].

Both operons functions in a similar manner[47] but one of the most important difference between them is that *van*A operon confers high level of resistance to both vancomycin and teicoplanin and *van*B confers variable resistance to vancomycin and does not confer resistance to teicoplanin[48,49].

**Table 1. Characteristics of the type of *van* operons found in *Enterococcus* species**

| Type of Van operon | Peptide precursor | Resistance | | Common carrier | Location | Expression |
|---|---|---|---|---|---|---|
| | | Vancomycin | Teicoplanin | | | |
| ***van*A** | D-Ala-D-Lac | High | High | *E. faecium* *E. faecalis* | Chromosome, transferable, plasmid | Inducible |
| ***van*B** | D-Ala-D-Lac | High (variable) | Susceptible | *E. faecium* *E. faecalis* | Chromosome, transferable | Inducible |
| ***van*C** | D-Ala-D-Ser | Low | Susceptible | *E. gallinarum* *E. casseliflavus* *E. flavescens* | Chromosome | Constitutive Inducible |
| ***van*D** | D-Ala-D-Lac | Variable | Variable | *E. faecium* | Chromosome | Constitutive Inducible |
| ***van*E** | D-Ala-D-Ser | Low to moderate | Susceptible | *E. faecalis* | Chromosome | Inducible |
| ***van*G** | D-Ala-D-Ser | Low | Susceptible | *E. faecalis* | Chromosome, transferable | Inducible |
| ***van*L** | D-Ala-D-Ser | Low | Susceptible | *E. faecalis* | Chromosome | Inducible |
| ***van*M** | D-Ala-D-Lac | High | High | *E. faecium* | Unknown, transferable | Inducible |
| ***van*N** | D-Ala-D-Ser | Low | Susceptible | *E. faecium* | Transferable, plasmid | Constitutive |

Information adapted from [38,45,49].

**Resistance to streptogramins**

Quinupristin-dalfopristin (Q/D) is a combined semisynthetic streptogramin antibiotic. It was the first drug approved by the Food and Drug Administration (FDA) to treat vancomycin resistant *E. faecium*[38,47].

This antibiotic has bactericidal activity[1]. It blocks the protein synthesis interacting with the 50S ribosomal subunit, inducing a conformational change in the ribosome and irreversible inhibition of the complex. *E. faecalis* is intrinsically resistant to both types of streptogramins, harbouring Lsa gene, an ATP-binding protein[1,47]. *E. faecium* presents in general lower level of resistance[48] and different mechanisms to reduce the activity of Q/D: (i) modification and ineffectiveness of dalfopristin through the acetyltransferases VatD and VatE, (ii) enzymatic cleavage by the lactonases VgbA and VgbB and (iii) the *msr*C gene that enconies ABC-efflux pump which removes Q/D from cell[47].

**Resistance to linezolid**

Linezolid is a synthetic antibiotic that has bacteriostatic activity, inhibiting protein synthesis by blocking the initiation complex[48], interfering the binding of the tRNA at the A site of the bacterial ribosome[1]. It is the second antibiotic against VRE approved by the FDA in 2000. Resistance to linezolid is rare and it still has low prevalence, less than 1 % in enterococci[49], although sporadic resistant cases using different mechanisms have been reported[48].

One of the resistance mechanisms arises through mutations in the copies of the rRNA genes. Depending on the number of rRNA genes mutated in each species (*E. faecalis* has four copies and *E. faecium* six) the resistance conferred is different[2]. For example, a mutation in a single *E. faecium* rRNA gene confers resistance to linezolid between 8-16 µg/ml and three rRNA genes mutated confer higher level of resistance: between 64-128 µg/ml[48].

**Resistance to daptomycin**

Daptomycin is used to treat VRE infections, mainly in the USA, although it has not been approved by the FDA. It is a cyclic lipopeptide with bactericidal activity that interacts with the cytoplasmic bacterial membrane[1,48]. It is inserted into the cell membrane causing "pores" and ion leakage, modifying the homeostasis. The gradient and concentrations of calcium and potassium are deregulated, causing cell death[2,47,48].

Due to the increasing use of daptomycin to treat VRE infections, *E. faecium* present higher rates of resistance compare to *E. faecalis*, which usually is treated with penicillins[48]. Depending on the species, there have been identified different types of mutations that confer daptomycin resistance. These mutations are related with alterations in the structure of the phospholipids of the cell membrane, preventing the interaction of the antibiotic with the membrane and its depolarization[1,2].

## 1.6 Epidemiology and genomics

In the last two decades, a large number of epidemiological studies have been performed to study the prevalence of enterococci infections and the clinical factors that increase the risk of infection[41]. The awareness of the enterococcal infections started with a first *E. faecalis* outbreak described in the late 1970s but during the last 20 years the outbreaks have been caused mainly by *E. faecium* strains, which are more resistant to antibiotics and more frequently found to date, being now more common as nosocomial infection than *E. faecalis*[1,50]. The causes that have led to this replacement are associated to the increase and extensive use of broad-spectrum antibiotics in hospitals besides the genomic characteristics that increase the capacity of *E. faecium* to adapt to hospital-associated environments.

*E. faecium* strains are characterized by having a clearly evolutionary lineage, being able to distinguish different clades, which are completely different from *E. faecalis*, whose

isolates are so close related that different strains cannot be separated into different groups[51].

One of the methods most commonly used to classify the diversity among *E. faecium* has been the multilocus sequence typing (MLST). This technique classifies *E. faecium* isolates based on the allelic variation of seven housekeeping genes (*atpA*, *adk*, *ddl*, *gdh*, *purK*, *gyd*, and *pstS*[52,53]).  A number is given to each allele of each of the MLST genes, resulting in a number with seven digits for each strain. This number is known as sequence type (ST)[51]. MLST approach has revealed that *E. faecium* isolated from hospitalised patients clusters in a specific group, designated as CC17 (clonal complex type 17) or ST17 (sequence type 17), being this group genetically different from commensal isolates. All the CC17 strains are enriched with antibiotic resistant and virulence genes[1,51]. Phylogenetic analysis that include whole-genome sequencing (WGS) have been used in combination to MLST methodologies to pool the entire information of the genome (comparative genomics studies) with the information given by the allelic variation of the seven housekeeping genes[45,54].

Several studies using MLST and/or comparative genomics of the whole genome have divided the *E. faecium* population in two different clades: clade A, associated to hospital isolates and clade B, related to commensal strains not causing disease[53,55,56].

One of the lately genomic studies that has been performed includes clinical, commensal and animal enterococcal strains from diverse environmental conditions to determine the origin of diversity, with a total of 73 strains[57] (Figure 5).
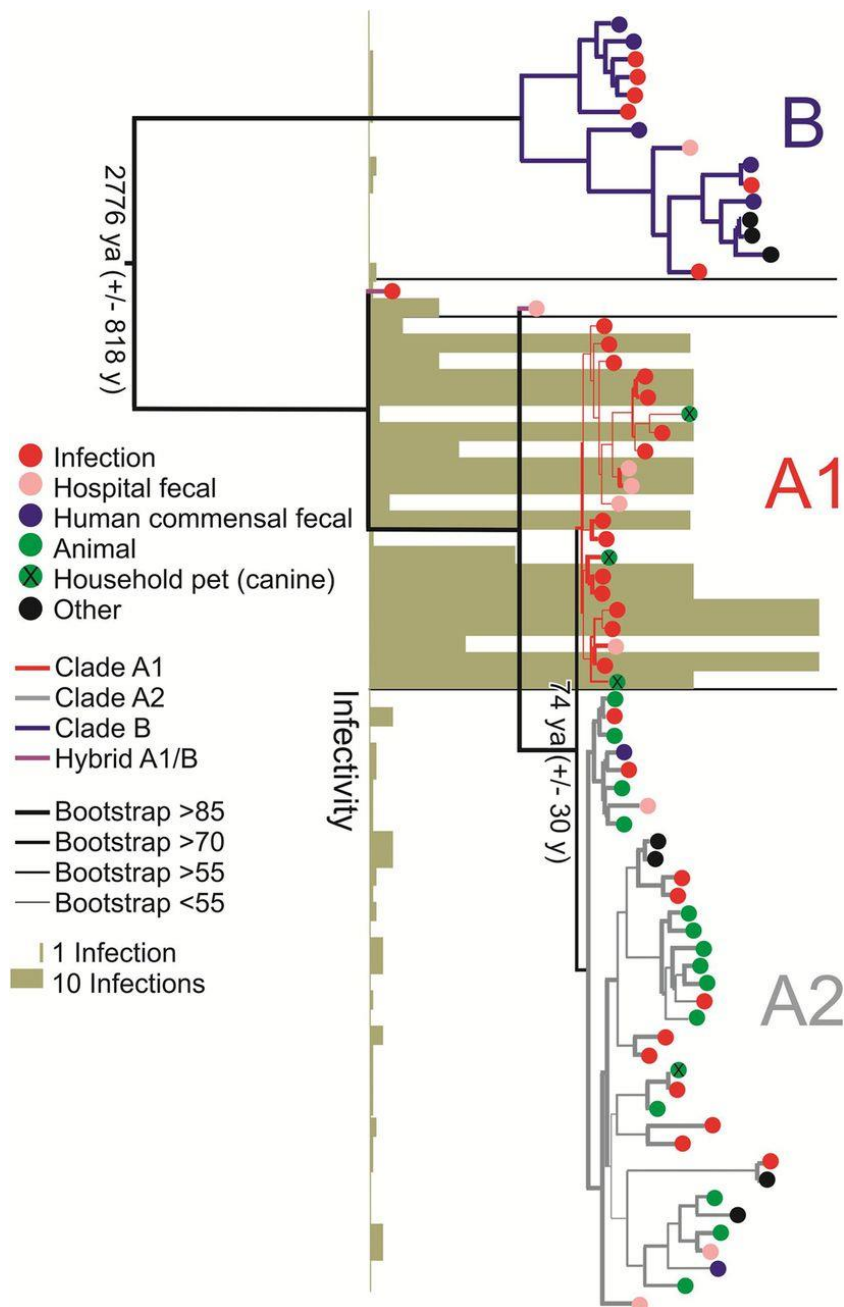
**Figure 5. Phylogenetic tree of *E. faecium* strains from different origins.** RAxML SNP-based of the alignments of the DNA of 1344 single-copy core genes in 73 *E. faecium* genomes. Boostrapping was performed with 1000 replicates. By BEAST analysis the split dates of the different clades are estimated and indicated (ya, years ago). The infectivity score (brown bars) indicates the strains of a particular ST isolated from infection. The colours used for each clade: clade B (commensals) in dark blue, clade A1 (hospital-associated strains) in red and clade A2 (animal-associated strains) in green. Figure reproduced from Lebreton el al[57].

This study confirmed the existence of the two previous identified clades but also identified a second split inside clade A, defining subclades A1 and A2. Subclade A1 contains the strains producing hospital infections, including sequence type ST17 strains, while animal strains are mostly classified into clade A2. Clade B contains mainly commensal human enterococci. Subclade A1 strains, the ones that produce hospital

infections, possess larger genomes and core genome size, more mobile genetic elements and higher mutation rate[57]. The divergence between clade A and B seems to correspond to the timing of urbanization period and domestication of animals, around 3000 years ago, related to the increase of hygienic measures and the insulation between the microbiota of humans and animals. The split between clade A1 and clade A2 has been calculated to occur 75 years ago approximately, according to the clinical use of antibiotics, although antibiotics have been used both for clinical and farming and other reason might have caused the separation of both subclades. These both A subclades share some traits like the higher ability to acquire mobile genetic elements, carbohydrate pathways and some resistances to antibiotics[57]. For example, vancomycin resistance has been found in clade A1 and A2 but it is completely absent from commensal strains from clade B[13]. Another recent study has confirmed the divergence between commensal *E. faecium* strains (clade B) with those from clinical strains (clade A1), and those from animals (clade A2). In addition this study found also sharing of a few traits and resistance genes between A subclades[58].

This classification is not absolute due to the fact that there have been strains isolated from bloodstream that cluster in clade B and strains isolated from healthy humans that has been classified as clade A[57], but in general their classification correlates with commensal (clade B) or hospital origin (clade A1). In other cases, it has been observed that some strains possess hybrid genomes from clade A1 and B, showing the overlapping of the niche between commensal and clinical strains[13].

Other important characteristic of the genome of *Enterococcus* clinical species, both *E. faecalis* and *E. faecium*, is the absence in the major part of the strains of the clustered regularly interspaced short palindromic repeats, known as CRISPR locus[59], which is associated with *cas* genes to performed the antiviral defence[60]. The CRISPR locus is present in approximately the 40-45 % of bacterial genomes, being associated to *cas* genes[60,61]. Nevertheless, in the case of *E. faecium* it has been found homologous sequences of the CRISPR, however the *cas* genes were absent and therefore the CRISPR has no function[60,61]. This absence of CRISPR functions could explain the diverse prophage elements integrated in some *E. faecium* strains that are absent in *E. faecalis* strains in which CRISPR-cas homologous sequences were present, acting as a barrier to acquire exogenous DNA[60]. The lack of CRISPR-cas system could explain the ability to acquire mobile genetic elements that encode antibiotic resistance and virulent genes[59,61].

## 1.7  Nutritional requirements

Enterococci used to be classified simply as lactic acid bacteria but actually they have versatile metabolism, which is responsible for their capacity to widely colonize and survive in extreme conditions, enhancing their resistance to antibiotics. Several studies have been recently performed using new molecular tools to better understand the different metabolic pathways and genes involve in the diverse enterococci functions[62].

The requirements of vitamins are diverse but all the enterococci cultures tested by Sherman and Niven[63], required pantothenic acid, nicotinic acid, pyridoxine and biotin. However, requirements for riboflavin and folic acid are variable depending on the strain[63]. All strains were able to synthesize thiamine[63] and the genes and the biosynthetic pathway for its production have been detected later in the enterococci genome deposited in the databases[62]. Enterococci also require cofactors such as lipoic acid, demethylmenaquinone and hematin[62].

Amino acid needs are different between individual enterococci although some are shared among them. Arginine can be hydrolysed by all enterococci being decarboxylated and used as a source of energy by several strains[62,64]. Only *E. faecalis* can use agmatine (an aminoguanidine naturally derived from arginine) and L-serine for growing, whereas only around 50 % of *E. faecium* can hydrolyse agmatine. The majority of enterococci can decarboxylate tyrosine and phenylalanine, although they cannot be used as a source of energy[64]. Moreover enterococci are auxotroph for various amino acids depending on the species studied. For example, all 23 different *E. faecalis* strains used in the study performed by Murray et al.[65], share to be auxotroph for four amino acids which are L-histidine, L-isoleucine, L-methionine and L- tryptophan. A high variability was observed among different strains for other amino acids[65].

An important pathway for gram-positive and lactic acid bacteria, like some enterococci, is glycerol metabolism to synthetize phospholipids[62,66]. Most enterococci species, including *E. faecalis* and *E. faecium*, are able to ferment glycerol to give lactate as the final product under aerobic and microaerophilic conditions[64].

Homeostasis and maintenance of the transport of ions are essential mechanisms for enterococci to maintain cytosolic concentrations and to adapt to extreme alkaline pH and different salt concentration environments[62]. Two of the important ions for cellular metabolism to maintain homeostatic concentrations, salt and pH are sodium and potassium[62,64]. Other important ions are calcium, although no specific uptake system has been described for enterococci. Cofactors, like cooper and manganese, with variable oxidation states play roles in regulatory metabolic pathways and stress[62,64].

Additionally, enterococci can use iron as a nutrient, although it is not an essential requirement. However, its acquisition has been related to play a potential role in virulence, as it has been shown when *E. faecalis* is cultivated in blood[67]. Several regulatory and iron uptake mechanisms have been described in the genome of *E. faecalis* V583, which are homologous to other bacteria. These mechanisms include ABC transporter, *feo*A and *feo*B operons, ferric uptake regulator-like sequences[67] and SUF machinery for [Fe-S] cluster biosynthesis[68]. The SUF system has been associated to the Firmicutes phylum, although it has been related to increase virulence in other microorganisms such as *Mycobacterium tuberculosis*. One of the subunits of the SUF system, the SufB protein, has been described to play an essential role in *M. tuberculosis* survival, conferring resistance to iron limitation[68].

Another important element for bacteria living is phosphorus in the form of phosphate[69,70]. It is a vital nutrient for different cell functions and processes in bacteria, being found in nucleotides, lipids, proteins or sugars. It is part of the membrane in phospholipids and incorporated into proteins through post-transcriptional modifications. It is also used as energy for bacteria cells, being acquired and manage by specific regulatory and transport systems for phosphate[69].

Carbohydrates sources are important nutrients for enterococci to survive in the GI. All enterococci species (with the exception of few of them) can produce acid from 15 carbohydrates which are N-acetyl glucosamine, amygdalin, arbutin, cellobiose, D-fructose, galactose, β-gentiobiose, glucose, lactose, maltose, D-mannose, methyl- β-D-glucopyranoside, ribose, salicin and trehalose[64]. Moreover, over 30 more carbohydrates are used by at least two species of the genus[62]. The increase of sequenced *Enterococcus* genomes are revelling additional carbohydrate sources that could be metabolized by enterococci[62]. The sequenced genomes from different *Enterococcus* species revealed that on average of 15.4 % of genes from the core-genome belonged to carbohydrate metabolism, although there are significant differences among strains[6]. Moreover, recent studies have suggested that the use of a broad range of carbon sources allows enterococci to grow and survive under competitive environments like the GI and that the ability to utilize new carbohydrates are the main drivers of speciation[7,62]. Moreover, the study previously mentioned[57], that confirmed the existence of the two clades A and B and subclades A1 and A2, showed that one of the main differences between clades is the gain of path and specialized operons for carbohydrates uptake. In addition, one of the main differences that allows to distinguish clades was the identification of new virulence genes and insertion sequences (IS) present in pathogenicity islands (PAI) in clade A, hospital-associated strains[13]. One of the group of genes identified to be inserted in mobile genetic elements were phosphotransferase systems (PTS), which sense and phosphorylate sugars in order to introduce them inside the cell. These systems have been enriched in strains of clade A1 and some of them have been detected in mobile elements[13,57]. Besides sugar uptake, the PTS systems play additional roles by regulating gene expression and also mutagenic processes[71]. PTS systems are formed by two cytoplasmic domains (IIA and IIB), which are hydrophilic subunits, specific for a concrete sugar and one or two integral proteins embedded in the phospholipid bilayer (IIC and IID). IIC subunit is a hydrophobic protein functioning as a sugar permease, while IID subunit could be present or not as second sugar-specific membrane protein for carbohydrate transport[71].

Referred to the presence of PTS in hospital isolates, it has been shown that a specific PTS system of *E. faecium* was present in strains of clade A1 (isolates from hospital origin) and it was absent in human commensal strains (clade B)[72]. The authors identified this PTS by comparing all the available *E. faecium* sequenced genomes, which allow them to detect the absence of a cluster of genes in the commensal strains, including the *pst*D gene, a protein that theoretically forms a membrane-associated Pst complex. Notably, deletion of *pst*D impaired the ability of *E. faecium* to colonize the gut of mice. However, the specific sugars uptaken by this particular PTS system have not been identified.

Besides this particular PTS system, no other studies have identified specific genes utilized by *E. faecium* to colonize the intestinal tract, a major objective in this thesis.

## 1.8 Virulence factors of enterococci

Besides the mechanisms of antibiotic resistance, enterococci also have other virulence factors that increase their pathogenic potential. Some of them are intrinsically present in their genomes but some others have been described to be in mobile genetic elements.

Most studies performed on virulence factors have focused mainly on the study of *E. faecalis* strains because historically this species was the main cause of infections[73] and although *E. faecium* has emerged recently as a leading cause of multi-drug resistant pathogen infections, its virulence factors are not extensively studied. Here, in the introduction of this thesis, we have mainly focused on the new virulence factors that have been characterised for *E. faecium* that are implicated in their infectivity. These properties contribute to adhesion and translocation through the intestine and its attachment to other organs and tissues[73].

### 1.8.1 Surface proteins

Opportunistic bacterial pathogens, including *Enterococcus*, have developed the capacity to adhere to host molecules to facilitate its colonization[74]. Especially, clinical strains are able to adhere to extracellular matrix of the cell host such as collagen, fibrinogen or fibronectins[74,75]. Here we describe some virulence factors, which are surface proteins of the enterococci that have been described mainly in *E. faecium* strains to bind to molecules encoded by the host (Table 2).

**Table 2. Enterococcal virulent factors related to the expression of surface proteins.**

| Surface proteins as virulent factors | Binding to |
|---|---|
| Acm | Collagen type I and IV |
| Scm | Collagen type V |
| EpbC$_{fm}$ | Polystyrene surface |
| Esp | Abiotic surfaces |
| SrgA | Fibrinogen and Nidogen |
| EcbA | Collagen type V and fibrinogen |

A first collagen-binding adhesin (Acm) was identified in *E. faecium* with the property to adhere to collagen. It is a new adhesin that belongs to the collective of MSCRAMMs (microbial surface components recognizing adhesive matrix molecules), which binds to the extracellular matrix of host cells. The *acm* gene is present in both clinical and commensal isolates but only the clinical ones bound to collagen type I and IV. *acm* gene was sequenced and whereas clinical isolated shared 100 % identity, some mutations, deletions and inserted sequences or codons were found in commensal strains, suggesting that it is essential to have the intact *acm* gene to acquire the ability to bind

to collagen[76]. Moreover, the deletion of *acm* gene in clinical strains confirmed the attenuated adhesion of the mutant *in vitro* and in experimental sera from patients with *E. faecium* endocarditis[77].

Recently, it has been characterized the major autolysin of *E. faecium*, $AtlA_{Efm}$, and its relation with the surface exposure of Acm, which is located at the septum and at the poles in the WT strain[78]. It was proved that $AtlA_{Efm}$ autolysin is involved in biofilm formation and stability. The mutant strain reduces its resistance to lysis, the capacity to attach to cells and the ability to produce biofilm. In addition, the mutant had lower capacity to bind to collagen types I and IV. This last characteristic is directly associated with the location of Acm in the mutant, which was exclusively detected at the septum[78].

A second adhesin that efficiently binds to collagen type V, which is expressed and located on the surface of clinical *E. faecium* isolates, is Scm. The secondary protein structure classified the adhesin in the collective of MSCRAMMS. The binding properties compare to Acm are different because the binding is less efficient for collagen type I and IV and fibrinogen but presents high adherence to collagen V[79].

Another enterococcal protein identified in clinical *E. faecium* strains, which has characteristics of the MSCRAMMs collective is the $EbpC_{fm}$ operon. This operon is a cluster of three genes that encodes structures swelling from the cell surface known as pili. The operon deletion reduced the adherence and biofilm formation and the mutant attenuated the infection cause in a urinary tract experiment with mice[80].

The enterococcal surface protein (*esp*) gene was hypothesized to be present in *E. faecium* before the presence of genes that confers resistance to some antibiotics such as ampicillin or glycopeptides[81]. It is related to clinical strains because it has been mainly detected in hospital outbreaks, being associated to increase the *E. faecium* virulence[81,82]. The *esp* gene is located in a pathogenicity island (PAI)[60,82] of different size, depending on the strain studied (between 64 and 104 kb)[60]. It has been demonstrated *in vitro* that the whole pathogenicity island carrying the *esp* gene is able to be transferred horizontally (through conjugation) between different strains and be inserted in an equivalent region of the genome of the new strain[60]. The deletion of the *esp* gene has confirmed its role in biofilm formation[83,84], and its implication in urinary tract infections[85] and endocarditis[86].Moreover the seven open reading frames that are flanking the *esp* gene are genes that enhance virulence and antibiotic resistance[81].

On the one hand, the cell surface protein SgrA that contains the conserved peptide motif LPxTG (formed by the amino acids leucine-proline-x-threonine-glycine) and on the other hand the EcbA proteins that belongs to MSCRAMM group, were enriched and conserved among strains related to hospital outbreaks. The SgrA is a surface adhesin that binds to fibrinogen and nidogens, components of the extracellular matrix of human intestinal epithelial cells, human bladder and kidney cells and is implicated in biofilm formation. The EcbA proteins binds to collagen type V and fibrinogen too. The expression of both surface proteins that have the ability to bind to cell matrix contribute to the increase of *E. faecium* virulence[87].

Nevertheless, the involvement of all these genes in intestinal colonization *in vivo* has not been studied. Only *in vivo* enhanced adherence to heart damage valves has been demonstrated for the Acm protein, playing a key role in the development of endocarditis[88].

In summary, adhesins and surface proteins play an essential role in cell attachment to the extracellular matrix of the host cells being an important first step for *E. faecium* adhesion and further colonization.

### 1.8.2 Virulence-associated genes

An hyaluronidase gene, designated as $hyl_{Efm}$, has been found in clinical *E. faecium* isolates resistant to vancomycin[89]. In a study performed by Rice et al.[90], the gene was described to be present in a transferable plasmid, pLRM23. The authors verified that the transconjugants harbouring the conjugative plasmid pLRM23 expressing the protein were enhanced to colonize the gut compare to those strains without the plasmid, suggesting that $hyl_{Efm}$ gene is important to colonize the intestine of mice[90]. Nevertheless, further studies are needed to elucidate that other genes encoded by the plasmid are not required to enhance intestinal colonization capacity of the strain containing the conjugative plasmid.

Genes involved in bile salt hydrolysis have been detected in *E. faecium* genome, being an important factor that contribute to growth and survival in the intestinal tract, and therefore are considered as virulence factors. Bile salts are part of bile, besides other compounds such as phospholipids and proteins for which bacteria have to be able to adapt to and tolerate them to survive in the gut[91]. The *bsh* gene encodes a protein that hydrolase bile salts and was first detected in *E. faecium* strains from food[92]. Lately, it has been identified in all the *E. faecium* genomes available. Other two genetic loci, encoding the Gls-like proteins, are also involved in resistance to bile. The homologous of these two gen clusters enhance resistance to bile salt in *E. faecalis*. The double mutant for the two loci decrease the tolerance to bile salts and the growth in a mouse peritonitis model, demonstrating that both loci play an important role in bile salt tolerance because neither single loci deleted showed decrease in resistance to bile salts[73,93]. In addition to these two loci, the gene *gltK*, which encodes a glutamate/aspartate bile transport system permease protein, was detected and confirmed to be a gene that confers resistance to bile[91].

Some other genes have been detected in some *E. faecium* strains such as *gelE* gene, which hydrolyses gelatine, collagen, casein and haemoglobin and contributes to biofilm formation[94,95].

### 1.8.3 Carbohydrate metabolism

It is clear the relevance of carbohydrates metabolism as nutritional requirement but at the same time carbohydrates are an enhancer of the colonization capacity and virulence of *E. faecium.* Some genetic determinants related to the use of carbohydrates or their transport have been detected as virulence factors.

Interestingly, some of the genes involved in sugar pathways in *E. faecium* are located in a pathogenicity island that can be transferred through horizontal transfer. Suggesting that these genes were acquired to adapt to new environments. One example is the location of a phosphotransferase system (PTS) and a glycoside hydrolase cluster that are involved in the inositol metabolism and are located in a hot spot of the *E. faecium* genome where is usually located the PAI encoding the *esp* gene[60]. Other example is the detection of an acquired genomic island that encodes for accessory functions which are enriched in clinical *E. faecium* isolates. This genomic island is involved in the uptake and the metabolism of carbohydrate sources[96].

Four carbohydrate-binding molecules, known as LysM domain, mainly present in the extracellular matrix of cells, have been identified, after performing transcriptomic analysis, as relevant molecules in the infection process of *E. faecium*[97]. The expression the genes of *E. faecium* was compared between two different conditions, growing the bacteria *in vitro* and after its isolation from inside the mouse peritoneum. The genes that encode for these proteins are organized in an operon and modulate the no covalent binding to the N-acetylglucosamine of the peptidoglycan of the matrix. To characterize them, corresponding mutants were constructed. Specifically, one of the proteins mutated seems to have a key role and its deletion led to lower capacity of *Enterococcus* binding to kidney cells and a growth defect in human urine. To demonstrate their effect *in vivo*, two mice model were used. First, the bacteria were inoculated intravenously and secondly, mice were infected with a catheter in the urethra. Using these models the authors demonstrated that the infection caused by the mutants was lower in kidneys, livers and bladder.

As previously mentioned, in the study performed by Zhang et al.[72], a PTS locus was identified to enhance the capability of *E. faecium* to colonize the intestine through the use of carbohydrates under antibiotic treatment. This PTS locus is present in the major part of hospital clinical strains but absent in commensal strains. After obtaining the mutant for one of the subunit proteins of the PTS, a mixture of the wild-type and the mutant strain was inoculated orogastrically to mice treated with antibiotics. The results corroborate the phenotype, the reduced capability of the mutant strain to colonize the murine gut compared to the wild type strain. However, the authors could not conclude if the defect of this mutant to colonize the gut was due to a defect in the uptake of sugars, and which particular sugars could be internalized by this PTS.

The capacity to form biofilms by *E. faecium* has been related to other carbohydrate phosphotransferase system (PTS), the gene *bepA*[98]. This gene is present only in clinical *E. faecium* strains and absent in commensal ones. Fitness competition experiments between the wild-type and mutant strain using a rat endocarditis model resulted in lower ability of the mutant to infect rats. To confirm the involvement of this gene in biofilm formation, an *in vitro* experiment was made using human serum. Using this assay, the authors could conclude that the biofilm formed was thicker in the WT than in the mutant strain. They found that the PTS was involved in the metabolism of β-glucoside, which is related to the metabolism of glycosaminoglycans. These sugars are

present in the heart valves. Thus, the deletion of the PTS impaired the metabolism of these sugars and the capacity to form biofilms[98].

Recently, though transcriptomics and a high-throughput transposon mutant library sequencing methodologies, it has been shown that three genes that encode subunits of PTS from *E. faecium* confer an important role to grow in the human serum. The mutant for one of these three genes was not able to grow in human serum and using an *in vivo* zebrafish model the authors demonstrated that the virulence of the mutant strain was reduced compare to wild type, reducing the mortality of the embryos. By homology with a PTS permease system in *Lactococcus lactis,* it was suggested that this PTS system was involved in glucose uptake. Interestingly, glucose is the unique sugar present in serum in higher amounts. Thus the inability of the mutant for this particular PTS to grow in serum may be related to its role in the uptake of the most abundant sugar in serum. It is noteworthy to mention that the PTS described in each study are different, demonstrating the existence of several PTS in *E. faecium* and their importance in different aspect of *E. faecium* virulence[99].

The importance of the carbohydrate transport and the PTS transporters has been demonstrated in *E. faecalis* too, with the identification of a PTS which enhances the growth in gluconate, enabling gut colonization and aggravating colitis[100].  Another relevant sugar for *E. faecalis* is maltotriose, the smallest sugar of the maltodrextrin family. This sugar could be transported by a PTS system, via MalT, that is a specific permease for maltose or by an ATP binding cassette (ABC) transporter. Although both transporters could internalize maltotriose, the ABC transporter had the major effect in reducing the ability of *E. faecalis* to colonize kidneys and liver[101]. Nevertheless, to what extend maltotriose utilization was important for *E. faecalis* colonization was not elucidate in this particular study.

An study performed by Lebreton et al.[57], conclude that PTS systems in *E. faecium* are one of the main drivers by which strains have been divided and classified in clade A or B because strains from clade A acquire more PTS systems on mobile elements and clade B strains eliminated the obsolete PTS from their genomes. Clade A1 strains have acquired mobile genetic elements that encode PTS for the theoretical use of carbohydrates such as fructose, sorbose and mannose, being one of these transporters the previous mentioned PTS[72].

The presence of *E. faecium* domination in the gut has been related to clinical diseases such as graft-versus-host disease (GVHD), a disease suffered by patients with haematological malignancies after allogeneic hematopoietic cell transplantation (allo-HCT)[102]. Importantly, the patients that develop GVHD after allo-HCT have a lower percentage of survival. Previous studies showed that the reduction of microbial diversity due to antibiotic therapies increase the risk of GvHD and mortality in allo-HCT patients. However, the role of particular bacterial species in the development of GvHD was not clear. Both in patients and using mouse models the authors demonstrated that antibiotic treatment promotes the expansion of *Enterococcus* after microbiota depletion, which promotes GvHD development[102]. Moreover, the authors

demonstrated that the expansion of *Enterococcus* was associated with the lactose present in the gut because the depletion of this sugar in the mouse food attenuated *Enterococcus* expansion, and reduced GvHD development. Furthermore, analysed patients were divided depending on their capacity to absorb or not lactose and those who could not absorb lactose had longer *Enterococcus* domination[102]. Thus, this study also pinpoint to carbohydrate metabolism as key in the *Enterococcus* gut colonization and development of the diseases in which *Enterococcus* is involved. Studying carbohydrate metabolism by *Enterococcus* and the microbiota will be also a major objective of this thesis.

Importantly, no systematic study has been performed to evaluate the effect of every gene from the *Enterococcus* genome in its intestinal colonization capability. Thus, just little is known about the genes that are required by *E. faecium* to colonize the gut and the nutritional requirements needed to expand, which is a major objective of this thesis.

### 1.8.4 Transcriptional regulator genes involved in pathogenicity

Some regulatory genes have been related to the adaptation and pathogenicity of *Enterococcus*, enhancing its ability to adapt to different environments, substrates or niches[103].

One example of this regulatory genes is a LacI family regulator gene, known as *mdx*R gene, which regulates the expression of eight genes (including the gene cluster mdxABCD-pulA predicted to encode maltose/maltodextrin ABC transporter proteins), necessary for maltodextrin utilization in *E. faecium*[104]. The maltodextrin are breakdown products from starch that are digested by the host. These products can reach the colon and can be metabolized by *Enterococcus*. The authors confirmed that the presence of the *mdx*R gene is required for the growth of the bacteria in the presence of maltodextrin, sugar that could be important for the growth of the *E. faecium* in the gut, although this hypothesis still needs to be tested[104]. In *E. faecalis*, it has also been identified a global transcriptional regulator, MafR, which affects the transcription of multiple genes, including PTS transporters, ABC membrane transporters and other genes related to the use of carbohydrates[103]. Interestingly, the deletion of MafR decreased the growth of the bacteria under difference carbon sources such as glycerol, maltose and mannitol and reduced the inflammatory response produced by *E. faecalis* in a mouse model of peritonitis. The authors hypothesized that this transcriptional regulator gene was key for the usage of different carbohydrates which enhance the ability of the bacteria to grow and increase *E. faecalis* virulence[103].

On the other hand, it has been identified an enterococcal biofilm regulator gene, *ebr*B (previously known as an AraC-like regulator), located upstream of the *esp* gene[105]. This gene was involved in the regulation of the expression of the *esp* gene (required as previously described for biofilm formation). Deletion of this regulator gene decreases the levels of the Esp protein in the surface of the cell. Moreover, abolishment of the Esp protein in the bacterial membrane impaired its ability to form biofilm in a flow system and conferred lower capability to colonize the small intestine of mice[105].

In summary, it has been described some regulatory genes that could be essential for the regulation of key genes involved in *Enterococcus* virulence but further studies are required to completely elucidate the multiple regulatory genes that could be involved in *Enterococcus* virulence and specifically in gut colonization.

## 2. INTESTINAL COLONIZATION BY ENTEROCOCCI

As it has been previously mentioned, *Enterococcus* pathogenic strains frequently start the infection through colonization of the intestinal tract and expansion after antibiotic treatment that causes an alteration of the microbiota. However, under normal conditions the clinical *Enterococcus* strains do not colonize the intestine because the microbiota plays a protective role inhibiting its growth. In the following section we will describe: (I) what is the microbiota and its roles, (II) mechanisms and bacteria from the microbiota that are key in the resistance to pathogen colonization.

### 2.1 The microbiota and their functions

As previously explained, antibiotics have a tremendous impact on the microbiota composition what leads to the loss of the protection capacity that healthy microbiota promote against intestinal colonization by pathogens. The microbiota plays an essential role of defence against external infectious bacteria but also microbiota alterations and poor gut health are related to several non-infectious diseases[106].

The human microbiome are all the microorganisms that colonize the human body including bacteria and other microorganisms such as fungi, viruses, and protozoa[107]. A study performed to characterize the microbiota of the entire body of western adult populations revealed remarkably differences among healthy individuals and among the different body microbiota present in the gut, skin, oral cavity and vagina[108].

An especial interest has been focused on the gut microbiota due to its involvement in numerous human diseases[107]. The human gut microbiota is complex, reaching around $10^{14}$ microorganisms in the whole intestine that are in a dynamic and stable equilibrium[106,109]. However, the bacterial load fluctuates along the intestinal tract (Figure 6), being lower in the stomach and at the beginning of the small intestine. The amount of microorganisms increases in the jejunum and ileum with around $10^4$-$10^7$ bacteria and the highest amount is reached in the colon[110]. It is estimated that there are more than 1000 bacterial species in the GI, most of them anaerobes.
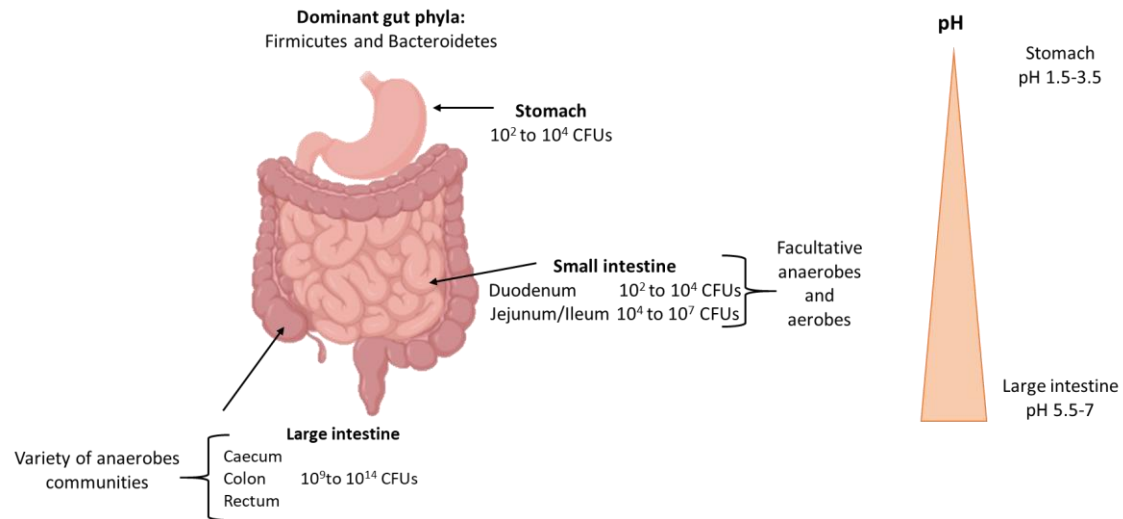
**Figure 6. Bacterial load distribution along the gastrointestinal tract.** The density and the amount of bacteria increases from the stomach to the jejunum/ileum and reach the highest level in the large intestine. pH limits the presence of bacteria in the stomach and the number of the bacteria increases with the increase in the pH. Figure created in Biorender.com.

The microbiota is acquired after birth, being the gut of the fetus sterile before that moment. The acquirement and the future development of the healthy bacteria depends on several factors: the delivery mode, the diet received as new-borns, the dietary habits during the adulthood and the use of antibiotics or the infections developed[106,107,111]. The most abundant bacteria found in the healthy gut microbiota are obligate anaerobes from Firmicutes and Bacteroidetes phyla[106,107] that contain around a 90 % of the bacteria present in gut microbiota[112]. Facultative anaerobic bacteria, mainly from the *Enterobacteriaceae* and *Enterococcaceae* families, are part of the microbiome but at much lower levels (more than 100 fold lower)[110].

The microorganisms of the GI take part of a global and complex ecosystem, cooperating with the host and its regulation. The study of its transcriptome, proteome and metabolome help to describe and understand the mutual interaction[106,111]. One of the main benefits to the host is the trophic organization of the microbiota to metabolize nutrients which reach the intestine and provide energy to the host, working as a bioreactor to break down polysaccharides to ferment monosaccharides by anaerobic bacteria[113].

Several studies have demonstrated the beneficial role of the normal gut microbiota and how microorganisms are involved in diverse processes that include drug metabolism, nutrient metabolism of peptides and proteins, fermentation of some carbohydrates and fibre from diet and structural functions such as barrier and immunomodulation with the production of constant signal molecules, induction of the immune and neuronal systems and the defence against pathogenic microbes[106,107,111]. This last function is also known as colonization resistance, which is described in detailed in the next subsection since one of the objectives of this thesis is to study mechanisms by which the microbiota confers protection against VRE.

## 2.2 Colonization resistance

Among the diverse roles of the microbiome previously mentioned, one of the main functions is the protection and defence against the establishment and growth of external pathogens, known as colonization resistance (CR). Nevertheless, this function is abrogated when antibiotic treatments deplete commensal bacteria, allowing the entry of the external pathogens. The colonization resistance term was first introduced by Waaij et al.[114], in 1971. After treating mice with antibiotics they observed the loss of protection against external organisms that was returned with the repopulation of the microorganisms after ending the treatment. The protection was mainly provided by anaerobic bacteria.

CR takes place through two major mechanisms[115,116]: (I) the microbiota induces a host response, known as indirect mechanisms. This indirect mechanism involves, for example, the production of antimicrobial peptides by epithelial cells, the maintenance of the epithelial barrier and the mucus layer besides the production of bile acids as toxic products to block the growth of external pathogens[116]; or (II) the microbiota directly confers protection mainly through secretion of inhibitory molecules or competition for nutrients[116]. Both types of mechanisms are impaired when the normal microbiota is absent or disturbed[117]. In the major part of the cases, both mechanisms are connected and can work together to inhibit the growth of pathogens, including *Enterococcus*[118].

### 2.2.1 Indirect mechanisms

### 2.2.1.1 Innate immune system

Toll-like receptors (TLR) from Paneth cells, present in the intestinal epithelium, are activated by components of the microbiota, such as lipopolysaccharide or flagellin. This activation induces the production of RegIII-γ, a C- type lectin that is effective killing gram-positive bacteria[119]. RegIII-γ acts by binding to the phospholipids of the bacteria membrane and forming pores which produce the bacterial death[120]. It has been demonstrated in mice that antibiotics, by depleting the microbiota, reduce the expression of RegIII-γ, which promotes VRE growth[121]. In contrast, administration of lipopolysaccharides to antibiotic-treated mice stimulates the TLR of the paneth cells and the production of RegIII-γ, reducing the capability of VRE to colonize the gut[121].

The maintenance of the colonic mucus is another factor that enhance CR[115]. The colonic mucus has a key role in the separation of the bacteria from the colon epithelia[122]. Commensal bacteria induces the production of mucus, thus in mice that received antibiotics, the mucus layer is thinner, which enhance *Citrobacter rodentium* attachment to epithelial cells and promotes a more severe infection caused by the pathogen[123]. In addition, RegIII-γ expressed by paneth cells remains in the mucus and also reduces the ability of bacteria to get in contact with the epithelium[124]. In fact, RegIII-γ deficiency favours *Listeria* infections by increasing the bacterial contact with the epithelium[124]. Thus, the maintenance of an intact mucosal barrier and expression of antimicrobial peptides contributes to inhibit GI colonization by pathogens.

## 2.2.1.2 Adaptive immune system

The interaction between the microbiome and the adaptive immune system involves the participation of B and T cells, crucial mediators in the defence against pathogens to maintain the gut homeostasis[125,126].

The microbiota plays an important role in the differentiation of T cells into T helper cells (Th) and T regulatory cells (Tregs), and the microbiota influence the amount of these cells present in the intestine. Moreover, the cytokines produce by these cells such as IL-17 and IL-22 by Th cells confer protection against pathogens and the production of IL-10, TGF-β or IL-35 by Tregs control inflammation[126].

The microbiota influences the development of B-cells, important mediators to maintain the gut homeostasis trough the production of IgA antibodies. These antibodies are key to avoid infection in the intestinal mucosa. It has been suggested that short chain fatty acids (SCFA) produce by the microbiota, might influence the production of IgA because mouse receiving a diet rich in fibre (major source to produce SCFA by microbiota) had higher levels of IgA. Importantly, mice with higher levels of IgA were more resistant to *C. rodentium* gut colonization.

## 2.2.2 Direct mechanisms

On the other hand, the microbiota is able to confer protection against pathogens without the need of the immune system, by direct mechanisms.

The microorganisms can produce and release substances such as bacteriocins to directly inhibit the expansion of a pathogen. For example the bacteriocin production by *L. lactis* to inhibit VRE growth[127].

Another important mechanism that confers protection is the production by the microbiota of SCFAs, which are produced as a result of the fermentation of complex carbohydrates that reach the large intestine. The more abundant SCFA present in the large intestine are acetate propionate and butyrate. The type of diet as well as the antibiotic treatment can affect the levels of SCFA in the large intestine through disruption of the intestinal microbiota[128]. SCFAs inhibit the growth of *E. coli* pathogenic strain O157:H7 by reducing the intracellular pH, being propionic acid the most effective in inhibiting *E. coli* growth[129].

Moreover, commensal bacteria are able to produce molecules that interfere with the expression of specific genes of the pathogen. One example is the production of the quorum-sensing molecule AI-2 by *Ruminococcus obeum* that decrease the colonization of *Vibrio cholera* because the molecule blocks the expression of the *tcp* operon needed to colonize the gut[130].

Additionally, the nutrients that are available in the gut are an important factor for pathogens to colonize the gut, and therefore the pathogens need to compete with commensal bacteria for the same resources. This nutrient competition could modulate and influence the expansion of the pathogen, being also a major mechanism of colonization resistance, especially competition for the main carbohydrates present in

the gut[126]. This mechanism has been studied for different *Enterobacteriaceae* pathogens such as *E. coli*[131]. The infection caused by pathogenic strain of *E. coli* can be prevented by colonizing the mouse gut with other two commensal strains, which use the sugars required for the pathogenic strain to grow[131].

With respect to VRE, an attempt to identify such mechanism was performed using *ex vivo* assays[132]. However, results from this particular study could not establish any relation between nutrient depletion by commensal microbes and VRE growth and concluded that inhibitory molecules produced by the microbiota rather than nutrient competition plays a role against VRE colonization.

The inhibition of pathogens through these diverse direct mechanisms could be produced as specified above by a healthy microbiota. In addition, these mechanisms of protection can be restored in individuals with an altered microbiota through the administration of a faecal microbiota transplantation (FMT), the administration of a defined bacterial consortia or by the administration of a single bacterium with the specific capabilities to block the expansion of a pathogen. Examples for different mechanisms of protection induce by the administration of different sources of commensal bacteria to restore colonization resistance against pathogen colonization are described in detail in the next sections.

### 2.2.3 Approaches to restore colonization resistance

**Faecal Microbiota Transplantation**

The Faecal Microbiota Transplantation (FMT) is the transfer of a healthy microbiota from one individual to another that has major alterations in its microbiota and subsequently suffers a specific infection or disease. FMT is used to restore the commensal bacteria of the gut and subsequently re-stablish the healthy state (Figure 7).
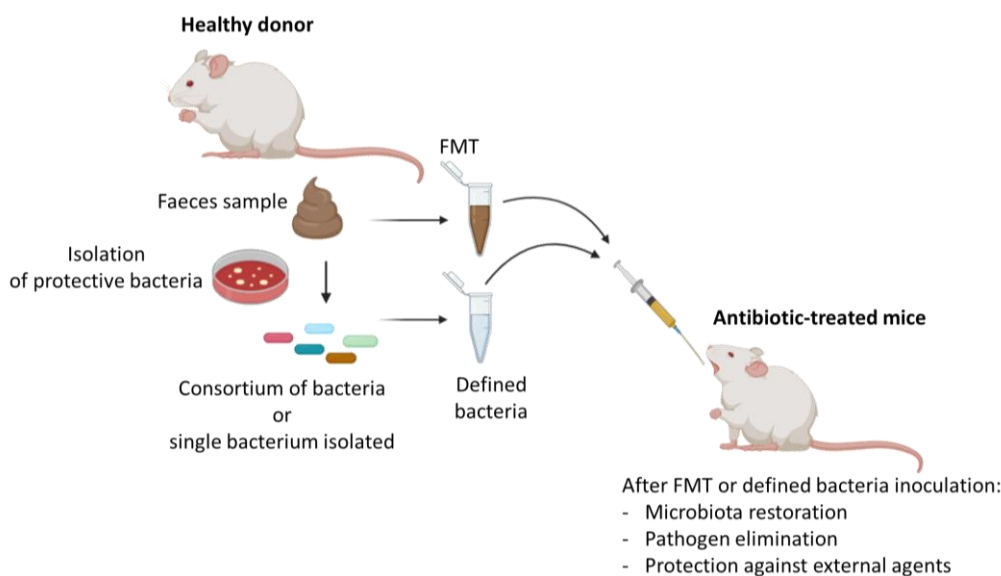


**Figure 7. Schematic representation of the oral administration of FMT or consortium of bacteria.** The administration of FMT from a healthy donor or a defined bacterial consortium to antibiotic-treated mice restores the microorganisms present in the gut and restores colonization resistance (CR). Figure created in Biorender.com.

*C. difficile* produces an intestinal infection that can be fatal for a patient's life. It is associated with the use of antibiotics that alters the microbiota, which promotes the expansion of the pathogen[133]. Generally, to treat these infections, repeated doses of oral metronidazole or vancomycin are administered (antibiotics that are effective against *C. difficile*)[133,134]. However, these antibiotics promote additional changes in the microbiota. For this reason, very frequently, there is a relapse of the infection after the antimicrobial therapy has been completed[133]. The use of faecal transplant has been reported to be an effective treatment against the recurrence of *C. difficile* infection. Thus the restoration of the intestinal microbiota through FMT is sufficient to inhibit *C. difficile* growth, avoiding the use of antibiotics and therefore precluding the infection relapse[133,134].

Several clinical trials have confirmed the effectiveness of the FMT to treat recurrent *Clostridium difficile* infection (CDI) with no adverse associated effects[133–135]. After FMT, the fecal microbiota of patients was studied. The patients contained similar microbiota to that of their healthy donor, with increased bacterial diversity as compare to samples obtained before the FMT administration[134]. Since this particular study, multiple clinical assays have proven to be effective against CDI[136]. Thus FMT has become an important treatment for CDI and in most cases it has been shown to be more effective than conventional antibiotic therapy.

Besides the application in CDI, FMT has been applied to treat other infections or co-infections with several pathogens. CDI and VRE infections could be spread in the hospital environment when patients are exposed to multiple antibiotic treatments. It has been observed that sometimes CDI induces a subsequent VRE infection due to the use of vancomycin to treat CDI, which causes a co-infection[137]. In a clinical study, patients who had recurrent CDI and also developed VRE infection were treated with FMT in order to clear both pathogens. From the 11 patients that were positive for VRE colonization (36 %) 8 eliminated VRE after the FMT. 2 patients were not able to clear either CDI nor VRE[137]. Although the results of this study are promising, the authors did not include a control group that did not receive the FMT for comparison, and the microbiota was not analysed. Thus, additional studies should be performed in order to clarify if the FMT also clears VRE gut colonization in humans. In mice, however, it has been clearly shown that FMT can restore the microbiota of antibiotic-treated mice and eliminate completely VRE gut colonization, which persists in the absence of FMT[138].

Patients who develop blood stream infections are usually colonized by more than one pathogen. In a study with 20 participants colonized by antibiotic resistant pathogens (ARPs), ranging from 1 to 4, FMTs were supplied and 75 % of the participants were decolonized. The resistant pathogens that colonized the gut of those patients include bacteria such as *Klebsiella pneumoniae*, *Escherichia coli*, *Pseudomonas aeruginosa* and VRE among others[139]. As in the previous study, there is not an adequate control group, being a limitation to interpret the results.

Other clinical study administrating FMT was performed in patients colonized with carbapenem-resistant *Enterobacteriaceae* (CRE) and VRE[140]. After three months, it was

observed a moderate effect with 2 of 8 patients (25 %) decolonized from CRE and 3 from 8 (38 %) from VRE. They could not identify differences to distinguish patients that were decolonized from those who had persistent colonization in terms of duration of carriage, comorbidities, hospitalization period, antibiotic therapy or antimicrobials after FMT. Although they did not observe any adverse incident and the FMT seems to be safe, the study has important limitations such as very small sample size, non-randomized, and patients followed a limited three-month period.

FMT technique is rapid and effective for some infectious diseases but this methodology also involves risks that need to be considered. It is important how the faeces are stocked for future faecal transplantations, not to lose part of the anaerobic microorganisms involved in protection and to study the properties of the microbiota not to expose the recipient to further diseases driven by the transplanted microbiota such as inflammatory or metabolic diseases[136].

To try to avoid the risk that the entire microbiota transplantation have, such as reported cases of bacteremia and even some deaths[141], there have been studies focused on the research of specific bacteria from the microbiome that confers protection to the infection. Therefore, the FMT could be replaced by an optimal combination of a bacterial consortium or a single bacterium, known as probiotics, which could confer the same protection than the faecal transplant.

**Optimal defined bacteria transplantation (studies performed by other groups)**

Several studies have verified that the supply of a defined bacterium or a bacterial consortium is able to impair VRE gut colonization.

Different species of *Lactobacillus* have been shown to be able to diminish VRE gut colonization after their administration to mice. The administration of *Lactobacillus rhamnosus* Lcr35 decreases VRE gut colonization and although VRE clearance was not complete, the density of VRE in the gut was reduced[142]. The supply of *Lactobacillus paracasei* CNCM I-3689 significantly reduced *E. faecalis* strain resistant to vancomycin in faeces, besides promoting a quicker recovery of members of the phylum Bacteroidetes[143]. Thus both a direct effect of *L. paracasei* or an indirect effect of the recovery of Bacteroidetes could be involved in the inhibition of VRE gut colonization. *Lactobacillus murinus* Y74 and *Lactobacillus plantarum* HT121 reduced VRE gut colonization after antibiotic treatment. In this particular study, the authors compared the administration of *Lactobacillus* with the administration of an FMT. *Lactobacillus* restored bacteria from Firmicutes Phylum faster than FMT, whereas FMT restored Bacteroides quicker, but in both cases microbiota diversity was restored faster than in the control group[144]. However, the specific mechanisms of protection against VRE were not identified in any of these studies.

On the other hand, it has been demonstrated that anaerobic bacteria eliminated with an antibiotic treatment could provide a robust defence against oxygen-tolerant bacteria, like enterococci. A study showed that obligate anaerobe bacteria, such as bacteria which belongs to *Barnesiella* genus, were associated with VRE clearance in the gut of mice

previously treated with ampicillin. Microbiota analysis of patients undergoing allo-HCT revealed that those patients who were colonized with *Barnesiella* were in general more resistant to subsequent VRE gut domination[138].

Another study demonstrated that the administration of a consortium of four anaerobic bacteria reduced the levels of VRE colonization in mice[145]. These four isolated bacteria were: *Blautia producta*, *Clostridium bolteae*, *Bacteroides sartorii* and *Parabacteroides distasonis* (CBBP). *In vitro*, it was shown that *B. producta* was the bacterium that directly inhibits the growth but *in vivo* all of them were required in a cooperative manner. The authors demonstrated two different levels of cooperation to confer resistant to VRE *in vivo*. First of all, the presence of *Bacteroides sartorii* and *Parabacteroides distasonis* was necessary since they secrete a β-lactamase that confers resistance to ampicillin. Since the mice were treated with ampicillin, these two bacteria were required for the intestinal colonization of *B. producta* and *C. bolteae,* which are sensitive to ampicillin. Secondly, the presence of *C. bolteae,* besides *P. distasonis*, and *B. sartorii,* is necessary for the intestinal colonization of *B. producta*, the bacteria required to directly inhibit VRE growth.

Subsequently, the same authors performed a follow-up study in which they focused on the mechanism by which *B. producta* inhibit VRE growth[146]. First, they demonstrated that exclusively the *B. producta* strain isolated from mice could inhibit VRE since other *B. producta* (what they called *B. producta* control) did not inhibit VRE intestinal colonization. Subsequently, the authors showed that the caecum content of the *B. producta* that could inhibit VRE contains a protein that is able to inhibit VRE growth. This inhibitor was found to be a lantibiotic expressed by *B. producta*. The results obtained in mice correlated with the results obtained from patients undergoing allo-HCT, in which samples with high lantibiotic abundances contained low abundances of *E. faecium*.

The production of bacteriocins to inhibit the growth of VRE have been detected in organisms such as *L. lactis*, that produce anti-enterococcal peptides, which are small molecules with antimicrobial activity[127]. Other bacteria expressing bacteriocins are certain strains of *E. faecalis,* which contain the conjugative plasmid pPD1, which encodes a bacteriocin that inhibit other *E. faecalis* strains, favouring competition for the same niche. This strategy was successfully used to eliminate antibiotic resistant *Enterococcus* from the intestine in mice[147].

In addition to bacteriocin production, it has been shown that other probiotics such as *Bifidobacteria* can confers protection against toxigenic *E. coli* infection through production of acetate. *Bifidobacterium* is able to metabolize carbohydrates through expression of an ABC sugar transporter. Fermentation of these carbohydrates produces acetate. Subsequently it was shown that acetate preserves the mucosal epithelia, which prevents dissemination of the Siga toxin (Stx) produced by the *E. coli*[148].

In addition to production of specific molecules that inhibit pathogen growth or preserve mucosal epithelia, nutrient competition has been shown to be a major mechanism of colonization resistance against *Enterobacteriaceae* pathogens, and especially

competition for the main carbohydrates present in the gut, as previously mentioned in the section above (section 2.2.2, Direct mechanisms). Competition for the same sugars was shown in mice to confer protection against *Citrobacter rodentium*[149]. The levels of the pathogen were reduced after *E. coli* administration, which consume similar simple sugars as *Citrobacter rodentium*. In contrast, the protection was not conferred by *Bacteroides thetaiotaomicron,* which consumes preferentially polysaccharides, leading monosaccharides available for *Citrobacter rodentium*[149]. Similar competition mechanisms have been studied in *E. coli* infections produced by a pathogenic strain[131]. The infection could be prevented in mice with the previous colonization of two commensal *E. coli* strains that consume all the different sugars that are mainly used by the pathogen. However, when only one commensal strain was administered, it was not able to consume all the sugars and the pathogen was still able to colonize the gut through consumption of the available sugars[131].

**Bacterial consortia capable of restricting VRE colonization (studies performed by our group)**

Previous results obtained in our laboratory by a previous PhD student also identified a bacterial consortium that is associated with protection against VRE. This bacterial consortium was identified by first studying how different antibiotics change the microbiota and promote VRE gut colonization. The rationality was that antibiotics that eliminate commensal bacteria relevant for protection against VRE will be the ones that will promote to a higher extend VRE gut colonization. While antibiotics that do not promote VRE gut colonization would be killing commensal bacteria dispensable for VRE gut colonization. By analysing the different changes promoted by the different antibiotics we should be able to identify commensal bacteria associated with protection against VRE. Therefore, to understand how microbiota changes induced by antibiotics promote the loss of colonization resistance and the VRE expansion, mice were treated one week with antibiotics of different spectrum (i.e. clindamycin, ceftriaxone, neomycin, ciprofloxacin, ampicillin or vancomycin). Subsequently, mice were inoculated with VRE and the levels of the pathogen were analysed 2 days after the inoculation. Different antibiotics induce different microbiota changes (Figure 8A), which were associated with a different capacity of VRE to colonize the gut of these mice (Figure 8B), being the effect greater for vancomycin and clindamycin treatments. Antibiotic-induced dysbiosis persisted 2 weeks after antibiotic cessation (Figure 8A, Antibiotic+recovery). Nevertheless, certain degree of microbiome recovery was detected for most antibiotics. Consistent with a partial recovery of the microbiota, the capacity of VRE to colonize the intestinal tract was in most cases reduced 2 weeks after antibiotic cessation.
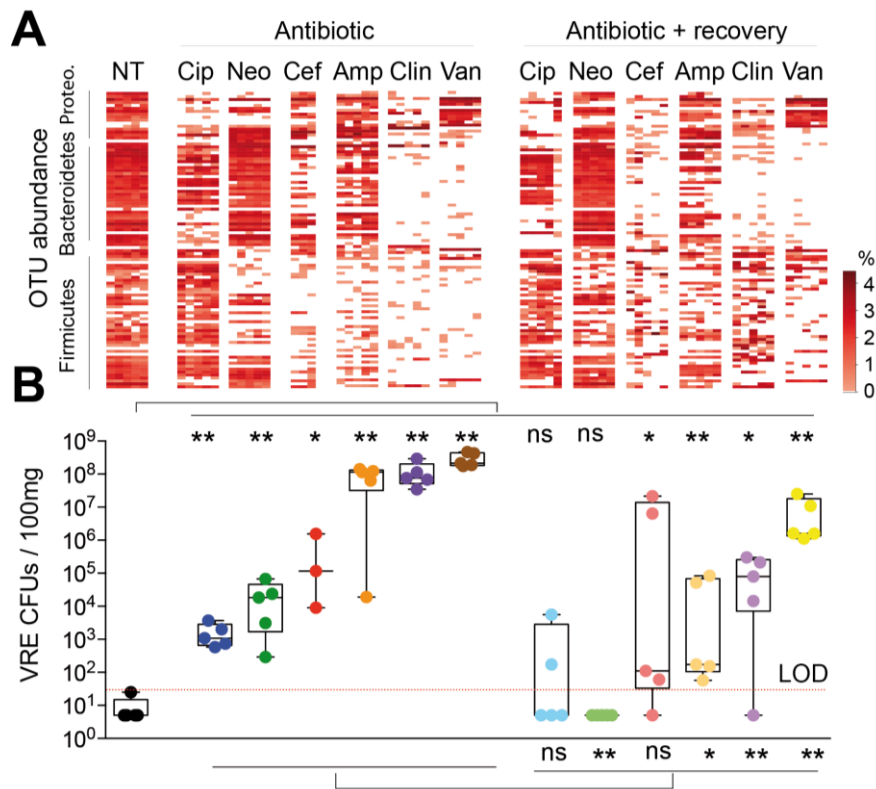
**Figure 8. Different antibiotics induce distinct dysbiosis states and grades of susceptibility to VRE colonization. (A)** Heatmap that shows the abundance of the most abundant Operational Taxonomical Units (OTUs) identified in the collected faecal samples from untreated mice (NT) or mice that received during one week Ciprofloxacin (Cip), Neomycin (Neo), Ceftriaxone (Cef), Ampicillin (Amp), Clindamycin (Clinda) or Van (Vancomycin). "+ recovery" indicates that after antibiotic administration the mice were allowed to recover for a couple of weeks before fecal sample collection and VRE inoculation. **(B)** VRE fecal levels 2 days post-VRE inoculation in mice receiving the different treatments. LOD = limit of detection. Points below the LOD indicate those mice in which it was not able to detect any VRE CFU after plating the faecal samples.

To identify the specific microbial populations that conferred resistance in untreated mice and whose depletion by antibiotics promote VRE gut colonization, changes in the different microbial populations were associated with the capacity of VRE to colonize the gut of mice. Spearman correlation analysis identified specific genera whose abundance was negatively associated with VRE colonization and extensive culturing of caecal contents was performed from untreated mice to isolate them (Figure 9A). Specifically, strains from the genera *Barnesiella*, *Oscillibacter*, *Olsenella and Alistipes* were isolated. In addition, an isolate defined as unclassified_Ruminococcaceae that was highly negatively correlated with VRE levels was also isolated. Consistent with a possible protective role of these taxa against VRE colonization, mice highly susceptible to VRE intestinal colonization contained lower levels of these specific taxa as compare to resistant mice (Figure 9B).
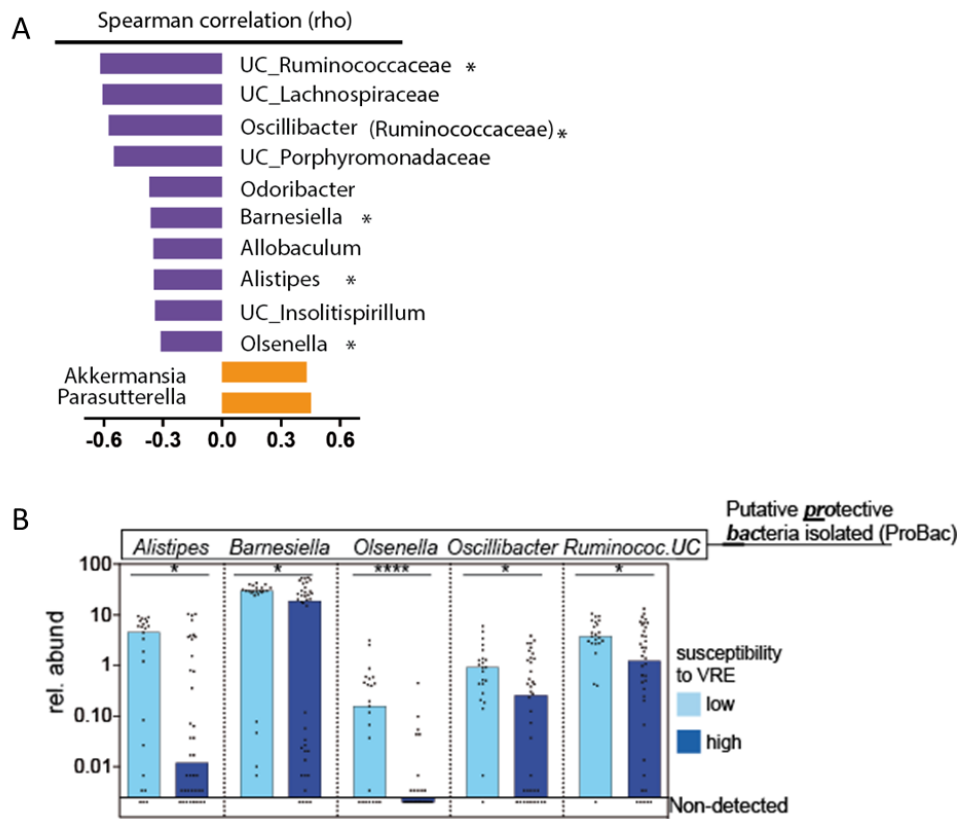
**Figure 9. Commensal bacterial taxa significantly associated with VRE levels. (A)** Spearman correlation (q<0.05) between the faecal levels of bacteria immediately before VRE inoculation and the fecal levels of VRE, 2 days post-VRE inoculation. The bars represent the rho values. Purple colors represent negative correlations while orange colors represent positive correlations. * denotes taxa that could be isolated. The correlation was done with the sequence classification at the genera level. If no genus could be assigned, the family classification was used preceded by UC (unclassified). **(B)** Relative abundance of the five bacterial taxa that could be isolated. Samples are divided in those collected from mice with high (>$10^4$ VRE CFUs / g) or low (<$10^4$ VRE CFUs / g) susceptibility to VRE intestinal colonization. ProBac (Protective Bacteria) is the acronym given to the bacterial mix containing the 5 isolates from the 5 taxa indicated.

To confirm the role of these putative protective bacteria (ProBac) in protection against VRE colonization, the bacteria were inoculated to mice that were previously treated with vancomycin. We selected this antibiotic due to the highest level of VRE colonization after antibiotic cessation, which is associated with permanent changes in the microbiota (Figure 8A), including the permanent loss of the 5 isolated bacteria (Figure 10).
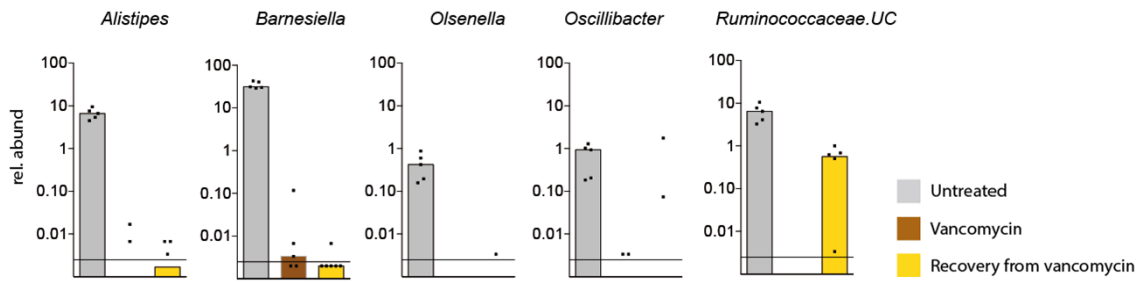
**Figure 10. Levels of the bacteria found to be associated with protection against VRE and that we were able to isolate.** The relative abundance of each bacterium in the group of untreated mice, mice treated with vancomycin and mice allowed to recover from vancomycin treatment is shown.

It was confirmed that, after two weeks of vancomycin cessation, mice that had received ProBac were significantly more resistant to VRE colonization that those mice that did not received the bacterial consortia (Figure 11).
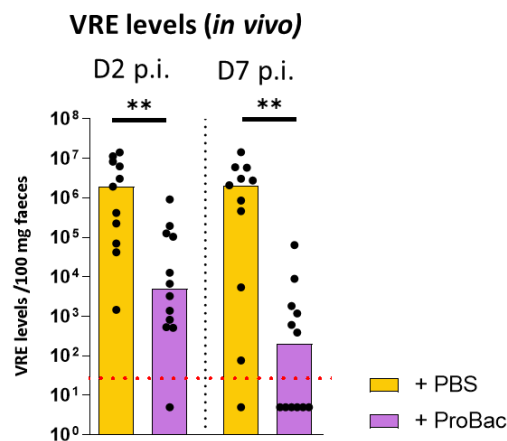


**Figure 11. ProBac consortium restores protection against VRE colonization during antibiotic-induced dysbiosis.** VRE levels in faecal samples collected 2 and 7 days after VRE inoculation (p.i.) from control group of mice that received the vehicle for bacteria administration (PBS-glycerol-cysteine) after vancomycin cessation and the ProBac group of mice receiving the five commensal bacteria after vancomycin cessation. Mice were inoculated with VRE, 2 weeks after vancomycin cessation.

The Probac protective effect was associated with restoration of the levels of the administered bacterial isolates (Figure 12), but not with an overall increase in microbiota richness or diversity. Unexpectedly, isolates from the taxa *Oscillibacter* or Ruminococcaceae_UC were also recovered in vancomycin treated mice that did not receive ProBac, suggesting that the recovery of these two taxa was not sufficient to confer protection.
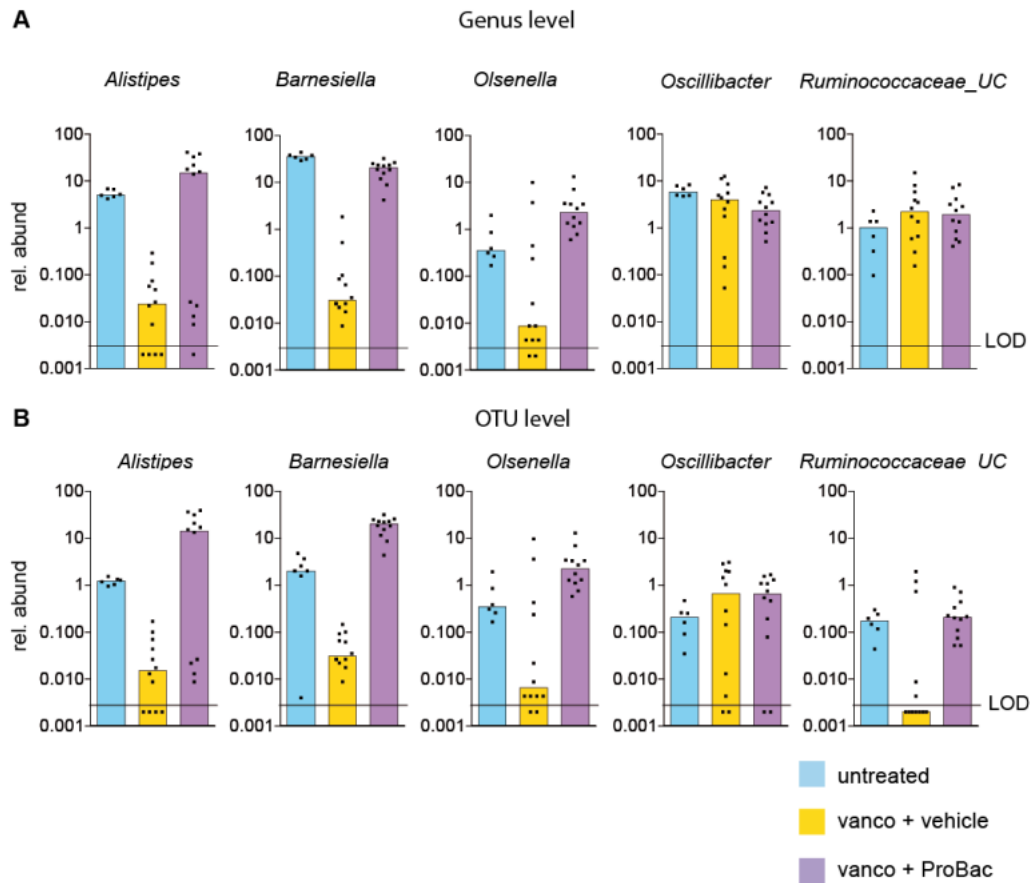
**Figure 12. Restoration of bacterial taxa lost after vancomycin treatment through the administration of ProBac.** Relative abundance at the genus **(A)** or OTU **(B)** level of those taxa depleted after vancomcyin treatment and whose recovery was induced by the administration of ProBac. The groups of mice indicated in each graph are mice that did not received any treatment (untreated), mice treated with vancomycin and allowed to recover 2 weeks after stopping the antibiotic treatment (vanco+PBS) and vanco+ProBac: same as vanco+PBS but in this case mice received ProBac during 3 consecutive days, starting one day after stopping the antibiotic treatment.

Subsequent studies performed by the previous PhD student showed that intestinal contents containing Probac were able to suppress VRE growth in the absence of host (*ex vivo*) suggesting that Probac was conferring protection through a direct mechanism. However, the specific mechanism by which ProBac confers protection was not elucidated. A major task that we have developed in this thesis

# II.   OBJECTIVES

The general goals of the present study are to characterize the genetic and nutritional requirement of Vancomycin-Resistant *Enterococcus* (VRE) required to colonize the intestine. In addition, we will investigate the mechanisms by which the ProBac consortium confers resistance against VRE.

The specific objectives that we propose in order to reach these two major goals are:

1. Construction and utilization of a transposon mutant library in VRE in order to identify genes that are necessary for gut colonization through high-throughput sequencing.

2. Transcriptomic analysis in order to identified genes most expressed by VRE during the process of intestinal colonization.

3. Verification of the genes detected as essential for gut colonization by direct mutagenesis.

4. Characterization of nutrients necessary for the growth of VRE using *in vitro* arrays under anaerobic conditions.

5. Functional characterization of genes that affect gut colonization, with special focus on those related to nutrient acquisition.

6. Characterization through transcriptomics of ProBac functions that could be relevant for conferring resistance against VRE.

7. Characterization through metabolomics of changes in metabolites induced by ProBac that could be relevant for conferring resistance against VRE.

8. *In vivo* and *ex vivo* experiments to demonstrate specific mechanisms (suggested by the omic studies) by which ProBac confers protection against VRE.

All these objectives will allow us to identify the molecular mechanisms by which VRE is able to colonize the intestine, as well as the nutrients consumed that promote its growth and mechanism by which specific commensal bacteria confer protection against VRE.

# III.   MATERIAL AND METHODS

## 1.    Bacterial strains, plasmids and growth conditions

Strains and plasmid used in this study are listed in Table 15 and Table 16 (section 32), respectively.

*E. faecium* strain E1162 was isolated from a bloodstream of a patient in France in 1996[60], and the genome has been sequenced and has been deposited in the GenBank (GenBank Accession Number ABQJ00000000, 2010). This strain was used in the first attempts to construct the mutant transposon library with the first approach (section 12.1) and it was tested to be used with the second approach (section 12.2).

*E. faecium* strain E1133 was isolated in the United States of America from a faeces of a patient during a VRE outbreak in 2002[150] (GenBank Accession Number AHWR00000000, 2013). This strain was tested to construct an optimal mutant transposon library with the second approach (section 12.2).

*E. faecium* strain Aus0004 was isolated from a bloodstream infection in Australia in 1998[61] and its genome has been sequenced (GenBank Accession NC_017022, 2017). This strain was used to construct the mutant transposon library in order to study the genes involved in intestinal colonization, since an optimal library was obtained with this strain.

*E. faecium* strain ATCC 700221 was isolated from human faeces in Connecticut, USA[151] (GenBank Accession Number CP014449, 2016). This strain was used for the experiments performed with the protective commensal bacteria (ProBac) since results obtained by a previous PhD student were obtained with this particular strain.

*E. faecium* strains were grown in Brain Heart Infusion medium (BHI, Pronadisa) or in Bile Esculin Azide Agar (BEA, Akralab) at 30 ºC or 37 ºC in aerobic or anaerobic conditions, as required and specified below.

*E. coli* strains DH5α and EC1000 were grown in Luria-Bertani medium (LB, pronadisa).

*E. coli* strains were routinely grown at 30 ºC, 37 ºC or 42 ºC in aerobic conditions.

*Lactococcus lactis* NZ9000[152] was grown in M17 (Oxoid) medium adding 0.5 % glucose (Sigma), GM17, at 30 ºC in aerobic conditions.

*Oscillibacter*, UC_Ruminococcaceae, *Alistipes*, *Olsenella*, *Barnesiella* and *Bacteriodes, the* commensal bacteria isolated from the mouse intestine named as ProBac in this work, were grown in Columbia Blood Agar (CBA, VWR) at 37º C in anaerobic conditions inside the anaerobic chamber (Whitlet DG250), with a gas atmosphere of 10 % $H_2$, 10 % $CO_2$, and 80 % $N_2$.

All strains were stocked in PBS with 20 % of glycerol at -80 ºC.

When necessary, antibiotics were added at different concentrations: ampicillin (AppliChem) 10 µg/ml for *E. faecium*; vancomycin (Alfa Aesar) 8 µg/ml for *E. faecium*; chloramphenicol (Panreac) 6 and 10 µg/ml for *E. faecium*, 10 ug/ml for *E. coli,* 5 µg/ml for *L. lactis*; gentamicin (Alfa Aesar) 300 µg/ml for *E. faecium* and 25 µg/ml for *E. coli*;

spectinomycin (Alfa Aesar) 300 µg/ml for *E. faecium* and 100 µg/ml for *E. coli*; kanamycin sulfate (AppliChem) 50 µg/ml for *E. coli*.

Growth of all strains was determined by measuring the optical density at 600 nm ($OD_{600}$).

## 2. Bacterial cells transformation

### 2.1 Electrocompetent cells transformation

**VRE**

To prepare electrocompetent VRE cells, an overnight BHI culture was diluted 1000 fold in BHI supplemented with 200 mM sucrose and 1 % of glycine and incubated overnight at 37 ºC and shaking at 150 rpm (revolutions per minute). Bacterial cells were centrifuged and resuspended in new pre-warmed 25 ml of BHI supplemented with 200 mM sucrose and grown for 1 hour at 37 ºC and shaking (150 rpm). After incubation, bacterial cells were washed three times with 0.5 M sucrose and 10 % ice-cold glycerol (wash buffer) and finally resuspended in 1.2 ml ice-cold wash buffer. 50 µl aliquots of bacterial competent cells were mixed with 10 µl of plasmid (approximately 50 ng/µl of plasmid concentration), transferred into an ice-cooled cuvette (1.5-mm gap) and kept on ice for 30 min. The suspension containing the cells and the plasmid was electroporated with these parameters: a potential of 2.5 kV, capacity of 25 µF and shunt resistance of 200 Ω. Bacterial cells were resuspended and incubated for 1 hour at the appropriate temperature for the replication of the plasmid in BHI with 0.5 M glucose. The transformants were grown in BHI agar supplemented with appropriate antibiotic/s and plates were incubated at corresponding temperature (depending on the plasmid replication requirements) for 1-3 days.

***E. coli***

One *E. coli* colony was inoculated into 20 ml of LB in a flask and grown overnight at 37 ºC and 150 rpm. 2 ml of bacterial cells were transferred into new 200 ml LB media and grown in a flask to obtain an initial $OD_{600}$ of 0.05 and incubated at 37 ºC for 2-3 hours with shaking at 150 rpm to reach $OD_{600}$ 0.6. After the incubation period, the bacterial culture was maintained 15 min on ice to stop the bacterial growth and centrifuged at 4 ºC, 4000 rpm for 10 min. Bacterial cells were washed twice with 10 % glycerol. Finally, bacterial cells were resuspended in 1 ml of 10 % glycerol, aliquoted in 50 µl and frozen in a bath of dry ice with ethanol to cool quickly bacterial cells and store them at -80 ºC. Bacterial cells were mixed with 0.5 µg of plasmid and electroporated at potential 1.8 kV, 25 µF of capacity and resistance of 200 Ω. After electroporation, bacterial cells were resuspended in 1 ml of SOC (Super Optimal broth with Glucose) medium for 2 hours and plated in LB agar adding the corresponding antibiotic/s at the appropriate temperature for 1 day.

## *L. lactis*

*L. lactis* was grown overnight at 30 ºC with shaking (150 rpm) in 40 ml of GM17 broth (M17 + 0.5 % glucose) culture. Subsequently, 40 ml of the overnight culture were inoculated into 400 ml of SGM17 broth (M17 medium + sucrose (0.5 M) + glycine (2%) + glucose (0.5%)) and incubated at 30 ºC with shaking (150 rpm) for 5 hours. After the incubation period, the culture was transferred to cold sterile centrifuge bottles (Sorvall) and bacterial cells were centrifuged at 4 ºC, 4500 rmp for 10 min. Bacterial cells were washed twice with 25 ml of ice cold 0.5 M sucrose with 10 % of glycerol solution. Finally, bacterial cells were resuspended in 1 ml of ice cold buffer, containing 0.5 M sucrose and 10 % of glycerol, and aliquoted to store bacterial cells at -80 ºC. Bacterial cells were mixed with 250 ng of plasmid and the electroporation was performed at 25 Mf, 200 Ω, 2 kV. After electroporation, bacterial cells were resuspended in 950 µl of GM17 broth and incubated at 30 ºC for 2 hours. Dilutions were plated on GM17 agar with the corresponding antibiotic/s and incubated aerobically at 30 ºC for 2 days.

### 2.2 Heat shock transformation

To prepare chemically competent bacterial cells, a colony from *E. coli* was grown overnight at 37 ºC with shaking (150 rpm) in 20 ml of LB broth. Next day, new 50 ml LB media was inoculated with 1/100 of inoculum from the overnight culture and incubated for 2 h at 37 ºC with shaking (150 rpm) to reach an $OD_{600}$ of 0.4. The bacterial cells were chilled on ice during 10 min and centrifuged at 4 ºC, 4000 rpm for 10 min. Subsequently, bacterial cells were washed with 1/5 volume of ice-cold 0.1 M calcium chloride ($CaCl_2$). This step was repeated two times. Subsequently, the supernatant was removed and bacterial cells were resuspended in 1/25 volume of ice-cold 0.1 M $CaCl_2$ containing 15 % of glycerol. Bacterial cells were left on ice overnight at 4 ºC and were subsequently aliquoted in 100 µl volume aliquots and stored at -80 ºC with a previous crash cool step, freezing the aliquots in a bath with a dry ice and ethanol to rapidly cool bacterial cells. To transform cells by heat shock, an aliquot was mixed with 0.5 µg of plasmid and maintained on ice during 30 min. Subsequently, heat-shock was induced by incubating bacterial cells for 45 s at 42 ºC. Next, bacterial cells were left on ice for 2 min. 900 µl of SOC medium (see section 31, Table 13) was added to the bacterial cells, which were incubated for 1 hour at corresponding temperature (depending on the plasmid replication requirements). Transformed cells were plated in LB agar plates containing the corresponding antibiotics and incubated for 1-2 days at the suitable temperature for plasmid replication. Temperatures for plasmid replication are indicated in each specific section in which a particular plasmid is cited.

### 3. Phenol-Chloroform-Isoamyl Alcohol DNA extraction

Pellets containing bacterial cells derived from *in vivo* or *in vitro* samples were resuspended in 500 µl of buffer extraction that contained: 200 mM of Tris pH8, 200 mM of sodium chloride (NaCl) and 20 mM of Ethylenediaminetetraacetic acid (EDTA) and added to 2 ml tubes containing 500 µl of glass beads (Sigma). 210 µl of SDS 20 % and

500 µl of Phenol:Chloroform:Isoamyl Alcohol (25:24:1) were added to the 2 ml tubes. Tubes were vortex for 5 min at 13000 rpm. Supernatant was collected and transferred into 1.5 ml tubes adding the same volume of Phenol:Chloroform:Isoamyl Alcohol to the 1.5 ml tubes. Tubes were vortexed until obtaining a homogeneous mixture and subsequently were centrifuged for 5 min at 13000 rpm. The upper phase was collected and transferred into 1.5 ml tubes and sodium acetate (3 M pH 5.2 $NaC_2H_3O_2$) was added in a 1/10 proportion and the tube was mixed gently. Two volumes of 100 % cold ethanol were added. The tubes were incubated overnight at -80 ºC to enhance DNA precipitation and subsequently centrifuged for 20 min at 4 ºC and 13000 rpm, to maintain the tubes cold during the centrifugation step to recover the DNA. The supernatant was removed and the pellet was washed twice with 70 % ethanol. Finally, the pellet was resuspended in 50 µl of $H_2O$ and 1 µl of 10 mg/ml of RNase was added to eliminate RNA present in samples.

## 4.  Plasmid extraction

Plasmid extraction was performed using QIAprep® Spin Miniprep Kit (Qiagen). Briefly, 1-5 ml of overnight bacterial culture were centrifuged at > 8000 rpm for 3 min. The pellet obtained was resuspended in 250 µl of buffer P1 and 250 µl of buffer P2 to lyse the cells. Subsequently, 350 µl of buffer N3 was added and tubes were centrifuged for 10 min at 13000 rpm. The pellet was discarded and supernatant was transferred to spin column and centrifuged. Two washed steps were performed, first adding 500 µl of buffer PB and then 750 µl of buffer PE containing ethanol. Finally, the wash buffer was removed by centrifugation and 50 µl of water were added to elute DNA retained in the column membrane.

In the case of the plasmid extraction from Gram-positive bacteria such as *Enterococcus* and *Lactococcus,* an extra step was added before starting the protocol. Lysozyme was resuspended in the buffer P1, 0.03 g of lysozyme in 1 ml of buffer P1. 250 µl of this mixture were added in the first step and tubes were incubated 30 min at 37 ºC to hydrolyse the peptidoglycan and increase the efficiency of plasmid recovery.

## 5.  PCR using KAPA HiFi polymerase

KAPA HiFi Hot Start Ready Mix (Kapa Biosystems) was used to amplify the PCR products to be cloned in the different plasmids used in the study. The reaction volumes used for the PCR are indicated in Table 3 and the conditions in Table 4.

This polymerase is a proofreading polymerase, designed to obtain high-fidelity PCR with an error 100 times lower than wild type Taq DNA polymerase.

**Table 3. KAPA HiFi Hot Start Ready Mix reaction volumes.**

| Component | Volume |
|---|---|
| 2X KAPA HiFi Hot Start Ready Mix | 12.5 µl |
| 10 mM Forward Primer | 0.75 µl |
| 10 mM Reverse Primer | 0.75 µl |
| Template DNA | Genomic DNA (10 – 100 ng) |
| | Less complex DNA (0.1 – 1 ng) |
| H$_2$O | As required |
| **Final Volume** | **25 µl** |

**Table 4. Conditions used with KAPA HiFi Hot Start Ready Mix polymerase.**

| Step | Temperature | Duration | Cycles |
|---|---|---|---|
| Initial denaturation | 95 ºC | 3 min | 1 |
| Denaturation | 98 ºC | 20 s | |
| Melting Tº (Tm) | 60-65 ºC | 15 s | 25 - 35 |
| Extension | 72 ºC | 1 min /kb | |
| Final extension | 72 ºC | 1 min/kb | 1 |

## 6. DNA digestion using restriction enzymes

Restriction enzymes used were from NEB (New England Biolabs). Digestions were performed at 37 ºC and prepared in a total volume of 25 µl adding: 0.25 µl of each enzyme (between 2.5 and 5 U), 2.5 µl of Buffer 10X (in the case of double digestion, compatible buffer with > 50 % activity for both enzymes), between 150 ng and 400 ng of DNA (plasmid or PCR product) and ultrapure autoclaved H$_2$O. The duration of the digestion depended on the enzyme compatibility and the complexity of DNA, if it was a plasmid or PCR product, ranging from 2 h (generally shorter periods for plasmids) to overnight incubations (for PCR products) at 37 ºC.

## 7. Ligation

Ligations were performed using T4 ligase from NEB. DNA concentrations were measured by Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific) and proportions were calculated adding 3 molecules of insert per one molecule of plasmid taking into account the size of each one. Following this proportion, ligations were prepared in a total volume of 20 µl adding: 1 µl of T4 ligase, 2 µl of T4 DNA ligase buffer 10X, between 20-60 ng of plasmid and the corresponding amount of insert and ultrapure autoclaved H$_2$O, if it was necessary to complete the 20 µl. Ligations were performed overnight at 16 ºC. In general, 10 µl of the ligation were used to transform bacterial cells.

## 8.  TOPO® Ta Cloning® Kit and transformation by electroporation

TOPO® TA Cloning® vector is an efficient cloning strategy for the direct insertion of products amplified with Taq polymerase. The TOPO plasmid is linearized with 3'-thymidine (T) overhangs to cloned PCR products amplified by Taq polymerase that add single deoxyadenosine (A) to the 3' ends containing 3'-(A) overhangs.

In our case, KAPA HiFi was used to amplify the fragments to be cloned in TOPO® TA Cloning® vector (see section 5). However, this proofreading polymerase, with low rate error, removes the 3'-A overhangs necessary for cloning. An additional step using DreamTaq polymerase (Thermo Fisher Scientific) was required to add the A overhangs to the 3' ends of the amplified product necessary to be inserted efficiently in the vector. PCR fragment amplified with KAPA HiFi polymerase was mixed with Taq polymerase (Table 5) and incubated for 15 min at 72 ºC to add the 3'-A overhangs.

**Table 5. Method for adding 3' adenines to clone into TOPO® vector.**

| Reagent | Volume |
|---|---|
| **10X Dream Taq Buffer** | 2.5 µl |
| **2 mM dATP** | 2.5 µl |
| **Taq polymerase 5 U/µl** | 0.45 µl |
| **PCR fragment, DNA** | Approximately 150 ng |
| **H$_2$O** | As required |
| **Final Volume** | **25 ul** |

After addition of the 3' A-overhangs, we set up the TOPO® Cloning reaction (Table 6).

**Table 6. Volumes to set up TOPO® cloning reaction.**

| Reagent | Volume |
|---|---|
| **PCR product (from the previous mix)** **20 – 30  ng** | 4 ul |
| **Salt Solution** | 1 ul |
| **TOPO ® Vector** | 1 ul |
| **Final Volume** | **6ul** |

The reaction was mixed gently and incubated for 5 minutes at room temperature (RT) to proceed with transformation by electroporation. To transform the TOPO® Vector by electroporation, prior salt concentration had to be reduced, diluting the reaction 4-fold with autoclaved ultrapure H$_2$O. Electroporation was performed as previously explained in section 2.1, and bacterial cells were grown overnight at 37 ºC in LB agar plates containing 50 µg/ml of kanamycin.

### 9. DNA clean-up and size selection by magnetic beads

To clean DNA and PCRs, we used the Nucleo Mag® NGS Clean-up and size select kit (Macherey-Nagel), commonly used in the library construction process for next generation sequencing (NGS). We followed the manufacturer protocol to have DNA free of contaminants, such as nucleotides, primers, salts or other smaller fragments not desired. Briefly, DNA samples were mixed with paramagnetic beads that bind DNA, volume 1:1. After magnetic separation using a specific magnet separation system, supernatant was removed and samples were washed with ethanol. To remove ethanol completely, a short drying step of 10 min was performed. Finally, samples were eluted with autoclaved ultrapure $H_2O$.

### 10. DNA Clean-up from agarose gel

The QIAquick Gel Extraction Kit (Qiagen) was used to remove primers, nucleotides and other impurities from agarose gel. The protocol is designed to extract and purify DNA of 70 bp to 10 kb from standard or low-melt agarose gel in TAE (Tris-acetate-EDTA) or TBE buffer (Tris/Borate/EDTA). The DNA fragments were excised from agarose gel with a sharp scalpel and weigh to add 3 volumes to 1 volume gel of high-salt buffer containing a pH indicator to dissolve them completely at 50 ºC. After the gel slices were dissolved completely, the yellow colour needs to be checked to ensure that the pH is around 7.5. The adsorption of the DNA to the membrane of the column is only efficient at pH <7.5. If the colour is orange or purple it is necessary to add 10 µl of sodium acetate ($NaC_2H_3O_2$, 3M pH5). Isopropanol was added in the proportion of 1 gel volume to the samples to increase the yield of DNA and washed with 750 µl of buffer containing ethanol. Finally, the samples were centrifuged and eluted with autoclaved ultrapure $H_2O$ to recover DNA from the membrane of the column.

### 11. Transposon mutant library construction and evaluation

A 50 µl aliquot of VRE competent cells was mixed with pZXL5/pGPA transposon vector and electroporated as described in section 2.1.

To construct the mutant library we used the pZXL5 plasmid for the first attempts, which contains a Gram-positive thermo-sensitive pWV01 replicon, a ColE1 replicon for plasmid replication in *E. coli*, a *mariner* transposase whose expression is induced by nisin, a gentamicin resistance *mariner* transposon with two outward-facing T7 promoters, allowing the generation of RNA products of the flanking regions of the genomic DNA where the transposon is inserted, and a chloramphenicol resistance (*cat*) gene (Figure 13). Note that the T7 promoters are used for the microarray-based strategy but are not relevant for our strategy. After the induction of the transposase by nisin, the gentamicin resistance transposon is randomly inserted into the genome of the bacterium. The system allows single and stable transposition events because the transposase activity is lost with the loss of the plasmid when the bacterium is grown at the restrictive

temperature. The sequence of pZXL5 is deposited in GenBank (GenBank Accession Number: JQ088279).

In a second strategy to obtain the mutant transposon library (Results, section 1.5), we used a modified pZXL5 plasmid, the pGPA plasmid. This plasmid allows the construction of the transposon mutant library coupled with high-throughput sequencing (Tn-seq) to analyze the sites where the transposon is inserted. Thus, the pZXL5 was adapted for being apt for Tn-seq (high-throughput transposon insertion site sequencing) by adding MmeI restriction sties in the repetitive sequences on both sides of the T7 cassette.
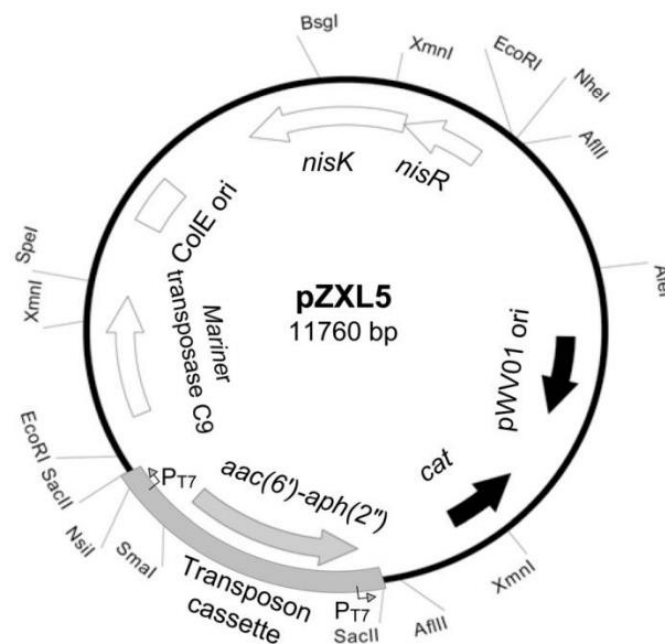


**Figure 13. pZXL5 plasmid map.** The plasmid contains a ColE1 replicon, a Gram-positive thermo-sensitive pWV01 replicon, a nisin inducible *mariner* transposase, including nisA promoter, a chloramphenicol aceyltranferase (cat) gene and a *mariner* transposon with the gene *acc(6')-aph(2")* conferring resistance to gentamicin and flanked by two T7 promoters. Restriction sites are indicated. The figure is taken from Zhang et al[153].

Bacterial cells transformed with the plasmids were plated in BHI plates containing gentamicin since pZXL5 and pGPA contain a mariner transposon carrying a gentamicin resistance gene, *acc(6')-aph(2")*. A gentamicin-resistant transformant was grown overnight in BHI broth adding 300 µg/ml gentamicin and 6 µg/ml chloramphenicol. The bacteria were grown at the permissive temperature, 28 ºC and with 150 rpm of shaking. Next day, 100 µl (approximately $10^8$ bacterial cells) were transferred to 200 ml pre-warmed BHI broth with 300 µg/ml of gentamicin and 25 ng/ml of nisin to induce the transposase. The culture was grown overnight at the non-permissive temperature (37 ºC) with shaking (150 rpm). Finally, 100 µl of the previous culture were inoculated in fresh pre-warmed BHI broth with 300 µg/ml of gentamicin at 37 ºC to select transformants with the transposon inserted in the genome of the bacteria. Culture was aliquoted at -80 ºC in BHI broth with 50 % glycerol. These three overnight steps were

used for the first attempts to construct the mutant library (1st methodology, Figure 14A). For the subsequent attempts, we modified the protocol by adding a fourth night incubation at 37 ºC with gentamicin. 100 µl of the third night culture were inoculated into new fresh pre-warmed BHI broth containing gentamicin, and the culture was incubated at 37 ºC. In this way, we increased the incubation time for the selection of the transformants with the transposon inserted (2nd methodology, Figure 14B).
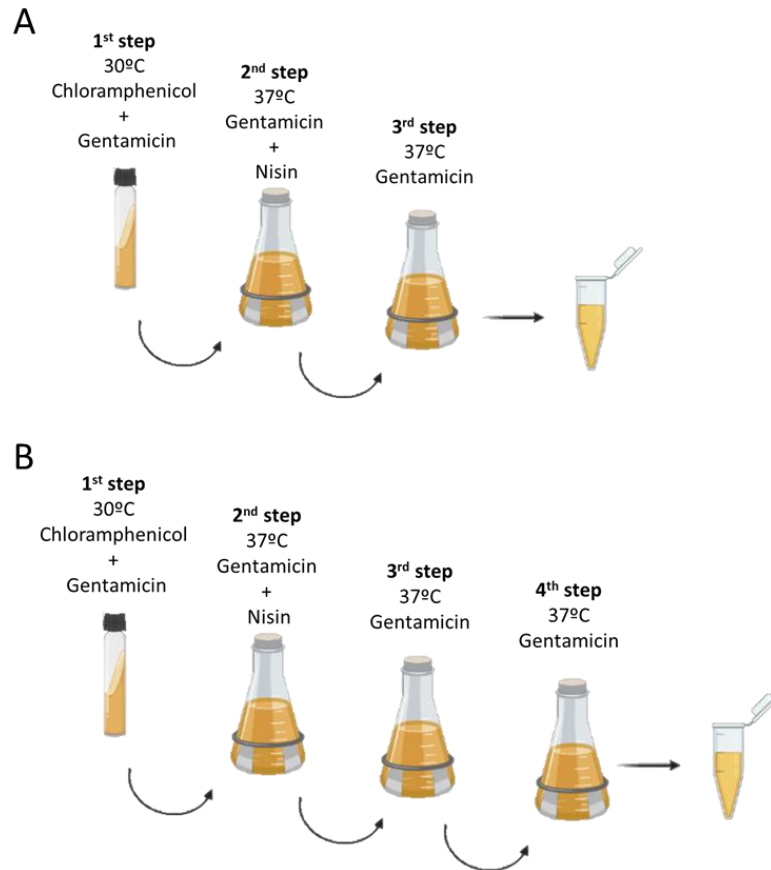


**Figure 14. Methodology used to construct the mutant transposon library.** Gentamicin resistant transformants were culture at the corresponding temperature and adding antibiotic to select transformants with the transposon inserted in the genome. **(A)** First methodology used with three overnight culture steps. **(B)** Second methodology used, adding an extra overnight culture step at 37 ºC to increase the incubation period for the selection of gentamicin resistant bacteria containing the transposon in their chromosome.

To evaluate the randomness of the transposon insertion in the genome, we performed an inverse PCR (iPCR) enrichment of the transposon adjacent sequences of 15 arbitrarily picked gentamicin-resistant colonies from the library, grown in 3 ml of BHI containing gentamicin. Briefly, Genomic DNA of the grown culture was isolated using the Wizard Genomic DNA Purification kit (Promega). 200 ng of the DNA were digested with HaeIII (New England Biolabs) during 3 hours at 37 ºC and then self-ligated during 15 minutes at RT in order to obtain circular DNA. Loci with the transposon inserted were amplified using specific primers iPCR HaeIII R/F with the following conditions: 94 ºC for 1 min; 35 cycles of 94 ºC for 18 sec, 54 ºC for 30 sec, 72 ºC for 8 min; and 70 ºC for 7 min. PCR

products were sequenced to confirm the exact position where the transposon was inserted using the primer iPCR HaeIII F (Figure 15).
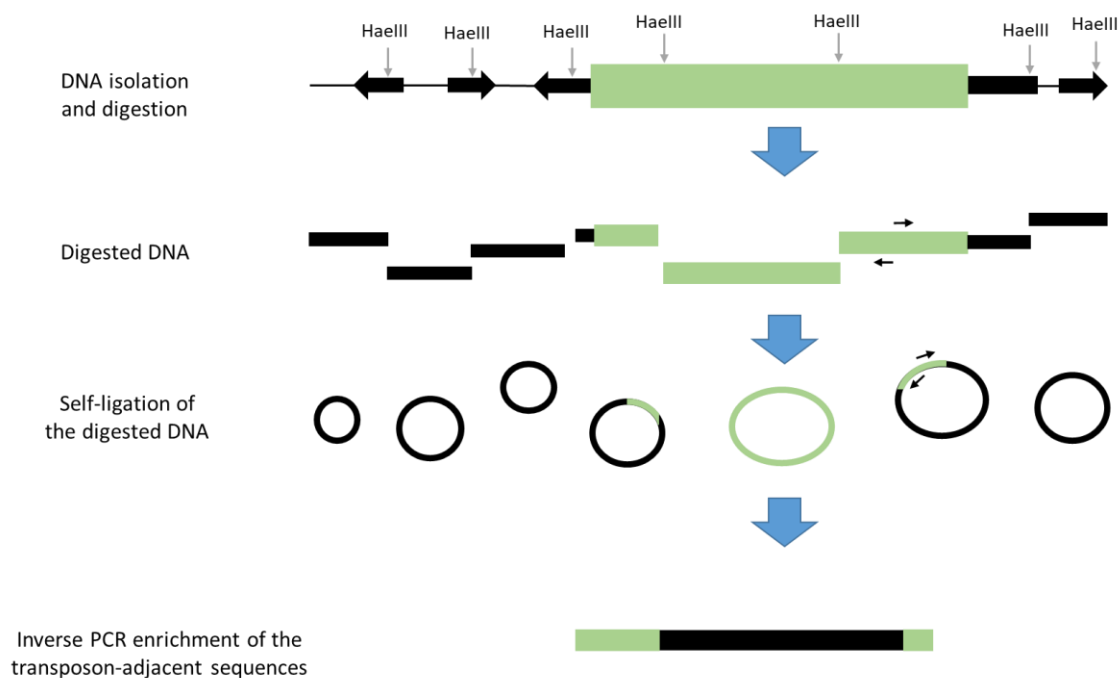


**Figure 15. Schematic diagram of iPCR methodology.** DNA is isolated from individual gentamicin-resistant colonies from *E. faecium* library and digested with the restriction enzyme HaeIII. After digestion, DNA is re-ligated and circularized. A fragment of the transposon and part of the chromosome are amplified by primers (black arrows) that hybridize with the transposon. The transposon is coloured in green and the chromosomal DNA of the bacteria in black.

## 12. High-throughput sequencing of the Tn-seq library

### 12.1 High-throughput sequencing preparation, 1st strategy, using pZXL5 plasmid for library construction

Libraries constructed with pZXL5 plasmid were prepared for high-throughput sequencing by adapting a protocol from Klein et al[154]. Fragments containing transposon and flanking chromosome sequences were amplified by two rounds of PCRs, adding poly-C tails to the fragmented DNA and adaptor sequences for sequencing (Figure 16).

Briefly, DNA from the library was extracted using Phenol:Chloroform Isoamyl alcohol (25:24:1) protocol (see section 3). Samples were maintained on ice for 15 min and DNA was fragmented using a sonicator, 10 min at 4 ºC. C-tails were added to 0.5 μg of fragmented DNA in a total of 20 μl reaction that contained deoxynuclotidyl transferase (TdT) enzyme (Promega), 475 μM deoxycytidine triphosphate (dCTP), 25 μM dideoxycytidine (ddCTP) and 4 μl of TdT 5x reaction buffer (Promega). The ddCTP limits the extension of the poly-C tails, working as terminator. The reaction was incubated 1 h at 37 ºC to obtain poly-C addition and it was heat-inactivated at 75 ºC for 20 min. 96-well plates from Exela Pure™, which contain an optimized ultrafiltration membrane,

were used to remove small molecules. MiSeq-Himar-2 and MiSeq-poliG-2 primers (Table 17) were used to perform a PCR to amplify fragments that contain the end of the transposon and the chromosome flanking region (gene were the transposon had been integrated). For this PCR, we used the KAPA polymerase as indicated in section 5. The PCR conditions used were: 95 ºC for 3 min; 28 cycles of 98 ºC for 30 s, 60 ºC for 30 s, 72 ºC for 2 min; and 72 ºC for 2 min. The amount of DNA (product from this PCR) was quantified using Qubit dsDNA HS Assay Kit. 25 ng of the obtained PCR product were used as template in a second PCR in order to selectively re-amplify the transposon-genomic fragments and to add additional adaptors for Illumina sequencing. This second PCR was performed with KAPA polymerase using primers MiSeq-Himar3 and MiSeq-N701 (Table 17). The PCR conditions used were exactly the same as in the first PCR but only 12 cycles were completed. Samples were ready to be cloned in TOPO® Cloning vector (Section 8).
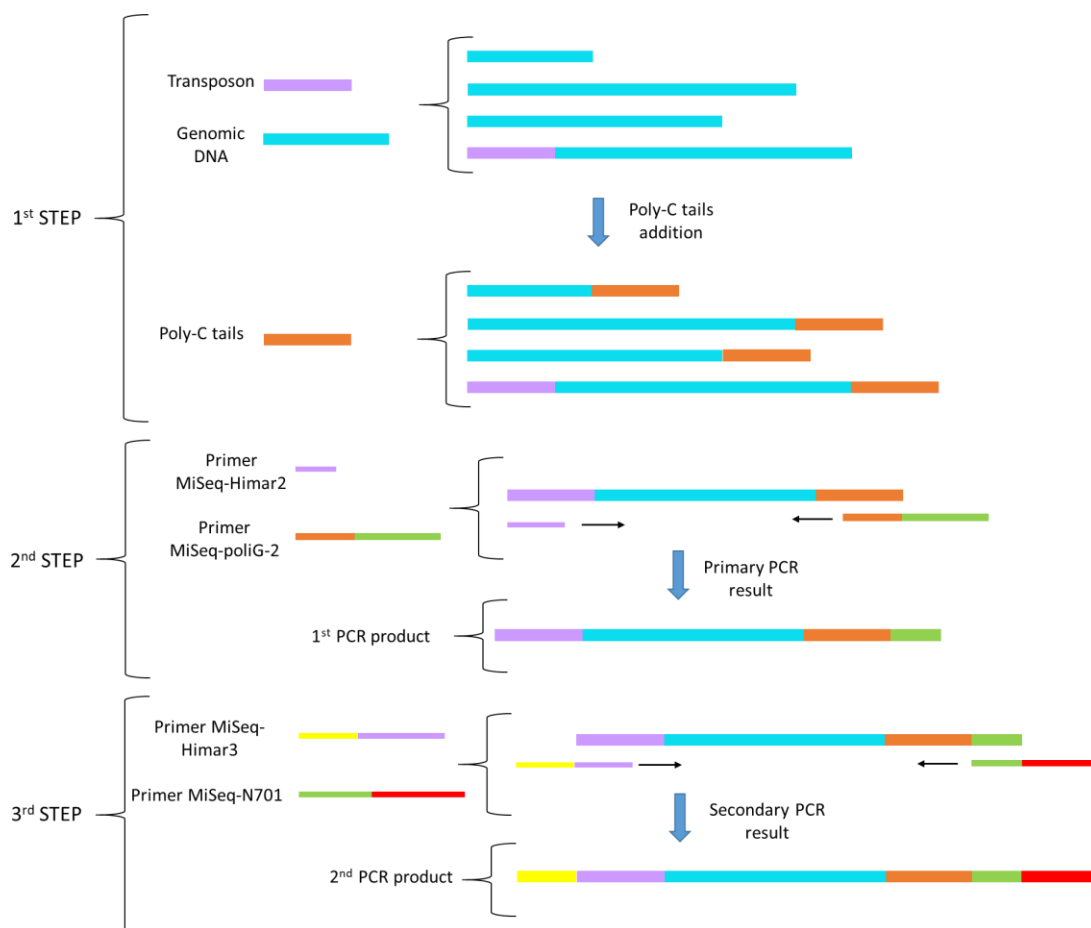


**Figure 16. Schematic representation of the 1st methodology used to prepare samples for sequencing.** First, DNA was fragmented and Poly-C tails were added. In a second step, PCR was performed to amplify fragments containing transposon sequences and to add an adapter to Poly-C tails. Finally, with a second PCR, adapters and barcodes for Illumina sequencing were added.

## 12.2 Tn-seq sample preparation, 2nd strategy, using pGPA plasmid for library construction

This protocol was performed in order to obtain PCR products from the mutant library for Illumina sequencing, which contain the part of the genome from *E. faecium* where the transposon was inserted. In the final part of the transposon sequence there is located a restriction site for MmeI enzyme that cuts between 16 and 18 nucleotides downstream/upstream. In this way, fragments are generated that contain the transposon and nucleotides of the flanking chromosomal region. Thus through high-throughput sequencing, one can identify the gene were the transposon was inserted. See Figure 17 and follow the detailed explanation of the protocol described below.

First, DNA from caecum, colon and 1ml aliquots of the library samples was extracted using phenol-chloroform isoamyl alcohol (25:24:1) protocol (see section 3).

Extracted DNA was cleaned up using magnetic beads (see section 9) and 250 ng were digested at 37 ºC for 4 h approximately using 1.25 U MmeI (New England Biolabs) with 1.2 µM S-adenosylmethionine. Next, 0.125 U of call intestine alkaline phosphatase (Invitrogen) was added to dephosphorylate the DNA and the total 25 µl of the mixture were incubated for 30 min at 50 ºC. After dephosphorylation, 175 µl of $H_2O$ were added to the previous volume. The DNA from the mixture was extracted with 200 µl chloroform-isoamyl alcohol (24:1). Upper layer was transferred into a new tube adding 20 µl of 3 M Sodium acetate (3 M pH 5.2 $NaC_2H_3O_2$), 0.25 µl of glycogen (20 mg/ml) and 500 µl of 100 % cold ethanol. Tubes were mixed and incubated at -20 ºC for 30 min, being subsequently centrifuged for 10 min at 13000 rpm and 4 ºC to precipitate the DNA. Supernatant was removed and pellet was washed again with 500 µl of 70 % ethanol. Samples were centrifuged for 5 min at 13000 rpm and 4 ºC. Supernatant was removed and the pellets were resuspended in 20 µl of ultrapure autoclaved $H_2O$.

Tn-seq adapters were prepared combining 5 nM of the primers that contained matching barcodes (PBGSF29 and PBGSF30 primers with different barcodes are in Table 18) in TE with 10 mM NaCl in 50 µl. Each matching barcode was added to a single sample, allowing to separate and distinguish, which sequences come from each sample. The matching barcoded combination was incubated for 10 min at 95 ºC to denature the DNA. A subsequent step of annealing the primers was performed by cooling the reaction slowly at RT. Annealed adapters were phosphorylated using T4 polynucleotide kinase (New England BioLabs) in T4 DNA ligase buffer (New England BioLabs) for 5 min at 37 ºC and immediately inactivated at 70 ºC for 10 min. 10 pmol/µl of phosphorylated adapters were ligated with 100 ng of dephosphorylated MmeI digested fragments with 2 U of T4 DNA ligase and 2 µl of 10x T4 DNA ligase buffer in a final volume of 20 µl. The mixture was incubated for 1 h at 16 ºC. 2.5 µl of this ligation reaction were used as template for a PCR that contained 20 pmol of primers P5 (Forward primer PBGSF31) and primer NextSeq (Reverse primer PBGSF40) to amplify fragments for Illumina sequencing. The PCR was performed with KAPA HiFi polymerase (see section 5) and conditions used in this case were: 72 ºC for 1 min and 98 ºC for 30 s; 30 cycles of 98 ºC for 30 s, 57 ºC for 30 s, and 72 ºC for 10 s; and 72 ºC for 5 min. The obtained PCR products were run on a

2.5 % agarose gel with a 100 bp ladder. PCR products of 130 bp were purified from the gel with the QIAquick Gel Extraction Kit (Qiagen) (see section 10). This 130 bp fragments contain (Figure 18): Primer P5 (PBGSF31), the adapter for Illumina sequencing and the specific barcode for each sample (6 bp), the 18 bp from *E.faecium* genome, the sequence of the transposon and primer Nextseq (PBGSF40). Samples were quantified using Qubit Fluorometric Quantification by Qubit dsDNA HS Assay Kit.

Before sequencing by Illumina, some of the samples were cloned in TOPO vector (see section 8) in order to check that the products obtained contained the different parts specified above. Primers from TOPO vector were used to sequence by Sanger the PCR product cloned in the vector to check all the parts present in the amplified fragments.

Once the PCR products were checked by Sanger, Tn-seq samples were prepared to be sequenced by Illumina NextSeq, using the 550 System High-Output Kit (1x75, single-end). Products were sequenced and the Sequencing core facility in Fisabio, Valencia.
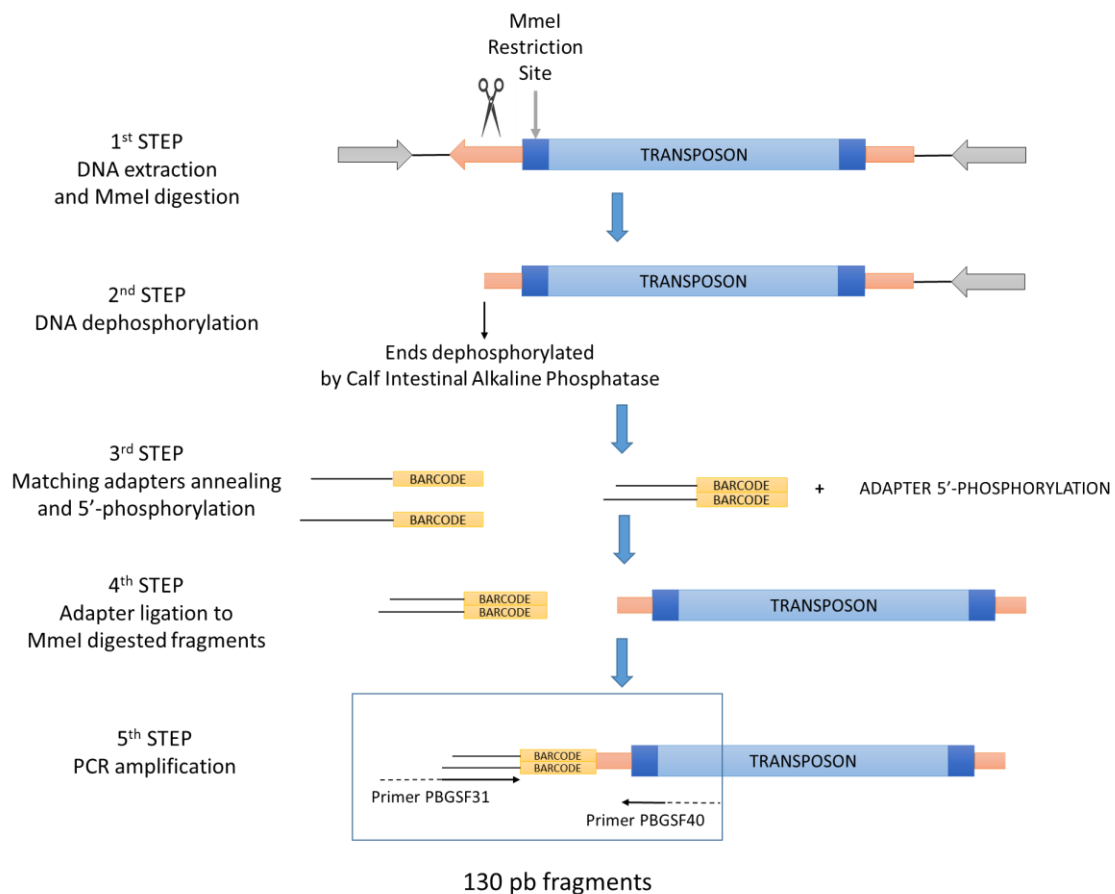


**Figure 17. Schematic representation of the Tn-seq methodology used to prepare PCR fragments for Illumina sequencing.** DNA extraction, digestion by MmeI and phosphorylation were performed in order to ligate fragmented DNA to annealed adapters. PCR amplification adding adapters for Illumina sequencing was performed to obtain 130 bp fragments. To simplify the methodology, only one scissors has been drawn to indicate the restriction site of MmeI enzyme at one end of the transposon. Nevertheless, another restriction site is present also in the other extreme of the transposon.

| 32 pb | 26 pb | 6 pb | 16 - 18 pb | 27 pb | 24 pb |
|---|---|---|---|---|---|
| Adapter P5, Illumina | Adapter | Barcode | *E. faecium* genome | Repeat region of the transposon | Adapter P7, Illumina NextSeq |
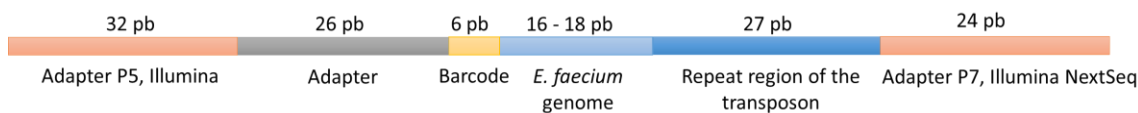
**Figure 18. Representation of the different parts of the amplified fragment of the Tn-Seq library.**

## 13. Tn-seq data analysis

### 13.1 Determination of the abundance of Tn-Seq mutants

To analyse samples obtained by Illumina platform, first of all sequences were split based on their barcodes using fastq-multx program, version 1.03, allowing one mismatch. Those unmatched sequences that did not contain a barcode were discarded. Subsequently, sequencing adapters were removed using cutadapt, version 1.9.1. We allowed a maximum of one mismatch, finding the adapter at the end of the reading with a minimum overlap of 15 bases (the complete adapter). Then, the sequences were filtered with PRINSEQ-lite 0.20.4. The base pairs from the end of the reading with quality lower than 20 were removed, those sequences whose length was below 16 base pairs were deleted and the reads with an average quality lower than 30 were deleted too.

Subsequently, the sequences with 16-17 nucleotides of average were compared first to the coding sequences (CDSs) and then to the entire genome of Aus0004 by BLAST. The alignments were performed with a minimum of 90 % of identity and coverage, allowing as maximum as 1 or 2 errors in the sequences. The reads whose best alignment was aligned to more than one gene were discarded. After the alignment, we obtained a table with the CDSs were the transposon has been integrated (here named as Tn-mutants, that are the genes that the transposon has interrupted). This table contained therefore the abundance of each transposon mutant in the library. Read counts per gene were then normalized to the total number of reads that mapped to the Aus0004 CDSs.

The sequences that did not align neither against CDSs nor to the entire genome were subsequently mapped against other databases of genomes: Mus musculus genome (assembly GRCm38.p6), Bacteria Refseq database (May 2019) and Fungi Refseq database (January 2018)[155]. Blastn (version 2.10.0+)[156] was applied to carry out these alignments, with the parameters set to maximum e-value 500 and 90 % of identity and coverage.

### 13.2 Statistical analysis of the Tn-Seq mutants and visualization of significant differences

In order to validate the reproducibility of Tn-seq data analysis, samples were compared by Pearson correlation in pairs of samples with a log-log plot of the normalized counts found for each gene. To analyse global differences in the relative abundance of mutants between groups of samples, we first calculated the bray-curtis distance between all pair of samples using the vegan library from R. Subsequently, a canonical correlation analysis (CCA), was performed using the generated distance matrix and the R package vegan to

study global differences in the Tn-mutants abundance among samples. On the other hand, we performed a Principal Coordinates of Analysis (PCoA) using the Bray-Curtis distance among pair of samples to separate samples based on their dissimilarity but without giving information about the sample origin. In addition, the PERMANOVA test was applied using the R package Adonis to evaluate if the detected differences in the CCA were statistically significant.

Relative abundances of genes from *in vivo* and *in vitro* conditions were compared by two different statistical tests: DeSeq2 and Wilcoxon tests. DeSeq2 test performs an internal normalization, calculates the mean for each gene in all samples and tests the differential gene presence based on the negative binomial distribution. Wilcoxon test is a non-parametric statistical test that compares the Tn-mutant mean ranks differences between samples, not assuming normal distribution in the data sample. To adjust for multiple hypothesis testing, we used the False Discovery rate (FDR) approach by Benjamini and Hochberg implemented in the fdr.R package[157]. The abundance of Tn-mutants was considered significantly different among groups of samples when the p-value was < 0.05 and the FDR q value was < 0.1.

To represent the differences between the abundance of the Tn-mutants, a heatmap was performed selecting those most significant mutants with FDR < 0.05. Tn-mutants abundances are represented as the difference of the relative abundance of each Tn-mutant with respect to the mean abundance of the corresponding Tn-mutant in all the samples. The Greens and Reds pallets in the RColorBrewer[158] library were selected to paint the abundances below and above the mean, respectively. The heatmap function in the R stats package[159] was used for plotting. Finally, from the significant genes, we selected those whose differences *in vivo* vs *in vitro* were more pronounce as the candidates to be involved in gut colonization.

The circular genome visualization for the transposon mutant insertion in the *in vitro* library samples was performed using the DNAplotter software[160]. This representation shows all the coding sequences (CDSs) from Aus0004 genome, displaying those CDSs with at least one mutant present for each specific gene. It is represented as the base 2 logarithm of the mean of the gene abundance in the library samples.

## 14.  Targeted mutagenesis

### 14.1   Construction of mutant strains of the candidate genes

In order to delete the candidate genes, a construction containing their flanking regions, including the beginning and the end of the candidate gene and a resistant marker within both flanking regions was obtained. The majority of the sequence of the candidate gene was replaced by the resistant marker through homologous recombination. Figure 19 represents the methodology procedure that is described in detail below.

The 450 bp of the upstream flanking region + 50 bp from the beginning of the gene and 450 bp of the downstream flanking region + 50 bp from the end of the gene were

amplified by KAPA PCR (see section 5) using primers from Table 17 that contain restriction sites for XhoI and XmaI, appropriate for cloning the PCR products into the plasmid pWS3. First, the upstream and downstream flanking regions were amplified separately. The primers that hybridize outside the gene contain the restriction enzymes for cloning. The primers that hybridize with the beginning of the gene contains and adaptor with a sequence that is complementary to the primer that hybridize with the end of the gene. In addition, it contains EcoRI or AvrII restriction site. In that way, the amplified products could be fused in a second PCR using the external primers. Therefore, by fusion PCR the two fragments were amplified in a single fragment, including an EcoRI or AvrII restriction site between both regions. These fused PCR products were cloned into the plasmid pWS3, using XmaI and Xho restriction sites and T4 ligase (see section 7). The fusion PCR products cloned in pWS3 were sequenced by Sanger to confirm that there were no mutations in the amplified fragments.

Subsequently, the gentamicin resistance gene *acc(6')-aph(2")* was amplified from pZXL5 plasmid adding EcoRI or AvrII (Primers pAT392 F/R EcoRI or AvrII, Table 17) to both extremes, using KAPA polymerase (see section 5) and an hybridization Tm of 60 ºC. The obtained PCR product was run and purified from agarose gel (see section 10) and digested with EcoRI or AvrII (see section 6). The pWS3 plasmid with the fusion PCR cloned was digested with the corresponding enzyme, EcoRI or AvrII. Then, digested gentamicin resistance gene was cloned by ligation (see section 7) in the digested pWS3 plasmid and transformed in *E. coli* by heat shock (see section 2.2). In this case we used the *E. coli* strain Ec1000, since it encodes a replicon that allows the replication of pWS3 plasmid in *E. coli*.
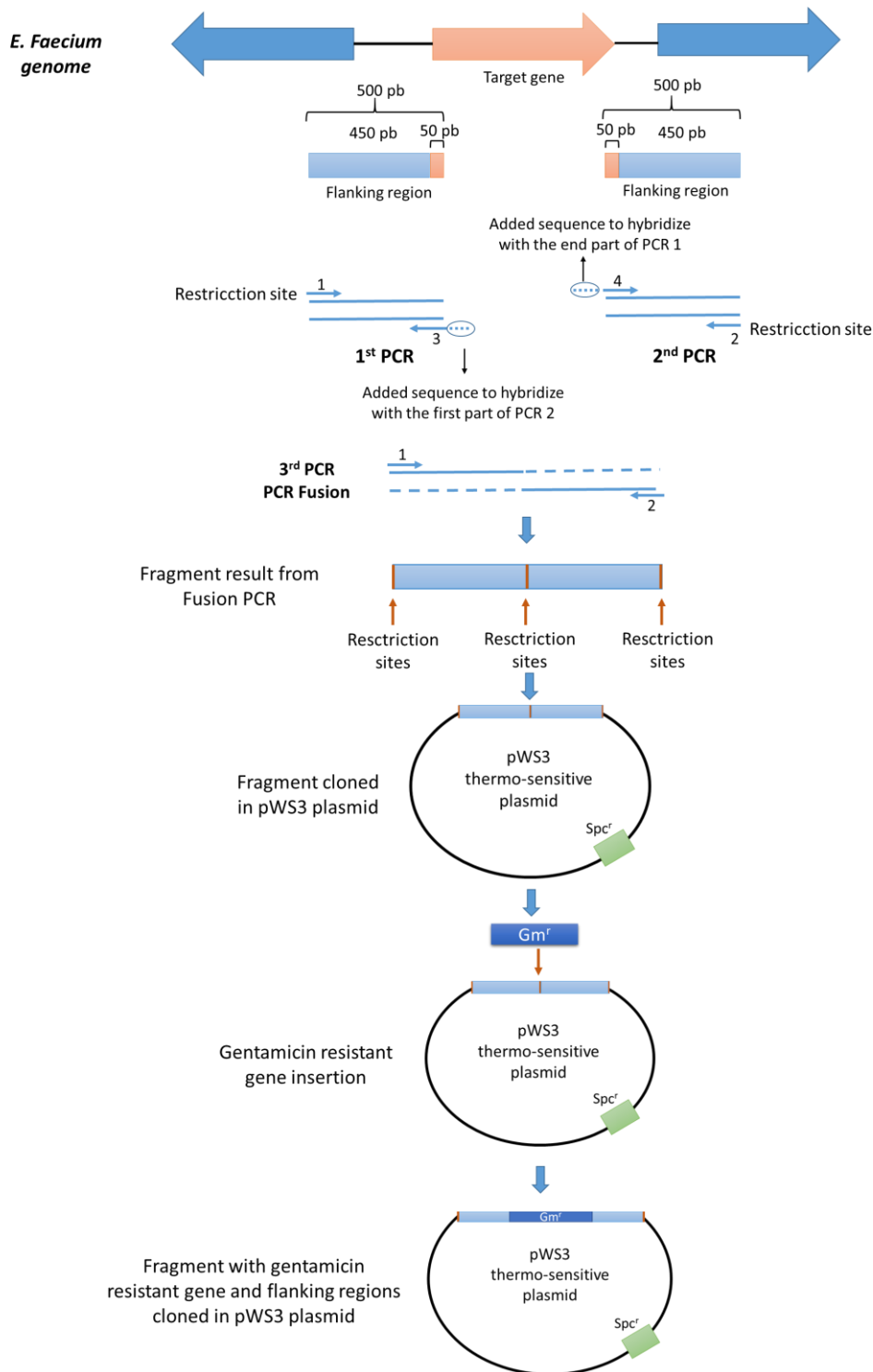
**Figure 19. Schematic representation of fusion PCR to obtain mutants by deleting a gene and inserting a gentamicin resistance marker.** Orange arrow represents the candidate gene to be deleted from VRE genome and the blue arrows the genes upstream and downstream. Blue and orange boxes below arrows represent flanking regions and the start/end of the gene respectively. In the 1st PCR and 2nd PCR, primers contain restriction sites and sequences that hybridize with the other flanking region to allow fusion of both PCRs. 3rd PCR represents the amplification of both flanking regions by fusion PCR. Blue arrows indicate the position where primers hybridize. Fusion PCR is cloned into pWS3 plasmid by restriction sites and then the gentamicin resistance gene, coloured in dark blue, is introduced by the restriction site present in the previous cloned fusion PCR. Finally, pWS3 plasmid contains the entire construction, the flanking regions with the gentamicin resistance marker between them.

The pWS3 plasmid with the cloned construction was extracted from *E. coli* (see section 4) and transformed by electroporation into *E. faecium* Aus0004 and plated in BHI plates containing 300 µl gentamicin. Subsequently, the protocol to achieve double-crossover recombination to obtain the targeted mutants (Figure 20) was followed. One transformant colony was grown overnight in BHI broth supplemented with gentamicin at 28 ºC. Next day, the cell culture was diluted 10000 fold in pre-warmed BHI broth without antibiotics and incubated overnight at 37 ºC to lose the thermosensitive plasmid and select transformants with the gene conferring gentamycin resistance inserted in the chromosome by recombination. The culture was plated on BHI agar supplemented with gentamicin and incubated again at 37 ºC (the restrictive temperature for plasmid replication). Colonies were restreaked on BHI agar plates supplemented with spectinomycin and on BHI agar plates containing gentamicin. The colonies that were resistant to gentamicin and susceptible to spectinomycin (conferred by the plasmid) were checked by PCR to confirm the generated deletions and introduction of the gentamicin resistance gene.
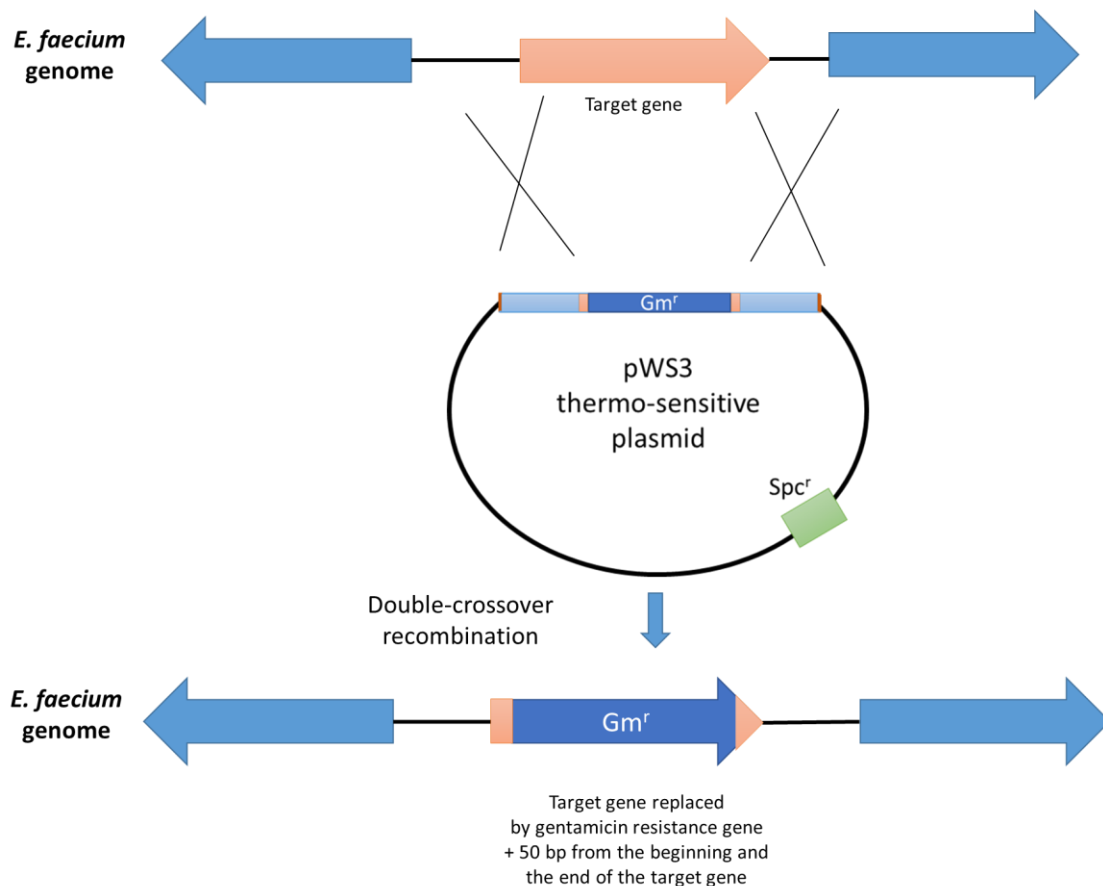


**Figure 20. Schematic representation of the methodology of targeted mutagenesis.** Orange arrow represents the gentamicin gene and blue arrows the genes upstream and downstream. pWS3 plasmid contains flanking regions (coloured in light blue) and gentamicin resistance gene (coloured in dark blue). In double-crossover mutants the candidate gene is replaced by the gentamicin marker. The resulting mutants are resistant to gentamicin and sensitive to spectinomycin after plasmid loss. (Gm[r]: gene conferring gentamicin resistance; Spc[r]: gene conferring spectinomycin resistance).

Next, in Table 7, we indicate the candidate genes of *E. faecium* Aus0004 strain replaced by gentamicin resistance gene and the name assigned for each of them throughout this work.

**Table 7. Candidate genes selected that were replaced by gentamicin resistance gene in Aus0004 *E. faecium* strain.**

| LocusTaq[a] | Annotation[b] | Gene Name[c] |
|---|---|---|
| EFAU004_RS03690 | PTS mannose transporter subunit EIIAB | *man*X |
| EFAU004_RS02800 | Fe-S cluster assembly protein SufB | *suf*B |
| EFAU004_RS07995 | Vat family streptogramin A O-acetyltransferase | *vat* |
| EFAU004_RS10170 | phosphate ABC transporter ATP-binding protein | *pst*B2 |
| EFAU004_RS12135 | hypothetical protein / Two-component system regulatory protein YycI | *yyc*I |
| EFAU004_RS14195 | LacI family DNA-binding transcriptional regulator | *ltr*R |
| EFAU004_RS03130 | N-acetylmuramic acid 6-phosphate etherase | *mur*Q |
| EFAU004_RS12420 | DsbA family protein | *dsb*A |
| EFAU004_RS01175 | bifunctional acetaldehyde-CoA/alcohol dehydrogenase | *adh*E |
| EFAU004_RS01355 | glycosyl transferase | *pgt* |
| EFAU004_RS02785 | Fe-S cluster assembly protein SufD | *suf*D |
| EFAU004_RS02790 | cysteine desulfurase | *csd* [Fe-S] |
| EFAU004_RS02795 | SUF system NifU family Fe-S cluster assembly protein | *nif*U cluster |
| EFAU004_RS02800 | Fe-S cluster assembly protein SufB | *suf*B |

[a] Indicates the gene ID number in the Aus0004 genome. [b] Indicates the function predicted. [c] Name assigned to each gene or group of genes deleted from Aus0004 genome base on their description of the genome deposited in NCBI or by similarity with other genes described with that function.

## 14.2 Construction of control strain for competition experiments by introducing the gentamicin resistance gene in an intergenic region of the genome

The manuscript of Van Schaik et al.[60], describes a large pathogenicity island (PAI) flanked by repetitive regions in *E. faecium*, which carries the *esp* gene. We decided to introduce the gentamicin resistance gene, *acc(6')-aph(2"*, in an intergenic region, upstream the PAI, expecting not to affect *E. faecium* growth. We reason that this was a good point for inserting the gene since the insertion of a large DNA fragment in that spot of the chromosome (the PAI) does not seem to affect the growth of *E. faecium*. The repetitive regions of the PAI were found in Aus0004 strain and we inserted the gentamicin resistance gene in a specific point of the *E. faecium* Aus0004 (nucleotide 2.762.777), which is located 20 bp upstream of the PAI. The methodology was exactly the same as used for targeted mutagenesis of the candidate genes (Figure 20), but instead of replacing a gene by the gentamicin resistance marker, we inserted the gentamicin resistance gene in an intergenic region, without deleting any gene.

The 500 bp flanking regions of the specific selected point of the genome were amplified performing 2 PCRs (see Figure 19) with primers Aus0004-1-R and Aus0004-3-F for one flanking region and Aus0004-2-R and Aus0004-4-F for the other flanking region (section 33, Table 17). Both flanking regions were amplified in a single fragment by fusion PCR

containing XmaI and Xho as a restriction sites to clone the fragment in the thermo-sensitive plasmid pWS3. The fusion PCR also contained, as described above, a EcoRI site between both flanking regions. Gentamicin resistance marker was amplified (Primers pAT392 F/R EcoR, section 33, Table 17) adding EcoRI sites to the extremes to be cloned between the two flanking regions.

The pWS3 plasmid with the construction obtained from *E. coli* was electropored into *E. faecium* Aus0004 and plated in BHI agar plates containing gentamicin. To select the double-crossover recombination mutant with the gentamicin resistance gene inserted, we followed the protocol explained in the section described above.

## 15. Growth curves determination

Growth curves were measured in aerobiosis in order to check and compare if the wild type (WT) and the mutant strains differed in their growth *in vitro.*

Spectrophotometer plate reader (Tecan Infinite F200) was used to measure the growth of the WT and each mutant strain. Each strain was previously grown overnight at 37 ºC in BHI plates. The grown colonies were recollected with a swab and resuspended in BHI broth to obtain an initial $OD_{600}$ of 0.1. Triplicates of 200 µl of each resuspended strain were added to a 96-well plate.  The plate was incubated overnight in the Tecan Infinite F200 instrument at 37 ºC. Before reading the absorbance at 570 nm, plate was orbital and linearly shaken (at 32 and 65 rpm correspondingly) for 5 s. Data was read using Magellan program.

## 16. Competition assays

With the purpose to check if the deleted gene affects the growth of the mutant when the WT strain is present in the same proportion, competition experiments were performed mixing WT with each individual mutant to measure the growth of both strains by plating them in specific media to differentiate them. The agar media contained gentamicin, so that we could count the colonies derived from the mutant strain. In addition, cultures were also plated in media not containing gentamicin, which allow us to count the colonies derived from the WT and the mutant strain. WT colonies were obtained by subtracting the number of colonies obtained in the media without gentamicin from the colonies obtained in the media with gentamicin. Note that a pure culture of a mutant strain led to similar number of colonies in agar plates containing or not containing gentamicin.

### 16.1    Competition assays *in vitro*

WT and mutant strains were grown in BHI plates overnight at 37 ºC. The bacteria were collected with a swab to be resuspended in BHI broth and adjust to have $10^4$ CFUs/µl (proportion 1:1*)* in a total of 10 ml of BHI broth. Culture tubes were incubated under both aerobic and anaerobic conditions for 24 h. Anaerobic incubation was performed in

anaerobic chamber (Whitlet DG250), with a gas atmosphere of 10 % $H_2$, 10 % $CO_2$, and 80 % $N_2$. After the incubation period, cultures were plated in selective media to differentiate both WT and mutants. On the one hand, BEA plates with ampicillin (BEA+A) were used to grow WT and mutants and, on the other hand, BEA plates with ampicillin and gentamicin (BEA+AG) were used to grow only mutants, which are resistant to gentamicin. After counting colonies from both types of plates, mutant colonies from BEA+AG were subtracted from the total amount of colonies from BEA+A plates, calculating the number of WT CFUs. CFUs from plates were count after two days of incubation at 37 ºC and the competitive index (CI) was calculated using the following formula: CI = nºCFUs WT / nºCFUs mut

## 16.2    Competition assays *ex vivo*

Colon samples from the murine intestine were collected after three days of clindamycin treatment (see section 28.3). Each sample was resuspended in the proportion of 100 mg colon sample in 1 ml of PBS pH7, centrifuged 2 min at 13000 rpm and the supernatant was collected in a new tube. 40 µl of the supernatant containing 1000 CFUs of the WT strain and 1000 CFUs of the mutant strain was incubated at 37 ºC and during 24 h in the anaerobic chamber. After the incubation period, samples were plated in selective media to differentiate both strains: BEA+A and BEA+AG. Competition Index (CI) was calculated as explained for previous competitions assays. Each competition assay was performed in triplicate.

## 17.  Characterization of chelator-sensitive growth

To ensure that the [Fe-S] cluster was necessary to acquire iron and grow under limited concentrations of iron, the WT and the mutant strain for these cluster were grown in iron limiting conditions.

Iron-limiting experiments were performed using BHI medium. The strains were previously grown overnight at 37 ºC in BHI plates. The grown bacteria were collected with a swab and resuspended in BHI broth to setup an initial $OD_{600}$ of 0.1, adding to the medium different concentrations of chelators. The three chelators used in the experiments were: dipyridyl (chelates the $Fe^{2+}$ ion form), ciclopirox olamine (chelated the $Fe^{3+}$ ion form) and phenantroline (chelates both the $Fe^{2+}$ and $Fe^{3+}$ ion forms).

First of all, the Maximal tolerable concentration (MTC) was checked on the WT strain by adding increasing concentration of each chelator. MTC was selected as the highestconcentration in which the chelating agent had no effect on WT growth. The growth curve of the WT adding the maximal concentration selected of each chelator was similar to the growth curve of the WT in BHI medium.

Once the MTC was determined, the experiment was prepared with cultures from the WT and the iron-sulfur [Fe-S] cluster mutant strain, adding the corresponding concentration of the three chelators to be tested. Subsequently, the cultures were

incubated under aerobic conditions during 24h at 37 ºC and the $OD_{600}$ was monitored as previously described in section 15.

## 18. Growth characterization under phosphate starvation conditions

In order to check if the phosphate ABC-transporter ATP binding protein was involved in the transport of phosphate under phosphate limiting conditions, we tried to set up different minimal medium to grow both strains, WT and mutant, under starvation conditions for phosphate ion, adding increasing concentrations to check growth differences between both strains.

Two different chemically defined media (CDM) that do not contain phosphate were used and potassium phosphate ($K_3PO_4$) was added in increasing concentrations to test the effect on the WT and the mutant strain growth.

The first medium used was a concentrated chemically defined medium (2x CDM), based on *Streptococcus cremoris* Defined Medium adjusted to pH7 and filtered (Otto et al.[161]) (see section 31, Table 11). To obtain the phosphate starvation medium, the phosphate compounds added to buffer the medium were replaced by MOPS (3-(*N*-morpholino) propanesulfonic acid) and adjusted to pH7. To the phosphate starvation medium buffered with MOPS, increasing concentrations of potassium phosphate stock, $K_3PO_4$ (pH7), ranging from 0.1 mM to 200 mM, were added to the medium.

The second medium tested was a phosphate starvation medium used in *Bifidobacterium breve UCC2003*[152]) (see section 31, Table 12). To remove phosphate from the medium the beta-blycerophosphate and the phosphate compounds to buffer the medium were replaced by MOPS. Both media, complete and phosphate starvation media were adjusted to pH7 and filtered. Then, increasing concentrations of $K_3PO_4$ (pH7), ranging from 0.1 mM to 20 mM, were added to the phosphate starvation medium.

In order to remove nutrients and phosphate traces present in BHI medium, the strains were grown in BHI plates overnight at 37 ºC, collected with a swab and resuspended in the chemically defined media without phosphate. From this medium the cultures were adjusted to $OD_{600}$ 0.1 and $K_3PO_4$ (pH7) was added to reach different concentrations. Growth experiments were carried out in 200 µl triplicates using spectrophotometer plate reader (Thermo-Scientific), measuring the growth at 37 ºC, every 30 min at wavelength of 600 nm.

## 19. Characterization of the carbon sources used by VRE for growing

To assess the capacity of VRE to grow in the presence of different carbon sources, we used Biolog plates with 95 carbon sources. The 96 well-plate contains a unique carbon source in each well to grow the bacterium in anaerobic conditions. The Biolog plates used for the analysis were the AN MicroPlate$^{TM}$ that contains the carbon sources presented in Figure 21.

| A1<br>Water | A2<br>N-acetyl-D-Galactosamine | A3<br>N-Acetyl-D-Glucosamine | A4<br>N-Acetyl-β-D-Mannosamine | A5<br>Adonitol | A6<br>Amygalin | A7<br>D-Arabitol | A8<br>Arbutin | A9<br>D-cellobiose | A10<br>α-Clyclodextrin | A11<br>β-Clyclodextrin | A12<br>Dextrin |
|---|---|---|---|---|---|---|---|---|---|---|---|
| B1<br>Dulcitol | B2<br>i-Erythritol | B3<br>D-Fructose | B4<br>L-Fucose | B5<br>D-Galactose | B6<br>D-Galacturonic Acid | B7<br>Gentibiose | B8<br>D-Gluconic Acid | B9<br>D-Glucosminic Acid | B10<br>α-D-Glucose | B11<br>Glucose-1-Phosphate | B12<br>Glucose-6-Phosphate |
| C1<br>Glycerol | C2<br>D,L-α-Glycerol Phosphate | C3<br>m-Inositol | C4<br>α-D-Lactose | C5<br>Lactulose | C6<br>Maltose | C7<br>Maltotriose | C8<br>D-Mannitol | C9<br>D-Mannose | C10<br>D-Melezitose | C11<br>D-Melibiose | C12<br>3-Methyl-D-Glucose |
| D1<br>α-Methyl-D-Galactoside | D2<br>β-Methyl-D-Galactoside | D3<br>α-Methyl-D-Glucoside | D4<br>β-Methyl-D-Glucoside | D5<br>Palatinose | D6<br>D-Raffinose | D7<br>L-Rhamnose | D8<br>Salicin | D9<br>D-Sorbitol | D10<br>Stachyose | D11<br>Sucrose | D12<br>D-Trehalose |
| E1<br>Turanose | E2<br>Acetic Acid | E3<br>Fromic Acid | E4<br>Fumaric Acid | E5<br>Glyoxylic Acid | E6<br>α-Hydroxybutyric Acid | E7<br>β-Hydroxybutyric Acid | E8<br>Itaconic Acid | E9<br>α-Ketobutyric Acid | E10<br>α-Ketovaleric Acid | E11<br>D,L-Lactic Acid | E12<br>L-Lactic Acid |
| F1<br>D-Lactic Acid Methyl Ester | F2<br>D-Malic Acid | F3<br>L-Malic Acid | F4<br>Propionic Acid | F5<br>Pyruvic Acid | F6<br>Pyruvic Acid Methyl Ester | F7<br>D-Saccharic Acid | F8<br>Succinamic Acid | F9<br>Succinic Acid | F10<br>Succinic Acid Mono-Methyl Ester | F11<br>m-Tartaric Acid | F12<br>Urocanic Acid |
| G1<br>L-Alaninamide | G2<br>L-alanine | G3<br>L-Alanyl-L-Glutamine | G4<br>L-Alanyl-L Histidine | G5<br>L-Alanyl-L-Threonine | G6<br>L-Asparagine | G7<br>L-Glutamic Acid | G8<br>L-Glutamine | G9<br>Glycyl-L-Aspartic Acid | G10<br>Glycyl-L-Glutamine | G11<br>Glycyl-L Methionine | G12<br>Glycyl-L-Proline |
| H1<br>L-Methionine | H2<br>L-Phenylalanine | H3<br>L-Serine | H4<br>L-Threonine | H5<br>L-Valine | H6<br>L-Valine plus L-Aspartic Acid | H7<br>2'-Deoxy Adenosine | H8<br>Inosine | H9<br>Tymidine | H10<br>Uridine | H11<br>Thymidine-5'-Monophosphate | H12<br>Uridine-5'-Monophosphate |

Figure 21. 95 carbon sources included in AN MicroPlate<sup>TM</sup> from Biolog Systems.

Each 96 well-plate was inoculated with 100 µl of the M1 medium without carbohydrates, based on the manuscript of Zhang et al.[162], in which they utilized this medium to identify carbon sources required for the growth of *Enterococcus*. M1 medium contains 10 g tryptone and 0.5 g yeast extract in 1 L of PBS (section 31, Table 14). Briefly, the bacterium was grown overnight in a BHI plate at 37 ºC in anaerobic conditions as described above. The M1 medium used in the experiment was previously reduced overnight in the anaerobic chamber. The bacterium was collected with a swab, resuspended in reduced M1 medium and adjusted to $OD_{600}$ 0.1. The AN MicroPlate$^{TM}$ was inoculated with 100 µl in each well. Since the AN Microplate had to be incubated for 24 h outside the anaerobic chamber in order to continuously measure the absorbance, in order to maintain the anaerobic conditions, 30 µl of autoclaved mineral oil was added on the top of the 96 well-plate inside the anaerobic chamber.

The lecture was performed in the Tecan Infinite F200 spectrophotometer, at 570 nm. The plate was incubated at 37 ºC, overnight. The absorbance was measured every 20 minutes with previous shaking.

The results were analysed subtracting the background of the first lecture. The experiment was performed once or twice, depending on the strain analysed.

## 20. Confirmation of VRE growth in the presence of individual sugars

To check the growth of the WT and mutant strains in the presence of selected sugars, both strains were grown overnight in BHI plates and resuspended in M1 medium (section 31, Table 14). Triplicates of 200 µl of each strain were distributed in a 96-well plate and 5 mM of the corresponding sugars were added to each well. The $OD_{600}$ was measured overnight in the spectrophotometer at 37 ºC in anaerobiosis, as it has been explained before, adding 30 µl of autoclaved mineral oil. The growth curves were represented subtracting the background of the M1 medium to each lecture point.

## 21. Gene complementation of VRE mutants

To complement the strains containing deletions in specific genes and check their role in gut colonization, the pNZ44 plasmid was used to clone the genes of interest and to express them. This plasmid carries a strong broad range constitutive promoter from *Lactococcus lactis* and a chloramphenicol resistance gene. *manX* and *murQ* genes were amplified from VRE genome using primers (section 33, Table 17) containing restriction sites to clone the PCR product into the pNZ44 plasmid. Melting temperature (Tm) used was 60 ºC and the rest of the parameters of the PCR are indicated in Table 4. PCR products and pNZ44 plasmid were digested 4 h at 37 ºC (see section 6). Restriction enzymes used are indicated in the primers used in Table 17. Then, both fragments were ligated at 16 ºC overnight (see section 7). Transformations were performed by electroporation into *Lactococcus lactis* competent cells (see section 2.1). Transformants were selected by plating in GM17 plates with chloramphenicol (5 µg/ml). Plasmid

extraction was performed (see section 4) and genes cloned were verified by sequencing prior to its transfer to VRE by electroporation.

To check if the overexpression of the previous deleted genes restores the function *in vitro*, the strains were grown with the corresponding carbon source, as described in the previous section, and the growth curves were measured by triplicates in the spectrophotometer at 37 ºC, overnight. On the other hand, to check if the expression of the gene cloned in pZN44 restores the function in the mutant strain under *in vivo* conditions, a competition experiment with both mouse models (section 28.3) was performed. Mice were orally inoculated by gavage with a mix of the WT and each complemented strain (1:1) and stool samples were collected and plated in selective media.

## 22. Comparative genomic analysis and phylogenetic tree construction

In order to compare the sequences of the genes involved in gut colonization among different *E. faecium* strains and study if they follow a similar distribution as the clades performed with the complete genomes, first each individual gene was compared by blastn with the selected 73 *E. faecium* strains[57] from the different clades A1, A2 and B. The alignment was performed with the nucleotide sequence of each gene using BLAST with 97 % of coverage. Those similar sequences (genes) found in these 73 strains for each gene were used to analyse the molecular evolution of each gene constructing a phylogenetic tree by MEGA X version 10.1. The neighbor-joining tree was constructed running 500 bootstraps.

## 23. RNA extraction protocol and VRE *in vivo* transcriptome analysis

To analyse the genes highly expressed by VRE *in vivo,* we performed a transcriptomic analysis comparing *in vivo* an *in vitro* conditions. Mice treated with clindamycin were infected with $10^8$ CFUs of VRE and samples from colon (N=3 samples) were collected at day 3 post-VRE inoculation (Section 28.3). On the other hand, VRE was grown in BHI medium during 24 h. Subsequently, 1 ml of the culture was centrifuged at 13000 rpm during 5 min and the supernatant was discarded (N=3 replicates). Colon samples (100 mg) and the VRE pellet obtained from the BHI medium were conserved in RNA later for 24 h at 4 ºC and then frozen and conserved at -80 ºC.

For RNA extraction, colon samples conserved in RNA later were thaw and approximately 100 mg were used for the RNA extraction. This 100 mg of colon sample was transferred to a new 2 ml tube and diluted with the same volume of cold DEPC (diethylpyrocarbonate) water to decrease the viscosity. In the case of the pellet of VRE grown *in vitro,* the RNA later was removed adding the same volume of DEPC water. Samples were centrifuged at 13400 rpm for 5 min at 4 ºC to remove RNA later.

To extract the RNA, we used the RNeasy PowerMicrobiome Kit (Qiagen) according to manufacturer protocol, with some modifications. In the first steps, to increase the

amount of RNA recovered, we added phenol:chloroform:isoamyl alcohol 25:24:1 (pH5) and beads to disrupt bacterial cells. Tubes containing phenol:chloroform:isoamyl alcohol and beads were vortex 10 min at 10000 rpm to lysate bacterial cells. Supernatant was transferred into spin columns to retain RNA. To completely eliminate the DNA from samples, the DNAse treatment step was increased from 15 min to 45 min. Finally, to increase the concentration, the RNA was eluted with 50 µl of PM8 buffer after 5 min of incubation instead of 100 µl and 1 min of incubation.

The most abundant RNA present in bacteria is ribosomal RNA (rRNA) but we were mainly interested in the messenger RNA (mRNA) to identify the genes expressed by VRE *in vivo* and *in vitro*. rRNA was removed before preparing samples for sequencing with Ribo-Zero RNA Removal Kit (Bacteria) (Illumina). The library for sequencing was constructed using the ScriptSeq Complete Kit Bacteria (Illumina) according to the manufacturer protocol. Libraries were sequenced using the NextSeq 500/550 High Output Kit v2.5 (150bp, pair-end) on a Nextseq platform according to the manufacturer protocol.

The paired-end fastq files were filtered by quality with the Prinseq[163] application (v0.20.4). Low quality ends were trimmed from each read, with a minimum quality of 20 per base. The reads shorter than 50 bases, with entropy less than 60, with more than 1 percentage of ambiguous N bases and with average quality lower than 30 were also discarded. Next, the FLASH[164] software (v1.2.11) was applied to join overlapping paired-ends, with a minimum length of 10 bases and a maximum of 10 % of mismatched bases in the overlapping region. Then, all the sequences passing the previous filters were mapped onto the *Enterococcus faecium* Aus0004 genome (NC_017022.1, NC_017023.1, NC_017024.1, NC_017032.1) by means of Bowtie2[165] (v2.3.4.1). Samtools[166] (v1.6) was used to convert the SAM alignments file into the corresponding compressed BAM file. Finally, the abundance of each gene was calculated with a custom R[159] (v3.4.0) script, counting the number of alignments inside the boundaries of the corresponding gene[167].

For one of the three *in vivo* samples the library was not of good quality so it was not included in the statistical analyses. The abundances of genes from *in vivo* (N=2) and *in vitro* (N=3) conditions were compared applying PERMANOVA and DeSeq2 test in order to find global differences among groups and genes whose expression significantly differ *in vivo* vs *in vitro*.

## 24. Transcriptome analysis of the VRE mutant for the LacI family DNA-binding transcriptional regulator

In order to analyse those genes affected by the deletion of the LacI family DNA-binding transcriptional regulator, we performed transcriptomic analysis comparing caecal samples collected from mice inoculated with the WT or the mutant strain separately. We used the clindamycin mouse model (Section 28.3) to inoculate $10^8$ VRE CFUs and three days after, the caecal contents were collected and maintained at 4 ºC during 24 h. Subsequently, the samples were frozen at -80 ºC and preserve until the RNA was extracted. We choose to analyse in this case caecal samples because we noticed that

the amount of material was relevant for extracting sufficient RNA por transcriptomic analysis and we could obtain more material from caecal than colon samples.

RNA extraction was performed as explained in previous section using the RNeasy PowerMicrobiome Kit.

The library for sequencing was constructed using TrueSeq Stranded mRNA (Illumina) according to the manufacturer protocol. Libraries were sequenced on a Nextseq platform (75bp, pair-end).

The fastq files were filtered by quality with the Prinseq[163] application (v0.20.4). Low quality ends were trimmed from each read, with a minimum quality of 20 per base. The reads shorter than 50 bases, with entropy less than 60, with more than 1 percentage of ambiguous N bases and with average quality lower than 30 were also discarded. Then, all the sequences passing the previous filters were mapped onto the *Enterococcus faecium* Aus0004 genome (NC_017022.1, NC_017023.1, NC_017024.1, NC_017032.1) by means of Bowtie2[165] (v2.3.5.1), discarding those that aligned against rRNA or tRNA genes, selecting only the coding sequences (CDSs). Samtools[166] (v1.10) was used to convert the SAM alignments file into the corresponding compressed BAM file. Finally, the abundance of each gene was calculated with a custom R[159] (v3.6.0) script, counting the number of alignments inside the boundaries of the corresponding gene[167].

Similar statistical analysis were performed as those indicated in the previous section in order to identify differences in gene expression in the WT and the mutant strain.

## 25. Transcriptomic analysis to identify a mechanism by which ProBac confers protection against VRE.

With the objective to determine mechanisms by which ProBac was conferring protection against VRE intestinal colonization, metatranscriptomic studies of murine cecal contents were performed in order to identify expressed bacterial functions that could be relevant for conferring protection against VRE. Metatranscriptomic analysis were performed on caecum samples collected after the administration of the ProBac bacteria, on a group of mice treated with vancomycin that did not received ProBac and on untreated treated. The samples collected from the same *in vivo* experiment were used both for transcriptomic and metabolomics analysis (section 25-26). In this particular experiment, the extraction and sequencing of the RNA was performed by a previous PhD student. However, the analyses of the RNA sequences performed in this thesis are novel and were not included in the previous PhD. We explain briefly below how the samples were collected and processed and how the sequences were obtained and analysed.

In order to collect the cecum samples, co-housed mice were treated during one week with vancomycin. Subsequently, the treatment was replaced by water and mice were co-housed in groups of two per cage during two weeks. Once the antibiotic was removed, mice were orally inoculated with the bacterial suspension containing ProBac or the vehicle to resuspend the bacteria (PBS with glycerol 20 % and cysteine 0.1 %).

After the recovery period, mice were separated individually and one of the mice from the same cage was inoculated with $10^6$ CFUs/200 µl of VRE to follow the colonization process whereas the other one was sacrificed to collect samples for metabolomics and transcriptomic studies. Note that we have previously shown that mice co-house in the same cage recover the same microbiota and are equally susceptible to VRE intestinal colonization[168].

The collected caecum was introduced into a microcentrifuge tube and was weighted. Subsequently, two volumes of 100 mM $Na_2HPO_4$ pH 7.4 buffer, referred to the caecum weight, were added. Samples were homogenised and centrifuged at 13200 rpm for 1 min, separating the supernatant for metabolomics studies and the pellet for transcriptomic analysis. 1 ml of RNA later was added to the pellet and mixed. These samples were maintained at 4 ºC for 24 h before being stored at -80 ºC for RNA extraction (see section 23).

As previously explained in section 23, rRNA was removed using Ribo-Zero rRNA Removal Kit (Bacteria) (Illumina) as we were interested in the mRNA to identify the functions expressed by commensal bacteria. The library for sequencing was constructed using the ScriptSeq Complete Kit Bacteria, following the instructions given by the manufacturer. Libraries were sequenced using the NextSeq 500/550 High Output Kit v2.5 (150bp, pair-end) on a Nextseq platform according to the manufacturer protocol.

Adaptor sequences and low-quality reads were removed using Cutadapt v. 1.10 and UrQt v.1.0.18. Although the RNA extraction method used contained a step to remove the ribosomal RNA, this step does not remove the 100 % of the ribosomal RNA. For this reason, metatranscriptomic data was mapped against a Short Ribosomal Subunit database (including 16S and 18S rRNA reference data) from SILVA[169], using bowtie v.2.2.9[165], and discarded. The remaining reads were re-mapped against the Long Ribosomal Subunit database from SILVA and the Mouse reference genome v.38 from the NCBI reference repository[170]. Hits mapping any of both databases were discarded in further analyses.

The remaining transcriptomic sequences were assembled using SPAdes v. 3.7.1 using the "rna" algorithm to improve the transcript reconstruction. A multi-kmer de Bruijn graph reconstruction was followed as it has been suggested to improve the assembly outcome. We used 6 different kmer lengths as described above. ORFs were identified and annotated using MetaGeneMark v.1.0.1[171]. Only ORFs with a minimum size of 50 aa and with at least one extreme complete were kept.

Subsequently, a catalogue of non-redundant ORFs (NR) was constructed. ORFs were clustered at 90 % identity, as described in Mirdita et al.[172], using vsearch and the "cluster smallmem" algorithm[173]. Non-redundant ORFs were annotated using KEGG and HMMer as previously described. To calculate the average coverage and the total number of mapped reads per ORF, the cleaned metatranscriptomic data was mapped against the formatted NR catalogue using bowtie2. Read counts were converted to transcripts per million (TPM), by first dividing the number of read counts by the length of each ORF

(RPK). Then we divided the sum of RPK numbers by 1000000 (thus resulting in the scaling factor) and finally dividing each RPK by such scaling factor. This normalization results in each sample summing up to 1000000 TPMs (transcripts per million), making them comparable.

In order to identify which transcripts were expressed by ProBac isolates, cleaned metatranscriptomic data was mapped against the bacterial genomes obtained from the ProBac isolates using bowtie2 with a percentage of identity greater or equal to 99 % and a minimal coverage length of 80 %. A similar normalization was performed as previously described (TPM calculated on the total number of reads per sample).

Transcriptomic analyses were performed in collaboration with Marc García-Garcera from the University of Lausanne, Switzerland.

## 26. Metabolomic analysis to analyse fructose levels

In order to demonstrate that ProBac decreases the concentration of fructose needed by VRE to colonize the intestine, metabolomics analysis were performed comparing the metabolome of caecum samples from mice treated with vancomycin that received ProBac or not. As previously mentioned, the same samples analysed through metabolomics were the same as those analysed through transcriptomics.

Cecum supernatants resuspended in phosphate buffer (see previous section) were utilized for metabolomic analysis. Analysis of fructose was performed by gas chromatography–mass spectrometry. Sample treatment consisted in protein precipitation and derivatization to their trimethylsilyl derivatives following a previous work from Medeiros et Simoneit[174].

To do so, 3 volumes of cold methanol were added to 50 uL of sample. After centrifugation the clean supernatant was collected and vacuum-dried. Dried samples were derivatized by addition of 20 μL of a BSTFA (N,O-bis-(trimethylsilyl)trifluoroacetamide) solution containing 1 % TMCS (trimethylchlorosilane) and 10μL of pyridine and maintained during 3 h at 70 ºC. The derivatized extracts were then vacuum-dried and solved in 100 μL of hexane for injection. A standard of fructose was derivatized the same way and used to identify the fructose peak.

GC-MS analysis was performed using an Agilent 7890A gas chromatograph coupled to an Agilent 7200 accurate mass high resolution GC/Q-ToF. Separation was performed using an Agilent DB-5ms + DG capillary column (30 m x 0.25 mm i.d., 0.25 μm film thickness + 10 m Duraguard) using Helium as carrier gas. Mass analysis was operated on EI conditions, recording data in full-scan mode at 70eV in a mass range of m/z 50 to 600. Ion source, quadrupole and transfer line temperatures were 250 ºC, 150 ºC and 290 ºC respectively. The integrated peak area of fructose peak at EIC (extracted ion chromatogram) m/z 217.1075 was used to qualitatively determine the monosaccharide.

Metabolomic analyses were performed in collaboration with Marina Lopez Nogueroles from the Analytical Unit Platform, Health Research Institute La Fe, Valencia, Spain.

**27. *ex vivo* nutrient competition experiments with ProBac**

To study that ProBac could be conferring protection against VRE through competition for nutrients, caecum samples obtained from mice treated with vancomycin that were allowed to recover for 2 weeks after stopping vancomycin treatment (section 28.5) were unfrozen and resuspended in the proportion of 100 mg of caecum content in 1 ml of PBS pH7. After resuspending the caecal content, a 2 min centrifugation was performed at 13000 rpm, collecting the supernatant in a new tube. To ensure that all bacteria present were removed, the supernatant was filtered in columns by centrifugation, firstly using columns with membranes of 0.4 μm pore size (Thermo Scientific) followed by a second column with a membrane of 0.2 μm (Thermo Scientific). 40 μl of the filtered supernatant was added to each well of a 96 well-plate for each condition tested. The plate containing the supernatants was reduced overnight in an anaerobic chamber to remove the oxygen.

The next day, stocks of the bacterial isolates (i.e. *Oscillibacter*, UC_Ruminococcaceae, *Alistipes*, *Olsenella* and *Barnesiella*) were unfrozen inside the anaerobic chamber. $10^8$ CFUs of each bacterium (bacteria were previously plated to calculate the number of CFUs/μl) resuspended in a final volume of 5 μl (containing the 5 bacterial isolates or individual isolates) were added to the 40 μl of caecum supernatant. The mixture was incubated at 37 ºC during 24 h. After the incubation period, 5 μl of PBS containing 500 CFUs of VRE, that was previously grown overnight in BHI medium, were added inside anaerobic chamber to the filtered intestinal content containing the ProBac mixture or the individual isolates. As a control, 500 CFUs of VRE were added to caecal filtered supernatant that was previously incubated as described above but that did not contain any commensal bacteria or that contained an isolate of the genera *Bacteroides* (not associated with protection against VRE). Besides the addition of VRE, when specified in the results, 5 μl of a sugar mixture (Table 8) was also added to the caecum filtered supernatants. The sugar mixture contained 0.5 M of each sugar indicated in Table 8 or just fructose. The sugars were previously resuspended in water and filtered (0.2 μm pore size). After VRE administration, the mixture was incubated for 6 h under anaerobic conditions at 37 ºC. Subsequently, dilutions of the mixture on PBS were plated on the selective media BEA+AV to quantify VRE growth.

**Table 8. Mix of eight sugars added to the filtered caecum.**

| Sugar Mix |
| --- |
| Fructose |
| Lactose |
| Mannose |
| Threalose |
| Cellobiose |
| N-acetyl galactosamine |
| D-glucosaminate |
| Maltose |

## 28. Mice experiments

Mice were maintained following the National guidelines (RD 53/2013), and the specific protocols were approved by the University of Valencia Animal Care committee. The procedures were performed in accordance with institutional protocol guidelines at the "Servei Central de Suport a la Investigació Experimental" at the University of Valencia. Experiments were performed with 7/8-week-old C57BL/6J bred in the animal facilities.

### 28.1 Antibiotic treatment set-up for the study of VRE genes involved in gut colonization

To evaluate the effect of different antibiotics on the mice gut microbiota and the capability of VRE to colonize the intestine after the treatment, mice were treated during seven days with oral vancomycin (0.5 g/l, Alfa Aesar), ampicillin (0.5 g/l, AppliChem), amoxicillin (0.25 g/l, AppliChem), amoxicillin with clavulanic acid (0.25 g/l, Argenol Laboratories) or a cocktail containing vancomycin (0.5 g/l, Alfa Aesar), ampicillin (0.5 g/l, AppliChem) and streptomycin (0.45 g/l, Sigma) in the drinking water. Clindamycin (1.4 mg/ml, Sigma) and metronidazole (1 mg/ml, Sigma) were administered twice daily punctured subcutaneously (Figure 22).

The fourth day of the treatment, mice were orally inoculated through oral gavage with $10^8$ CFUs of VRE (*E. faecium* strain Aus0004). Before inoculation, a faecal sample was collected and plated on BEA+AV plates to check that no bacteria from the mouse microbiota were able to grow in the medium and there was no previous contamination with VRE.

At the seventh day of treatment (third day post-VRE inoculation) a faecal sample was retrieved to plate serial dilutions on BEA+AV. VRE CFUs were quantified two days after incubating the plates at 37 ºC to calculate the amount of VRE present in the murine gut, referred as VRE CFUs/10 mg faeces.
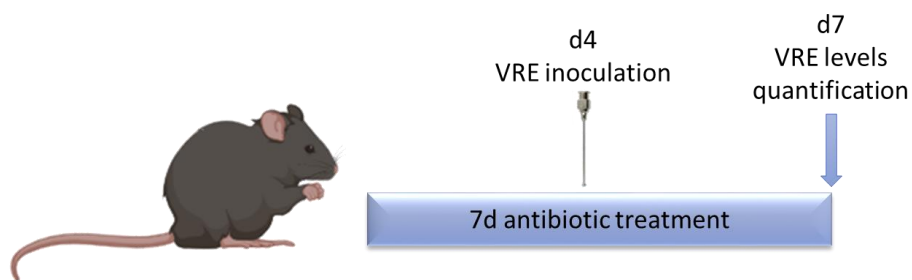
**Figure 22. Design of mouse model experiments with C57BL6 strain.** Mice were treated during seven days with antibiotics and were orally infected with VRE at the fourth day of treatment. Three days post-VRE inoculation a faecal sample was obtained to quantify VRE levels.

A summary of antibiotics used in these experiments is indicated in the Table 9 and the overview of their spectrum of action in the Figure 23. Antibiotics were chosen according to their different spectrum and their frequent usage in the clinic.

**Table 9. Antibiotic tested in the mouse model.** Different treatments administered to mice to select the antibiotic treatment that cause higher VRE colonization levels. The concentration and route of administration for each antibiotic are indicated.

| Antibiotics | Concentration/daily dose | Administration route |
|---|---|---|
| Clindamycin | 1.4mg/ml – 1ml | Subcutaneous |
| Vancomycin | 0.5g/l | Drinking water |
| Ampicillin | 0.5g/l | Drinking water |
| Metronidazole | 1mg/ml – 1ml | Subcutaneous |
| Amoxicillin | 0.25g/l | Drinking water |
| Amoxicillin + clavulanic acid | 0.25g/l | Drinking water |
| Vancomycin + ampicillin + streptomycin | 0.5g/l 0.5g/l 0.45g/l | Drinking water |
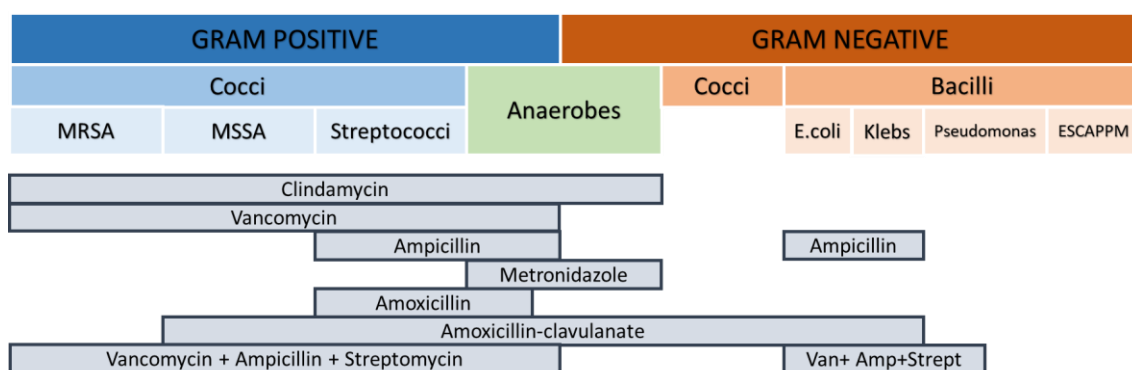


**Figure 23. Overview of the spectrum of the antibiotics used in the mouse model.** Figure adapted from the Wellingtonicu.com drug manual. MRSA: Methicillin Resistant Staphylococcus aureus, MSSA: Methicillin Sensitive Staphylococcus aureus, ESCAPPM: bacteria with β-lactamase activity.

## 28.2 Mutant transposon library oral inoculation to identify genes required for gut colonization

To identify genes involved in gut colonization by VRE, mice were treated for 1 week with clindamycin punctured subcutaneously (1.4 mg/ml, twice daily). The fourth day of the treatment, mice were inoculated with the VRE library inoculum. This inoculum was prepared by inoculating the previous day 100 µl from a mutant transposon library stock into a flask containing 20 ml of BHI broth supplemented with 300 µg/ml gentamicin. Bacterial cells were grown for 6 h at 37 ºC. Part of the culture was stored for further DNA extraction, sequencing and another part of the culture was used to infect mice with the mutant transposon library by oral gavage. Mice were orally gavage with $10^8$ CFUs/200 µl. The inoculum was plated to determine viable counts.

A faecal sample was collected before the inoculation and plated in BEA+AV to check the absence of VRE. Three days after inoculation, samples were collected from different parts of the intestine: ileum, caecum, colon, faeces and wall scrapings (Figure 24). Part of the sample was plated in the BEA+AV plates to assess the levels of VRE colonization at different intestinal parts and the other part of the sample was stored at -80 ºC for Tn-Seq as described below.
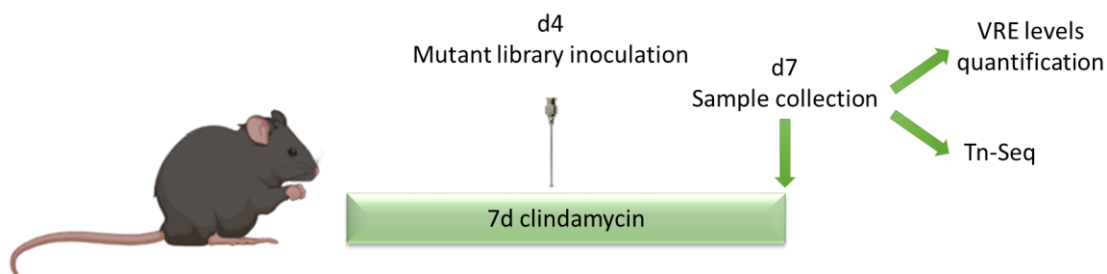


**Figure 24. Clindamycin mouse model for mutant transposon library inoculation.** Mice were treated during seven days with subcutaneous clindamycin (1.4 mg/ml). The fourth day of the treatment, mice were orally inoculated with the mutant transposon library. Intestinal samples were collected three days post-library inoculation.

## 28.3 *In vivo* competition assays

**Clindamycin mouse model**

To verify that the genes identified by the Tn-Seq approach were required for VRE gut colonization, we designed competition experiments between WT strain and mutant strains (targeted mutagenesis explained in section 14) to check if their ability to colonize the gut differs between them. Mice were treated with clindamycin for seven days and orally gavage with the WT and mutant strains on the fourth day of the treatment, as indicated above (section 28.1). To prepare the mix for inoculation, the strains were previously grown overnight in BHI plates at 37 ºC. Bacteria were collected from BHI plates with a swab and resuspended in PBS and were mix in the same proportion so that the mixture contained $10^8$ CFUs of each strain/200 µl. Dilutions of each individual bacterium were plated to count the number of CFUs inoculated to mice. Faeces samples

were collected before inoculation to ensure no VRE presence before oral gavage of the strains. The third day post-VRE inoculation, faeces samples were collected and plated in a selective medium with the corresponding antibiotics (BEA+A and BEA+AG) to determine the number of CFUs of each strains and calculate the competition index as explained in section 16.1.

**Vancomycin mouse model**

To assess the involvement of genes in the intestinal colonization of VRE under different conditions and to evaluate long-term effects in VRE intestinal colonization, the antibiotic treatment given to the mice was vancomycin (0.5 g/l), supplied in the drinking water. Mice were treated a total of seventeen days with vancomycin. As in the clindamycin model, the fourth day of vancomycin treatment, mice were inoculated with a mix of WT and each mutant strain by oral gavage with a total of $10^8$ CFUs/200 µl. The inoculum contained an equal proportion of WT and each mutant strain, which were individually grown in BHI plates as prepared for the clindamycin model. Dilutions of each bacterium were plated before inoculation to count the number of CFUs. Stool samples were collected before inoculation to verify no presence of VRE in the mice gut intestine. Faeces samples were collected three days post-VRE inoculation and plated in selective medium to differentiate both strains, BEA+A and BEA+AG. In addition, samples were collected six and thirteen days post-VRE inoculation to follow the proportion of both strains in the mice intestine over time (Figure 25). The competitive index (CI) for each time point collected was calculated as explained before.
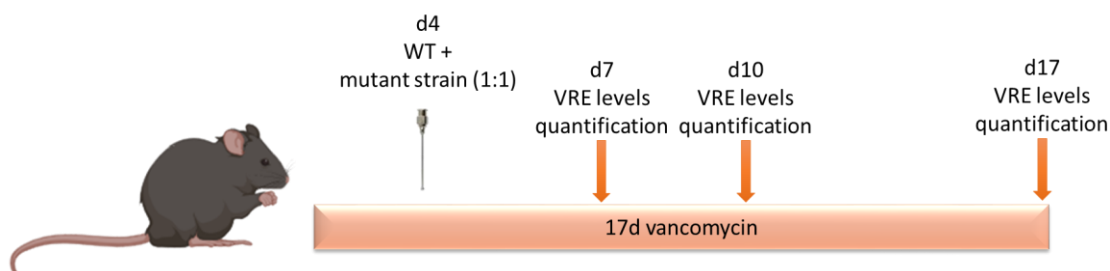


**Figure 25. Design of the vancomycin mouse model to study the effect of a particular gene on VRE intestinal colonization.** Mice are treated during seventeen days with vancomycin in the drinking water (0.5 g/L) and co-inoculated at day fourth of treatment. Faecal samples to evaluate VRE levels were collected on days three, six and thirteen post-VRE inoculation. Samples are plated in selective medium to evaluate the levels of VRE.

## 28.4 *In vivo* model adding mannose to the drinking water

A mouse model adding mannose to the drinking water was set up to verify if adding an excess of this sugar in the drinking water would increase the growth of VRE in the intestine (Figure 26).

Vancomycin treatment was administered to co-housed mice (3 per cage) during three days in the drinking water (0.5 g/l). The fourth day, antibiotic was replaced by water and mice were separated individually, administering $10^6$ CFUs of VRE in 200 µl. Three days

after the administration of VRE, stool samples were collected in order to check the levels in the intestine and mannose was added to two of the three mice that were previously co-housed during the antibiotic treatment. Faeces samples were collected one and two days after the addition of the mannose and were plated in selective media (BEA+AV) to check the levels of VRE colonization in the different groups of mice.
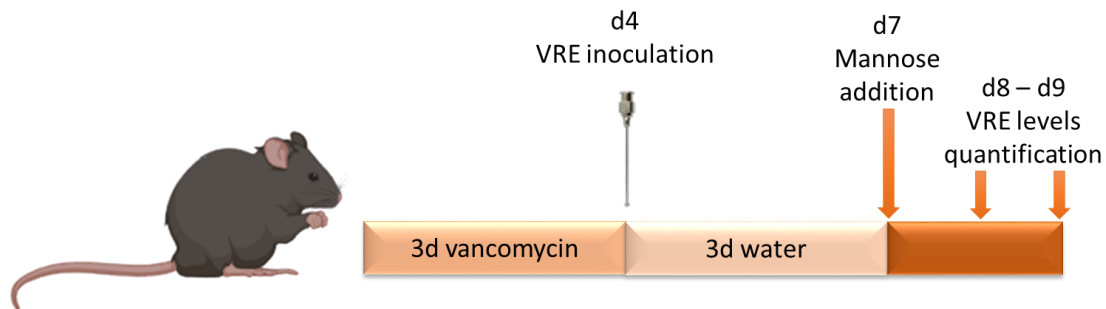


**Figure 26. Design of vancomycin mouse model adding mannose.** Mice were treated three days with vancomycin (0.5 g/l). Subsequently, antibiotic is replaced by water and VRE was inoculated. Three days after inoculation, mannose is added to the drinking water. Stool samples are collected one and two days after mannose addition to measure VRE levels in faeces.

## 28.5    Mouse model with probiotics administration

In order to evaluate the capacity of specific bacteria to protect against VRE, mice were treated with vancomycin in drinking water (0.5 g/l) for seven days and then the antibiotic was replaced by water (Figure 27). One day after the antibiotic withdrawal, mice were co-housed (two mice per cage) and bacteria to be tested (ProBac or *Olsenella*), bacterium control (*Bacteroides*) or the vehicle were the bacteria were resuspended (PBS with 20 % glycerol and 0.1 % cysteine) were administered through oral gavage to mice for three consecutive days.

To decrease the contact of the anaerobic bacteria with the atmosphere, avoiding an exchange of oxygen during their administration, mix aliquots or individual bacterium were maintained in dry ice and unfrozen immediately before administration.

We let the co-housed mice to recover their microbiota for two weeks since the day that the antibiotic was replaced by water (see legend in Figure 27 for the explanation of the co-housing strategy). After the recovery period, mice were housed individually and two stool samples were collected from each mouse. The first one to be plated in selective BEA+AV medium to check that no bacteria from the mouse microbiota were able to grow in the selected medium used to quantify VRE levels and that there was no previous contamination with VRE. The second sample was conserved at -80 ºC for further analysis of the microbiota composition. Subsequently, mice were orogastrically inoculated with $10^6$ CFUs of VRE in 200 µl. Stool samples were collected two days after the inoculation of VRE and were plated in the selective media BEA+AV to assess the levels of VRE present in the gut.
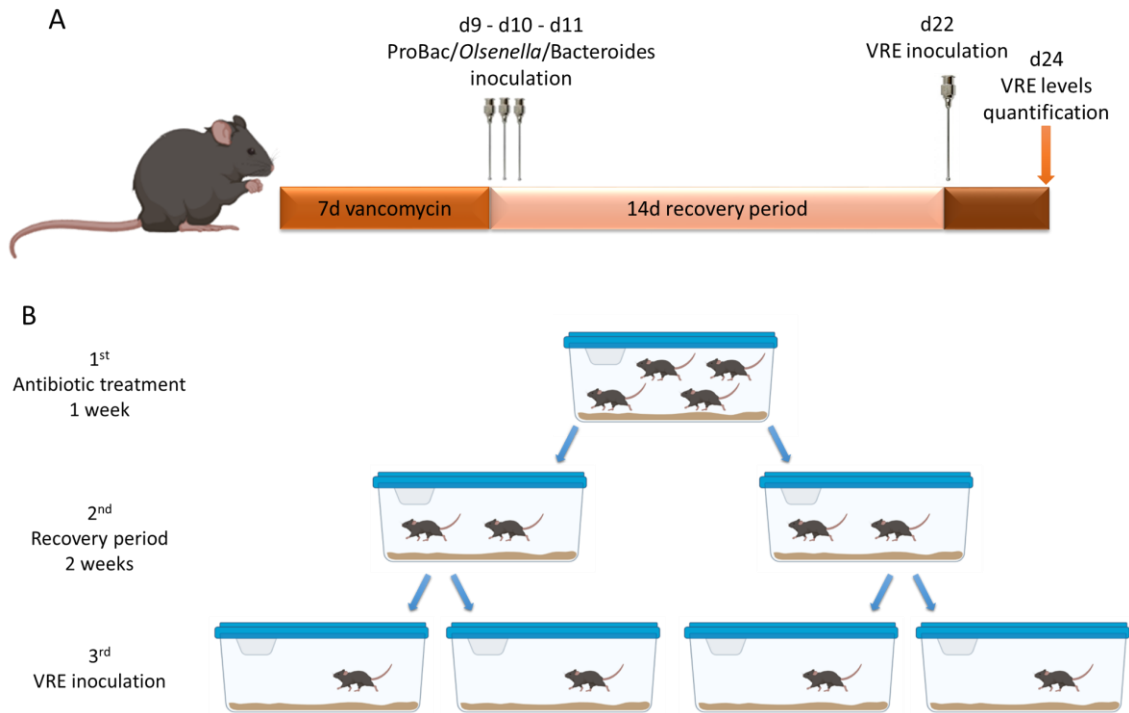
**Figure 27. Mouse model used to study the ProBac protective effect against VRE intestinal colonization. (A)** Mice were treated during seven days with vancomycin applied in the drinking water. One day after antibiotic withdrawal, mice were inoculated with the commensal bacteria (i.e. ProBac, *Olsenella* or *Bacteroides*) and were kept for two additional weeks to allow the recovery and stabilization of the microbiota. VRE was inoculated after the recovery period and stool samples were collected two days after VRE inoculation and were plated in selective media in order to quantify VRE levels. **(B)** Mice co-housing during the development of the experiment. First, mice are co-housed during the administration of vancomycin treatment. After antibiotic withdrawal mice were co-housed in pairs during the recovery period. Finally, mice are placed in individual cages when VRE is administrated by oral gavage. The reason of this co-housing scheme is: (i) the co-house during the recovery period is done using 2 mice per cage for increasing the number of replicates (i.e. cages), since mice from the same cage behave very similar in terms of which microbiota they recover and the ability of VRE to colonize those mice. For these reason we do not cohouse the maximum mice (N=6) per cage during these period. Nevertheless, we include 2 mice per cage and not house them individually in order to favour the recovery of the bacteria that we administer: if the probiotic has efficiently colonized one mouse, it will be transmitted to the other mouse through faecal shedding and coprophagy. (ii) when we inoculate VRE, we house the mice individually so that there is not VRE transmission from one mouse to the other.

## 28.6    Mouse model with probiotics administration adding fructose

With the aim of verifying if fructose facilitates VRE growth *in vivo*, we followed the mouse model explained in the previous section: mice treated with vancomycin but modifying the last part, after the VRE administration. As previously explained, the day of VRE inoculation, mice were separated individually. Two days post-VRE inoculation, faecal samples were collected to quantify VRE levels and fructose (100 mM) was added in the drinking water to one of the two mice that were co-housed together during the recovery period. We collected samples four days after the inoculation of VRE (two days after the administration of fructose) to quantify the levels of VRE and compared samples

from mice that were receiving fructose during two days and those that were receiving water without fructose that were previously co-housed in the same cage (Figure 28).

The advantage for VRE to grow in vivo with the extra fructose administration was calculated as follows:

Log2 (**(VRE CFUs D4 p.i. / VRE CFUs D2 p.i.**[mouse receiving fructose] **)** / (*VRE CFUs D4 p.i. / VRE CFUs D2 p.i*))

In bold (in the numerator of the fraction) are included the VRE CFUs of the co-housed mouse that received fructose. In italic (in the denominator of the fraction) are included the VRE CFUs of the co-housed mouse that did not receive fructose.
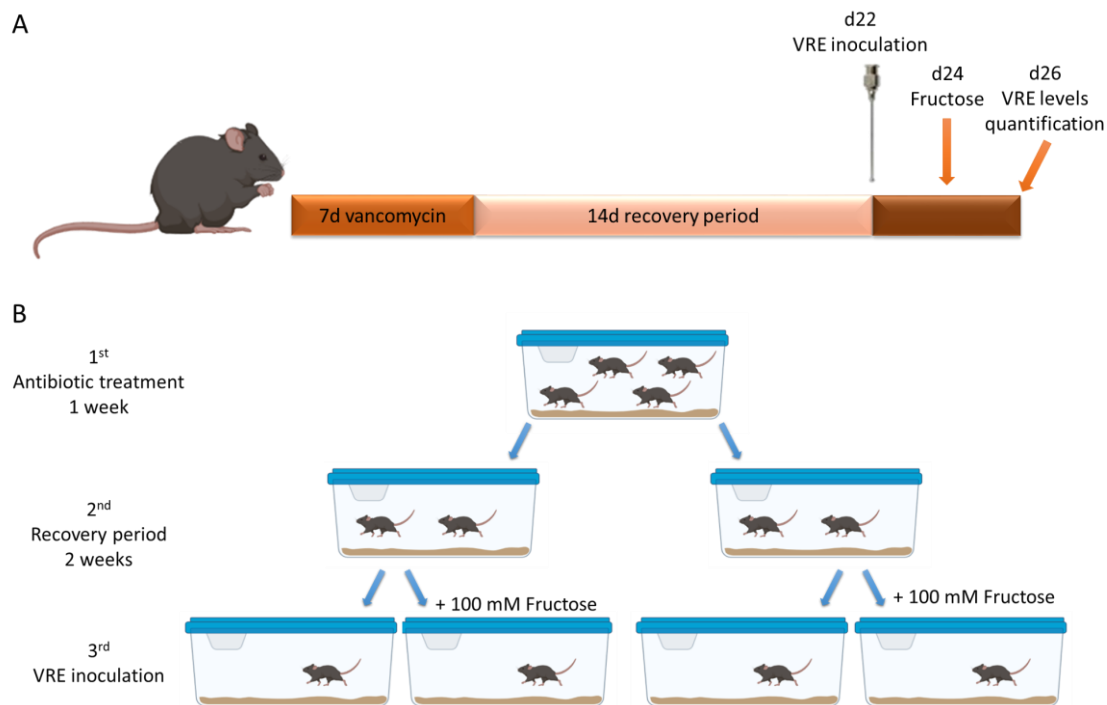


**Figure 28. Design of mouse model adding fructose after VRE inoculation. (A)** Mice were treated with vancomycin and two weeks after antibiotic withdrawal mice were inoculated with VRE. Two days after VRE inoculation, fructose was added to the drinking water. Faeces samples were collected the day the fructose was added and two days after the fructose addition to measure VRE levels. **(B)** Mice co-housed during the development of the experiment adding fructose to the drinking water. First, mice are co-housed during the administration of vancomycin treatment. After antibiotic withdrawal, mice were co-housed in pairs during the recovery period. Finally, mice are placed in individual cages when VRE is administrated orally and two days after its inoculation, 100 mM of fructose is added to the water to one of the two mice that were housed in the same cage.

## 28.7   Mouse model with *Olsenella* administration after VRE inoculation

In order to check if *Olsenella* was capable to suppress VRE growth *in vivo* once VRE has been stablished in the gut, we designed a mouse model, administering VRE before the inoculation of *Olsenella* (Figure 29)*.* Mice were co-housed in groups of two and treated with vancomycin during seven days, inoculating VRE at day four of the treatment. As previously commented, mice that are in the same cage have comparable microbiota and the susceptibility to be colonized with VRE is similar too.

At day seven, antibiotic treatment was replaced by water. Next day mice were separated individually and *Olsenella* or the vehicle to resuspend the bacteria (PBS with 20 % glycerol and 0.1 % cysteine) was administered respectively to each mouse of the cage. *Olsenella* or the vehicle was administered during three consecutive days.

Mice were let for two weeks since the day that the antibiotic was replace by water. Then, faeces samples were collected to be plated in selective media BEA+AV to assess if the presence of *Olsenella* favoured the VRE elimination compared to those mice previously co-housed that were receiving the vehicle.
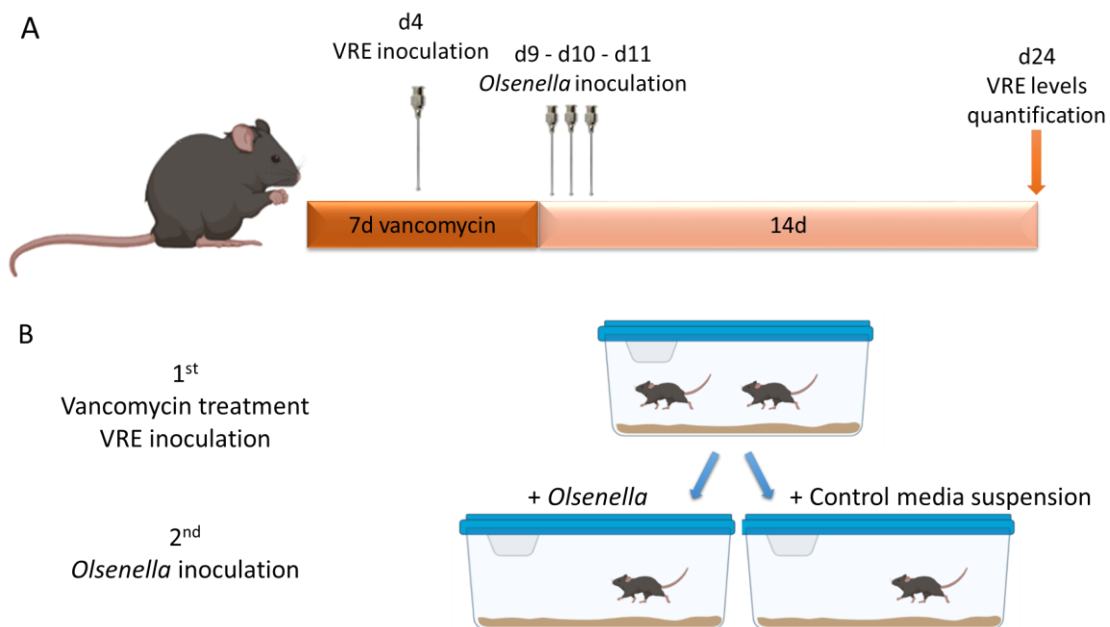


**Figure 29. Mouse model designed to evaluate if *Olsenella* was able to eliminate VRE previously stablished in the gut. (A)** Mice were treated during seven days with vancomycin treatment in drinking water, supplying orogastrically VRE at day four. Antibiotic was replaced by water at day eight and the three next consecutive days mice were inoculated with *Olsenella* or the vehicle used to resuspend the bacterium. Faeces samples were collected immediately after antibiotic cessation and two weeks after antibiotic withdrawal and plated in selective media to quantify the levels of VRE. **(B)** Mice were co-housed in pairs during the administration of the antibiotic treatment in water and the inoculation of VRE. Then, mice were housed individually before the administration of *Olsenella* or the vehicle used to resuspend the bacterium. In this particular case, mice were housed invidually before VRE inoculation to increase the number of replicates (one mouse per cage). *Olsenella* colonize the gut very efficiently and for this reason we did not house the mice in pairs to facilitate *Olsenella* colonization.

## 28.8    Mouse model used to collect sample for *ex vivo* experiments

With the objective to collect mouse samples for *ex vivo* experiments, we followed the mouse model with two weeks of recovery (see section 25.5) but without administering either probiotics or VRE (Figure 30). Mice were treated with vancomycin during one week and treatment was substituted by water at day eight. The next day, mice were separated and co-housed in groups of two mice. After the two weeks of recovery period, mice were housed individually. Next day, mice were sacrificed to collect samples from caecum, which were frozen directly at -80 ºC for further *ex vivo* experiments.
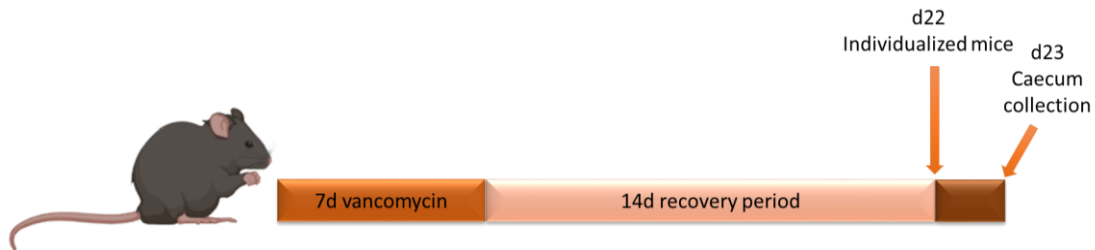


**Figure 30. Mouse model to collect caecum samples for *ex vivo* experiments.** Mice were treated with vancomycin for seven days followed by a recovery period of two weeks. Subsequently, mice were individually housed and the day after were sacrificed in order to collect caecum samples.

## 29.  Protective bacteria preparation to be administered to mice

Each individual bacterium that constituted the ProBac mixture was previously isolated by another student from our group and stored at -80 ºC[168]. Each bacterium stock was taken out from the -80 ºC freezer, left it in dry ice and plated, scraping the surface of the stock with a loop inside the anaerobic chamber to maintain anaerobic conditions. Each bacterium was plated in one Columbia blood agar (CBA) plate and incubated under anaerobic conditions at 37 ºC[168], during the time required to grow each bacterium (Table 10).

**Table 10. Growth time of each bacterium isolate tested in a mouse model.**

| Genus | Growth Medium | Growth time (days) |
|---|---|---|
| *Oscillibacter* | CBA | 6 |
| unclassified_Ruminococcaceae | CBA | 6 |
| *Alistipes* | CBA | 3 |
| *Olsenella* | CBA | 3 |
| *Barnesiella* | CBA | 3 |

Once each bacterium was grown, they were replated in several CBA plates to obtain enough bacteria to perform the experiments. Each individual bacterium was collected with a swab and resuspended in PBS with 20 % of glycerol and 0.1 % of cysteine previously reduced in the anaerobic chamber. $OD_{600}$ was measured and each isolate was adjusted to an $OD_{600}$ of 2.5 approximately, corresponding to ~ $10^7$ CFUs/ml of each

individual bacterium. A mixture of the five isolates was prepared by adding into the same tube an equal volume of each of them. In addition to the mix preparation, each individual bacterium was stocked in 100 µl aliquots for other *ex vivo* experiments.

In order to confirm that the bacteria grown was not a contaminant from the anaerobic chamber, a PCR of the 16s rRNA gene using KAPA polymerase (section 5, Material and Methods) was performed and sequenced through Sanger as previously described.

## 30. Statistical tests utilized to analysed different set of data

Parametric statistical tests were used when the groups followed a normal distribution (t-test) and non-parametric (Wilcoxon) when the normal distribution was not followed.

All unpareid t-test and Wilcoxon were made two-sided except on one occasion: to test *in vivo* the protection of *Olsenella.* We applied a one-side test because the experiment performed was to test the hypothesis that *Olsenella* could recapitulate the ProBac protective effect against VRE colonization and therefore the only expected result, which we confirmed, was a lower level of VRE CFUs in the *Olsenella* group as compare to the PBS-glycerol-cysteine group.

In some occasions (competition between WT and mutant strains – competitive index) we applied one sample t-test in order to detect if the mean of the logFC values obtained in one group of samples was significantly different from 0. Thus, if the result was not significant it means that both strains (WT and mutant) have the same capacity to grow.

Pearson correlation analysis was applied to investigate the correlation between the abundance of the Tn-mutants detected in library replicates among samples from the same and different experiments and the correlation between the fructose caecal levels and the $\log_{10}$ VRE faecal CFUs in co-housed mice.

All results were considered significant when p values were lower than 0.05.

## 31. Composition of prepared media used

**Table 11. Preparation of 2x CDM and 2x phosphate starvation media based on Cremoris Defined Medium.**

| 2x Chemically defined medium | | 2x Phosphate Starvation Medium | |
|---|---|---|---|
| Dissolve in 600ml $H_2O$ | | | |
| $K_2HPO_4$ | 2g | MOPS | 10g |
| $KH_2PO_4$ | 10g | | |
| Sodium acetate.$3H_2O$ | 3.32 | Sodium acetate.$3H_2O$ | 3.32 |
| Tri-ammonium citrate | 1.2g | Tri-ammonium citrate | 1.2g |
| Ascorbic acid | 1g | Ascorbic acid | 1g |
| Glucose-monohydrate | 15g | Glucose-monohydrate | 15g |
| Tyrosine | 0.5g | Tyrosine | 0.5g |

**Add mix solutions and adjust volume to 1L (see below for composition):**

| | |
|---|---|
| 10x Amino acid solution | 20ml |
| 100x DNA precursor mix | 20ml |
| 100x Metal solution | 20ml |
| 100x Vitamin Solution | 20ml |

| 100x Vitamin solution per Litre, pH10 adding 10M NaOH for proper dissolving | |
|---|---|
| Pyridoxamine-HCl | 500mg |
| D-biotin | 250mg |
| 6,8-thioctic acid | 250mg |
| Pyridoxine-HCl | 200mg |
| Nicotinic acid | 100mg |
| Ca($D^+$)pantothenate | 100mg |
| Riboflavin | 100mg |
| Thiamine-HCl | 100mg |
| Folate | 100mg |
| p-aminobenzoic acid | 1g |
| Orotic acid | 500mg |
| Thymidine | 500mg |
| Inosine | 500mg |

| 100x Metal solution per Litre | |
|---|---|
| $MgCl_2.6H_2O$ | 20g |
| $CaCl_2.2H_2O$ | 5g |
| $MnCl_2.4H_2O$ | 1.6g |
| $FeCl_3.6H_2O$ | 0.3g(10ml 17 %HCl) |
| $FeCl_2.4H_2O$ | 0.5g(200ml $H_2O$ ) |
| $ZnSO_4.7H_2O$ | 0.5g |
| $CoSO_4.7H_2O$ | 0.25g |
| $CuSO_4.5H_2O$ | 0.25g |
| $(NH_4)_6Mo_7O_{24}.4H_2O$ | 0.25g |

| 100x DNA precursor mix, dissolved in 10ml 0.1M NaOH | |
|---|---|
| Adenine | 10mg |
| Guanine | 10mg |
| Uracil | 10mg |
| Xanthine | 10mg |

| 10x Amino acid solution per Litre, pH6.8 for proper dissolving of all aminoacids | |
|---|---|
| Alanine | 2.40g |
| Arginine | 1.25g |
| Aspartic acid | 4.20g |
| Cysteine-HCl | 1.30g |
| Glutamic acid | 5g |
| Histidine | 2.10g |
| Isoleucine | 4.75g |
| Leucine | 4.40g |
| Phenylalanine | 2.75g |
| Proline | 6.75g |
| Serine | 3.40g |
| Threonine | 2.25g |
| Tryptophan | 0.50g |
| Valine | 3.25g |
| Glycine | 1.75g |
| Methionine | 1.25g |

**Table 12. Complete and phosphate starvation medium previously used in *Bifidobacterium breve UCC2003*.**

| Chemically defined medium | | Phosphate Starvation Medium | |
|---|---|---|---|
| Basal Medium – Dissolve in 850ml $H_2O$ | | | |
| B-glycerophosphate | 19g | MOPS | 10g |
| $K_2HPO_4$ | 3g | | |
| $KH_2PO_4$ | 1g | | |
| Sodium acetate | 1g | Sodium acetate | 1g |
| Tri-ammonium citrate | 0.6g | Tri-ammonium citrate | 0.6g |
| Ascorbic acid | 0.5g | Ascorbic acid | 0.5g |
| Tyrosine | 0.25g | Tyrosine | 0.25g |
| Tween 80 | 0.1 % | Tween 80 | 0.1 % |
| Glycerol | 0.1 % | Glycerol | 0.1 % |

| Trace metals per litre: | |
|---|---|
| $MgSO_4.7H_2O$ | 50mg |
| $MnSO_4.7H_2O$ | 10mg |
| $FeSO_4.7H_2O$ | 5mg |
| $CuSO_4.7H_2O$ | 5mg |
| $CoCl_2$ | 0.5mg |
| $ZnSO_4.7H_2O$ | 0.15mg |

| Bases (in 0.1N NaOH) per litre: | |
|---|---|
| Adenine | 0.1mg |
| Guanine | 0.1mg |
| Uracil | 0.1mg |
| Xanthine | 0.1mg |

| | |
|---|---|
| Casamino acids | 5g/L |

| | |
|---|---|
| Sugar (glucose) | 1 % final conc. |

| | |
|---|---|
| Cys-HCl | 0.05 % final conc. |

| | |
|---|---|
| Fatty acids | 10µg/ml |

| Vitamins per litre: | |
|---|---|
| Inosine | 5mg |
| p-aminobenzoic acid | 10mg |
| Thymine | 5mg |
| 6,8-thiotic acid | 2.5mg |
| Biotin | 2.5mg |
| Pyridoxine-HCl | 2mg |
| Folic Acid | 1mg |
| Ca-($D^+$)-pantothenate | 1mg |
| Riboflavin | 1mg |
| Thiamine-HCl | 1mg |
| Cobalamin | 1mg |

**Table 13. Ingredients added to prepare SOC medium.**

| Ingredient | Per Liter |
|---|---|
| Yeast extract | 0.5 % |
| Tryptone | 2 % |
| NaCl | 10 mM |
| KCl | 2.5mM |
| $MgCl_2$ | 10 mM |
| $MgSO_4$ | 10 mM |
| Glucose (added after autoclaving the solution with the remaining ingredients) | 20 mM |

**Table 14. Ingredients added to prepare M1 medium.**

| Ingredient | Per Liter of PBS |
|---|---|
| Yeast extract | 0.5 g |
| Tryptone | 10 g |
| Carbon source (if required) | 5 mM |

## 32. Strains and plasmids

**Table 15. Strains used in these studies.**

| Strain | Name used in the thesis | Description | Reference |
|---|---|---|---|
| *E. faecium* | | | |
| E1162 | E1162 | *Enterococcus faecium* clinical isolate from bloodstream infection, France, 1996 | [60] |
| E1133 | E1133 | *Enterococcus faecium* clinical isolate from patient faeces during a VRE outbreak, USA, 2002 | [150] |
| ATCC 700221 | ATCC 700221 | *Enterococcus faecium* clinical isolate from human faeces in Connecticut, USA | [151] |
| Aus0004 | Aus0004 | *Enterococcus faecium* clinical isolate from bloodstream infection, Australia, 1998 | [61] |
| CU737 | IR::*gen* | Gentamicin marker inserted next to the esp PAI in Aus0004 strain | This study |
| CU738 | Δ*man*X::*gen* | Deletion of *man*X gene replaced by gentamicin marker in Aus0004 strain | This study |
| CU739 | Δ*suf*B::*gen* | Deletion of *suf*B gene replaced by gentamicin marker in Aus0004 strain | This study |
| CU740 | Δ*vat*::*gen* | Deletion of *vat* gene replaced by gentamicin marker in Aus0004 strain | This study |
| CU747 | Δ*pst*B2::*gen* | Deletion of *pst*B2 gene replaced by gentamicin marker in Aus0004 strain | This study |
| CU743 | Δ*yyc*I::*gen* | Deletion of *yyc*I gene replaced by gentamicin marker in Aus0004 strain | This study |
| CU742 | Δ*ltrR*::*gen* | Deletion of LacI family DNA-binding transcriptional regulator (*ltrR*) replaced by gentamicin marker in Aus0004 strain | This study |
| CU741 | Δ*mur*Q::*gen* | Deletion of *mur*Q gene replaced by gentamicin marker in Aus0004 strain | This study |
| CU745 | Δ*dsb*A::*gen* | Deletion of *dsb*A gene replaced by gentamicin marker in Aus0004 strain | This study |
| CU744 | Δ*adh*E::*gen* | Deletion of *adh*E gene replaced by gentamicin marker in Aus0004 strain | This study |
| CU746 | Δ*pgt*:: *gen* | Deletion of *pgt* gene replaced by gentamicin marker in Aus0004 strain | This study |
| CU748 | Δ*suf*DΔ*csd*Δ*nif*UΔ*suf*B::*gen* | Deletion of a cluster of genes *suf*D, *csd*,*nif*U,*suf*B replaced by gentamicin marker in Aus0004 strain | This study |
| CU951 | Δ*man*X::*gen*+*manX* | *manX* gene cloned into the pNZ44 plasmid and transformed into Δ*man*X::*gen* strain | This study |
| CU952 | Δ*mur*Q::*gen*+*murQ* | *murQ* gene cloned into the pNZ44 plasmid and transformed into Δ*mur*Q::*gen* strain | This study |
| *E. coli* | | | |
| DH5α | DH5α | *E. coli* host strain for routine cloning | [153] |
| EC1000 | EC1000 | *E. coli* host strain for routine cloning with pWS3 plasmid | [153] |

| | | | |
|---|---|---|---|
| **L. lactis** | | | |
| NZ9000 | L. lactis | L. lactis strain for routine cloning | 152 |
| **ProBac isolates** | | | |
| Oscillibacter | Oscillibacter | Oscillibacter murine isolate | 168 |
| unclassified_ Ruminococcaceae | unclassified_ Ruminococcaceae | unclassified_Ruminococcaceae murine isolate | 168 |
| Alistipes | Alistipes | Alistipes murine isolate | 168 |
| Olsenella | Olsenella | Olsenella murine isolate | 168 |
| Barnesiella | Barnesiella | Barnesiella murine isolate | 168 |
| Bacteroides | Bacteroides | Bacteroides murine isolate | 168 |

**Table 16. Plasmids used in these studies**

| Plasmid | Name used in this thesis | Description | Reference |
|---|---|---|---|
| pCR[TM]4 TOPO® | TOPO | TOPO® TA Cloning vector, Kn[r], Amp[r] | Invitrogen |
| pZXL5 | pZXL5 | Transposon delivery plasmid. Nisin induced mariner transposase, Gm[r] mariner transposon, gram-positive thermosensitive origin of replication and Cm[r] cassette | [153] |
| pGPA | pGPA | Modified pZXL5 plasmid, contains MmeI restriction sites at two outward-facing T7 promoters of the Gm[r] mariner transposon | [99] |
| pWS3 | pWS3 | Gram-positive thermosensitve origin; Spc[r] | [153] |
| pCU957 | pWS3-flanking regions *manX + gen* | pWS3 containing flanking regions of *manX* gene and Gm[r] cloned between both | This study |
| pCU958 | pWS3-flanking regions *sufB + gen* | pWS3 containing flanking regions of *sufB* gene and Gm[r] cloned between both | This study |
| pCU959 | pWS3-flanking regions *vat + gen* | pWS3 containing flanking regions of *vat* gene and Gm[r] cloned between both | This study |
| pCU960 | pWS3-flanking regions *murQ + gen* | pWS3 containing flanking regions of *murQ* gene and Gm[r] cloned between both | This study |
| pCU961 | pWS3-flanking regions *ltrR + gen* | pWS3 containing flanking regions of LacI family DNA-binding transcriptional regulator (*ltrR)* and Gm[r] cloned between both | This study |
| pCU962 | pWS3-flanking regions *yycI + gen* | pWS3 containing flanking regions of *yycI* gene and Gm[r] cloned between both | This study |
| pCU963 | pWS3-flanking regions *adhE + gen* | pWS3 containing flanking regions of *adhE* gene and Gm[r] cloned between both | This study |
| pCU964 | pWS3-flanking regions *dsbA + gen* | pWS3 containing flanking regions of *dsbA* gene and Gm[r] cloned between both | This study |
| pCU965 | pWS3-flanking regions *pgt + gen* | pWS3 containing flanking regions of *pgt* gene and Gm[r] cloned between both | This study |
| pCU966 | pWS3-flanking regions *pstB2 + gen* | pWS3 containing flanking regions of *pstB2* gene and Gm[r] cloned between both | This study |
| pCU967 | pWS3-flanking regions *suf*D-*csd-nif*U-*suf*B + *gen* | pWS3 containing flanking regions of genes *sufD* and *sufB* genes and Gm[r] cloned between both | This study |
| pNZ44 | pNZ44 | Broad range origin, Cm[r] | [175] |
| pCU955 | pNZ44-manX | Complementation plasmid for manX; pNZ44 carrying gene manX | This study |
| pCU956 | pNZ44-murQ | Complementation plasmid for murQ; pnZ44 carrying gene murQ | This study |

## 33. Primers and adapters

**Table 17. Primers used in these studies.**

| PRIMER NAME (a) | SEQUENCE 5' – 3' (b) |
|---|---|
| **Primers used to detect the gene conferring resistance to gentamicin** | |
| genta_probe_F | ACAGAGCCTTGGGAAGATGA |
| genta_probe_R | TGCCTTAACATTTGTGGCATT |
| **Primers used to detect the gene conferring resistance to chloramphenicol** | |
| CmR-1F | CGACGGAGAGTTAGGTTATTGGG |
| CmR-2R | GCCAGTCATTAGGCCTATCTGAC |
| **Primers used to add C-tails to fragments with the transposon** | |
| MiSeq-Himar-2 | CTTACCTATCACCTCAAATGGTTC |
| MiSeq-poliG-2 | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGGGGGGGGGGGGGGGGG |
| **Primers used for high-throughput sequencing of transposon and C-tails fragments** | |
| MiSeq-Himar-3 | AATGATACGGCGACCACCGAGATCTACACGCTGGGTTTTAATACGACTCAC |
| MiSeq-N701 | CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGGAGATG |
| **Primers used to amplify a fragment from E1162 Contig 11** | |
| E1162-C11F | GAATAAACTGCTTACATGCTAAG |
| E1162-C11R | GAATATCGGAAAAAAGCACGAG |
| **Primers used to amplify a fragment containing part of E1162 Contig 11 and the transposon** | |
| E1162-C11F | GAATAAACTGCTTACATGCTAAG |
| MiSeq-Himar-3 | AATGATACGGCGACCACCGAGATCTACACGCTGGGTTTTAATACGACTCAC |
| **Primers used for iPCR to check transposon insertion in libraries constructed** | |
| iPCR_HaeIII_F | AAACAGGAATTTATCGAAAATGGT |
| iPCR_HaeIII_R | CCCCCTGAAATCCTTACAT |
| **Primers used to add barcodes for Illumina Sequencing in the Tn-Seq strategy** | |
| PBGSF29 | TTCCCTACACGACGCTCTTCCGATCTACACNN |
| PBGSF30 | GTGTAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| **Primers used to add Adapters for Illumina Sequencing in the Tn-Seq strategy** | |
| PBGSF40 | CAAGCAGAAGACGGCATACGAGATAGACCGGGGACTTATCCAACCTG |
| PBGSF31 | AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT |
| **Primers used to introduce the gentamicin resistance marker within the flanking regions** | |
| Aus0004-1-R-XmaI | TCCC<u>CCCGGG</u>GAGAATGGTTGCAAGGACG |
| Aus0004-2-F-XhoI | CCG<u>CTCGAG</u>CTTGTATATGCGTTGGAACTTCTC |
| Aus0004-3-F-EcoRI | <u>GAATTC</u>AACTCGAACGATAATCGCAATTAAC |
| Aus0004-4-R-EcoRI | GCGATTATCGTTCGAGTT<u>GAATTC</u>CACGAGTTACGAGAATCGTAC |

| **Primers used to amplify gentamicin gene from pGPA, adding restriction sites** | |
|---|---|
| pAT392_EcoRI_lox66_ genta_F | GAGG<u>GAATTC</u>TACCGTTCGTATAGCATACATTATACGAAGTTATGATAAACCCAG CGAACCATTTGAGG |
| pAT392_EcoRI_lox71_ genta_R | CTCC<u>GAATTC</u>TACCGTTCGTATAATGTATGCTATACGAAGTTATTCAATCTTTATA AGTCCTTTTATAA |
| pAT392_AvrII_lox66_ genta_F | GAGG<u>CCTAGG</u>TACCGTTCGTATAGCATACATTATACGAAGTTATGATAAACCCA GCGAACCATTTGAGG |
| pAT392_AvrII_lox71_ genta_R | CTCC<u>CCTAGG</u>TACCGTTCGTATAATGTATGCTATACGAAGTTATTCAATCTTTATA AGTCCTTTTATAA |
| **Primers used to amplify flanking regions of *man*X, adding restriction sites** | |
| PTSmannose-1-F-XmaI | TCCC<u>CCCGGG</u>GAAGCTGCACCGCCGGGG |
| PTSmannose-2-R-XhoI | TTCCG<u>CTCGAG</u>ACATTTCAACCTTACGGATGTTTC |
| PTSmannose-3-R-EcoRI | CAAAATATCATCCATGT<u>GAATTC</u>AAGATTCCTTCAGCAAATTGTCC |
| PTSmannose-4-F-EcoRI | GCTGAAGGAATCTT<u>GAATTC</u>ACATGGATGATATTTTGAAAAAAGC |
| **Primers used to amplify flanking regions of *suf*B, adding restriction sites** | |
| Fe-Scluster-1-F-XmaI | TCCC<u>CCCGGG</u>TATACCGTCAAGTGATCTTGG |
| Fe-Scluster-2-R-XhoI | TTCCG<u>CTCGAG</u>TCCTCAACAACTGGAAATTAC |
| Fe-Scluster-3-R-EcoRI | CAATCGATTGAGTTCGA<u>GAATTC</u>ACATCATCGTGAAACCCGAATTT |
| Fe-Scluster-4-F-EcoRI | CGGGTTTCACGATGATGT<u>GAATTC</u>TCGAACTCAATCGATTGATCAG |
| **Primers used to amplify flanking regions of *vat*, adding restriction sites** | |
| Vatstrep-1-F-XmaI | TCCC<u>CCCGGG</u>TAACAGCAATGCTAGTGATCTG |
| Vatstrep-2-R-XhoI | TTCCG<u>CTCGAG</u>TAAGTGCTATGTATCGTGGAG |
| Vatstrep-3-R-EcoRI | GAATGAAAATATAAAGAC<u>GAATTC</u>GGTTAGAGGAATTAATGAATAAAT |
| Vatstrep-4-F-EcoRI | CATTAATTCCTCTAACC<u>GAATTC</u>GTCTTTATATTTTCATTCGGATA |
| **Primers used to amplify flanking regions of *mur*Q, adding restriction sites** | |
| Nacetylm-1-F-XmaI | TCCC<u>CCCGGG</u>AGATCTATTAGCAAGTAACACG |
| Nacetylm-2-R-XhoI | TTCCG<u>CTCGAG</u>ATCATTGGCACAAAGATATTAG |
| Nacetylm-3-R-EcoRI | CTTCAGCTAATTTTGCTTGAG<u>GAATTC</u>TCTTGGTTTCTGCGTTCCG |
| Nacetylm-4-F-EcoRI | CGGAACGCAGAAACCAAGA<u>GAATTC</u>CTCAAGCAAAATTAGCTGAAG |
| **Primers used to amplify flanking regions of *ltr*R, adding restriction sites** | |
| LacIfam-1-F-XmaI | TCCC<u>CCCGGG</u>ATAAGTAATCAAGAAAATATCTTC |
| LacIfam-2-R-XhoI | TTCCG<u>CTCGAG</u>CAAAAGTAAAGATAGATCATTTTTTG |
| LacIfam-3-R-EcoRI | GATTCTTTTTTCTTCA<u>GAATTC</u>TTAATACTCTGGAGACAGTTG |
| LacIfam-4-F-EcoRI | GTCTCCAGAGTATTAA<u>GAATTC</u>TGAAGAAAAAGAATCATTCTTTC |
| **Primers used to amplify flanking regions of *yyc*I, adding restriction sites** | |
| Hyphotprot-1-F-XmaI | TCCC<u>CCCGGG</u>AGCCGCTGCGTCATGGGAG |
| Hyphotprot-2-R-XhoI | TTCCG<u>CTCGAG</u>AGTTTCCCCTATATTAGCAAATTAG |
| Hyphotprot-3-R-EcoRI | CGTCGCTTTTTTAGGGTT<u>GAATTC</u>GCAATACGATCGTGACAAACAA |

| | |
|---|---|
| Hyphotprot-4-F-EcoRI | GTCACGATCGTATTGC<u>GAATTC</u>AACCCTAAAAAAGCGACGAAAAAT |
| **Primers used to amplify flanking regions of *adh*E, adding restriction sites** | |
| BifuncOH-1-F-XmaI | TCCC<u>CCCGGG</u>CAATCAAAACGGTCAAACCAG |
| BifuncOH-2-R-XhoI | TTCCG<u>CTCGAG</u>GCGAGGATTTTTTGTCATCTG |
| BifuncOH-3-R-EcoRI | GTTTCAATTCAGAGATCAATG<u>GAATTC</u>GCTTGTACGTCAATTGTCTTC |
| BifuncOH-4-F-EcoRI | GACAATTGACGTACAAGC<u>GAATTC</u>CATTGATCTCTGAATTGAAACAG |
| **Primers used to amplify flanking regions of *dsb*A, adding restriction sites** | |
| DsbAfam-1-F-XmaI | TCCC<u>CCCGGG</u>TTTATAACAGAAGAAGCAGAAGC |
| DsbAfam-2-R-XhoI | TTCCG<u>CTCGAG</u>GGTTAGGGAAAAAATTGAAGCG |
| DsbAfam-3-R-AvrII | CTTCTTTCTGTTCTGTCCA<u>CCTAGG</u>TCGAAACAGTATTCCCCTAAAG |
| DsbAfam-4-F-AvrII | GGGGAATACTGTTTCGA<u>CCTAGG</u>TGGACAGAACAGAAAGAAGAAAAC |
| **Primers used to amplify flanking regions of *pgt*, adding restriction sites** | |
| Glyctrans-1-F-XmaI | TCCC<u>CCCGGG</u>GTGCTCGCGTAATGGAACC |
| Glyctrans-2-R-XhoI | TTCCG<u>CTCGAG</u>CAAGTACTTGAAAATATCAAGTC |
| Glyctrans-3-R-EcoRI | CCTTGTTACAATGAGGA<u>GAATTC</u>ATGAATTGAATCTATATCATGTTTA |
| Glyctrans-4-F-EcoRI | CATGATATAGATTCAATTCAT<u>GAATTC</u>TCCTCATTGTAACAAGGAATC |
| **Primers used to amplify flanking regions of *pst*B2, adding restriction sites** | |
| PhosABCt-1-F-XmaI | TCCC<u>CCCGGG</u>TCACATCATCCCAAACAGACG |
| PhosABCt-2-R-XhoI | TTCCG<u>CTCGAG</u>AAGCGGTCCATTACACACAAC |
| PhosABCt-3-R-EcoRI | GAAGTAGCACTATATA<u>GAATTC</u>AACTAACAAGATATTTACCAGAC |
| PhosABCt-4-F-EcoRI | GTAAATATCTTGTTAGTT<u>GAATTC</u>TATATAGTGCTACTTCTTTCTC |
| **Primers used to amplify flanking regions of 4 proteins from Fe-S cluster, adding restriction sites** | |
| Fe-SC-1-F-XmaI | TCCC<u>CCCGGG</u>GACAAAATTTCTGTTATGGCATT |
| Fe-SC-2-R-XhoI | TTCCG<u>CTCGAG</u>TCCTCAACAACTGGAAATTAC |
| Fe-SC-3-R-EcoRI | GATCAATCGATTGAGTTCGA<u>GAATTC</u>GAAAAATCATTTACCGCGTC |
| Fe-SC-4-F-EcoRI | GACGCGGTAAATGATTTTTC<u>GAATTC</u>TCGAACTCAATCGATTGATCAG |
| **Primers used to clone manX in the pNZ44 plasmid** | |
| PTS-Fw-NcoI | CCACGT<u>CCATGG</u>ATGGTAGGGATTATCCTAGC |
| PTS-Rev-XbaI | CGACGT<u>TCTAGA</u>TTACGCTTTTGCTAATTCAGC |
| **Primers used to clone *mur*Q in the pNZ44 plasmid** | |
| N-acetyl-Fw-XbaI | TTACGT<u>TCTAGA</u>ATGAGCAAAATCAATTTAGAGAACC |
| N-acetyl-Rev-HindIII | ATTCAA<u>AAGCTT</u>TCATAAGGTCCCTTTTACAAATCCG |
| **Primers used to amplify 16rRNA gene** | |
| F27 | ACGAAGCATCAGAGTTTGATCMTGGCTCAG |
| R1492 | CGGTTACCTTGTTACGACTT |

(a) Enzymes for cloning are indicated in the name of the primer

(b) Restrictions sites sequences for the enzymes required for cloning are underline

**Table 18. i7 Adapters used to sequence library samples.**

| Name | i7 Index Adapters | Primer Name (a) | Sequence (5'-3') (b) |
|------|-------------------|-----------------|----------------------|
| A | ATCACG | PBGSF29 ATCACG | TTCCCTACACGACGCTCTTCCGATCT<u>ATCACG</u>NN |
|   |        | PBGSF30 ATCACG | P-<u>CGTGAT</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| B | CGATGT | PBGSF29 CGATGT | TTCCCTACACGACGCTCTTCCGATCT<u>CGATGT</u>NN |
|   |        | PBGSF30 CGATGT | P-<u>ACATCG</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| C | TTAGGC | PBGSF29 TTAGGC | TTCCCTACACGACGCTCTTCCGATCT<u>TTAGGC</u>NN |
|   |        | PBGSF30 TTAGGC | P-<u>GCCTAA</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| D | TGACCA | PBGSF29 TGACCA | TTCCCTACACGACGCTCTTCCGATCT<u>TGACCA</u>NN |
|   |        | PBGSF30 TGACCA | P-<u>TGGTCA</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| E | ACAGTG | PBGSF29 ACAGTG | TTCCCTACACGACGCTCTTCCGATCT<u>ACAGTG</u>NN |
|   |        | PBGSF30 ACAGTG | P-<u>CACTGT</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| F | GCCAAT | PBGSF29 GCCAAT | TTCCCTACACGACGCTCTTCCGATCT<u>GCCAAT</u>NN |
|   |        | PBGSF30 GCCAAT | P-<u>ATTGGC</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| G | CAGATC | PBGSF29 CAGATC | TTCCCTACACGACGCTCTTCCGATCT<u>CAGATC</u>NN |
|   |        | PBGSF30 CAGATC | P-<u>GATCTG</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| H | ACTTGA | PBGSF29 ACTTGA | TTCCCTACACGACGCTCTTCCGATCT<u>ACTTGA</u>NN |
|   |        | PBGSF30 ACTTGA | P-<u>TCAAGT</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| I | GATCAG | PBGSF29 GATCAG | TTCCCTACACGACGCTCTTCCGATCT<u>GATCAG</u>NN |
|   |        | PBGSF30 GATCAG | P-<u>CTGATC</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| K | GGCTAC | PBGSF29 GGCTAC | TTCCCTACACGACGCTCTTCCGATCT<u>GGCTAC</u>NN |
|   |        | PBGSF30 GGCTAC | P-<u>GTAGCC</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| L | CTTGTA | PBGSF29 CTTGTA | TTCCCTACACGACGCTCTTCCGATCT<u>CTTGTA</u>NN |
|   |        | PBGSF30 CTTGTA | P-<u>TACAAG</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| M | ACTGAT | PBGSF29 ACTGAT | TTCCCTACACGACGCTCTTCCGATCT<u>ACTGAT</u>NN |
|   |        | PBGSF30 ACTGAT | P-<u>ATCAGT</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| N | ATGAGC | PBGSF29 ATGAGC | TTCCCTACACGACGCTCTTCCGATCT<u>ATGAGC</u>NN |
|   |        | PBGSF30 ATGAGC | P-<u>GCTCAT</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| O | ATTCCT | PBGSF29 ATTCCT | TTCCCTACACGACGCTCTTCCGATCTATTCCTNN |
|   |        | PBGSF30 ATTCCT | P-AGGAATAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| P | CAAAAG | PBGSF29 CAAAAG | TTCCCTACACGACGCTCTTCCGATCT<u>CAAAAG</u>NN |
|   |        | PBGSF30 CAAAAG | P-<u>CTTTTG</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| Q | CAACTA | PBGSF29 CAACTA | TTCCCTACACGACGCTCTTCCGATCT<u>CAACTA</u>NN |
|   |        | PBGSF30 CAACTA | P-<u>TAGTTG</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| R | CACCGG | PBGSF29 CACCGG | TTCCCTACACGACGCTCTTCCGATCT<u>CACCGG</u>NN |
|   |        | PBGSF30 CACCGG | P-<u>CCGGTG</u>AGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |

| S | CACGAT | PBGSF29 CACGAT | TTCCCTACACGACGCTCTTCCGATCTCACGATNN |
| | | PBGSF30 CACGAT | P-ATCGTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| T | CACTCA | PBGSF29 CACTCA | TTCCCTACACGACGCTCTTCCGATCTCACTCANN |
| | | PBGSF30 CACTCA | P-TGAGTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |
| U | CAGGCG | PBGSF29 CAGGCG | TTCCCTACACGACGCTCTTCCGATCTCAGGCGNN |
| | | PBGSF30 CAGGCG | P-CGCCTGAGATCGGAAGAGCGTCGTGTAGGGAAAGAGT-P |

(a) Index adapters are indicated in the name of the primer. There are two pairs of primers for each index adapter corresponding to primer forward and reverse.

(b) Index adapters are underlined.

# IV.  RESULTS

# 1. IDENTIFICATION AND CHARACTERIZATION OF GENES ENCODED BY VRE FOR INTESTINAL COLONIZATION

## 1.1 Mutant library construction

To identify genes necessary for gut colonization, we followed a strategy previously used to identify genes that are essential for *E. faecium* to grow in the human serum and in the presence of ampicillin[153]. This strategy involves the generation of random mutants inserting a single transposon in the genome of the pathogen to have a high-density transposon mutant library, followed by Microarray-based Transposon Mapping (M-Tram). To perform this study we used the methodology to construct the mutant transposon library. However, to detect the genes involved in gut colonization we used a high-throughput sequencing approach instead of microarray hybridization. We adapted the protocol from Klein et al.[154]. In this study, essential genes of the periodontal pathogen *Porphyromonas gingivalis* were identified using high-throughput sequencing technology.

To construct the mutant library, we used the pZXL5 plasmid, containing a ColE1 Gram-positive thermo-sensitive replicon, a mariner transposase induced by nisin, a gentamicin resistance *mariner* transposon and a chloramphenicol resistance (*cat)* gene (Material and Methods, section 11, Figure 13). Using this strategy, the transposon is integrated in one site of the genome and since the bacteria is grown at the restrictive temperature for plasmid replication, the plasmid is lost.

The *E. faecium* strain selected to construct the library was the clinical E1162 strain, as previously used in the study performed by Zhang et al[153]. This strains was isolated from a bloodstream infection in France in 1996[60] and the genome is deposited in GenBank (GenBank Accesion Number ABQJ00000000).

The pZXL5 plasmid was electrotransformed into E1162 strain and transformants were selected by plating in gentamicin plates. One of the gentamicin-resistant transformants was picked to follow the mutant library construction protocol (Materials and methods, section 11) adding 300 µg/ml of gentamicin and 4 µg/ml of chloramphenicol for plasmid replication in the first step of the protocol (Material and Methods, section 11, Figure 14A, 1st method). Once the library was generated, it was aliquoted and maintained at -80 ºC with 50 % glycerol.

To evaluate the randomness, coverage of transposition and plasmid loss in the library, one aliquot was diluted and plated in BHI plates with 4 µg/ml of chloramphenicol and also on BHI plates containing 300 µg/ml of gentamicin in order to analyze separated colonies. We expected to have a 10 % of plasmid presence or even less (i.e. colonies resistant to chloramphenicol), according to results from Zhang et al[153]. Nevertheless, we obtained an unexpected high number of colonies resistant to chloramphenicol (theoretically conferred by the plasmid if it has not been lost). Indeed, the number of gentamicin (contain the transposon) and chloramphenicol resistant colonies (contain the plasmid) was basically the same. We repeated the protocol with other transformants and we observed similar results for most of them (Figure 31). However, for one of the

generated libraries, we were able to obtain a smaller number of chloramphenicol resistant colonies, a chloramphenicol resistance proportion around 1/1000.

**E1162 libraries**
**Gm$^r$ vs Cm$^r$ CFUs**
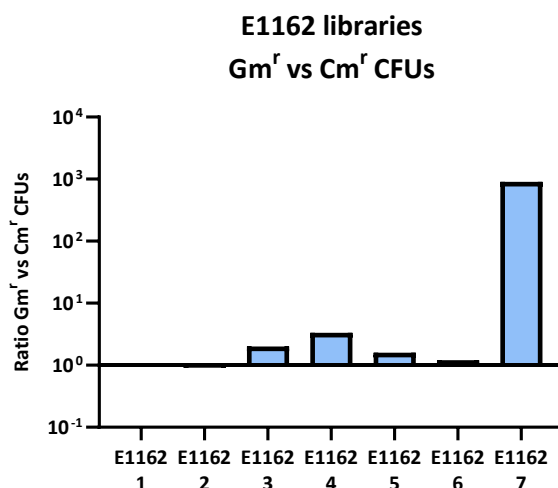


**Figure 31.  Ratio of gentamicin and chloramphenicol resistant colonies from E1162 transposon libraries generated.** Representation of the ratio between the numbers of CFUs detected in BHI plates containing gentamicin (Gm$^r$ ) or chloramphenicol (Cm$^r$) after plating stocks from E1162 transposon libraries.

Besides the original protocol, we repeated the protocol modifying some of the conditions at different steps of the protocol to test if we could get an improvement in plasmid loss and more reproducibility in the generated libraries. These modifications consisted on:

- Inoculating different culture volumes and number of CFUs between the different steps (see steps in Material and Methods, section 11)
- Removing the antibiotics from the first step in order to decrease the number of resistant clones to chloramphenicol
- Starting the protocol directly at step 2, in which gentamicin is directily added, to avoid the chloramphenicol resistance and directly allow transposon insertion and plasmid loss at the restrictive temperature (37 ºC).
- Increasing the restrictive temperature to 42 ºC to enhance plasmid loss.

Despite the changes made, none of the protocol modifications increased the loss of the plasmid, neither the reproducibility (not shown). Indeed, the library with the lower number of chloramphenicol resistant colonies obtained was library 7 (Figure 31). Thus we decided to continue with this library and verify several aspects of the library as specified below.

### 1.2 Verification of E1162 mutant library

### 1.2.1 Plasmid presence

We checked the presence of the plasmid from the individual colonies plated from library stocks at -80 ºC. We picked 20 colonies from gentamicin BHI plates (colonies 1 to 20) and 10 colonies from chloramphenicol BHI plates (colonies 21 to 30) to perform PCRs (Material and Methods, genta_probe_F/R and CmR-1F/CmR-2R primers in section 33, Table 17) of these 30 colonies in total. The PCR results from the colonies picked from gentamicin plates showed a positive result for the PCR using primers specific for the gene conferring resistance to gentamicin but not for the gene conferring resistance to chloramphenicol, suggesting that those colonies had lost the plasmid that contains the chloramphenicol resistance gene and that the transposon was inserted in the genome of each bacterium (Figure 32A-B). Chloramphenicol resistant colonies gave positive results for both PCRs (Figure 32C-D), suggesting that those 10 colonies still contain the plasmid. Note that the amplification product obtained for both PCRs had exactly the same size as the amplification products obtained with a positive control (i.e. plasmid pZXL5, used as template). As expected no amplification products were detected in the negative control (water) and when a colony derived from the original strain E1162 was used as a template.

A

Gentamicin resistant colonies
PCR to detect gentamicin resistance gene



B

Gentamicin resistant colonies
PCR to detect chloramphenicol resistance gene



Chloramphenicol resistant colonies

C

PCR to detect gentamicin resistance gene

D

PCR to detect chloramphenicol resistance gene
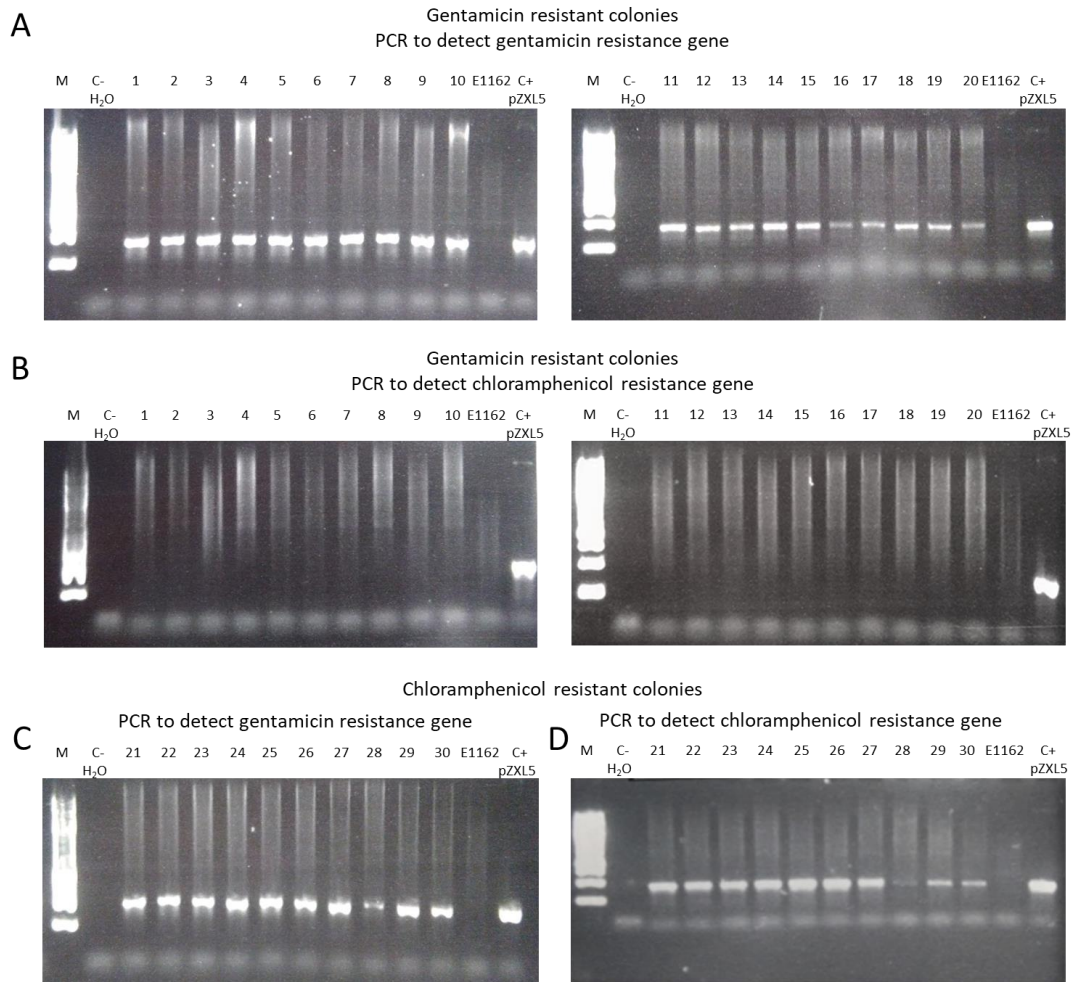


**Figure 32. Evaluation of the chloramphenicol resistance gene and gentamicin resistance gene presence in the E1162 transposon mutant library.** PCR from 30 randomly selected colonies (20 from gentamicin BHI plates **(A-B)** a 10 from chloramphenicol BHI plates **(C-D)**) derived from E1162 library. PCR of the 20 colonies isolated from plates containing gentamicin (Gm$^r$) using specific primers for the gene conferring resistance to gentamicin **(A)** or the gene conferring resistance to chloramphenicol **(B)**. **(C-D)** PCR of the 10 colonies isolated from plates containing chloramphenicol (Cm$^r$) using specific primers for the gene conferring resistance to gentamicin **(C)** and the gene conferring resistance to chloramphenicol **(D)**. In addition, a colony from a culture of the strain E1162 was used as a control (theoretically no amplification should be detected), water instead of a resuspended colony was also used as a negative control (C-), and the plasmid pZXL5 was used as template as a positive control (C+).

Therefore, the PCR results obtained suggested that this library had less than 10 % of colonies containing the plasmid as we did not detect the gene conferring resistance to chloramphenicol in the 20 colonies picked from plates containing gentamicin. Therefore, we decided to continue checking the random insertion of the transposon to verify if the library was optimal for our studies.

**1.2.2 Checking the random transposon insertion in the genome of the E1162 obtained mutant library**

To confirm that the transposon was randomly inserted in the VRE genome in the generated E1162 library, we extracted DNA from an aliquot of the library and performed two rounds of PCRs (Material and Methods, section 5) to amplify fragments that contained the flanking regions of the transposon (i.e. the genes where the transposon was inserted). Subsequently we cloned this 150-500 pb fragments into the TOPO vector (Material and Methods, section 8) to obtain sequences of several clones to verify that the transposon was inserted in different locations in the genome. To sequence the fragments cloned in TOPO we used MiSeq-Himar-3 primer (Material and Methods, section 33, Table 17), which hybridises with the transposon sequence. Sanger sequencing was performed and the seven sequences achieved were aligned to the E1162 genome. The results obtained showed that the transposon was inserted in the same point of the E1162 genome in all the colonies tested, in the nucleotide 13.528 of the 11th Contig (Table 19). The size of the fragments was different but all shared the same insertion site on the genome, indicating that the insertions did not occurred randomly, as expected.

**Table 19. Verification of the transposon insertion point in the genome of E1162.** Fragments amplified contain the end of the transposon and the sequence of the genome where the transposon has been inserted. The obtained sequence was aligned with E1162 genome.

|  | Size of the genome sequenced (bp) | Contig | Specific genome point (bp) |
|---|---|---|---|
| **1** | 147 | Contig 11 | 13.392 – 13.538 |
| **2** | 41 | Contig 11 | 13.498 – 13.538 |
| **3** | 145 | Contig 11 | 13.394 – 13.538 |
| **4** | 110 | Contig 11 | 13.429 – 13.538 |
| **5** | 74 | Contig 11 | 13.465 – 13.538 |
| **6** | 145 | Contig 11 | 13.394 – 13.538 |
| **7** | 438 | Contig 11 | 13.538 – 13.101 |

In order to verify the insertion of the transposon in a larger number of colonies, an aliquot was grown from library stocks to obtain single colonies. Taking into account the results shown in Table 19, we designed primers (Material and Methods, C11F and C11R primers in section 33, Table 17) to amplify the fragment where the transposon was found to be inserted in Contig 11, using a primer specific for the transposon and a primer specific for the point in the genome where the transposon was found to be inserted (Material and Methods, C11F and MiSeq-Himar-3 primers in section 33, Table 17; Figure 33A). In this way, if we obtained a PCR fragment from these colonies with the designed primers, we could verify that the transposon had been inserted in that specific point in the Contig 11 of the E1162 genome. As a negative control, we used the wild type strain E1162. PCR from twenty colonies isolated from the library was performed using the designed primers. The results showed that a PCR amplification product was obtained

when the combination of transposon and Contig 11 primers were used (Figure 33B), confirming that the transposon was inserted at the same point in the genome in all tested colonies.  When we used both primers that hybridize specifically in Contig 11 we only obtained a PCR amplification for the control sample, DNA from E1162 strain (Figure 33C). The fact that we did not obtain an amplification product of a bigger size with the colonies derived from the library was probably due to the short elongation time used that only allowed the amplification of the product in the strain that did not contain the transposon.
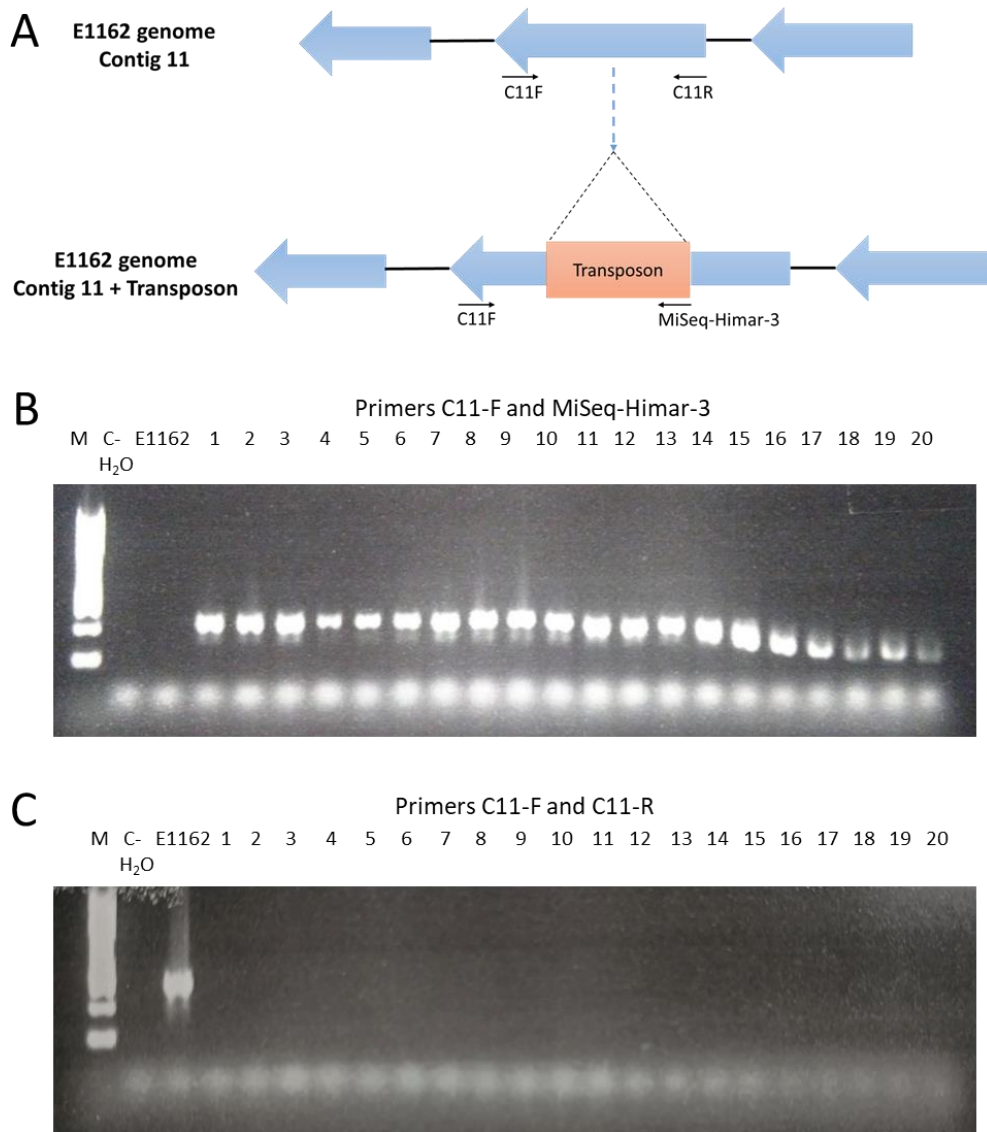


**Figure 33. Evaluation of the transposon insertion in the transposon mutant library of E1162 strain.** 20 colonies were checked by PCR with specific primers of the insertion point in the genome. **(A)** Schematic representation of the combination of primers used to detect insertion of the transposon in a specific point of the genome (i.e. contig 11, position 13.528). The figures show both cases: top – the transposon was inserted in another part of the genome, bottom - the transposon was inserted in that particular point of the genome. **(B)** PCR result using primers hybridizing in Contig 11 and the end of the transposon (C11-F, MiSeq-Himar-3). **(C)** PCR result using both primers from Contig 11 (C11-F, C11-R). A colony from a culture of the strain E1162 was used as a control, water instead of a resuspended colony was also used as a negative control (C-).

IV. Results

Thus, the results obtained confirmed that the mutant transposon library constructed was not optimal to continue with the study because the transposon was inserted in the same point in the genome, at least in the tested colonies.

**1.3 Selection of *E. faecium* Aus0004 strain as the clinical strain to construct the transposon mutant library**

Despite the changes and multiple attempts made to construct an optimal library, we could not obtain a library with low prevalence of colonies maintaining the plasmid and with a transposon inserted randomly. To solve these problems, the construction of the library was carried out in the laboratory where the technique was developed and published, in Utrecht (the Netherlands), under the supervision of Dr. Willem Van Schaik.

To create the mutant transposon library and to detect the genes where the transposon is inserted by massive sequencing, we used a modified pZXL5 plasmid, known as pGPA plasmid[99]. This plasmid was designed to construct a high-density transposon mutant library coupled with high-throughput sequencing to analyse the insertion sites of the transposon in the genome. The pZXL5 plasmid was modified for being suitable for Tn-seq by adding MmeI restriction sites in the inverted repeats on both sides of the transposon cassette T7, carrying a gentamicin resistance gene (Material and Methods, section 11).

We first attempt to obtain an appropriate transposon mutant library by repeating the protocol that we described above but plating the culture after each overnight step to evaluate the resistant colonies to chloramphenicol and gentamicin. The pGPA plasmid was electroporated into the E1162 strain and gentamicin resistant transformants were selected. We followed the protocol explained previously with different gentamicin transformants, adding one extra night at 37 ºC at the end of the protocol with gentamicin to force the selections of the mutants and the plasmid loss (Material and Methods, section 11, Figure 14B, 2nd method). The results obtained were similar to the results obtained in the first attempts, having a final proportion of 40 % of colonies resistant to chloramphenicol or higher (results not shown).

As we had similar results, we decided to increase the chloramphenicol concentration to 6 µg/ml instead of 4 µg/ml for the first step of plasmid replication at 28 ºC. Moreover, we decided to use new clinical *E. faecium* strains to construct the library, in order to check if the protocol could work better with other strains. We checked for clinical *E. faecium* strains sensitive to both antibiotics present in the pGPA plasmid, chloramphenicol and gentamicin, and we selected the E1133 and Aus0004 strains. The Aus0004 strain was isolated from the bloodstream of a patient in Australia in 1998 (GenBank Accession NC_017022, 2017). The E1133 strain was isolated from patient faeces during a VRE outbreak in 2002 in USA (GenBank Accession NZ_AHWR00000000, 2019). These clinical strains are resistant to several antibiotics such as ampicillin, aminoglycosides and vancomycin, being commonly known as VRE (Vancomycin-Resistant *Enterococcus*).

We repeated the protocol with the new clinical *E. faecium* strains selected and with the previous E1162 strain, constructing two libraries for each clinical strain. After counting resistant colonies to both antibiotics, we had a mutant library of E1162 strain with approximately 10 % of Cm$^r$ colonies and a mutant library with around 0.00002 % of Cm$^r$ colonies using Aus0004 strain (Figure 34). Results obtained with the E1133 strain were similar to those obtained for E1162 but we decided to continue with E1162 transposon library because this strain had been previously used by the group of Dr. van Schaik to construct a mutant library. In addition, we also continue with the characterization of the library obtained in the Aus0004 strain.
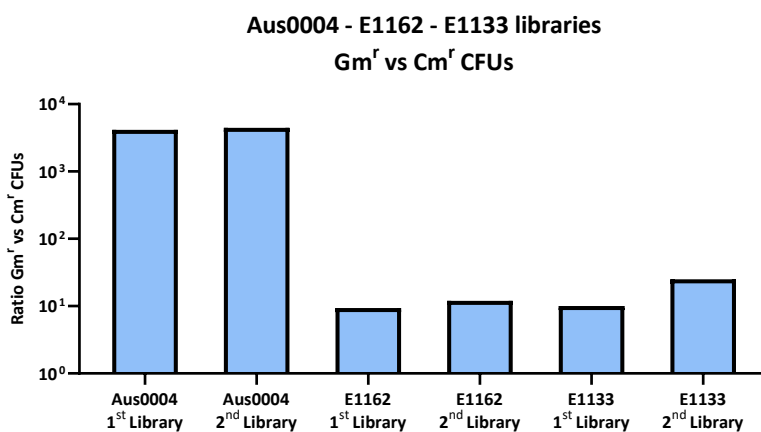


**Figure 34. Ratio of the gentamicin vs chloramphenicol resistant CFUs in the transposon mutant libraries constructed with different *E. faecium* strains.** Representation of the ratio between the number of Gm$^r$ and Cm$^r$ colonies in the two libraries constructed with the clinical *E. faecium* strains Aus0004, E1162 and E1133.

Afterwards, we evaluated the transposon insertion of one of the libraries from each strain. The randomness of the transposon insertion was determined by randomly isolating 20 Gm$^r$ colonies, extracting and digesting DNA with HaeIII and finally carrying out an inverse PCR (iPCR) to amplify the part of the genome where the transposon was inserted (Material and Methods, section 11, Figure 15). The PCR products were sequenced to confirm the position in the genome where the transposon was inserted. The results suggest that each colony tested carried a single transposon inserted (one brighter PCR product) (Figure 35A and C). For some colonies we did not obtain a PCR product due to lower DNA levels for some of the samples. Although we observed for some colonies two different PCR fragments we clearly visualized one brighter PCR product, which corresponds to the transposon. The lighter band may be the product of non-specific amplification. Moreover, the electrophoresis of the PCR products suggested that the transposon was randomly distributed along the genome on both strains, the bands appeared at different positions in the gel.

The PCRs were sequenced through the Sanger technique to analyse where the transposon had been inserted in the genome. Finally, we obtained 16 sequences from E1162 library and 14 from Aus0004 library from the 20 initially picked colonies due to

lower quality sequences for some of the sequenced colonies.  In the case of the strain E1162 (Figure 35B), we detected the presence of the plasmid in two of the isolates analysed and the transposon was inserted in the same point of the genome in two of the isolates tested. The pGPA plasmid was not detected in the Aus0004 colonies analysed and, as in the E1162 library, the transposon was inserted in the same point of the genome in only two of the isolates (Figure 35D). Both libraries were stocked at -80 ºC with glycerol.

In addition to the iPCR and the sequencing of fragments by Sanger, we performed a PCR of the chloramphenicol resistance gene (Material and Methods, section 33, primers in Table 17) to check the presence of the Cm$^r$ colonies in those 20 Gm$^r$ selected colonies. By doing this PCR we corroborated the results previously obtained by sequencing (i.e. a PCR product of the gene conferring Cm$^r$ could only be detected in two colonies derived from the E1162 library).

| Sample | Location in E1162 genome (bp) |
|---|---|
| 1 | Contig7, 6-636 |
| 3 | Plasmid, Cm$^r$ colony |
| 6 | Contig22, 1 398-2 562 |
| 7 | Contig46, 7 871-8 695 |
| 8 | Contig8, 2-225 |
| 9 | Contig8, 32 488-33 603 |
| 11 | Contig48, 36 616-37 742 |
| 12 | Plasmid, Cm$^r$ colony |
| 13 | Contig53, 11 669-12 382 |
| 14 | Contig53, 11 765-12 382 |
| 15 | Contig52, 22 424-23 443 |
| 16 | Contig41,109 662-110 310 |
| 17 | Contig136, 3 023-4 028 |
| 18 | Contig11, 24 319-25 336 |
| 19 | Contig102, 19 047-19 270 |
| 20 | Contig121, 6 425-7 221 |

| Sample | Location in Aus0004 genome (bp) |
|---|---|
| 1 | 893 853-894 924 |
| 3 | 2 856 879-2 857 037 |
| 4 | 70 560-71 628 |
| 5 | 2 535 251-2 536 292 |
| 8 | 901 125-901 959 |
| 9 | 242 159-243 025 |
| 10 | 681 705-682 761 |
| 11 | 919 605-920 734 |
| 12 | 438 079-438 992 |
| 13 | 1 618 870-1 619 659 |
| 14 | 438 079-438 941 |
| 15 | 2 455 609-2 455 884 |
| 16 | 852 079-853 196 |
| 17 | 683 255-684 304 |

**Figure 35. Evaluation of the E1162 and the Aus0004 transposon mutant libraries. (A-C)** Inverse PCR (iPCR) of randomly selected *E. faecium* transposon insertion mutants from the two different libraries. **(B-D)** Sequencing analysis of the iPCRs obtained from randomly mutants. The numbers represent the base pairs (bp) where the transposon is inserted. Sequences that aligned with the pZXL5 plasmid and were positive for the PCR against the chloramphenicol resistance gene are indicated in red. Colonies with the transposon inserted at the same point of the genome are coloured in blue.

Therefore, as Aus0004 mutant library had less Cm$^r$ colonies (around 1/1000) we decided to continue the study using the transposon mutant library of the strain Aus0004, which is resistant to the most common antibiotics used to treat *E. faecium* infections (i.e. ampicillin and vancomycin), also known as VRE for vancomycin-resistant-*Enterococcus*.

### 1.4 Setting-up a mouse model to identify genes required for VRE gut colonization

### 1.4.1 Clindamycin and vancomycin promote higher levels of VRE colonization

Once we constructed the Aus0004 mutant library, we set up a mouse model mimicking the VRE infections develop in patients, which usually begin by colonizing the intestinal tract. Under normal conditions, the intestine is colonized by hundreds of bacterial species, the microbiota, preventing the colonization by VRE, what is known as "Colonization Resistance" (CR)[114]. However, the antibiotics alter the microbiota composition allowing VRE to colonize the gut. Once established, the bacterium is able to reach very high levels in the gut.

Thus, different antibiotic treatments were tested to disrupt the microbiota. Subsequently, mice were orally inoculated with VRE to check the levels of the pathogen in the gut. Mice were treated with antibiotics of different spectrum of action against (1) Gram-positive bacteria (vancomycin), (2) Gram-positive and anaerobes (clindamycin), (3) Gram-positive and negative anaerobes (metronidazole) (4) a broad-spectrum (ampicillin, amoxicillin, amoxicillin with clavulanic acid and a cocktail containing vancomycin, ampicillin and streptomycin) (Material and Methods, section 28.1, Figure 23).

Antibiotic treatment was provided during seven days to mice and the fourth day mice were infected orally with $10^8$ CFUs/200 µl of VRE (wild type Aus0004 strain). Stool samples were collected three days post-VRE inoculation and plated in selective BEA medium adding vancomycin and ampicillin (BEA+AV). After two days of incubation at 37 ºC, CFUs were counted and the levels of VRE were calculated per 10 mg of faeces.

In Figure 36 it is shown that the highest and homogeneous levels of VRE colonization were reached with clindamycin and vancomycin treatments. Besides that, we checked that control mice, which were not treated with antibiotics, were not colonized with VRE (Figure 36), probably due to the fact that an undisturbed microbiota prevents VRE colonization. The rest of antibiotics promote to a different extend the intestinal colonization by VRE, probably because they produce different alterations in the microbiome composition, as shown previously by our research group[168] (Introduction, section 2.2.3, Figure 8). Notably, the cocktail containing three antibiotics did not allow VRE colonization. This could have been the cause of a synergistic effect of antibiotics. We notice *in vitro* that although the VRE strain used is resistant to the three antibiotics individually, the strain grows poorly when the 3 antibiotics are included in the same media.

Therefore, we next focus on the two antibiotic treatments that allow a higher VRE growth (i.e. vancomycin and clindamycin) and checked the levels of colonization in other parts of the intestine after the inoculation of the transposon mutant library.
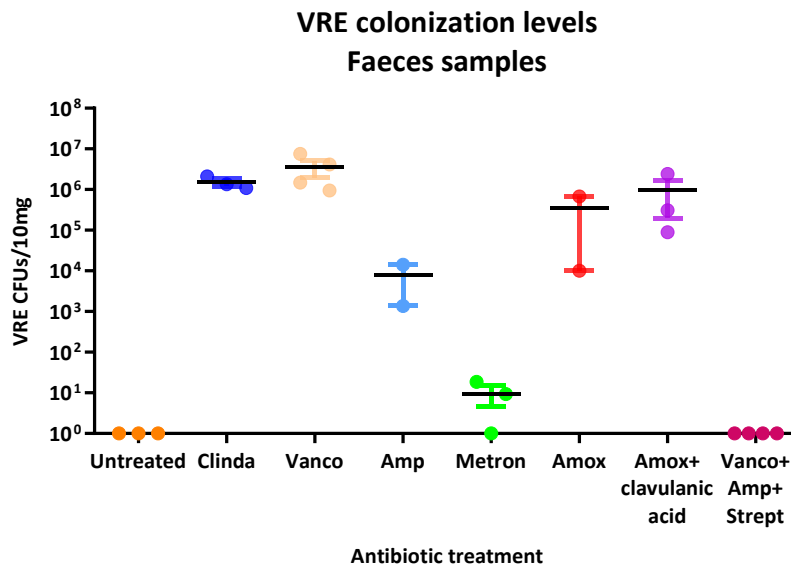


**Figure 36. Antibiotic treatments promote to a different extend VRE intestinal colonization.** Mice were treated during seven days with different antibiotic treatments and inoculated at day four with VRE. Three days post-VRE inoculation, faeces samples were collected and plated in specific medium for VRE detection in order to evaluate the levels of colonization. Antibiotics were administered during the 7 days that the experiment last. As control, a group of mice did not receive any treatment. N=2-4 mice per group. No treatment (Untreated), Clindamycin (Clinda), Ampicillin (Amp), Vancomycin (Vanco), Metronidazole (Metron), Amoxicillin (Amox), Amoxicillin with clavulanic acid (Amox + clavulanic acid), Vancomicin with ampicillin and streptomycin (Vanco + Amp + Strept). Lines represent the mean and whiskers represent the SEM.

### 1.4.2   Clindamycin induce higher and more homogeneous VRE intestinal colonization levels as compare to vancomycin

Once we had constructed and checked the mutant library, we performed another *in vivo* experiment using the mouse model previously explained in Figure 24 (Material and Methods, section 28.2) and the antibiotics vancomycin and clindamycin. In this case, instead of using the WT strain, we used the Aus0004 mutant library obtained. In addition, besides collecting faeces we also collected different parts of the intestinal tract. Thus, mice were infected with $10^8$ CFUs/200 µl at the fourth day of the treatment. At day seven, samples from different parts of the intestine were collected, including ileum, caecum, colon and the intestinal walls of each part. Samples were obtained to plate and check the levels of colonization and for further sequencing by Illumina to analyse the genes involved in gut colonization.

The levels of colonization of the transposon mutant library were different along the intestinal tract, being higher and more homogeneous in the large intestine and faeces. The levels decreased in the small intestine and in the intestinal walls. Two experiments

were performed, with six mice per experiment. The results obtained in both experiments were similar, which confirmed the reproducibility of the mouse model used (Figure 37).
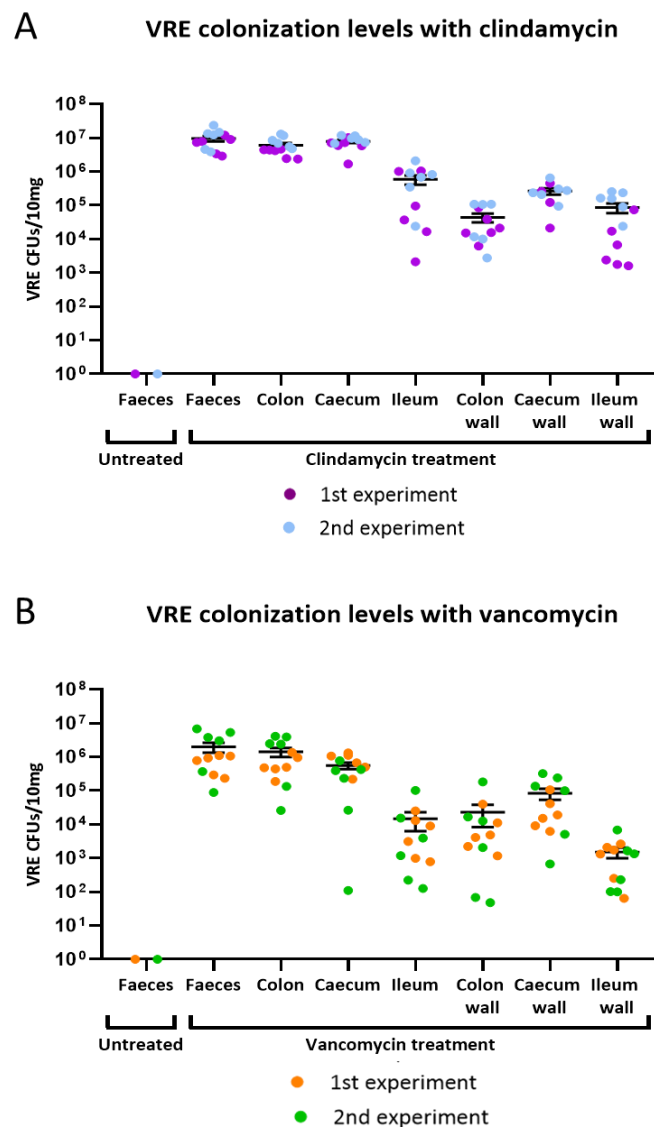


**Figure 37. VRE colonization levels in different parts of the intestine.** The experiment performed in duplicate following the protocol of Figure 24 (Section 28.2, Material and Methods) (N=6 mice per experiment and antibiotic treatment). Briefly, mice were treated with either clindamycin or vancomycin for seven days and inoculated with VRE library the day four. Levels of colonization were analysed three days post-VRE inoculation by plating the samples in selective media (see methods). **(A)** VRE levels reached using clindamycin treatment. **(B)** VRE levels reached using vancomycin. Lines represent the mean and whiskers represent the SEM.

VRE was able to colonize the gut reaching high levels with both antibiotics treatments but the levels were higher and more homogeneous (Table 20) when clindamycin was supplied to the animals, especially in the large intestine. As the highest and more homogeneous levels of VRE colonization were reached in the large intestine after clindamycin treatment, samples from caecum and colon from clindamycin treated mice

were selected for further sequencing and analysis. Although similar levels were obtained in faeces, we did not select these samples due to the limitation in the amount of the sample available. In addition, we did not analysed samples from the small intestine or intestinal wall due to the low level of VRE obtained.

Table 20. Comparison between the levels of VRE quantified in the same types of samples obtained from mice treated with either vancomycin or clindamycin.

| Vancomycin vs Clindamycin samples | Average of VRE CFUs/10mg | | Differences between treatments | Variance of VRE CFUs/10mg | | Differences between treatments |
|---|---|---|---|---|---|---|
| | Vancomycin | Clindamycin | Wilxocon test | Vancomycin | Clindamycin | F test to compare variances |
| | | | p-value | | | p-value |
| Faeces | $1 * 10^6$ | $8.43 * 10^6$ | *** | 0,08 | 0,34 | * |
| Colon | $7.33 * 10^5$ | $4.75 * 10^6$ | **** | 0,05 | 0,44 | ** |
| Caecum | $4.67 * 10^5$ | $7.34 * 10^6$ | **** | 0,05 | 1,29 | **** |
| Illeum | $3.56 * 10^3$ | $5.21 * 10^5$ | *** | 0,89 | 0,75 | ns |
| Colon wall | $4.53 * 10^3$ | $1.86 * 10^4$ | * | 0,29 | 1,07 | * |
| Caecum wall | $3.03 * 10^4$ | $2.51 * 10^5$ | ** | 0,17 | 0,64 | ns |
| Illeum wall | $1.34 * 10^3$ | $4.92 * 10^4$ | *** | 0,72 | 0,44 | ns |

* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$, ns:not significant $p>0.05$. Variance was obtained using the log10 VRE CFUs value.

## 1.5 Sequencing sample preparation

Samples from *in vivo* (clindamycin treated mice; N=12 from colon and ceacum, 6 from each experiment) and *in vitro* conditions (mutant library used to inoculate mice; 5 replicates per aliquot uses in each of the 2 experiments, N=10) were prepared for Illumina sequencing as recommend by Zhang et al.[99] (Material and Methods, section 12.2, Figure 17) in order to identify the genes that are essential for gut colonization.

Finally, after reproducing the prototocol with the repetition and modificacions explained below, we lost part of the DNA for some of the *in vivo* samples and we reduced the number of samples to be sequenced. Thus, we sequenced 11 samples from colon (5 from experiment one and 6 from experiment two) and 9 samples from caecum (5 from experiment one and 4 from experiment two) with the N=10 stored aliquot samples from library (5 from each experiment). Below, I explain the different results obtained in each step for processing the library before the high-throughput sequencing.

First the DNA was extracted from the samples of the first experiment (N=5 from *in vitro* library and N=12 containing caecum and colon samples) (Figure 38) using the phenol-chloroform procedure following the protocol of Zhang et al.[99] (Material and Methods, section 3). Then, 2 µg of DNA were digested using the restriction enzyme MmeI, located at the end of the transposon which cuts 18 nucleotides from the site of recognition, cutting inside the gene where the transposon was inserted (Material and Methods, section 12.2). Subsequently, digested DNA was dephosphorylated at 50 ºC. DNA was

recovered again adding chloroform-isoamyl alcohol (24:1). Phosphorylated adapters (containing matching barcodes to distinguish each sample after sequencing) were ligated with 100 ng of dephosphorylated MmeI digested fragments using T4 DNA ligase.
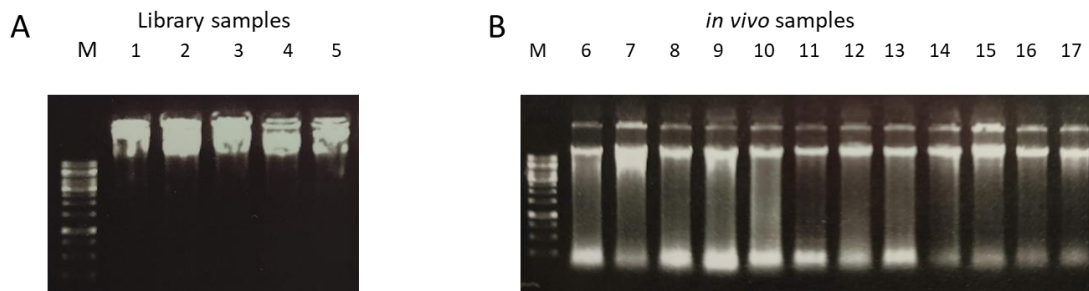


**Figure 38. DNA extraction by phenol-chloroform methodology. (A)** DNA extracted from library samples. **(B)** DNA obtained from *in vivo* samples.

At that point, a PCR was prepared to amplify the ligation fragments that contain the final part of the transposon and 16-18 nucleotides from the genome of Aus0004 strain, adding adapters to both extremes for further Illumina sequencing (Material and Methods, section 12.2, Figure 17). The obtained PCR products were run on an electrophoresis gel. The expected size of the PCR products was 130 bp, which was obtained for all the PCR products obtained from the library cultures (Figure 39A). However, we observed that the PCR products obtained from the intestinal samples contained a smear and not all of them had the same size (Figure 39B). These PCR products were purified from the gel (result not shown), but we still observed, even clearer, size differences between the fragments from the *in vivo* samples.
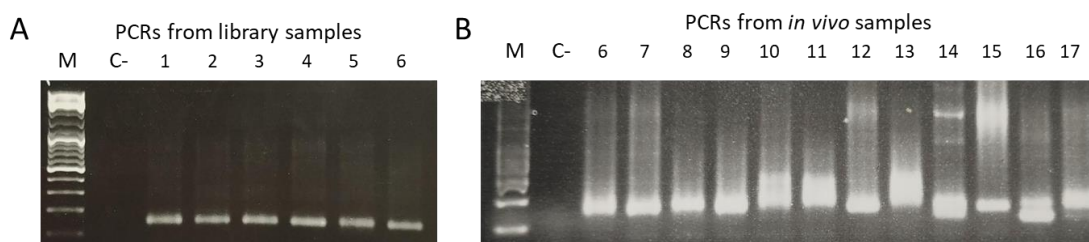


**Figure 39. PCR amplification of fragments containing the end of the transposon and the flanking sequence of the Aus0004 chromosome for Illumina sequencing. (A)** Fragments obtained from library samples present the expected size of 130 bp. **(B)** Amplified fragments obtained from *in vivo* samples present variable sizes. Water was added, instead of DNA, as a negative control.

We noticed that the DNA was degraded in samples from *in vivo* compared to samples from library after the DNA extraction using phenol-chloroform protocol (Figure 38). This smear could be affecting negatively to the different steps performed to obtain the PCR fragments for DNA sequencing. We reasoned that adapters could bind to small fragments from the smear, besides the fragmented DNA with MmeI, resulting in

unspecific amplification in the last step having fragments with the different sizes observed.

To solve it, the protocol was repeated to clean up the extracted DNA from the caecum and colon by magnetic beads. By doing this, we eliminated small fragments derived from DNA degradation that could give rise to non-specific amplification from *in vivo* samples (Figure 38B). Consequently, we followed the protocol but with some adaptations from the original protocol. Instead of using 2 µg of DNA, we used 0.5 µg, adjusting the volume of the other reagents to this concentration of DNA. We reduce the amount of DNA used because we lost some DNA during the purification process with clean-up beads. Nevertheless, from the step of adapter ligation to the step of dephosphorylation of MmeI digested fragments, we used the recommended concentrations given by the original protocol. Applying these changes in the methodology, we were able to obtain 130 bp PCR fragments of the same size from *in vitro* and *in vivo* samples (Figure 40). However, we lost the DNA of two of the samples, one from colon and one for caecum, resulting in N=10 PCR fragments from *in vivo* samples. Sample nº5 from the library had a smaller size than expected. Thus, the protocol was repeated for this particular sample. In this second attempt, we were able to obtain a fragment of the expected size (not shown). The obtained PCR fragments were purified from the gel (Material and Methods, section 10) to eliminate small fragments such as primer dimers (result not shown). The purified PCR products were sequenced by Illumina NextSeq 550 System High-Output Kit (1x75, single-end) (Fisabio, Valencia).
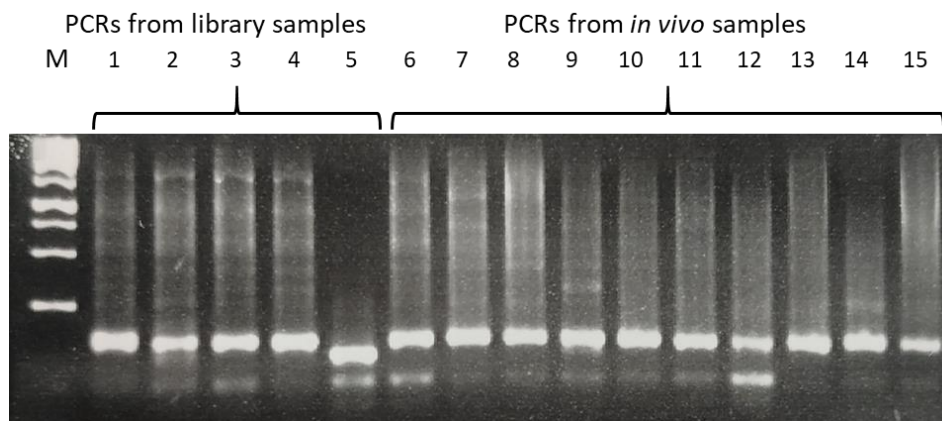


**Figure 40. PCR amplification of fragments containing the end of the transposon and the flanking sequence of the Aus0004 chromosome for Illumina sequencing using purified DNA as template.** PCR fragments from library and from *in vivo* samples present the expected size (130 bp). Negative control for the PCR was performed adding water instead of DNA, although the result is not shown.

## 1.6 Transposition mutagenesis analysis

After obtaining the sequences using the Illumina platform, a bioinformatics pipeline was developed to align the fragment of 16-18 nucleotides (chromosomal region flanking the transposon – where the transposon has been inserted) with the genes from the entire genome, to obtain the number of transposon mutants (Tn-mutants) present in each

111

condition. First of all, samples were separated by the barcodes using fastq-multx, allowing one mismatch. 9.2 % of the total sequences were discarded in this step because they did not contain a barcode sequence matching the list of utilized barcodes (Material and Methods, section 33, Table 18). Subsequently, adapters for Illumina sequencing were removed from the sequences obtained with cutadapt, allowing again one mistmatch as maximum. Sequences that did not contain a recognized illumina adapter were excluded from the analysis. After adapter trimming, sequences were filtered and cut with PRINSEQ-lite (minimum length 16, minimum quality mean 30, trimming quality right 20)[163], remaining those sequenced with more than 16 base pairs (cleaned sequences, Table 21).

**Table 21. Nº of sequences obtained from each sample after separating samples by barcodes and removing adaptors for Illumina sequencing.**

| Experiment | Sample type | Nº of sequences separated by barcodes | Sequences after removing adapters | | Cleaned sequences | |
|---|---|---|---|---|---|---|
| | | | Nº | % | Nº | % |
| 1 | Library | $6,71 \times 10^6$ | $5,01 \times 10^6$ | 74,69 | $2,72 \times 10^6$ | 40,50 |
| 1 | Library | $7,80 \times 10^6$ | $6,46 \times 10^6$ | 82,77 | $3,61 \times 10^6$ | 46,25 |
| 1 | Library | $7,97 \times 10^6$ | $6,36 \times 10^6$ | 79,78 | $3,47 \times 10^6$ | 43,54 |
| 1 | Library | $1,35 \times 10^7$ | $1,15 \times 10^7$ | 84,81 | $9,98 \times 10^6$ | 73,88 |
| 1 | Library | $1,11 \times 10^7$ | $9,08 \times 10^6$ | 81,84 | $8,03 \times 10^6$ | 72,41 |
| 2 | Library | $8,00 \times 10^6$ | $5,97 \times 10^6$ | 74,66 | $3,05 \times 10^6$ | 38,20 |
| 2 | Library | $1,45 \times 10^7$ | $1,26 \times 10^7$ | 87,12 | $7,27 \times 10^6$ | 50,08 |
| 2 | Library | $9,39 \times 10^6$ | $7,90 \times 10^6$ | 84,12 | $4,15 \times 10^6$ | 44,21 |
| 2 | Library | $1,23 \times 10^7$ | $9,95 \times 10^6$ | 80,74 | $8,67 \times 10^6$ | 70,32 |
| 2 | Library | $1,55 \times 10^7$ | $1,23 \times 10^7$ | 78,95 | $1,07 \times 10^7$ | 68,65 |
| 1 | Colon | $9,41 \times 10^6$ | $6,66 \times 10^6$ | 70,8 | $3,93 \times 10^6$ | 41,72 |
| 1 | Colon | $8,31 \times 10^6$ | $6,73 \times 10^6$ | 81 | $3,97 \times 10^6$ | 47,75 |
| 1 | Colon | $5,47 \times 10^6$ | $3,60 \times 10^6$ | 65,76 | $1,94 \times 10^6$ | 35,48 |
| 1 | Colon | $7,49 \times 10^6$ | $5,58 \times 10^6$ | 74,55 | $3,25 \times 10^6$ | 43,45 |
| 1 | Colon | $1,01 \times 10^7$ | $6,65 \times 10^7$ | 66,14 | $5,41 \times 10^6$ | 53,82 |
| 2 | Colon | $1,06 \times 10^7$ | $6,48 \times 10^6$ | 60,97 | $5,52 \times 10^6$ | 51,86 |
| 2 | Colon | $1,10 \times 10^7$ | $7,02 \times 10^6$ | 63,98 | $5,92 \times 10^6$ | 53,96 |
| 2 | Colon | $9,19 \times 10^6$ | $5,86 \times 10^6$ | 63,75 | $5,04 \times 10^6$ | 54,78 |
| 2 | Colon | $1,00 \times 10^7$ | $4,91 \times 10^6$ | 48,9 | $4,19 \times 10^6$ | 41,70 |
| 2 | Colon | $9,65 \times 10^6$ | $5,76 \times 10^6$ | 59,68 | $4,86 \times 10^6$ | 50,29 |
| 2 | Colon | $9,63 \times 10^6$ | $6,18 \times 10^6$ | 64,2 | $5,31 \times 10^6$ | 55,16 |
| 1 | Caecum | $1,26 \times 10^7$ | $8,24 \times 10^7$ | 65,68 | $7,19 \times 10^6$ | 57,31 |
| 1 | Caecum | $1,56 \times 10^7$ | $1,19 \times 10^7$ | 76,04 | $1,03 \times 10^7$ | 65,65 |
| 1 | Caecum | $1,43 \times 10^7$ | $9,11 \times 10^6$ | 63,69 | $7,92 \times 10^6$ | 55,32 |
| 1 | Caecum | $1,64 \times 10^7$ | $7,98 \times 10^6$ | 48,52 | $6,63 \times 10^6$ | 40,32 |
| 1 | Caecum | $1,37 \times 10^7$ | $9,25 \times 10^6$ | 67,75 | $8,05 \times 10^6$ | 58,94 |
| 2 | Caecum | $1,38 \times 10^7$ | $8,50 \times 10^6$ | 61,39 | $7,16 \times 10^6$ | 51,71 |
| 2 | Caecum | $1,31 \times 10^7$ | $9,16 \times 10^6$ | 69,68 | $7,95 \times 10^6$ | 60,49 |
| 2 | Caecum | $1,37 \times 10^7$ | $8,97 \times 10^6$ | 65,45 | $7,60 \times 10^6$ | 55,50 |
| 2 | Caecum | $3,70 \times 10^6$ | $2,45 \times 10^6$ | 66,28 | $2,03 \times 10^6$ | 54,84 |

% from the total number of sequences obtained in a particular sample

The remaining sequences were compared using blastn against the genes encoded by Aus0004 strain, with a minimum of 90 % of identity and coverage. We name those genes as coding sequences (CDSs), being all regions from the genome that possibly encode for proteins (the minimum size of CDS is 90 nucleotides (30 amino acids). As the average length of the readings was 16-17 nucleotides, 90 % of identity and coverage corresponded to a maximum of 1 incorrect nucleotide. Table 22 shows the number of sequences used in this analysis (cleaned) and the percentage of them that aligned to CDSs from the Aus0004 genome.

**Table 22. Sequences per sample matching VRE CDSs**. Cleaned sequences were aligned against the CDSs of Aus0004 strains.

| Experiment | Sample | Nº of cleaned sequences | Sequences matching VRE | |
|---|---|---|---|---|
| | | | Nº | % |
| 1 | Library | $2,72 \times 10^6$ | $1,92 \times 10^6$ | 70,52 |
| 1 | Library | $3,61 \times 10^6$ | $2,64 \times 10^6$ | 73,17 |
| 1 | Library | $3,47 \times 10^6$ | $2,51 \times 10^6$ | 72,41 |
| 1 | Library | $9,98 \times 10^6$ | $7,46 \times 10^6$ | 74,77 |
| 1 | Library | $8,03 \times 10^6$ | $5,89 \times 10^6$ | 73,27 |
| 2 | Library | $3,05 \times 10^6$ | $2,22 \times 10^6$ | 72,53 |
| 2 | Library | $7,27 \times 10^6$ | $5,41 \times 10^6$ | 74,40 |
| 2 | Library | $4,15 \times 10^6$ | $3,07 \times 10^6$ | 74,08 |
| 2 | Library | $8,67 \times 10^6$ | $6,48 \times 10^6$ | 74,70 |
| 2 | Library | $1,07 \times 10^7$ | $7,75 \times 10^6$ | 72,62 |
| 1 | Colon | $3,93 \times 10^6$ | $1,46 \times 10^6$ | 37,21 |
| 1 | Colon | $3,97 \times 10^6$ | $2,29 \times 10^6$ | 57,74 |
| 1 | Colon | $1,94 \times 10^6$ | $5,87 \times 10^5$ | 30,26 |
| 1 | Colon | $3,25 \times 10^6$ | $1,29 \times 10^6$ | 39,67 |
| 1 | Colon | $5,41 \times 10^6$ | $2,86 \times 10^6$ | 52,87 |
| 2 | Colon | $5,52 \times 10^6$ | $2,34 \times 10^6$ | 42,42 |
| 2 | Colon | $5,92 \times 10^6$ | $2,08 \times 10^6$ | 35,15 |
| 2 | Colon | $5,04 \times 10^6$ | $2,18 \times 10^6$ | 43,31 |
| 2 | Colon | $4,19 \times 10^6$ | $2,06 \times 10^6$ | 49,19 |
| 2 | Colon | $4,86 \times 10^6$ | $1,88 \times 10^6$ | 38,67 |
| 2 | Colon | $5,31 \times 10^6$ | $1,90 \times 10^6$ | 35,85 |
| 1 | Caecum | $7,19 \times 10^6$ | $3,68 \times 10^6$ | 51,22 |
| 1 | Caecum | $1,03 \times 10^7$ | $6,15 \times 10^6$ | 59,85 |
| 1 | Caecum | $7,92 \times 10^6$ | $3,70 \times 10^6$ | 46,73 |
| 1 | Caecum | $6,63 \times 10^6$ | $3,05 \times 10^6$ | 46,06 |
| 1 | Caecum | $8,05 \times 10^6$ | $4,05 \times 10^6$ | 50,31 |
| 2 | Caecum | $7,16 \times 10^6$ | $3,34 \times 10^6$ | 46,69 |
| 2 | Caecum | $7,95 \times 10^6$ | $4,28 \times 10^6$ | 53,84 |
| 2 | Caecum | $7,60 \times 10^6$ | $3,50 \times 10^6$ | 45,99 |
| 2 | Caecum | $2,03 \times 10^6$ | $9,56 \times 10^5$ | 47,10 |

% from the cleaned sequences obtained in a particular sample

Results obtained showed that library samples contained higher percentage of sequences matching the CDSs of Aus0004 (73.25 %) whereas lower percentage of matching sequences were detected in samples coming from *in vivo* samples (45.5 % on average) (Figure 41A). In principle, since the library samples are derived from pure cultures of Aus0004, we would expect a higher number of sequences matching the Aus0004 genome. We reasoned that most of the sequences that did not match Aus0004 CDSs were matching intergenic regions, rather than being contaminants or sequences with errors. To verify this, we aligned the sequences against the whole chromosomal DNA of VRE, rather than with just the CDSs. Effectively, the number of sequences matching VRE genome increased (Figure 41B). In the case of the *in vitro* samples, 95.81 % of sequences matched the VRE genome. In addition, the percentage of the sequences derived from *in vivo* samples aligning VRE genome was also increased (from 45.5 % to 54.4 %).
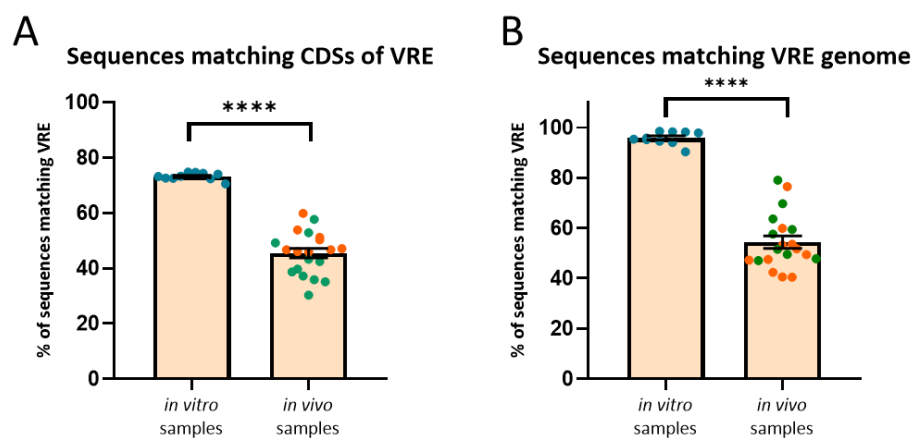


**Figure 41. Percentage of the filtered sequences that match CDSs encoded by Aus0004 or the entire genome.** Higher number of sequences coming from *in vitro* samples (library) aligned to the VRE strain Aus0004 compared to *in vivo* samples, both when comparing against coding sequences (CDSs) **(A)** or with the entire genome **(B)**. As expected, higher percentage of sequences matched the entire VRE genome compared to the alignment performed against the CDSs. Samples from the cecum are coloured in orange, while samples from the colon are coloured in green. **** $p < 0.0001$, Unpaired t-test. Bars represent the mean and whiskers represent the SEM.

On the other hand, we compared the *in vivo* derived sequences that did not match against the VRE genome with other databases, including the mouse genome, fungi database (RefSeq Fungi) and bacteria database (RefSeq Bacteria) in order to check if we were able to classify those sequences. Although the sequences were short, between 16 and 18 nucleotides, which increase the probability of obtaining random alignments, we observed that most of the sequences from *in vivo* samples that did not match against VRE aligned against mouse genome (average of 87.15 %) and in lower percentage to the bacterial database (12.14 %). A small percentage of sequences aligned against fungi database or no alignment was found (unknown) (Figure 42). Thus, we reasoned that some remnants of mouse tissue and other bacteria should have been extracted and also amplified, obtaining a percentage of sequences that do not come from the VRE genome in the *in vivo* derived samples.

**in vivo derived sequences
that did not match with the VRE genome**
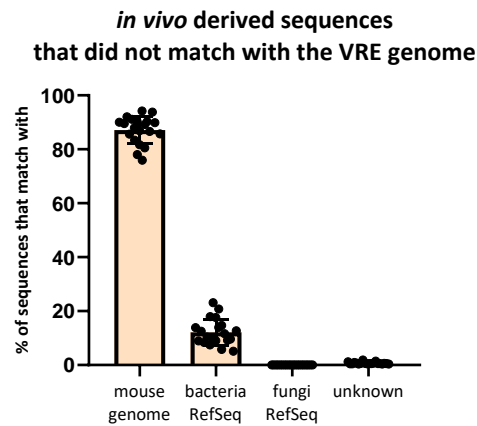


**Figure 42. Percentage of sequences from *in vivo* derived samples that aligned against different databases.** Sequences not matching VRE genome were aligned against different databases observing that the majority of these sequences align against the mouse genome and in a lower percentage to the bacteria database.

After checking the number of sequences matching CDSs, we followed the study analysing the distribution of the mutants in the transposon mutant library grown *in vitro* in order to check the randomness and the amount of mutants present for each CDSs. As it can be observed in the DNA plotter (Figure 43), mutations were spread across the entire bacterial genome (dark blue) confirming the capacity of the transposon to be inserted randomly.
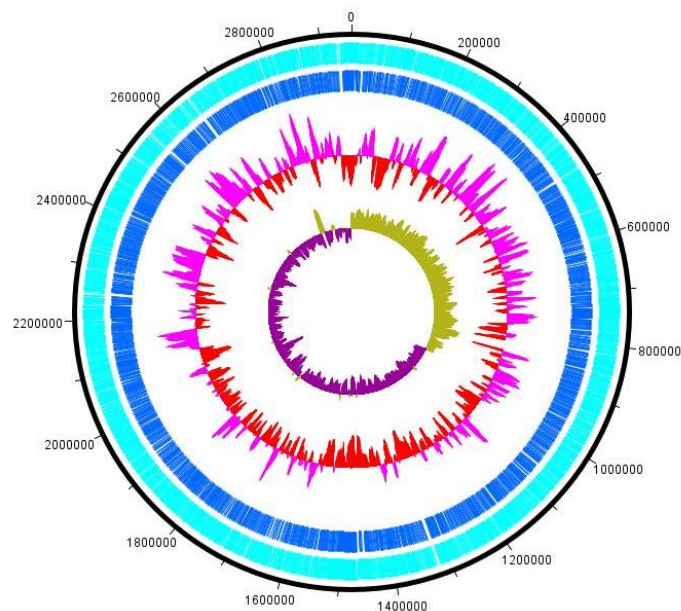


**Figure 43. Circular representation of the distribution of CDSs from the *E. faecium* Aus0004 genome in which the transposon was inserted and detected.** The light blue (track 1, outermost) represents all CDSs present in the genome. The dark blue (track 2) shows those CDSs in which at least one sequence from the library match that CDS. Pink and red peaks (track 3) represent the base 2 logarithm of the mean abundance of sequences matching a specific CDS in the library divided by average of sequences matching CDSs of the entire genome, pink above the average and red below the average. Track 4 (violet-yellow, innermost) shows a GC skew plot. The plot was generated with the DNAPlotter software.

115

Then, we analysed the number of CDSs of Aus0004 that could be detected in at least one sample, including both *in vivo* and *in vitro* samples. Table 23 represents the CDSs detected in each type of sample. 2808 CDSs from the total of 2919 encoded by Aus0004 could be detected in at least one of the analysed samples. In other words, sequences matching 111 CDSs (3.8 %) were absent in any of the analysed samples (Appendix, Table 35). Table 24 shows an example of 10 of those 111 genes not found in the analysis. We have mainly found genes without assigned functions designated as "hypothetical protein", different families of transposase genes, transcriptional regulators and 30S ribosomal protein S14. This result suggests that these genes were essential for the growth of Aus0004 strain both *in vitro* and *in vivo*. Alternatively, since most of these genes are from mobile genetic elements, it is possible that these elements were present in the original strain but were absent in the clone that we utilize. Re-sequencing of the genome would be necessary to discern these two possibilities.

**Table 23. Number of sequences matching CDSs of VRE from each type of sample.**

| Sample | CDSs matching VRE genome |
|---|---|
| Library | 2723 |
| Colon | 2783 |
| Caecum | 2729 |
| All samples | 2808 |

**Table 24. Example of genes absent from the Tn-seq analysis performed.**

| LocusTaq[a] | Annotation[b] |
|---|---|
| EFAU004_RS07230 | Autolysin |
| EFAU004_RS07350 | hypothetical protein |
| EFAU004_RS04180 | holin |
| EFAU004_RS11850 | IS256-like element IS16 family transposase |
| EFAU004_RS03960 | helix-turn-helix transcriptional regulator |
| EFAU004_RS07105 | N-acetylmuramoyl-L-alanine amidase |
| EFAU004_RS07220 | thymidylate synthase |
| EFAU004_RS08765 | DUF3042 domain-containing protein |
| EFAU004_RS01090 | 30S ribosomal protein S14 |
| EFAU004_RS09320 | group II intron reverse transcriptase/maturase |

[a] Indicates the gene ID number in the Aus0004 genome. [b] Indicates the predicted function.

We also analysed those CDSs not very abundant in all the conditions tested, assuming that these genes are candidate to be essential or nearly essential genes. There are other studies in which, appart from including as essential genes those not detected in the mutant transposon library, genes with very few insertions are also included[176,177]. Thus, we decided to search for those genes with few insertions, with less than five read counts and we identified 688 Tn-mutants. It means that 23.6 % of genes are candidates to be essential genes or key genes required for VRE growth both *in vivo* and *in vitro*. Among them, we have detected several genes enconding proteins supposed to be essential for

the growth of the bacterium, such as different subunits of the ribosome, membrane proteins o cell division proteins. Further studies will be needed to elucidate the essensiality of these low abundant genes.

Once genes where the transposon has been inserted (named as Tn-mutants in this work) were detected through Tn-Seq in every sample, we analysed the reproducibility of the Tn-mutants detected in each sample. We first focus on replicates from the library used in the same experiment. These could be considered as technical replicates because they all come from the same *in vitro* culture of the library. As expected, we detected a highly significant and positive correlation between the abundance of the Tn-mutants detected in library replicates from the same experiment, obtaining an $R^2$ average of 0.889 (p<0.0001, Pearson correlation test), being the highest $R^2$ of 0.9812 for all pair-comparisons (Figure 44A and B). Notably, we obtained a similar correlation when we compared the abundance of Tn-mutants between pairs of samples from the caecum collected from the same experiment, and was slightly reduced when samples from the colon were compared (Figure 44B). We next compared the variability obtained in samples collected in different experiments (inter-experiment) for the library, colon and caecum samples (i.e. the abundance of the mutants detected in samples collected in the first performed experiment as compared to samples collected from the second experiment). Comparing replicates from both libraries experiments we obtained similar $R^2$ data than intra-experiments. However, comparisons between colon and caecum samples from different experiments resulted in lower $R^2$ for some of the compared replicates. This last result was not unexpected taking into account the higher variability that is usually detected *in vivo* vs *in vitro* experiments, since some variables cannot be control *in vivo* (e.g. different changes in the microbiota of mice in different experiments).

Altogether the results presented in Figure 44 indicate that (i) the TnSeq technique was highly reproducible, (ii) the two performed experiments led to reproducible results, (iii) the abundance of mutants is also very similar in samples collected from different mice, although a slightly more variability was detected in samples collected from the colon than from the cecum.

117

A



B
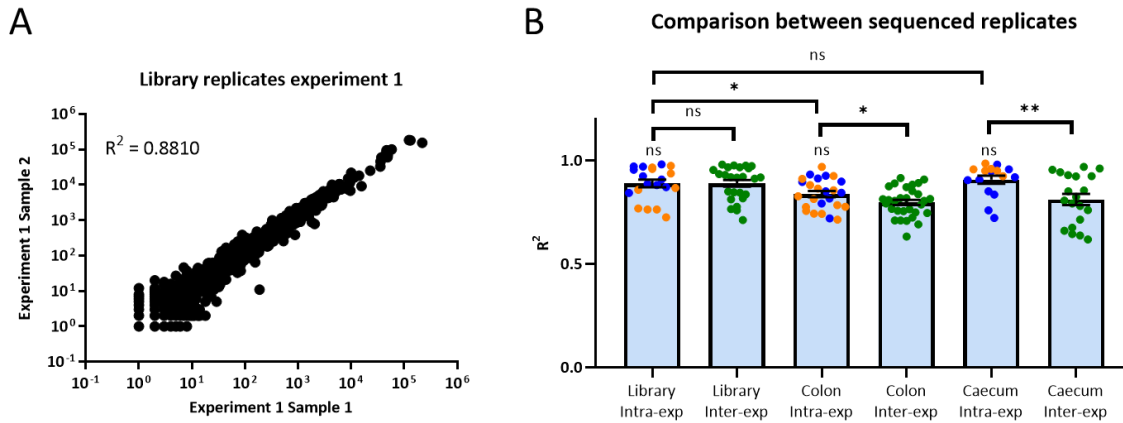


**Figure 44. Reproducibility of the Tn-seq methodology. (A)** Example of the comparisons performed between pair of samples to evaluate similarity among samples in the abundance of Tn-mutants. Each black dot represents the relative abundance of transposon insertions in a specific gene (Tn-mutant) in each of the two samples compared (sample 1 – X axis; sample 2 – Y axis). The R-squared value ($R^2$) of the regression line was calculated and indicated as a measure of similarity between samples. **(B)** $R^2$ values obtained for every pair of samples comparison. Comparisons were performed between (i) *in vitro* samples obtained from the same library aliquot (Library Intra-exp), (ii) *in vitro* samples obtained from different library aliquots use in the two different experiments (Library Inter-exp), (iii) *in vivo* samples from the same experiment (colon or cecum), (iv) *in vivo* samples from different experiments (colon or cecum). Dots coloured in blue represent comparison between replicates from the first experiment. Dots coloured in orange represent comparison between replicates from the second experiments. The comparisons between replicates from the same experiments are denominated as intra-experiments. Dots coloured in green represent comparisons between replicates from both experiments performed (inter-experiments). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns:not significant $p > 0.05$, Unpaired t-test. Bars represent the mean and whiskers represent the SEM.

Following this analysis, we performed a canonical correlation analysis (CCA) to study global differences in the relative abundances of Tn-mutants from *in vivo* an *in vitro* samples. Canonical correlation analysis showed that the abundance of Tn-mutants differed *in vitro* vs *in vivo* conditions (Figure 45).
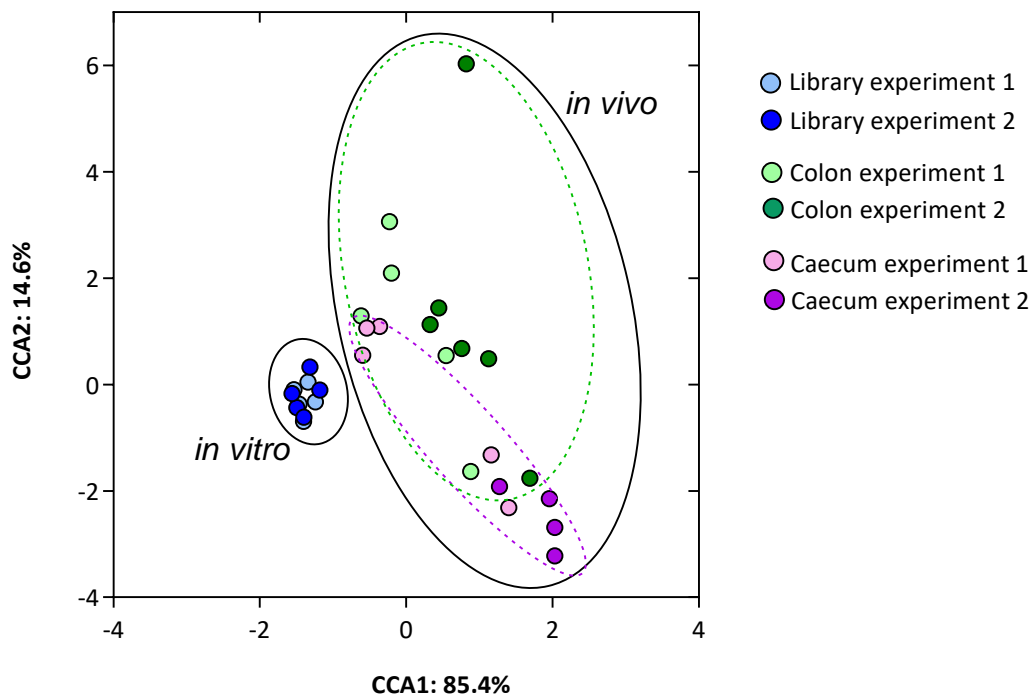
**Figure 45. Canonical correlation analysis (CCA) of Tn-seq samples.** The samples are separated in different groups based on the abundance of the Tn-mutants present in each condition. CCA1 and CCA2 are coordinates 1 and 2 respectively, showing the percentage of the variability explained by each one.

On the other hand, we performed a Principal Coordinates of Analysis (PCoA) using the Bray-Curtis distance among pair of samples in order to visualize similarities or dissimilarities of the abundance of Tn-mutants from library, colon and caecum samples. As oppose to the CCA analysis, this is an unbiased analysis in which the samples are separated by their dissimilarity but no information about the sample origin is used to perform the PCoA. Results revealed that *in vitro* samples clustered separately from *in vivo* samples (Figure 46), containing both colon and caecum samples, suggesting that the abundance of the mutants detected *in vitro* is different from that one detected in colon and caecum. In addition, there was not a clear separation between the samples from the colon and caecum suggesting that Tn-seq mutants overall behave similarly in both intestinal sites. In contrast, we detected a separation among *in vitro* samples analysed from the two performed experiments. This last result suggests stochastic variations in the abundance of Tn-seq mutants after an overnight culture.
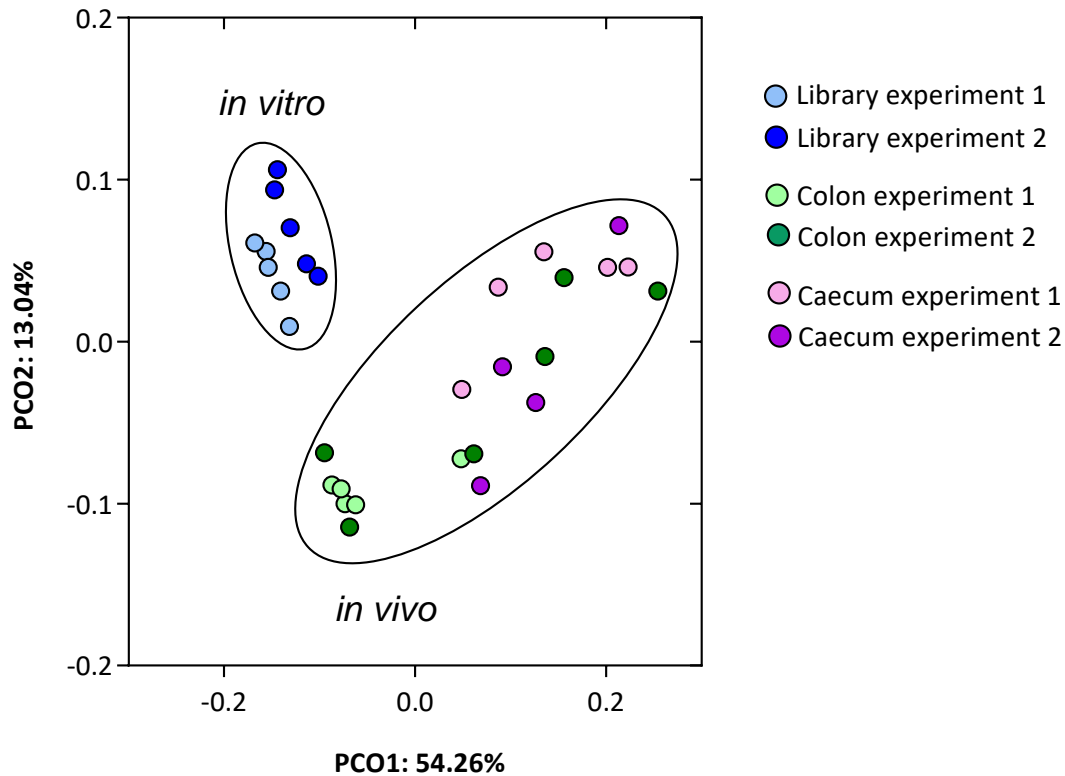
**Figure 46. Principal coordinate of analysis (PCoA) of Tn-Seq samples.** The samples are separated in different groups based on the Bray-Curtis distance among pair of samples that reflect similarities/dissimilarities on the abundance of Tn-seq mutants. PCO1 and PCO2 are coordinates 1 and 2 respectively, showing the percentage of the variability explained by each one.

These results suggested that VRE requires different genes for growing *in vitro* and *in vivo*. In addition, both the CCA and the PCoA analysis showed that there was larger variability in the samples obtained *in vivo* compare to those obtained *in vitro*. This higher variability could be produced by the variable intestinal environment and distinct nutrient availability in the gut of different mice caused by the variability in the remaining mouse microbiota upon antibiotic treatment.

Although PCoA shows clear differences among groups of samples, we subsequently apply a statistical test (PERMANOVA) to confirm that the detected differences were statistically significant (Table 25). PERMANOVA test showed no statistical differences between the samples obtained from the colon and the caecum, confirming the high similarity of the abundances of Tn-mutants between both *in vivo* samples. However, when the comparisons were performed between the *in vitro* and colon samples or the *in vitro* and caecum samples, the results were significant, either if all samples from both experiments were analysed together or the experiments were compared separately. We also notice a significant difference when we compared the abundance of Tn-Seq mutants of cecum or colon samples obtained from the two different experiments, indicating certain level of variability between experiments *in vivo*.

**Table 25. P-value obtained applying PERMANOVA test between different sample comparisons.**

| Sample comparison | PERMANOVA p-value |
|---|---|
| **Caecum vs colon, samples from both experiments** | 0.26 |
| *in vitro* **vs caecum, samples from both experiments** | 0.0017 |
| *in vitro* **vs colon, samples from both experiments** | 0.0017 |
| *in vitro* **vs caecum (experiment 1)** | 0.0083 |
| *in vitro* **vs colon (experiment 1)** | 0.013 |
| *in vitro* **vs caecum (experiment 2)** | 0.015 |
| *in vitro* **vs colon (experiment 2)** | 0.0017 |
| **Library, experiment 1 vs 2** | 0.088 |
| **Colon, experiment 1 vs 2** | 0.012 |
| **Caecum, experiment 1 vs 2** | 0.025 |

The previous analysis identified overall differences between groups of samples, based on the abundance of Tn-seq mutants. Subsequently, we studied differences in the abundance of each specific Tn-mutant between groups of samples. To this end, we applied two statistical tests: Wilcoxon test (a non-parametric test) frequently used to identify quantitative differences between two groups of samples and DeSeq2 test that was designed to identify differences between groups of samples in RNA-seq data.

First, we compared *in vitro* samples versus colon samples. We detected 927 Tn-mutants whose abundance significantly differs between *in vitro* and colon samples by Wilcoxon test (p<0.05, FDR<0.1, log2 Fold Change (FC) > |1|) and 958 by DeSeq2 (p<0.05, FDR<0.1, log2 Fold Change (FC) > |1|). 914 of those genes were detected by both types of analysis (Table 26). On the other hand, we compared *in vitro* samples versus caecum samples, obtaining 716 significant Tn-mutants by Wilcoxon test and 839 by DeSeq2, sharing 700 of those significant Tn-mutants between both statistical tests.

**Table 26. Tn-mutants whose abundance significantly differ between groups of samples.**

| Comparisons | Statistical test | Nº Tn-mutants significantly different | Common Nº significant Tn-mutants using both tests |
|---|---|---|---|
| *in vitro* **vs** | Wilcoxon | 927 | |
| **colon samples** | DeSeq2 | 958 | 914 |
| *in vitro* **vs** | Wilcoxon | 716 | 700 |
| **caecum samples** | DeSeq2 | 839 | |

Subsequently, we compared *in vitro* samples with *in vivo* (colon and caecum) from each experiment independently to compare the reproducibility of the results obtained in each experiment. As it is shown in Figure 47A, Tn-mutants whose abundance was significantly different in both experiments were higher compared to significant Tn-mutants detected only in one of the experiments, both comparing *in vitro* samples to colon or to caecum. A similar result was obtained when the data was analysed using Wilcoxon test (Figure 47A) or DeSeq2 (Figure 47B).

A

**CAECUM**

*In vitro* vs caecum

FDR < 0.1

FC > |1|

**COLON**

*In vitro* vs colon

FDR < 0.1

FC > |1|



B

**CAECUM**

*In vitro* vs caecum

FDR < 0.1

FC > |1|

**COLON**

*In vitro* vs colon

FDR < 0.1

FC > |1|
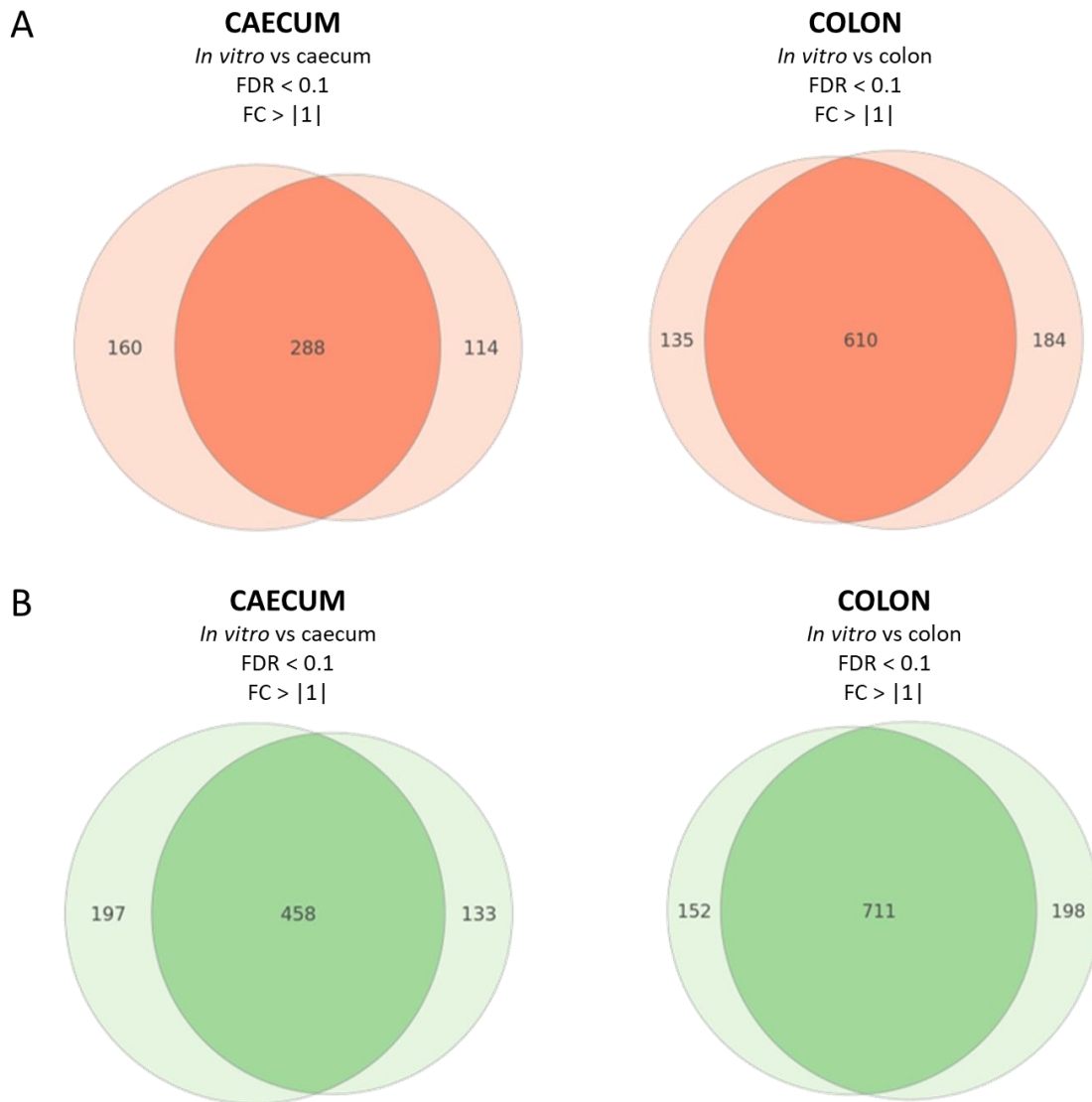


**Figure 47. Venn diagram representing the number of Tn-mutants whose abundance significantly differ** *in vivo* **(caecum and colon) as compared to** *in vitro* **samples in the two performed experiments.** Number of genes (Tn-mutants) whose abundance is significantly different *in vivo* (caecum or colon) vs *in vitro* ($p < 0.05$; FDR < 0.1; Log2 Fold change > |1|). The number within the circle indicates Tn-mutants whose abundance is significantly different *in vivo* vs *in vitro* in both experiments performed. Numbers outside the circle show Tn-mutants whose abundance is significantly different (*in vitro* vs *in vivo*) only in one of the experiments. **(A)** Results obtained applying Wilcoxon test. **(B)** Results obtained applying DeSeq2 test.

Furthermore, we compared to what extend the Tn-mutants whose abundance was found to be significantly different between *in vitro* and *in vivo* samples were similarly detected in both intestinal sites. As it is shown in Figure 48, the majority of the Tn-mutants whose abundance was found to be significantly different between *in vivo* and *in vitro* samples were identified in the two analysed parts of the large intestine, suggesting that the majority of the genes involved in the intestinal colonization are common for both cecum and colon. And the same result was obtained using both statistical analysis.
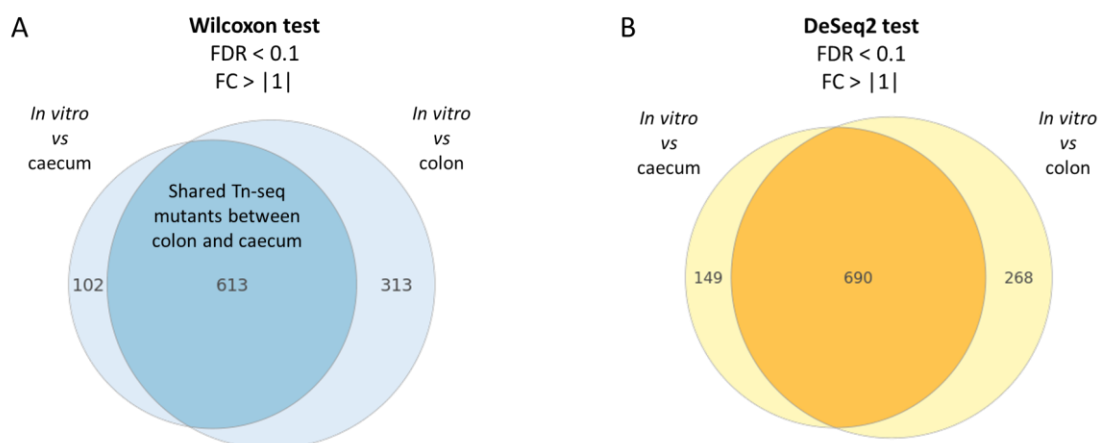
**Figure 48. Venn diagram comparing the Tn-seq mutants significantly different between *in vitro* and *in vivo* samples in both cecum and colon.** Circles represent the Tn-mutants whose abundance is significant different *in vivo* vs *in vitro* analysing all samples from both experiments together. The number within the circle indicates the number of Tn-mutants whose abundance significantly differ between *in vitro* vs caecum samples and also between *in vitro* vs colon samples. Digits outside the circle are referred to those Tn-seq mutants that are significant different only for caecum or for colon samples vs *in vitro* samples. **(A)** Results obtained applying Wilcoxon test. **(B)** Results obtained applying Deseq2 test.

After all these comparisons between the different statistical test performed and between different experiments and conditions, we represented in a heatmap the abundance of the Tn-mutants whose abundance significantly differ using the Wilcoxon test. This heatmap was applied to globally visualise the differences between the groups of samples but at the same time showing the abundance obtained in each individual sample. We have selected those genes with p<0.05; FDR<0.05 with a Log2 Fold-Change (FC) higher than |1|to be more restrictive and reduce the number of genes to be represented in the heatmap. Table 27 summarizes the number of Tn-seq mutants significantly more abundant *in vivo* than *in* vitro using these new parameters.

**Table 27. Number of Tn-seq mutants more abundant *in vitro* or *in vivo* samples.**

| Tn-seq mutants more abundant: | *in vitro* vs caecum samples | *in vitro* vs colon samples |
|---|---|---|
| - *in vitro* | 135 | 82 |
| - *in vivo* | 488 | 793 |

Moreover, from these significant genes, we selected for showing in the heatmap those whose abundance was significantly different both in the colon and caecum, as compare to the *in vitro* samples reducing the number of mutants more abundant *in vitro* to 71 and those more abundant *in vivo* to 485 (Appendix, Table 37). In the heatmap (Figure 49), we observed that the Tn-mutants less abundant *in vivo* behave similar both in colon and caecum in most of the samples. However, marked differences were detected in the abundances of the Tn-mutants that were more abundant *in vivo* in the samples collected from both colon and cecum.
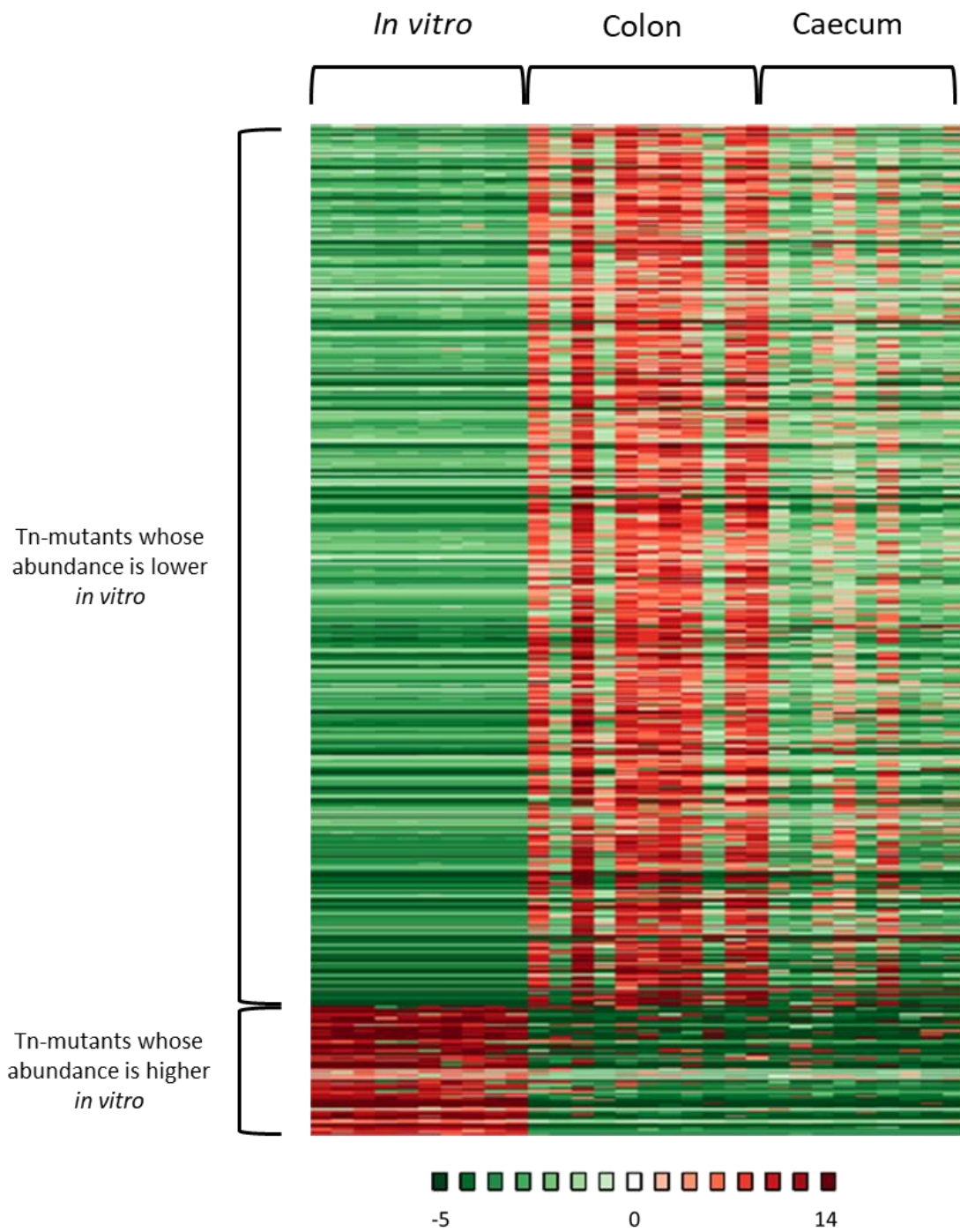
**Figure 49. Differential abundance of Tn-mutants in the *in vitro* an *in vivo* (colon and caecum) conditions.** The heatmap shows the 556 statistically significantly Tn-mutants (p<0.05; FDR<0.05 Log2FC > |1|) comparing *in vivo* vs *in vitro* conditions. Colours in the heatmap represent the difference of the relative abundance of each Tn-mutant respect the mean abundance of the corresponding Tn-mutant in all the samples, those above the mean in read and those below in green.

From all Tn-mutants whose abundance is higher *in vivo* than *in vitro* we show in Table 28 the top 10 that most affect growth (i.e. higher log2FC). Most of them are hypothetical proteins that do not have a clear function. The relative abundance of these Tn-mutants is represented in Figure 50, showing the lower or even non-detectable presence of those mutants *in vitro* whereas their relative abundance *in vivo* is higher. The higher abundance of these genes *in vivo* could indicate that this gene mutation could be favouring intestinal colonization. Alternatively, and more likely, taking into account that these Tn-seq mutants were almost absent *in vitro*, these gene mutations could be essential for VRE to grow in BHI medium *in vitro.* Although the characterization of these genes could be interesting, this is not the main focus of this thesis and therefore should be investigated in subsequent studies. Here, we focus on the Tn-mutants whose abundance is higher *in vitro* than *in vivo,* which could therefore have a direct role in the intestinal colonization of VRE, as indicated in the next section.

**Table 28. Selected *E. faecium* genes detected by Tn-Seq analysis whose abundance was higher *in vivo* as compare to *in vitro*.**

| LocusTaq[a] | Annotation[b] | Colon vs *in vitro* Log2(FC)[d] | Caecum vs *in vitro* Log2(FC)[d] |
|---|---|---|---|
| **EFAU004_RS00830** | 30S ribosomal protein S4 | 11,57 | 12,60 |
| **EFAU004_RS07510** | hypothetical protein | 10,63 | 12,029 |
| **EFAU004_RS08210** | hypothetical protein | 10,67 | 11,86 |
| **EFAU004_RS08070** | hypothetical protein | 12,49 | 11,73 |
| **EFAU004_RS04720** | hypothetical protein | 10,14 | 11,07 |
| **EFAU004_RS00165** | hypothetical protein | 8,93 | 10,32 |
| **EFAU004_RS06780** | cold-shock protein | 9,06 | 9,57 |
| **EFAU004_RS13175** | primosomal protein N' | 10,43 | 9,21 |
| **EFAU004_RS11215** | tRNA-specific adenosine deaminase | 8,87 | 8,04 |
| **EFAU004_RS01955** | hypothetical protein | 9,44 | 7,90 |

[a] Indicates the gene ID number in the Aus0004 genome. [b] Indicates the function predicted. [c] Log2 Fold-change (FC) derived from the ratio of the relative abundance of the Tn-mutant present in *in vivo* (colon and caecum) as compare to *in vitro* samples. A positive FC indicates that the mutant was found to be in higher levels *in vivo*. The mutants are ordered by average FC of both comparisons (colon *vs in vitro* and caecum vs *in vitro*). Only the top 10 genes with a highest significant difference (highest log2FC) are shown.
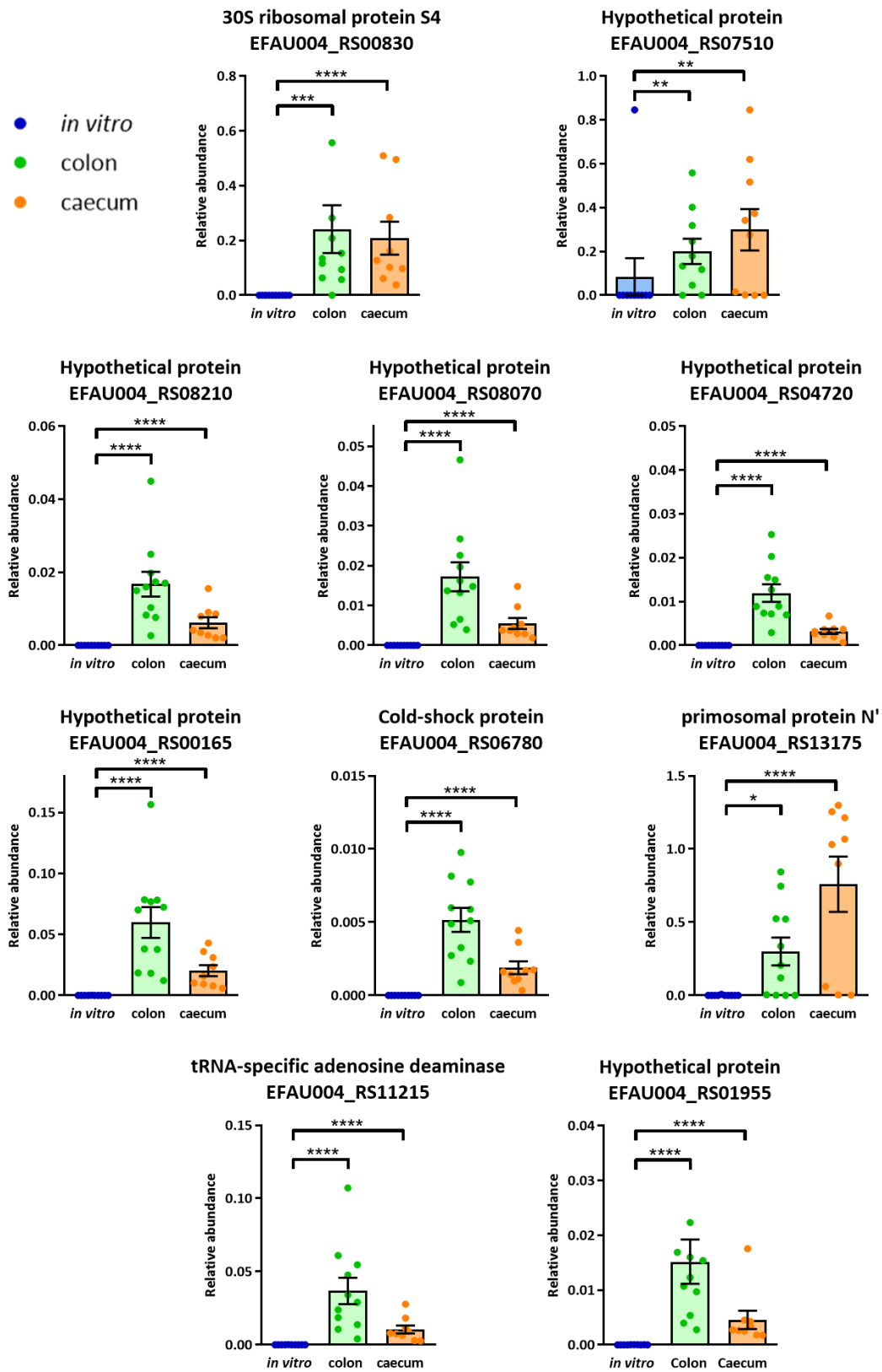
**Figure 50. Relative abundance of Tn-mutants whose abundance is significantly higher *in vivo* than *in vitro*.** The graph shows the 10 Tn-mutants with the highest fold change difference between *in vitro* and *in vivo* samples ($p < 0.05$; FDR < 0.05; Log2FC > 2). * p<0.05, ** p<0.01, *** p<0.001, **** p<0.0001, Wilcoxon test. Bars represent the mean and whiskers represent the SEM.

126

### 1.6.1 Selection of candidate genes that could be key for VRE gut colonization

From the list of VRE genes that potentially could have a direct role in gut colonization (i.e. those whose abundance was higher *in vitro* than *in vivo*), we selected those genes whose differences *in vitro vs in vivo* were more pronounced, 10 candidate genes in total, all of them containing a Log2 Fold-Change (FC) higher than 2, both comparing with the caecum and the colon samples (Table 29), and in all of them the phenotype was detected in the two performed experiments. In this table, it is also shown the functions predicted by the reference Aus0004 genome deposited in GenBank, inferring the possible function of the gene.

We focused on those genes because the relative abundance of mutants *in vivo* is lower or even non-detectable and the relative abundance *in vitro* is higher (Figure 51). These 10 candidate genes had a significantly lower abundance both in the caecum and in the colon compare to *in vitro*, and the abundance of these Tn-seq mutants was found to be significantly different (*in vitro* VS *in vivo*) in both experiments performed. Therefore, these 10 genes could be important for gut colonization of VRE in the entire large intestine.

**Table 29. Candidate *E. faecium* genes detected by Tn-Seq analysis as potentially important for gut colonization**

| LocusTaq[a] | Annotation[b] | Gene Name[c] | Colon vs *in vitro* Log2(FC)[d] | Caecum vs *in vitro* Log2(FC)[d] |
|---|---|---|---|---|
| EFAU004_RS03690 | PTS mannose transporter subunit EIIAB | *man*X | -6,23 | -5,52 |
| EFAU004_RS02800 | Fe-S cluster assembly protein SufB | *suf*B | -4,64 | -6,61 |
| EFAU004_RS07995 | Vat family streptogramin A O-acetyltransferase | *vat* | -4,60 | -4,70 |
| EFAU004_RS10170 | phosphate ABC transporter ATP-binding protein | *pst*B2 | -2,84 | -6,15 |
| EFAU004_RS12135 | hypothetical protein / Two-component system regulatory protein YycI | *yycI* | -3,23 | -4,8 |
| EFAU004_RS14195 | LacI family DNA-binding transcriptional regulator | *ltrR* | -3,24 | -3,60 |
| EFAU004_RS03130 | N-acetylmuramic acid 6-phosphate etherase | *mur*Q | -3,55 | -2,84 |
| EFAU004_RS12420 | DsbA family protein | *dsb*A | -2,59 | -3,37 |
| EFAU004_RS01175 | bifunctional acetaldehyde-CoA/alcohol dehydrogenase | *adh*E | -2,66 | -2,88 |
| EFAU004_RS01355 | glycosyl transferase | *pgt* | -2,55 | -2,75 |

[a] Indicates the gene ID number in the Aus0004 genome. [b] Indicates the function predicted. [c] Name assigned to each gene base on their description of the genome deposited in NCBI or by similarity with other genes described with that function. [d] Log2 Fold-change (FC) derived from the ratio of the relative abundance of the Tn-mutant present in *in vivo* (colon and caecum) as compare to *in vitro* samples. A negative FC indicates that the mutant was found to be in lower levels *in vivo*. The mutants are ordered by average FC of both comparisons (colon vs *in vitro* and caecum vs *in vitro*).

**PTS mannose transporter subunit EIIAB**
**EFAU004_RS03690**

**Fe-S cluster assembly protein SufB**
**EFAU004_RS02800**

- in vitro
- colon
- caecum

**Vat family streptogramin A O-acetyltransferase**
**EFAU004_RS07995**

**Phosphate ABC transporter ATP-binding protein**
**EFAU004_RS10170**

**Hypothetical protein / YycI protein**
**EFAU004_RS12135**

**LacI family DNA-binding transcriptional regulator**
**EFAU004_RS14195**

**N-acetylmuramic acid 6-phosphate etherase**
**EFAU004_RS03130**

**DsbA family protein**
**EFAU004_RS12420**

**bifunctional acetaldehyde-CoA/alcohol dehydrogenase**
**EFAU004_RS01175**

**Glycosyl transferase**
**EFAU004_RS01355**

**Figure 51. Relative abundance of Tn-mutants that most affected the intestinal colonization by VRE.** The graph shows the 10 genes with a higher FC difference in abundance *in vivo* vs *in vitro* (i.e. the mutant was more abundant *in vitro*) (p < 0.05; FDR < 0.05; Log2FC > 2) **** p<0.0001. Wilcoxon test. Bars represent the mean and whiskers represent the SEM.

## 1.7 Targeted mutagenesis

To validate the analysis of the sequencing results and to characterize the role of the genes detected that could be involved in gut colonization, we used a targeted mutagenesis methodology from Zhang et al[153]. The candidate gene was removed from the genome of the wild type strain by directed mutagenesis and replaced by a gentamicin resistance marker (Material and Methods, section 14).

### 1.7.1 Construction of a gentamicin resistant strain as control strain for experiments performed with targeted mutants

Since the gene conferring resistance to gentamicin is the marker used to select the targeted mutants by substitution of the candidate gene, we constructed a gentamicin resistant strain as a control mutant, to check that the insertion of the antibiotic resistance gene did not affect the growth of the bacterium.

First of all, a bibliographic search was carried out to choose the insertion point where the marker did not affect the growth of the bacteria. As the mobile genetics elements are inserted at certain points in the genome and usually do not interfere with the growth of the bacterium, we checked the VRE genome searching for insertion points of mobile genetics elements, bacteriophages or pathogenicity islands. In the manuscript of Van Schaik et al.[60], it is described that the *esp* gene is encoded by a large PAI flanked by repetitive regions. In the study, it is described that the PAI is found in three different species of VRE and it varies in size between 64 kb (in E1162), 104 kb (in E1679) and 68 kb (in U0317). We checked the genome of *E. faecium* strain Aus0004 and the PAI flanked by repetitive regions was found in the inverted orientation compared to E1162, being around 115 kb. We reasoned that introducing the gene conferring resistance to gentamicin next to the PAI, in an intergenic region, would not affect Aus0004 growth, taking into account the natural insertion of 104 kb (the PAI) already present in Aus0004.

In order to introduce the marker, we selected a specific point in the Aus0004 genome, the nucleotide 2.762.777, which is located 20 bp upstream of the PAI (Material and Methods, section 14.2). Primers were designed to amplify the marker and the flanking regions of the insertion point in the VRE genome (Material and Methods, section 33, Table 17). The obtained cassette was cloned in the thermo-sensitive plasmid pWS3 and VRE was electrotransformed. By recombination of the flanking regions the marker was inserted in the genome of the bacterium in the desired point and the insertion was checked by PCR amplification of the gentamicin resistance gene and by sequencing the entire gene and the flanking region (result not shown). The bacterial isolate obtained was grown in BHI plates adding gentamicin to verify its resistance.

#### 1.7.1.1 The gentamicin resistance marker does not affect the bacterial growth *in vitro*

In order to examine if the gentamicin resistance gene introduced affects the bacteria growth, we measured the growth of the gentamicin resistant strain (IR::*gen,* IR stands for intergenic region and gen for gentamicin*)* and the WT strain, using Tecan spectrophotometer, as mentioned in Material and Methods, section 15. The growth

curves obtained for each strain were basically identical (Figure 52A). To calculate if there were significant differences between strains in the *in vitro* growth, we calculated the area under the curve and as expected we obtained no significant differences ($p>0.05$; Unpareid t-test) (Figure 52B).
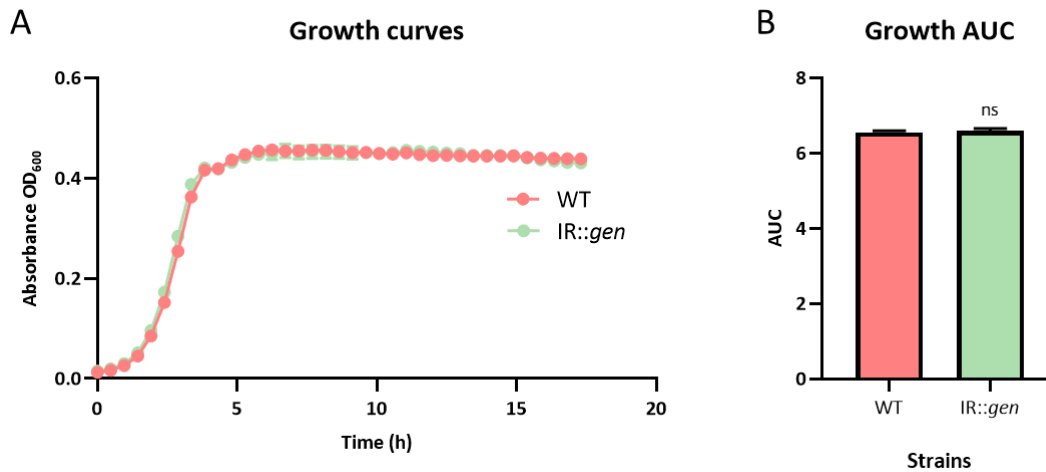


**Figure 52. Growth of the WT strain and the IR::*gen* mutant *in vitro*.** Growth curves **(A)** and their corresponding AUC **(B)** show no significant differences between the growth of the WT strain and the strain with the gentamicin marker inserted in an intergenic region. The optical density (OD) was quantified at a wave length of 600 nm. ns: not significant $p<0.05$, unpaired t-test. Bars represent the mean and whiskers represent the SEM.

### 1.7.1.2 The gentamicin resistance marker does not affect the fitness of the bacteria under *in vivo* and *in vitro* conditions

After showing that both strains grew to the same extend *in vitro* when they were grown separately, we check that both strains would grow to the same extend also when they are grown *in vitro* together in the same media (i.e. we performed a competition assay to check if when both strains were grown together, any of them would outcompete the other one).

In order to perform the *in vitro* competition experiment, we mixed both strains in an equal proportion of CFUs to inoculate BHI medium with approximately $10^4$ CFUs/µl (5 x $10^3$ CFUs WT + 5 x $10^3$ CFUs IR::*gen)*. The cultures were incubated in aerobic and anaerobic conditions to verify no phenotypic differences between both strains at different conditions. Cultures were plated after 24h of incubation in the previous selective media at 37 ºC. The results confirmed that there was no growth difference between WT and IR::*gen* control strain, resulting in a similar number of colonies after 24h of incubation (i.e. competitive index - CI for the WT strain not different from 1, see Material and methos, section 16 for the definition of CI) (Figure 53A).

To perform the competition assays *in vivo*, both strains were collected from BHI plates and resuspended in BHI broth to be mixed in the equal amount to inoculate mice by oral gavage with a total of $10^8$ CFUs/200 µl (5 x $10^7$ CFUs of each strain). Mice were previously

treated with clindamycin. As shown in Figure 53B, VRE was able to colonize the intestinal tract of mice reaching high levels three days post-VRE inoculation, with no significant differences between both strains. In Figure 53B (mice experiment) is represented the competitive index (CI) between WT and IR::*gen,* which is not significantly different from 1*.*

This result demonstrated that the WT and the control strain colonized the intestine to the same extend.
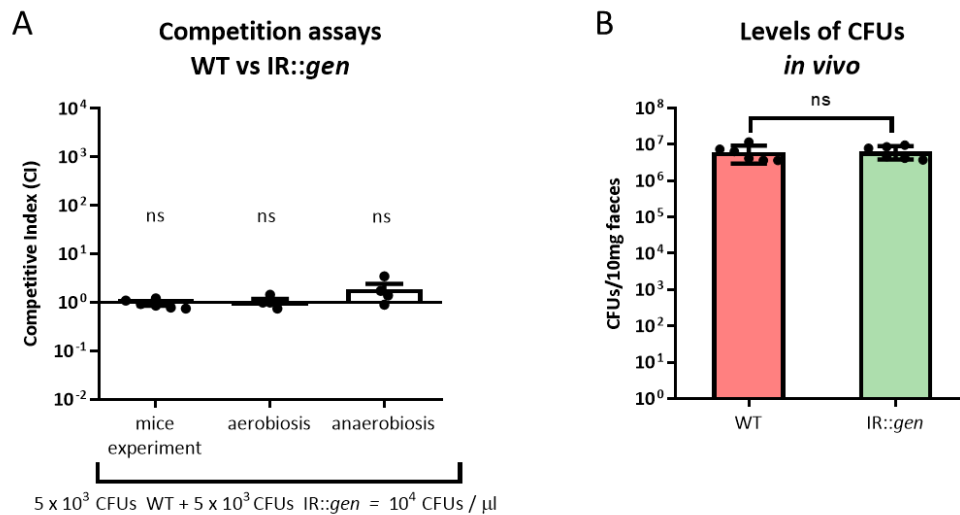


**Figure 53. Gentamicin resistance marker insertion does not affect the growth under competition conditions with the WT strain both *in vivo* and *in vitro*. (A)** For *in vitro* experiments, cultures were mixed 1:1 and plated after 24 hours of incubation in aerobic and anaerobic conditions. Two experiments with two replicates (N=4) were performed for each condition. Both experiments were plated in selective media to differentiate both types of strains. Results of the *in vivo* competition experiment are also included in the graph. The graph represents the Competitive Index (CI) in each mouse sample and *in vitro* replicates, obtaining similar levels of both strains in all conditions tested. **(B)** For the *in vivo* assays, six mice (N=6) were orally infected with a 1:1 mix of WT and IR::*gen* strain. Faeces samples were collected three days post-VRE inoculation. Both strains were able to colonize the gut, reaching high and similar levels. ns: not significant $p < 0.05$, one sample t-test calculated from competitive index (CI) data as logarithm values (log10), introducing as the hypothetical value 0 to compare differences inside each group (A-B). Note that 0 is the log10 of a FC=1, which is the result of having the same number of colonies in both the WT and mutant strain. Thus, if the result is not significantly different from log2FC=0, then both strains have the same growth capacity. Wilcoxon test was used to compare differences between two groups, ns: not significant $p < 0.05$. Bars represent the mean and whiskers represent the SEM.

In summary, the gentamicin resistance marker did not cause a fitness cost for the bacteria neither *in vivo* nor *in vitro*. Thus, the gene conferring resistance to gentamycin can be used as a marker to replace the genes that we want to delete from the VRE genome.

## 1.7.2 Targeted mutagenesis of the candidate genes

To characterize the genes involved in intestinal colonization we have used the *E. faecium* Aus0004 strain but to simplify the reading of the text we will call it VRE. Nevertheless, we consider that some differences could exist between different strains of vancomycin-resistant *E. faecium.*

After the identification of the genes that could affect most the intestinal colonization of VRE (Section 1.6.1, Figure 51), their deletion was carried out from the genome of Aus0004 strain to verify their relevance in the intestinal colonization process. Primers used to construct mutants are indicated in Material and Methods, section 31, Table 17. A total of 10 mutant strains were generated. The methodology used was targeted mutagenesis (Material and Methods, section 14.1) followed by the replacement of the candidate gene by a gentamicin marker.

## 1.7.2.1 Candidate genes deletion does not affect the bacterial growth *in vitro*

Once the genes were deleted from the WT strain, we first check their growth profile *in vitro*. The different strains were resuspended from BHI agar plates into BHI broth to adjust the OD $_{600}$ to 0.1. 200 µl of the culture were added to 96-well plate in triplicates and the absorbance was measured overnight in the spectrophotometer. The results verified that the deletion of the genes does not diminish the growth of VRE *in vitro*, comparing the growth curves (Figure 54A) and the area under the curve (AUC) (Figure 54B) of WT strain and mutant strains, except for one particular mutant (Δ*pgt*::gen) in which the growth of the mutant was slightly decreased compare to the WT.

A

**Growth curves**



Legend:
- WT
- $\Delta manX::gen$
- $\Delta sufB::gen$
- $\Delta vat::gen$
- $\Delta pstB2::gen$
- $\Delta yycI::gen$
- $\Delta ltrR::gen$
- $\Delta murQ::gen$
- $\Delta dsbA::gen$
- $\Delta adhE::gen$
- $\Delta pgt::gen$
- IR::gen

B

**Growth AUC**
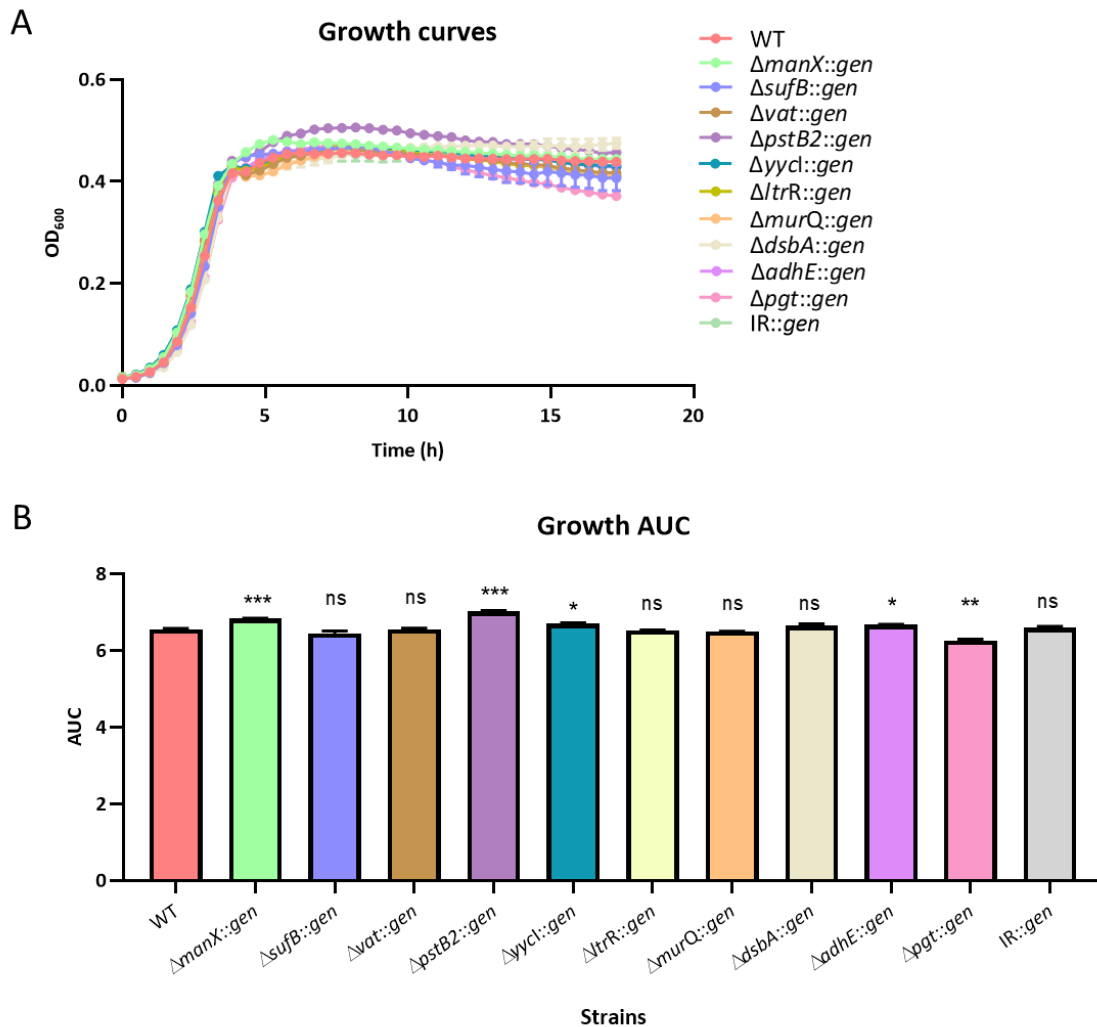


**Figure 54. Growth comparison of mutant and WT strains.** No significant differences were found in the growth curves profile **(A)** and its respective AUC **(B)** of WT and mutant strains. Each strain was grown per triplicate. Absorbance was measured overnight every 20 minutes in the spectrophotometer. The optical density (OD) was quantified at a wave length of 600 nm. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, ns: not significant > 0.05, unpaired t-test. Bars represent the mean and whiskers represent the SEM.

### 1.7.2.2 *In vivo and in vitro* competition assays show an *in vivo* growth advantage of the WT over the mutant strains

The clindamycin mouse model explained above was used to test the growth of the mutants *in vivo*, co-inoculating the WT strain with each of the mutant strains at the fourth day of the antibiotic treatment. Stool samples were collected three days post-VRE inoculation and plated in selective media to differentiate both strains. Two independent experiments were performed (N=3 mice per experiment). The competitive index (CI) was calculated as indicated for the previous competition assays.

The results indicate that, in the major part of the competitions assays performed, WT strain had a significant advantage over mutant strains, since a higher number of WT strain derived colonies were identified for most of the competition assays performed

133

(Figure 55). In the graph is included the control competition experiment performed previously, the mix of WT versus IR::*gen* (control strain). The results indicated that the deleted genes have a relevant role in the intestinal colonization of VRE. Nevertheless, we obtained a different result for *Δvat*::*gen* mutant because in this particular case, we identified a higher number of colonies derived from the mutant that from the WT strain. Although in the previous library analysis the mutant was in a lower relative abundance in the large intestine, we did not obtain the same result for the *in vivo* experiment with the gene deleted by targeted mutagenesis.
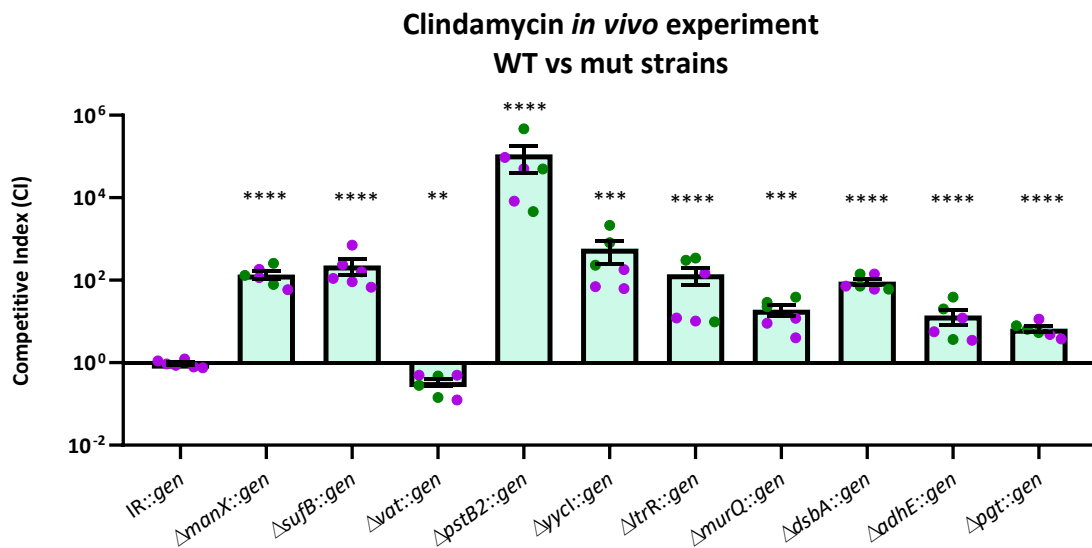


**Figure 55. *In vivo* competition assays verify the growth disadvantage of specific mutants over the WT strain.** Antibiotic treated mice were co-inoculated with wild type strain and each of the mutants. Two replicates for the experiment were performed (N=3 mice per experiment, total N=6). The competitive index was calculated for each mutant tested as previously described. Symbols coloured in purple represent mice from first experiment and in green from the second one. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Unpaired t-test between the control group (IR::*gen*) and each of the mutant group using data in logarithm scale (log10). Bars represent the mean and whiskers represent the SEM.

On the other hand, we performed competition assays *in vitro* to verify that there were not differences between WT and mutant strains *in vitro*. The experiments were performed as in the *in vivo* assays, mixing both strains in the same proportion, but then inoculating them in BHI. Cultures were incubated for 24 h in aerobic and anaerobic conditions. After plating in selective media and counting the colonies, we observed that both the WT and mutant strains grew to the same extend with one exception: there was a clear advantage of the growth of the WT compare to the growth of one of the mutants both in aerobiosis and anaerobiosis. This mutant strain was the Fe-S cluster assembly protein SufB mutant (Figure 56).
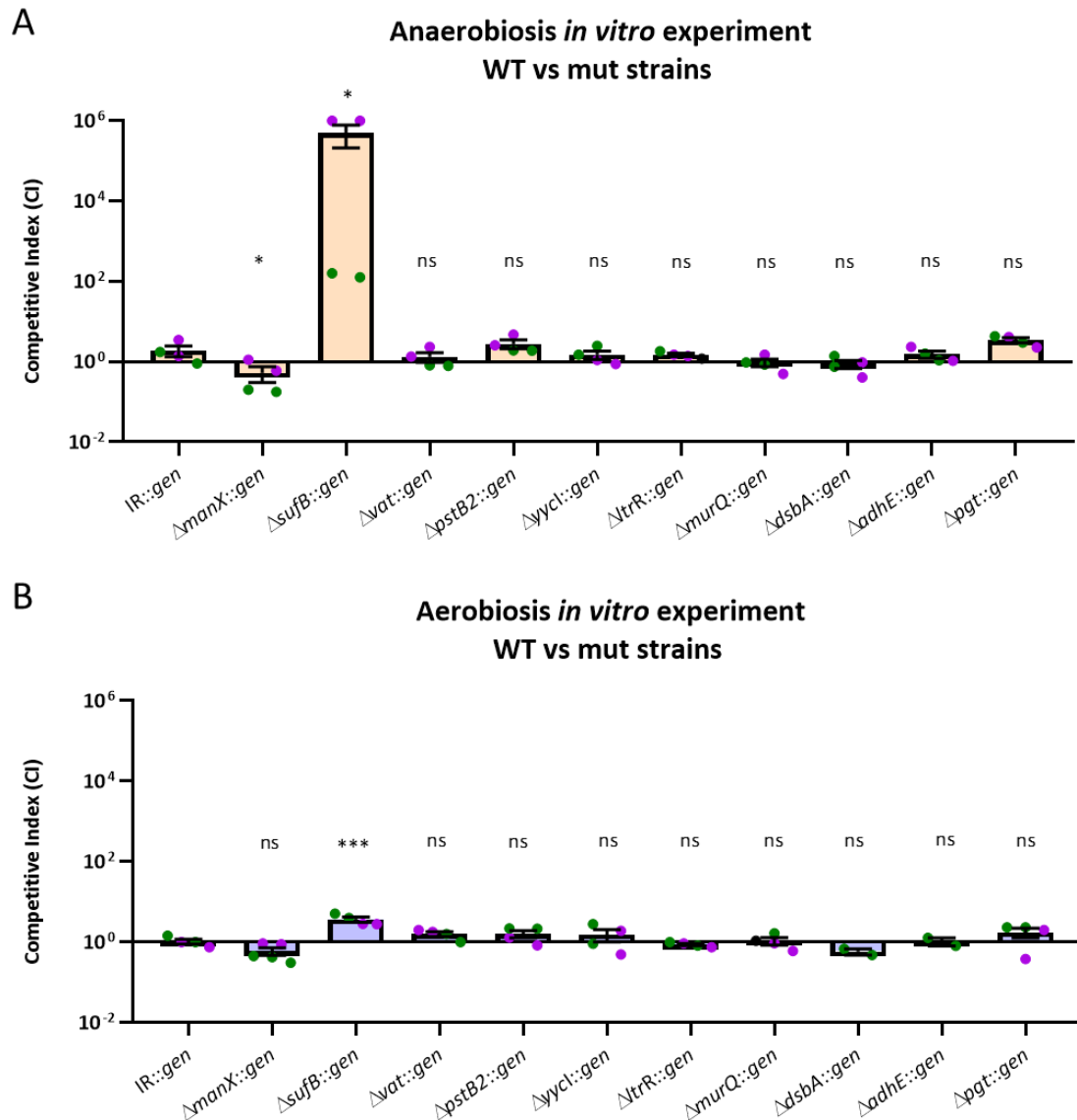
**Figure 56. *In vitro* growth differences between the WT and the mutant strains.** Competition assays *in vitro* co-culturing in BHI medium each mutant strain with the WT strain, plating after 24h of incubation in aerobiosis **(A)** and anaerobiosis **(B)**. The experiment was performed twice, two replicates per experiment. The competitive index for each mutant was calculated as previously described. Symbols coloured in purple represent the first experiment and in green the second one * $p < 0.05$, *** $p < 0.001$, ns:not significant $p > 0.05$. Unpaired t-test between the control group (IR::*gen*) and each of the mutant group using data in logarithm scale (log10). Bars represent the mean and whiskers represent the SEM.

The SufB gene is encoded within an operon containing another 3 genes that together theoretically form a Fe-S cluster. We hypothesize that the deletion of only one of the genes from the operon could generate a Fe-S complex that could be toxic for the bacteria. For this reason, we decided to delete four genes from the [Fe-S] operon to check their phenotypic effect *in vitro* and *in vivo.* In this case, we observed an advantage of the WT strain *in vivo* (Figure 57) as compared to the mutant of the whole Fe-S cluster operon*,* although less pronounce that the previous effect observed with the mutant of

135

the Fe-S cluster subunit SufB protein (Figure 55). However, compared to the previous results (Figure 56), the mutant for the whole operon does not confer any phenotypic deffect *in vitro,* neither in aerobiosis nor in anaerobiosis (Figure 57), indicating that the deletion of those genes are involved in gut colonization without affecting *in vitro* growth.
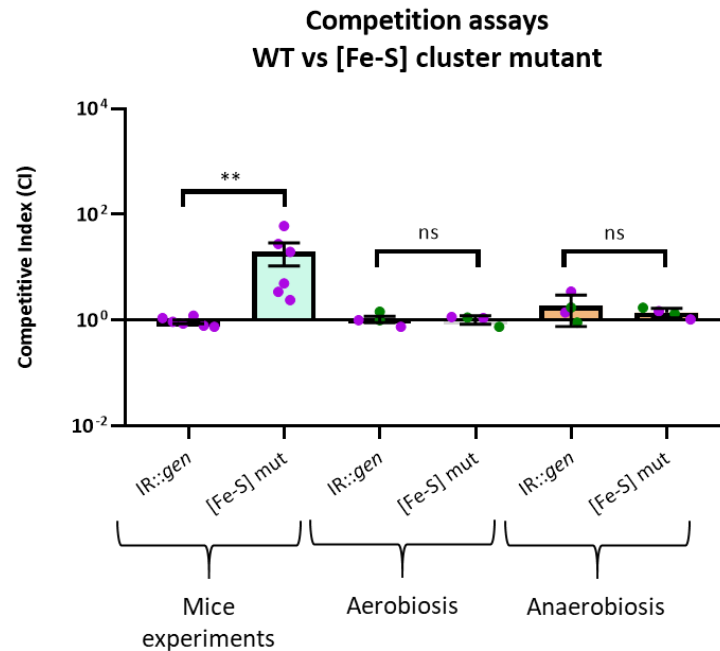


**Figure 57. *In vivo* and *in vitro* competition experiments with the WT and the mutant of the four genes of the [Fe-S] cluster.** Mice treated with clindamycin were co-inoculated with both strains (N=6 mice). *In vitro* assays were performed co-culturing both strains in BHI medium and plated after 24h. The *in vitro* experiments were repeated twice with two replicates per condition. The competitive index was calculated for each condition tested. Symbols coloured in purple represent samples from first experiment and in green from the second one. ** $p<0.01$, ns: not significant $p>0.05$. Unpaired t-test between the control group (IR::*gen*) and each of the mutant group in each different condition using data in logarithm scale (log10). Bars represent the mean and whiskers represent the SEM.

The results obtained in the competition experiments confirm the role of the genes selected (Table 29) in gut colonization, discarding those two that showed unexpected differences *in vivo* (Δ*vat*) or *in vitro* (Δ*SufB*). In addition to these two mutants, we also decided to eliminate for the subsequent analysis the gene with a hypothetical function that was further assigned to two-component system regulatory protein YycI (Δ*YycIH*), due to the extensive work to characterize the role of the two component system. Finally, we also decided to discard the glycosyltransferase gene (Δ*pgt*) for subsequent analysis because this enzyme belongs to a numerous and diverse family and there are plenty of bacterial cell activities that could be modified by a glycosyltransferase.

We reasoned to continue the study with the selected seven mutants after checking that under *in vitro* conditions the WT and the mutants growth equally, but *in vivo* there is an advantage of the WT strain and this suggests that those genes are involved the colonization of the intestine by VRE.

**1.8 *In vivo* competition assays under different conditions confirms the previous results**

In addition to the clindamycin mouse model, we performed the *in vivo* competitions assays using another mouse model in which we treated mice with vancomycin. As shown in Figure 8 (Introduction, section 2.2.3.), both antibiotics promote different microbiota changes so we wonder if the mutants would have a growth defect under different intestinal environments. On the other hand, since vancomycin can be supplied in the drinking water and not subcutaneously twice a day (the procedure for clindamycin treatment), it would allow us to study the dynamics of the WT and mutant strains at longer time-points.

The experiment was performed with the preceding mutants selected from the results obtained in the *in vivo* and *in vitro* competition assays, adding the mutant control strain with the gentamicin resistance marker inserted in the intergenic region. In this mouse model, besides collecting samples on day three post-VRE inoculation as in the previous clindamycin model, we also collected samples on day six and thirteen post-VRE inoculation (Material and Methods, section 28.3, Figure 25).

The outcome of the experiment showed that WT strain had a significant growth advantage over the mutant strains (Figure 58), exactly as happened with the previous model used (Figure 55). We obtained similar results on the third day of the treatment, compared to the clindamycin mouse model. The collection of samples on days six and thirteen post-VRE inoculation allowed us to observe that levels of WT strain increased ($10^8$ CFUs/g) and some of the mutants were completely outcompete by the WT strain at day thirteen (samples coloured in orange in Figure 58). However, in the case of the competition between WT and the control strain IR::*gen*, no differences were detected, corroborating that the insertion of the gentamicin resistance marker per se does not affect the *in vivo* fitness of VRE. This assay confirmed that the deleted genes were key in the intestinal colonization of VRE under the effects of a different antibiotic treatment.
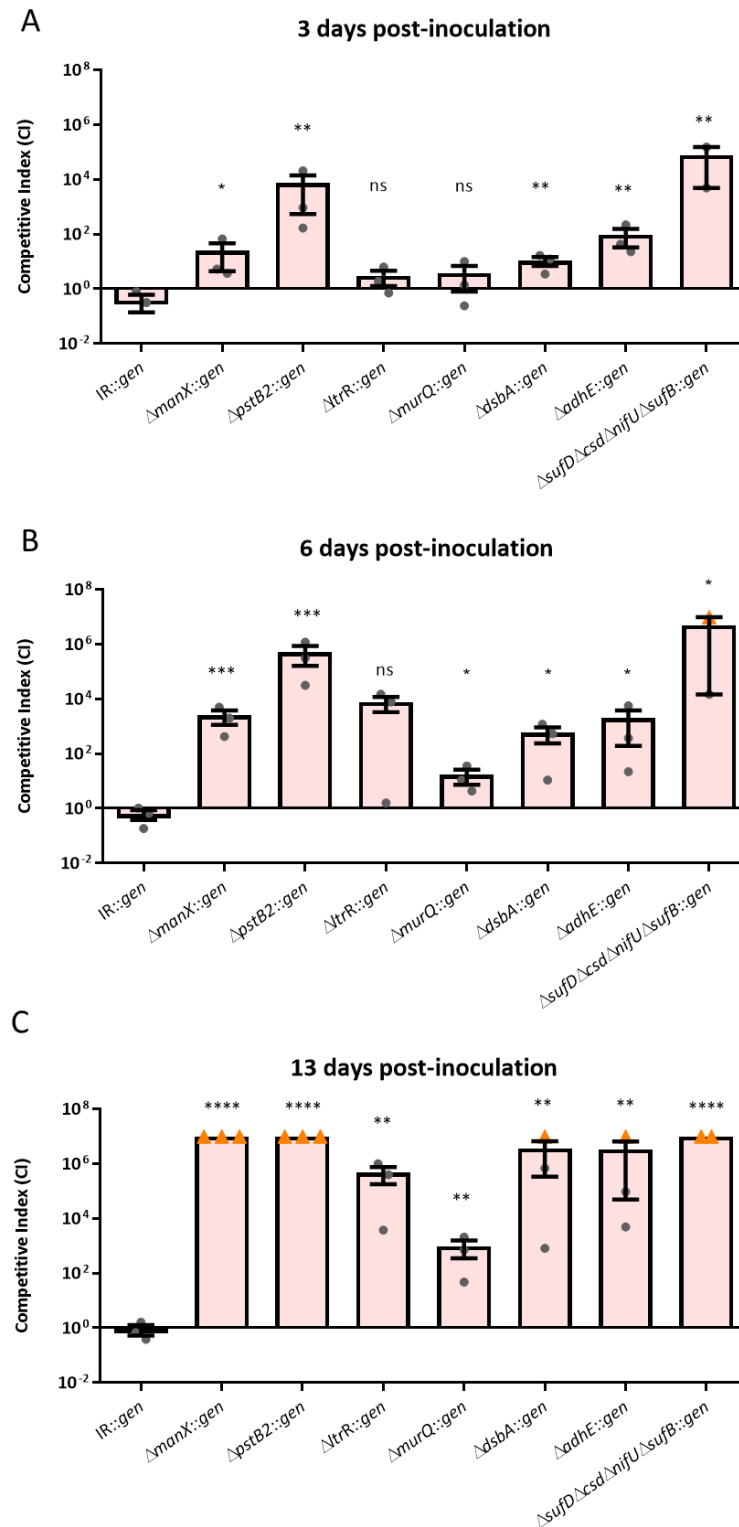
**Figure 58. Competition assays using vancomycin antibiotic treatment confirm the role of the studied genes in gut colonization by VRE.** Antibiotic-treated mice are co-inoculated with the wild type strain and each mutant strain. Stool samples are collected three **(A)**, six **(B)** and thirteen days after inoculation **(C)**. Levels of colonization are represented as the proportion of wild type and mutant strain (competition index). N=3 mice used per mutant tested. When the mutant is not detected, the limit of detection is used to calculate the CI (10 CFUs).* $p<0.05$, ** $p<0.01$, *** $p<0.001$, **** $p<0.0001$, ns:not significant $p>0.05$. Unpaired t-test between the control group (IR::*gen*) and each of the mutant group using data in logarithm scale (log10). Bars represent the mean and whiskers represent the SEM.

138

## 1.9 Characterization of mutants with putative functions related with the use of nutrients

For some of the candidate genes confirmed to be involved in gut colonization, their role in the intestinal colonization was difficult to infer based on their protein sequence. These genes were those that theoretically encoded a bifunctional acetaldehyde-CoA/alcohol dehydrogenase (*adh*E) and DsbA family protein (*dsb*A). Their role in gut colonization could be part of a cascade of signals that may involve multiple genes and due to the limitation of time to finish the PhD we decided not to continue with the characterization of these particular genes.

However, other identified genes seemed to have a more straightforward role in VRE intestinal colonization, since these genes could be involved in the acquisition or utilization of specific nutrients required for VRE to colonize the intestine. We therefore decided to expand our study to genes related to the use/acquisition of nutrients, including ions such as phosphate and iron, coding the phosphate ABC transporter ATP-binding protein (*pstB*2) and Fe-S cluster that involves *sufD*, *csd*, *nifU*, *sufB* genes, respectively.

On the other hand, there were genes related to the use/acquisition of some sugars, which was the case of the PTS mannose transporter subunit EIIAB (*manX*) , predicted to be involved in the transport of mannose and the N-acetylmuramic acid 6-phosphate etherase (*murQ*), predicted as an enzyme required for the utilization of the N-acetylmuramic acid.

Finally, LacI family DNA-binding transcriptional regulator gene (*ltrR*) could be a transcriptional regulator of genes involved in the use of some nutrients, favouring the adaptation and growth in the gut environment.

### 1.9.1 Fe-S cluster may be necessary for the growth of VRE under limited conditions of both $Fe^{2+}$ and $Fe^{3+}$

In the case of the study of the proteins associated to Fe-S cluster (Δ*sufD*Δ*csd*Δ*nifU*Δ*suf*B::*gen* mutant) we used chelators to verify the role of these proteins in the transport of iron ions. We reasoned that these cluster of genes could be relevant for gut colonization because the intestinal environment has limited concentrations of free iron. Thus, we tried to mimic an environment poor in iron *in vitro* by using chelators specific for iron. We set up the experiments based on a previous study published by Lanigan N. et al[178]. In that study, authors characterised genes required by *Bifidobacterium breve* to grow under limiting conditions of iron using three chelators to chelate the different forms of iron ions (Material and Methods, section 17). These experiments were performed in a training stay that I performed in the laboratory of Douwe van Sinderen, the Principal Investigator of the Lanigan et al. study.

Iron-limiting experiments were carried out in BHI medium. First, the Maximal Tolerable Concentration (MTC) was determined for the WT strain. It means the highest concentration of chelator that we can add to the medium affecting growth to a lesser

extent. We used different chelators concentrations, ranging from 2000 µM to 0 µM for dipyridyl (chelator of $Fe^{2+}$ ions), from 1200 µM to 0 µM for phenanthroline (chelator of $Fe^{2+}/Fe^{3+}$) and 400 µM to 0 µM for ciclopirox olamine (chelator of $Fe^{3+}$), based on the concentrations used in the study performed by Lanigan N. et al[178]. The results showed that the WT strain was more susceptible to the ferric ion chelator, $Fe^{3+}$ (MTC 50 µM) and it was more resistant to $Fe^{2+}/Fe^{3+}$ chelator (MTC 200 µM, approximately) (Figure 59). Ferrous chelator ($Fe^{2+}$), hardly affected the growth of the WT strain, even when the highest concentration was added (500 µM), being the chelator that least affected VRE growth. The MTC determined for each chelator used is indicated in Table 30.
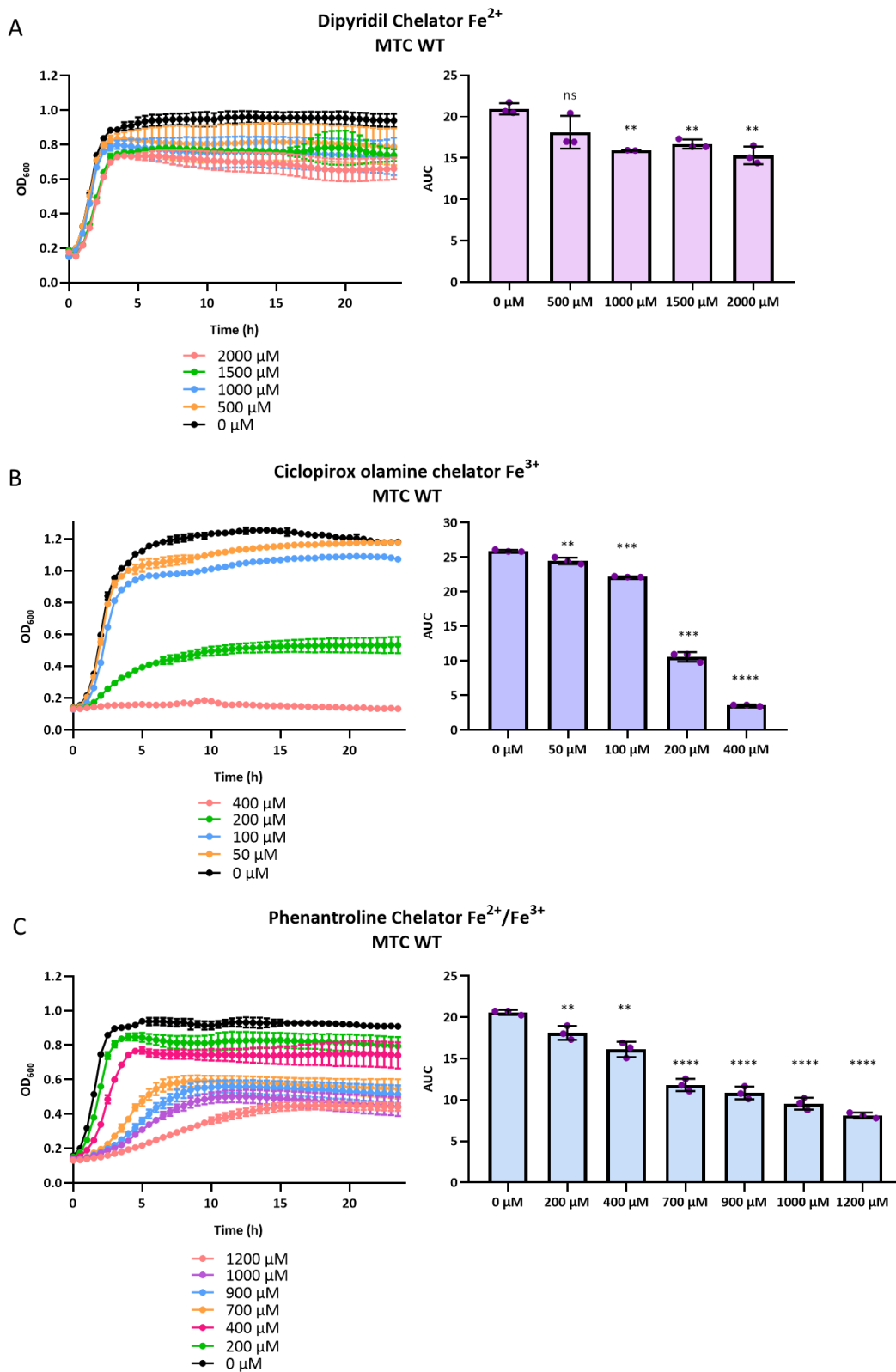
**Figure 59. Growth assay to determine de Maximal Tolerable Concentration (MTC) of the VRE WT strain to Fe chelators.** The curves measure the growth of the WT strain in BHI broth containing different concentrations of the three different chelators tested: **(A)** Dipyridil, which chelates $Fe^{2+}$; **(B)** Ciclopirox olamine, which chelates $Fe^{3+}$, and **(C)** Phenantroline, which chelates $Fe^{3+}$ and $Fe^{2+}$ ions. The optical density (OD) was quantified at a wave length of 600 nm. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$, ns:not significant $p > 0.05$. Unpaired t-test between the control group (0 µM) and each chelator concentration. Bars represent the mean and whiskers represent the SEM.

**Table 30. Maximal Tolerable Concentration determined for each chelator.**

| Chelator Name | Chelator Ion | MTC |
|---|---|---|
| Dipyridyl | $Fe^{2+}$ | 500 µM |
| Ciclopirox olamine | $Fe^{3+}$ | 50 µM |
| Phenantroline | $Fe^{2+}$ and $Fe^{3+}$ | 200 µM |

Once we obtained the MTC information, the WT strain and the $\Delta sufD\Delta csd\Delta nifU\Delta sufB::gen$ mutant were resuspended in BHI broth from BHI plates previously grown overnight. The $OD_{600}$ was adjusted to 0.1 and triplicates of 200 µl with the corresponding chelator were added to a 96-well plate and incubated overnight at 37 ºC in the spectrophotometer to measure the growth curves.

As a result of these experiments, we observed that by adding the MTC of each chelator to the WT and mutant strain, we did not observe differences in growth for the chelator of $Fe^{2+}$ and the chelator of $Fe^{2+}/^{3+}$ (p>0.05; comparing AUC; not shown). However, a small difference, still significant, was observed when adding ciclopirox olamine chelator, chelator for $Fe^{3+}$, at 50 µM, the MTC observed for this chelator (Figure 60B). We next decided to increase the concentration of each chelator to check if higher concentrations reduced growth mostly in the mutant strain compared to the WT strain. Notably, the phenanthroline chelator (chelator for $Fe^{2+}$ and $Fe^{3+}$ ions) in higher concentrations significantly reduced more the growth of the mutant compared to the WT strain (Figure 60C). However, no significant differences between both strains were observed when adding the chelator for $Fe^{2+}$ at higher concentrations. Altogether, these results suggested that the iron cluster was relevant for VRE growth under limited concentrations of iron, mainly for the $Fe^{3+}$ form.
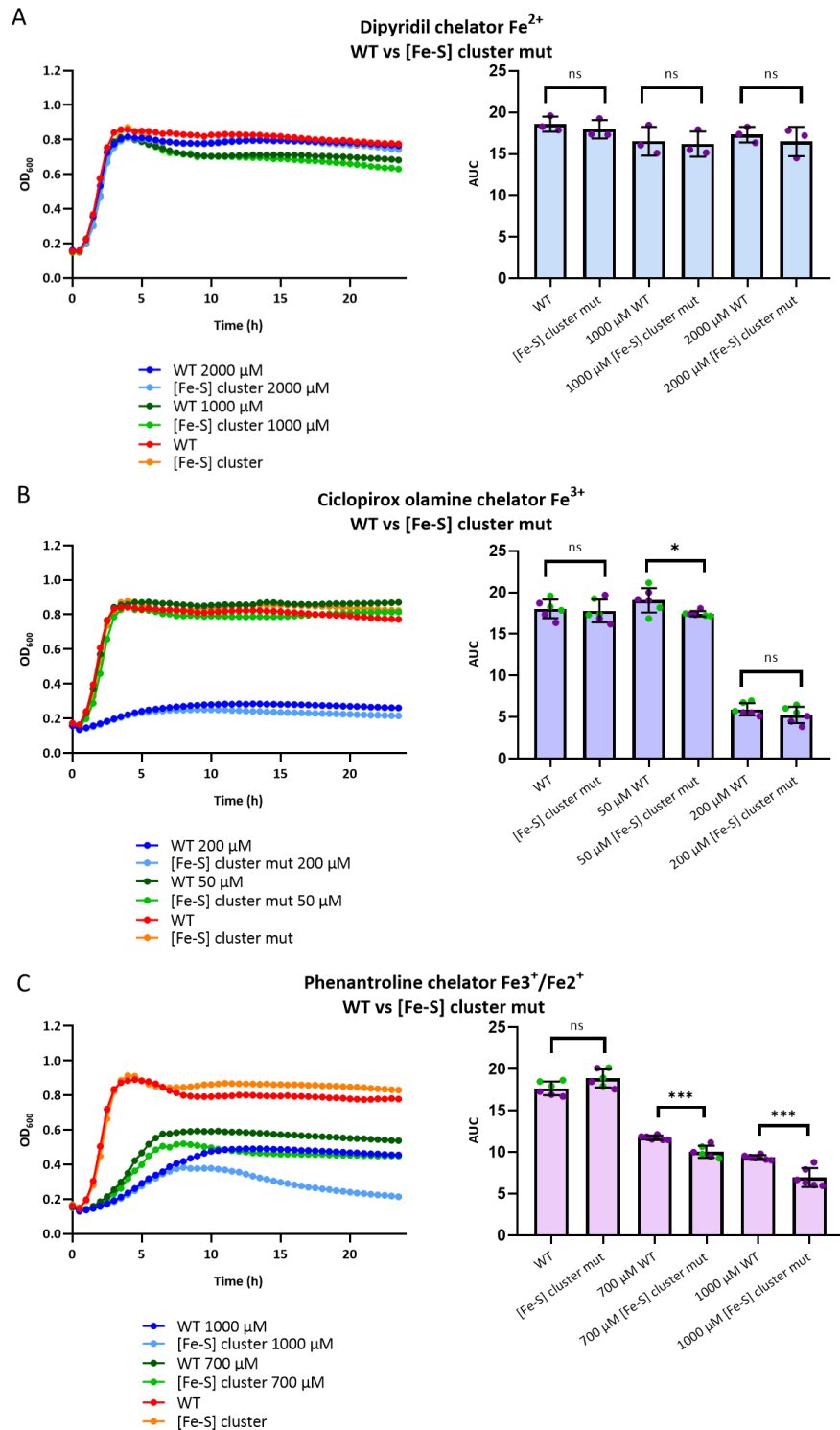
**Figure 60. Growth assay with BHI medium adding different chelators.** Curves with their corresponding area under the curve (AUC) show the growth of wild type and mutant strain Δ*sufD*Δ*csd*Δ*nifU*Δ*sufB*::*gen* in the presence or absence of different chelators: **(A)** dipyridil, **(B)** ciclopirox olamine and **(C)** phenatroline. The concentration of each chelator is indicated. Growth decrease was observed in the mutant strain mainly when phenantroline (chelator for $Fe^{2+/3+}$) was added to the media. Experiments have been performed once (B) or twice (A and C) with three replicates per experiment. Symbols coloured in purple represent samples from first experiment and in green from the second one. The optical density (OD) was quantified at a wave length of 600 nm. * p<0.05, *** p<0.001, ns:not significant p>0.05. Unpaired t-test. Bars represent the mean and whiskers represent the SEM.

143

## 1.9.2 Determination of a suitable minimal medium to study the phosphate ABC transporter ATP-binding protein mutant strain

In order to study the role of the phosphate ACB transporter ATP-binding protein mutant (Δ*pst*B2::*gen*) to phosphate starvation, we tested two different minimal medium without phosphate and we compared the growth of the WT and the mutant strain individually during 24h. We reasoned that there might be phosphate limitation in the intestinal environment, so the uptake and transport of ions could be important for the growth of the pathogen in the gut. Therefore, we used a defined media to control the concentration of phosphate and its effect in the growth of WT and the mutant strain.

The first medium used was a chemically defined medium (2x CDM) based on *Streptococcus cremoris* Defined Medium (Otto et al.[161]) (Material and Methods, section 31, Table 11). The phosphate compounds were substituted by 3-(N-morpholino)propanesulfonic acid (MOPS) buffer. Both strains were grown in the CDM medium and the phosphate starvation medium adding increasing concentrations of potassium phosphate ($K_3PO_4$) stock (pH7) ranging from 0.1 mM to 200 mM. When $K_3PO_4$ was added from 0.1 to 10 mM, we did not observe growth (Figure 61). Higher concentrations added ranging from 50 mM to 200 mM produced the same growth effect for both strains after 17 hours of incubation. When both strains were grown in the complete CDM medium, we could observe that the growth curves were not the typical growth curve observed in media used to grow VRE, such as BHI. With the CDM medium, the bacteria reach an initial stationary phase, then the absorbance decreased and after a few ours the absorbance increased again.

The delay in the exponential growth phase and the absence of differences between the strains suggested that this minimal medium was not adequate for the study of VRE growth under phosphate limiting conditions.
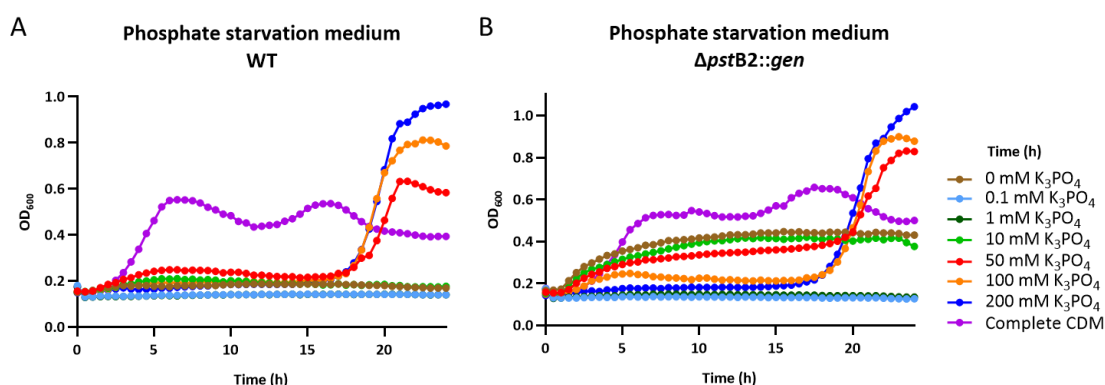


**Figure 61. Growth assays testing phosphate starvation effect on Δ*pstB2*::*gen* mutant strain compared to the WT strain.** Data represents the mean of three replicates. Each graph represents the individual growth of the WT strain **(A)** and the phosphate ABC transporter ATP-binding protein mutant **(B)**. The concentration of phosphate added to the medium is indicated with different colours as shown in the colour legend. The optical density (OD) was quantified at a wave length of 600 nm.

A second minimal medium was tested. In this case was a chemically defined medium previously used to study how a two-component system from *Bifidobacterium breve* UCC2003 respond to phosphate starvation[152] (Material and Methods, section 31, Table 12). As in the previous medium, the phosphate compounds were substitute by MOPS to buffer the medium and potassium phosphate (pH7) was added from 0.1 mM to 100 mM concentrations. Growth curves of the WT and the mutant strain were measured overnight at 37 ºC in the spectrophotometer but no growth was detected in this medium (results not shown).

Therefore, we were not able to found a suitable medium for VRE to characterize the gene that was predicted to be involved in phosphate transport.

## 1.9.3 Characterization of the impact of the LacI family DNA-binding transcriptional regulator on the VRE transcriptome

In order to characterize the LacI family DNA-binding transcriptional regulator (*ltr*R), we performed an *in vivo* experiment using the clindamycin mouse model (section 28.3, Material and Methods). To this end, mice were orally inoculated with the WT or the Δ*ltrR*::*gen* strains ($10^8$ CFUs/200 µl) (N=6 mice per strain). Samples from caecum were collected after 3 days of inoculation and maintained with RNA later for further RNA extraction (Material and Methods, section 23).

Transcriptomic analysis (Material and Methods, section 24) were performed comparing the expression of each gene between the two mice groups, WT and Δ*ltrR::gen*. First of all, we made a global analysis applying PERMANOVA test for the purpose of observing differences between both groups. However non-significant differences were detected (p<0.05). This results suggest that the LacI family DNA-binding transcriptional regulator was not a global regulator. Nevertheless, we studied if the deletion of this regulator could be influencing the expression of specific genes.

To this end, we applied DeSeq2 test, which was designed to study differences between samples using RNA-seq data. From the total of 30 genes identified, whose expression significantly differ between both strains (p<0.05, FDR<0.1, log2 Fold Change (FC) > |2|), 26 genes were overexpressed in the strain mutant for the transcriptional regulator. Thus, this result indicates that the transcriptional regulator LacI may be a repressor and its elimination promoted the expression of this set of genes. Notably, among the most overexpressed genes, it means genes with higher log2FC (Table 31), we found several that are transcribed in the chromosome, contiguously, but in the opposite direction as the LacI family DNA-binding transcriptional regulator, more specifically 3 of them are upstream, which are indicated in bold on the Table 31. Figure 62 represents these three upstream genes that are overexpressed along with a fourth gene, a sucrose phosphorylase, which is also overexpressed but not included in the Table 31 due to a lower log2FC (+3.77).

This result strongly supports the hypothesis that *ltrR* is a repressor of an operon of genes related with the transport of a sugars. Although the nature of the sugar transported by these operon is unknown.

145

**Table 31. Significant genes overexpressed in the mice group infected with Δ*ltrR*::*gen.***

| Locus tag | Product | Log2FC (a) |
|---|---|---|
| EFAU004_RS15640 | hypothetical protein | 5,95 |
| EFAU004_RS15695 | MFS transporter | 5,83 |
| EFAU004_RS15205 | XkdX family protein | 5,82 |
| EFAU004_RS07265 | antitoxin | 5,74 |
| EFAU004_RS12060 | hypothetical protein | 5,71 |
| EFAU004_RS15110 | phenylalanyl-tRNA synthetase subunit beta | 5,71 |
| EFAU004_RS14140 | peptide-methionine (S)-S-oxide reductase | 5,37 |
| **EFAU004_RS14175** | **carbohydrate ABC transporter permease** | **5,08** |
| **EFAU004_RS14180** | **sugar ABC transporter permease** | **5,08** |
| EFAU004_RS14155 | glycosyl hydrolase family 65 protein | 4,96 |
| **EFAU004_RS14185** | **sugar ABC transporter substrate-binding protein** | **4,96** |

[a]Log2FC obtained dividing the average gene expression (transcripts per million) detected *in* the mutant vs the WT. A positive value means a higher expression in mice inoculated with Δ*ltrR*::*gen* strain vs mice inoculated with WT strain. In bold are those genes that are transcribed in the chromosome, contiguously, but in the opposite direction as the LacI family DNA-binding transcriptional regulator.
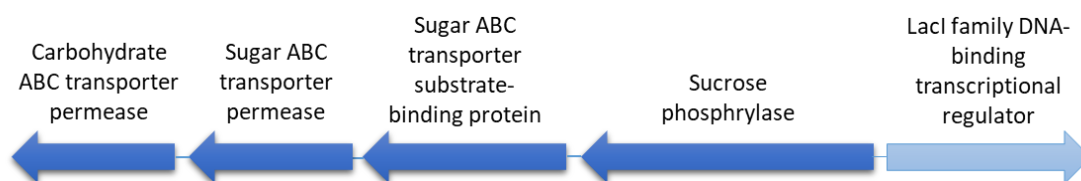


**Figure 62. Upstream genes overexpressed due to the deletion of the LacI family DNA-binding transcriptional regulator.** The deleted gene is represented in light blue and the four genes affected by its deletion are coloured in dark blue. Carbohydrate ABC transporter permease corresponds to locus tag EFAU004_RS14175, sugar ABC transporter permease to EFAU004_RS14180 and sugar ABC transporter substrate-binding protein to EFAU004_RS14185, which are indicated in Table 31. Sucrose phosphorylase with locus tag EFAU004_RS14190 is not included in Table 31 due to lower value of log2FC (+3.77).

## 1.9.4 Detection of the carbon sources utilized for the growth of the wild type and the mutant strains

We next decided to expand our study on the role of two genes that we show to be relevant for gut colonization and that *in silico* studies show to be related to the utilization/acquisition of sugars (i.e. *man*X and *mur*Q). In order to determine the function of genes related to the transport or metabolism of sugars, we used Biolog technology, a 96-well plate, in which each well contains a unique carbon source, including sugars (Material and Methods, section 19, Figure 21). The WT and the two mutant strains related to the use of sugars were previously grown overnight under anaerobic conditions in BHI and resuspended in the M1 medium without carbohydrates (Material and Methods, section 31, Table 14), which has been previously used for identifying genes required for the utilization of sugars in *Enterococcus*[162]. The M1 medium containing VRE was inoculated into the 96 well-plates of Biolog and the plate was grown under anaerobic conditions to monitor the growth of VRE using the different carbon sources.

First of all, we inoculated the WT VRE strain into the 96-well plate to check the most important carbon sources used by VRE. Unfortunately, Biolog plates do not contain the N-acetylmuramic acid to check the importance of this sugar but this was subsequently tested individually as indicated in Figure 65. The results are expressed as percentage of the area under the curve (AUC) referred to the sugar in which VRE reached the maximum growth (i.e. threalose). The figure 63 shows the carbon sources in which VRE grew more than 10 % (compare to the maximum growth) in both replicates. As it can be noted, mannose is one of those sugars in which VRE grew more, besides galactose, glucose and fructose, some of the carbohydrates that are more abundant in the large intestine.
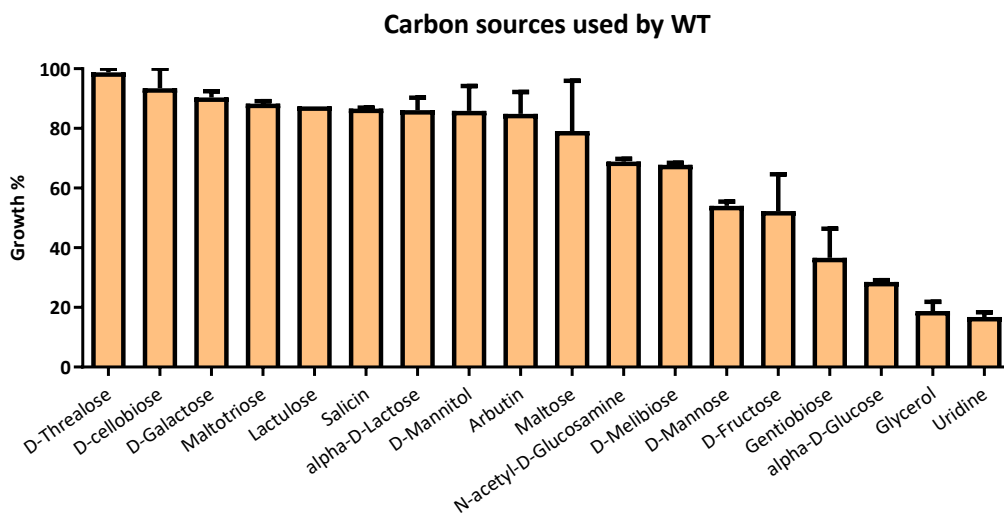


**Figure 63. Biolog assay shows the main carbon sources used by VRE to grow.** WT strain was grown in Biolog 96 well-plate containing unique carbon sources. Two replicates were performed. The growth is represented as the percentage of growth referred to the area under the curve (AUC) of the carbon source that allowed the highest growth (i.e. threalose, which is considered as 100 %). Only carbon sources in which VRE grew more than 10 % are shown. Bars represent the mean and whiskers represent the SEM.

Then, the same experiment was repeated twice with both mutants whose function was theoretically related to the use of mannose and N-acetylmuramic acid, Δ*manX*::*gen and* Δ*murQ*::*gen,* respectively. In the Figure 64 is represented the growth of WT and both mutants for each carbon source. The results obtained are expressed as the area under the curve (AUC) in order to analyse the differences between the growth of WT VRE and the mutants using the different carbon sources. As previously explained, the growth is presented as the percentage referred to the sugar in which the WT strain reached the maximum growth, considering only those nutrients in which the strains grew more than 10 % in both replicates.

The results obtained verified that the deleted *man*X gene was relevant for mannose uptake and transport due to the fact that the growth of this mutant was clearly decreased when mannose was the unique carbon source available (Figure 64). In the case of the deleted *mur*Q gene, we could not use the biolog plate to determine its growth defect using N-acetylmuramic because this carbon source was not present in the

147

biology plate (as described above). Through the Biolog technique we show that these two individual gene deletions did not affect the growth of the VRE for the majority of sugars. Indeed, we only detected a clear growth decrease with mannose and exclusively for the Δ*manX*::*gen* strain and not for the Δ*murQ*::*gen* strain. However, the growth of both mutants was increased in the presence of glucose as compare to the WT, although we do not know the exact reason for this difference. It is important to mention that in this particular experiment we only included two replicates per strain, it will be important to increase these number of replicates in order to better detect, through statistical analysis, subtle growth difference between strains in the presence of different carbon sources. Nevertheless, even with just two replicates, we could confirm that the growth differences between the Δ*manX*::*gen* strain and the WT strain in the presence of mannose were statistically significant. A result that has been confirmed in another experiment, as described below.
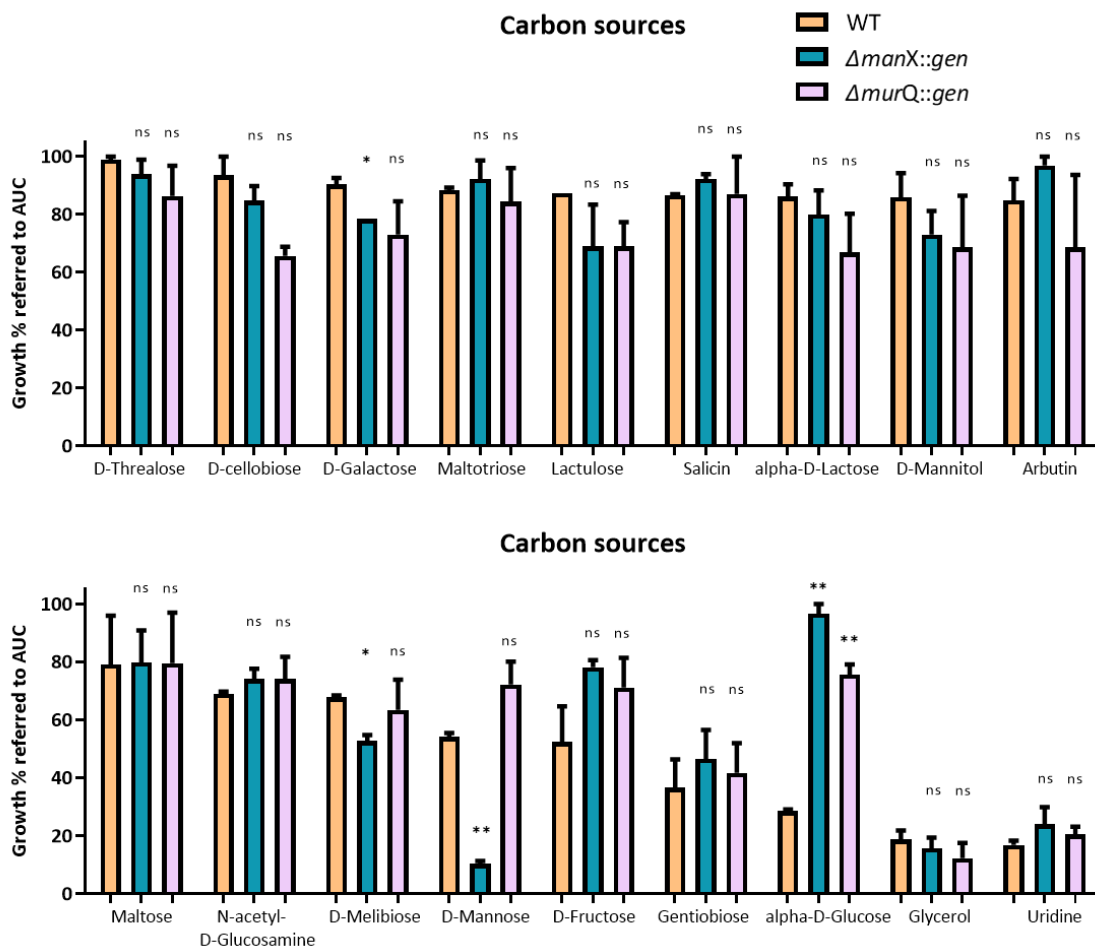


**Figure 64. Biolog assay shows that the *manX* gene is necessary for the VRE growth in the presence of mannose.** WT and mutants Δ*manX*::*gen and* Δ*murQ*::*gen* were grown in a 96 well-plate that contains a unique carbon source in each well. The growth is represented as the percentage of growth refer to the area under the curve (AUC) of the carbon source that allowed the highest growth (i.e. threalose, which is considered as 100 %). The experiment has been done twice for each strain. Only those carbon sources in which we detected a growth higher than 10 % are shown, referred to the sugar in which VRE reached the maximum growth. * p<0.05, ** p<0.01, ns: not significant p>0.05. Unpaired t-test compared to the WT strain. Bars represent the mean and whiskers represent the SEM.

### 1.9.4.1 The *man*X and *mur*Q genes are necessary for VRE to grow in the presence of mannose and N-acetyl muramic, respectively

To confirm the results obtained using the Biolog methodology, we decided to repeat the experiment in regular 96 well plates in which we measure the growth of VRE adding exclusively mannose or N-acetyl muramic acid to the M1 medium. We added the N-acetylmuramic acid that was no present in the Biolog plate to test how mutation of the *mur*Q gene affected the growth of VRE when this particular sugar was the unique carbon source added to the medium. In addition, we could also verify the growth defect of the *man*X mutant in the presence of mannose as the unique carbon source.

Figure 65 shows the results obtained, which demonstrated that the studied genes were necessary to the use of the specific sugars tested. The WT strain grew efficiently in the presence of both sugars. As observed in the Figure 65, the results obtained for Δ*man*X::*gen* were similar to the Biolog results (Figure 64). This mutant strain grew less in the presence of mannose and the growth curve shape change completely from the typical sigmoid growth curve to a linear growth. The growth of this mutant was almost not affected when N-acetyl muramic acid sugar was added to the medium as a carbon source.

In the case of the Δ*mur*Q::*gen* mutant, the growth deficiency was clear and more evident. The deletion of the *mur*Q gene, theoretically required for the metabolism of N-acetyl muramic acid, did not allow VRE to grow in the presence of N-acetyl muramic acid. However, in the presence of mannose, the growth of the Δ*mur*Q::*gen* mutant was similar to the WT strain, which was able to grow in the presence of both sugars.

Thus, the results obtained demonstrated that *man*X and *mur*Q were relevant for the growth of VRE in the presence of mannose and n-acetyl-muramic acid, respectively.

**M1 Medium
+ 5 mM mannose**

**M1 Medium
+ 5 mM N-acetylmuramic acid**
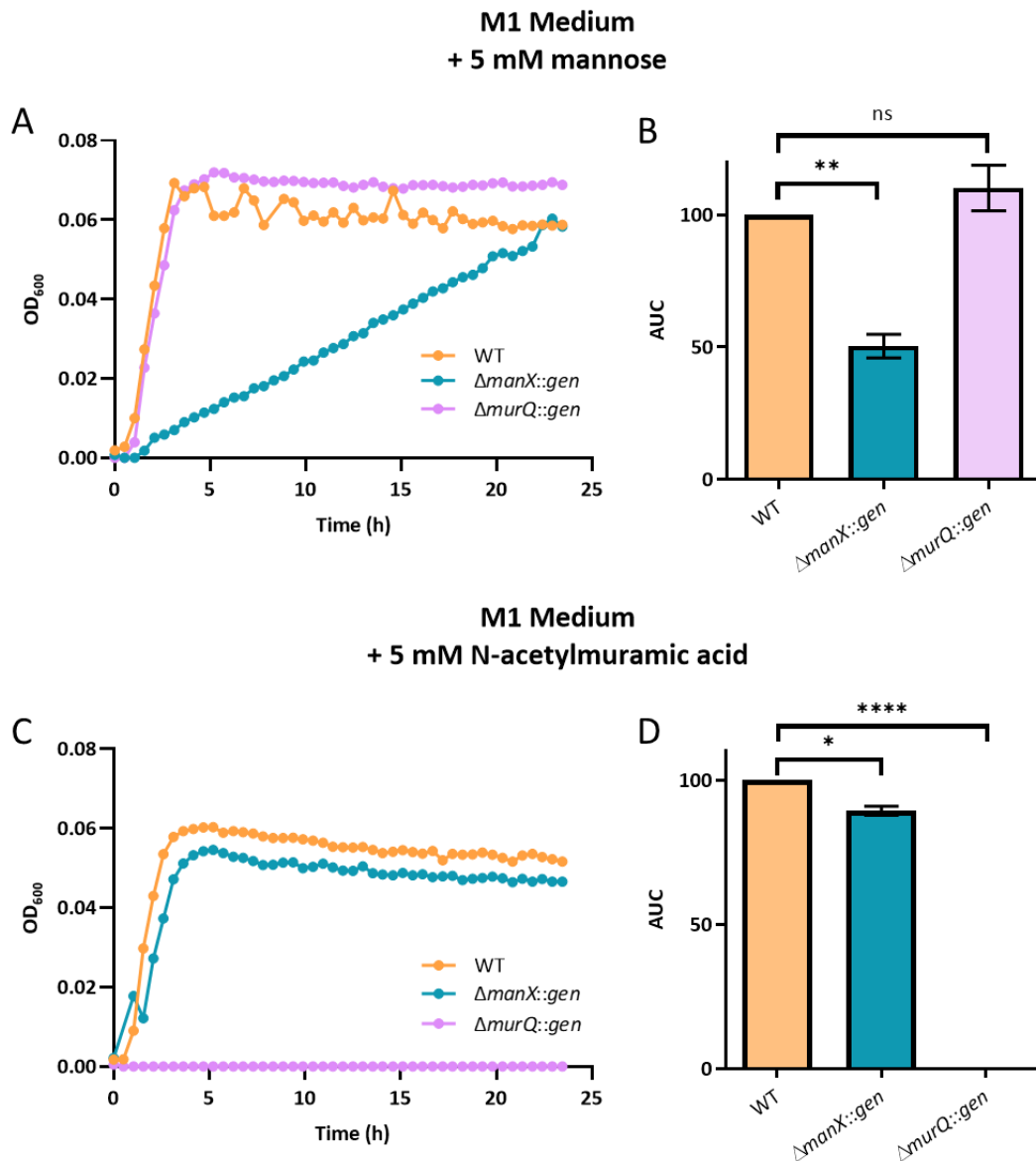
**Figure 65.** *In vitro* **cultures with specific carbon sources confirmed the role of two genes (***manX* **and** *murQ***) in the utilization of sugars mannose and N-acetylmuramic acid.** Growth assays with M1 medium containing mannose or N-acetylmuramic acid sugars as unique carbon sources. Graphs **A** and **C** represent the growth curve of the strains tested (WT, Δ*manX*::*gen* and Δ*mur*Q::*gen*) as the average from the two replicates for each condition. Graphs **B** and **D** represent the area under the curve (AUC) refer as the growth percentage of each of the replicates in M1 medium. The growth percentage is referred to the growth of the WT, whose AUC is considered 100%. The optical density (OD) was quantified at a wave length of 600 nm. * $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$, ns: not significant $p > 0.05$, unpaired t-test compared to the WT strain. Bars represent the mean and whiskers represent the SEM.

### 1.9.4.2 Complementation of mutant strains involved in the use of specific carbon sources confirms the role of specific genes in utilization of specific sugars

In order to confirm the role of the *man*X and *mur*Q genes in the use of mannose and N-acetylmuramic acid sugar, we decided to introduce copies of each gene to the mutant strains using the pNZ44 plasmid, which was suggested and provided by Dr. Van Sinderen. This plasmid, which contains a broad range promotor from *Lactococcus lactis,* can drive the expression of genes in different Gram-positive bacteria, and is used routinely to perform complementation experiments. Cloning was carried out to introduce the two deleted genes into the plasmid (Material and Methods, section 21). After verifying the sequences of the generated constructions, the plasmids were electrotransformed into each mutant strain.

To check if the expression of these genes encoded by the plasmid could restored the function of the previous deleted genes from the genome, the WT, mutant and complemented mutant strains were grown in the previous M1 medium adding 5 mM of sugars, mannose and N-acetyl muramic acid. As shown in Figure 66, when each gene was expressed in the pNZ44 plasmid, the mutant strains could grow again using N-acetyl muramic acid and mannose as carbon sources. The growth curve of the complemented strains was completely restored being similar to the growth curve of the WT.

These results indicate that the function of the previously deleted genes could be restored by expression of the genes in trans and the phenotypic effect was due to the deletion of these specific genes and not to polar effects of the deletion.

151
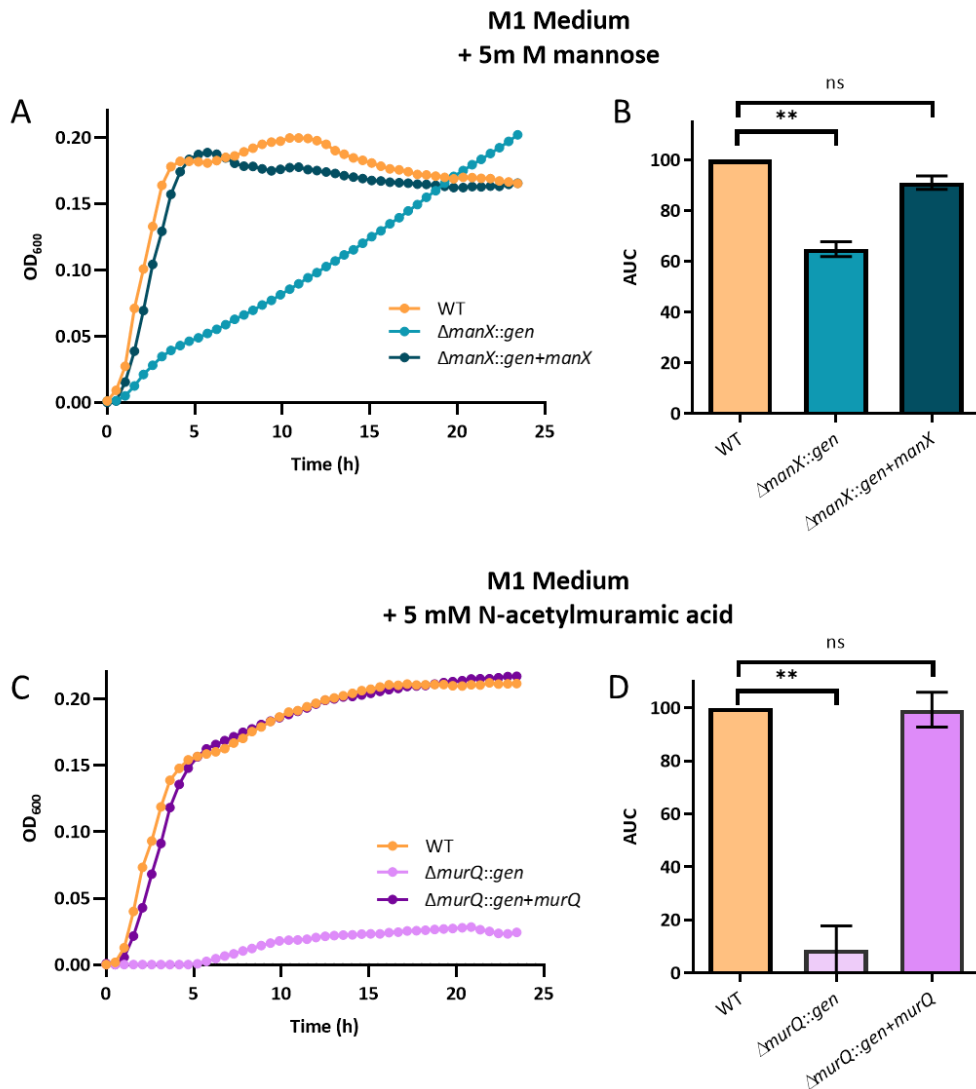
**Figure 66. Gene complementation to verify function of the mutants Δ*manX*::*gen* and Δ*murQ*::*gen* in vitro.** Growth tests were performed in M1 medium containing the sugars of interest (mannose or N-acetylmuramic acid) and the different strains (WT, mutant and complemented mutant). The results show that the expression in trans of the specific gene in the mutant strain restores the function and the growth of the Δ*manX*::*gen* and the Δ*murQ*::*gen* mutants. Graphs **A** (mannose) and **C** (N-acetyl muramic acid) represent the growth curve of the strains during 24h as the average from the two replicates for each condition. Graphs **B** and **D** represent the area under the curve (AUC) refer as the growth percentage of each of the replicates in M1 medium. The growth percentage is referred to the growth of the WT, whose AUC is considered 100%. The optical density (OD) was quantified at a wave length of 600 nm. ** $p < 0.01$, ns:not significant $p > 0.05$. Unpaired t-test compare to the WT strain. Bars represent the mean and whiskers represent the SEM.

**1.9.4.3 Gene complementation of mutant strains partially restore VRE growth *in vivo***

Our previous results indicated that gene complementation of the mutant strains involved in the use of carbon sources restored the phenotype detected in the WT strain *in vitro*. In order to check if this also occurred *in vivo*, we performed mouse experiments co-inoculating the WT and the complemented strains. For comparisson, we also co-inoculated other mice with the WT and the mutant strains. *In vivo* assays were carried out in both mouse models previously explained and tested: the clindamycin and the vancomycin mouse model.

**1.9.4.3.1    Gene complementation in the clindamycin mouse model partially restores mutant growth *in vivo***

First, to verify if the gene function is restored *in vivo,* we used the clindamycin mouse model and gavage mice orally with a mix of WT and each of the complemented mutant strains (Δ*manX*::*gen+manX* and Δ*mur*Q::*gen+mur*Q). Strains were grown overnight in BHI plates, mixed in a 1:1 ratio and orally gavage to 6 mice per group at the 4$^{th}$ day of the clindamycin treatment. We included 2 mice as a control group that were co-inoculated with the WT and each of the mutants (Δ*man*X::*gen* and Δ*mur*Q::*gen*). Stool samples were collected three days post-infection.

In Figure 67, the results obtained from Δ*man*X::*gen+manX* mixed with the WT showed that expression of the *man*X gene favours the growth of the complemented strain. In fact, there was higher growth of the complemented strain in most of the mice, resulting in a ratio close to 1 comparing the growth of the complemented strain (Δ*manX*::*gen+manX)* and the WT strain. On the other hand, co-inoculation of WT and Δ*mur*Q::*gen+mur*Q provided similar results, detecting a growth increase of the complemented mutant and obtaining a WT/complemented strain ratio close to 1. With regard to the control groups, we observed results similar to the previous competitions assays with the *murQ* mutation. However, for an unknown reason, in this particular experiment we did not detect a growth defect in the *manX* mutant.

**Complementation *in vivo* experiment**
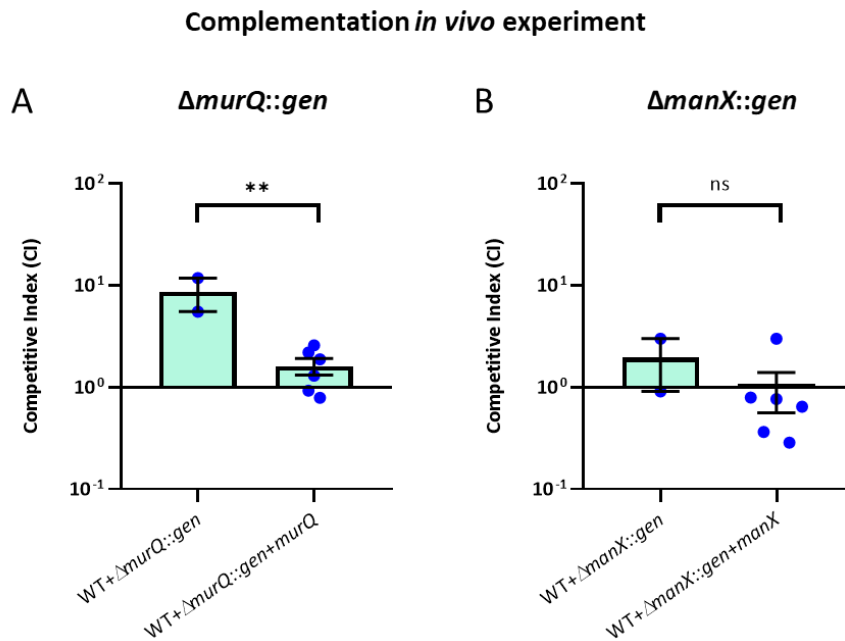


**Figure 67. *In vivo* competition assays of the complemented mutants Δ*manX*::*gen+man* and Δ*murQ*::*gen+murQ*.** Clindamycin-treated mice are co-inoculated with WT strain and Δ*mur*Q::*gen* mutant or its complemented strain **(A)** or co-inoculated with WT and Δ*manX*::*gen* mutant or its complemented strain **(B)**. The competitive index for each mutant was calculated as previously described and plotted. ** p<0.01, ns:not significiant p>0.05. Unpaired t-test. Bars represent the mean and whiskers represent the SEM.

### 1.9.4.3.2 Gene complementation in the vancomycin mouse model restores the growth of the mutant strains progressively

In order to validate the results obtained with the clindamycin mouse model and to follow the progression of both strains in the mouse intestine for a longer period of time, we performed the competition assay in the vancomycin mouse model (Material and Methods, section 28.3, Figure 25). Mice were treated with vancomycin in the drinking water and orally gavage at day 4$^{th}$ of the treatment. WT, mutants and complemented mutants were collected from BHI plates and mixed in equal proportion in order to inoculate mice with them. As in the previous experiment, for comparison we added another group, WT mixed with each mutant strain. Faeces samples were collected at 3$^{rd}$, 6$^{th}$ and 13$^{th}$ post-infection.

The results from the infection with the mix of WT and the complemented mutant: Δ*man*X::*gen+manX* (Figure 68A) indicate that the expression in trans of the previous deleted gene did not restore the growth of the complemented mutant strain three days post-infection. However, at day 13$^{th}$ post-infection we observed an increase of Δman*X*::*gen+manX* levels, resulting in a CI closer to 1 (compare to the WT), whereas the control group, WT mixed with Δ*manX*::*gen*, maintained the higher WT levels during the entire experiment. We notice however, that for some reason, the defect growth of the *manX* mutant in this particular experiment was not so pronounce as in previous experiments. This result indicates that there is some degree of variability in the effect of

154

this mutation on the *in vivo* growth, depending on the experiment. Reasons for this variability will be discussed in the discussion section.

Regarding the experiment with the co-inoculation of the WT and the complemented mutant Δ*mur*Q::*gen*+*mur*Q (Figure 68B), we observed the same pattern as explained for the other mutant. We did not observe the restorage of the growth of the mutant when the gene was complemented in trans three days post-VRE inoculation. However, at day six post-VRE inoculation, we detected a clear difference between both groups because the complemented strain restored the growth and the levels between the complemented strain and the WT were very similar, resulting in a CI of 1. This result was maintained thirteen days post-VRE inoculation. However, in the case of the co-inoculation of the WT and the mutant strain, we detected both at day 6 and 13 an *in vivo* growth defect of the mutant Δ*mur*Q::*gen*, although more pronounce at day 13.

As commented before, the results of this experiment suggest that the complementation of this two specific genes restored the growth of the mutants under *in vivo* conditions.
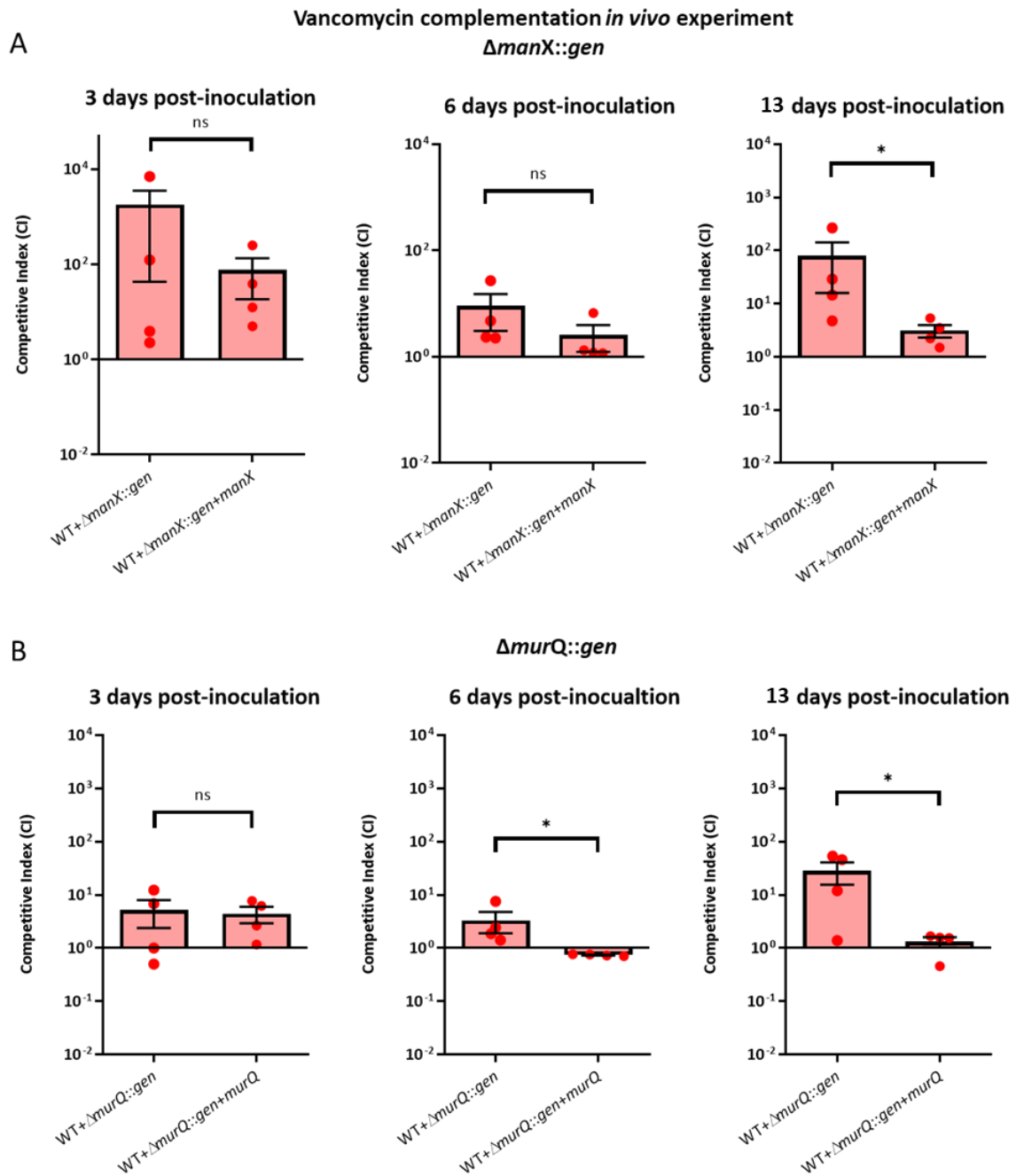
**Figure 68. Vancomycin *in vivo* competition experiments of the complemented mutant strains Δ*man*X::*gen*+*manX* and Δ*mur*Q::*gen*+*murQ* and the WT VRE strain.** Mice were co-inoculated with the WT and the mutant or complemented mutant strains, **(A)** Δ*manX* and its complemented strain **(B)** Δ*murQ* and its complemented strain. Levels of colonization are represented as competitive index (CI) between the WT and each mutant strain. * $p < 0.05$, ** $p < 0.01$, ns:not significant $p > 0.05$. Unpaired t-test. Bars represent the mean and whiskers represent the SEM.

**1.9.4.4 Administration of mannose to antibiotic treated mice seems to slightly increase the levels of VRE in the intestine**

After verifying the growth deficiency of the Δ*man*X::*gen* compared to the WT strain in the intestine of mice treated with antibiotics and the reasoning that mannose could be a limited nutrient, we performed an *in vivo* experiment adding mannose. Mannose was added to mice water to check if an excess of mannose in the intestine enhances VRE growth. Mice were treated with vancomycin in the drinking water during three days, inoculating at the fourth day with the WT VRE strain and replacing the antibiotic with water. Three days post-VRE inoculation we collected faeces to check VRE levels (Material and Methods, section 28.4, Figure 26). The idea of stopping the antibiotic treatment was to allow a decrease in VRE levels due to a partial microbiota recovery (we have observed this in previous results, Introduction, Figure 8). By doing this, we expected that the boosting effect of mannose on VRE growth would be easier to detect.

However, as shown in Figure 69A, VRE levels remained very high three days after removing the antibiotic (3rd day post inoculation in the graph). These levels were similar to those observed in previous experiments with vancomycin in which the antibiotic was not removed. This result was probably due to the fact that three days of antibiotic removal may not represent enough time for the microbiota to partially recover and decrease VRE levels. In any case, after quantifying VRE levels, one group of mice received mannose while the other group was maintained with regular drinking water. One day after mannose addition, no differences were detected in the levels of VRE comparing mice that received or not the sugar (Figure 69B). However, after two days of mannose addition, we detected a tendency of increasing the VRE levels in the group receiving mannose, although not statistically significant as compare to the control group (Figure 69C). A similar experiment could not be performed with the N-acetyl muramic acid since this compound is provided in very low levels by companies.

157
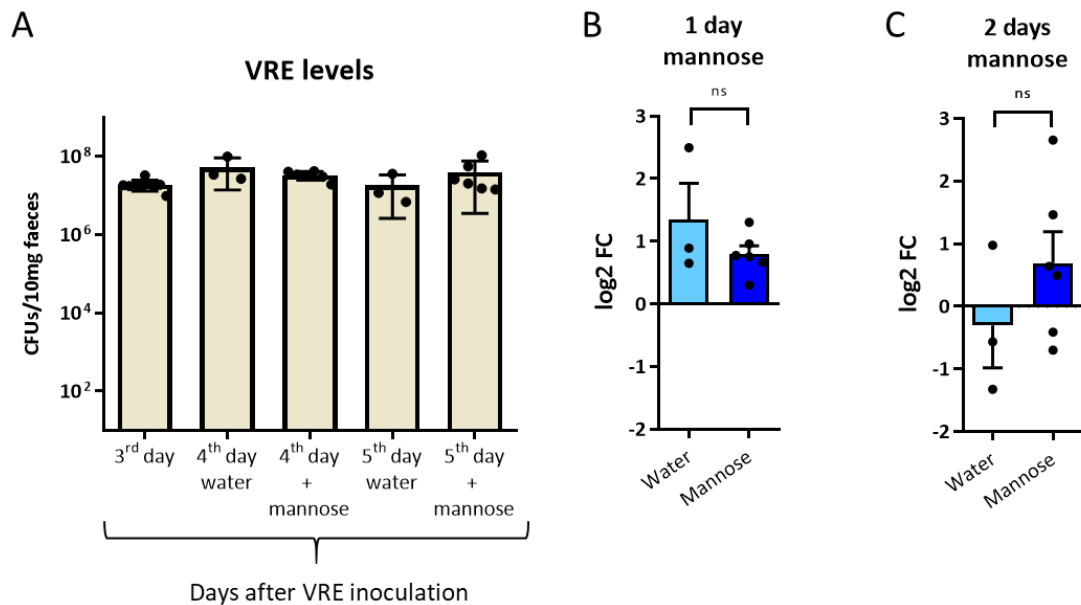
**Figure 69. Study the effect of mannose supplementation on VRE intestinal growth *in vivo*. (A)** Mice were treated during 3 days with vancomycin in their drinking water. The fourth day mice were inoculated with VRE. Three days post-VRE inoculation the levels of VRE colonization (CFUs/10mg) were analyzed in faeces. A group of mice (N=6) received mannose, while the other group of mice (N=3) was maintained with regular drinking water. The levels of VRE in faeces were analysed one and two days post-mannose administration (four and five days post-VRE inoculation). **(B)** Change in the levels of VRE (log2FC) one day after the addition of mannose or **(C)** two days after the addition of mannose. ns:not significant p>0.05, Unpaired t test. Bars represent the mean and whiskers represent the SEM.

## 1.10    Prevalence of the five selected candidate genes in clinical *E. faecium* isolates

After analysing the possible function of five genes (*man*X, *mur*Q, *ltr*R, *pst*B2, proteins from [Fe-S] cluster) in the gut colonization process of VRE, we decided to expand the study by analysing if the presence of those genes was enriched in clinical strains or if on the contrary, they were encoded by both commensal and clinical strains.

For this purpose, we searched by blast alignments the nucleotide sequence of these genes in the 73 strains described in the previous study of Lebreton et al.[57], in which they determined the classification of these 73 strains in the different clades of *E. faecium* (15 strains from clade B (the majority being human commensal isolates), 2 strains with recombinant genome between clade A and B and 56 strains from clade A: A1 (mainly clinical isolates) an A2 (mainly animal derived isolates). Figure 70 represents the presence or absence of the five candidate genes in each of the 73 strains used for this study.

**Genes**

| Clade | Strain | Category | manX | murQ | ltrR | pstB2 | [Fe-S] cluster |
|---|---|---|---|---|---|---|---|

**Category legend:**
- Human commensal
- Human infection
- Hospital faecal
- Animal
- Other
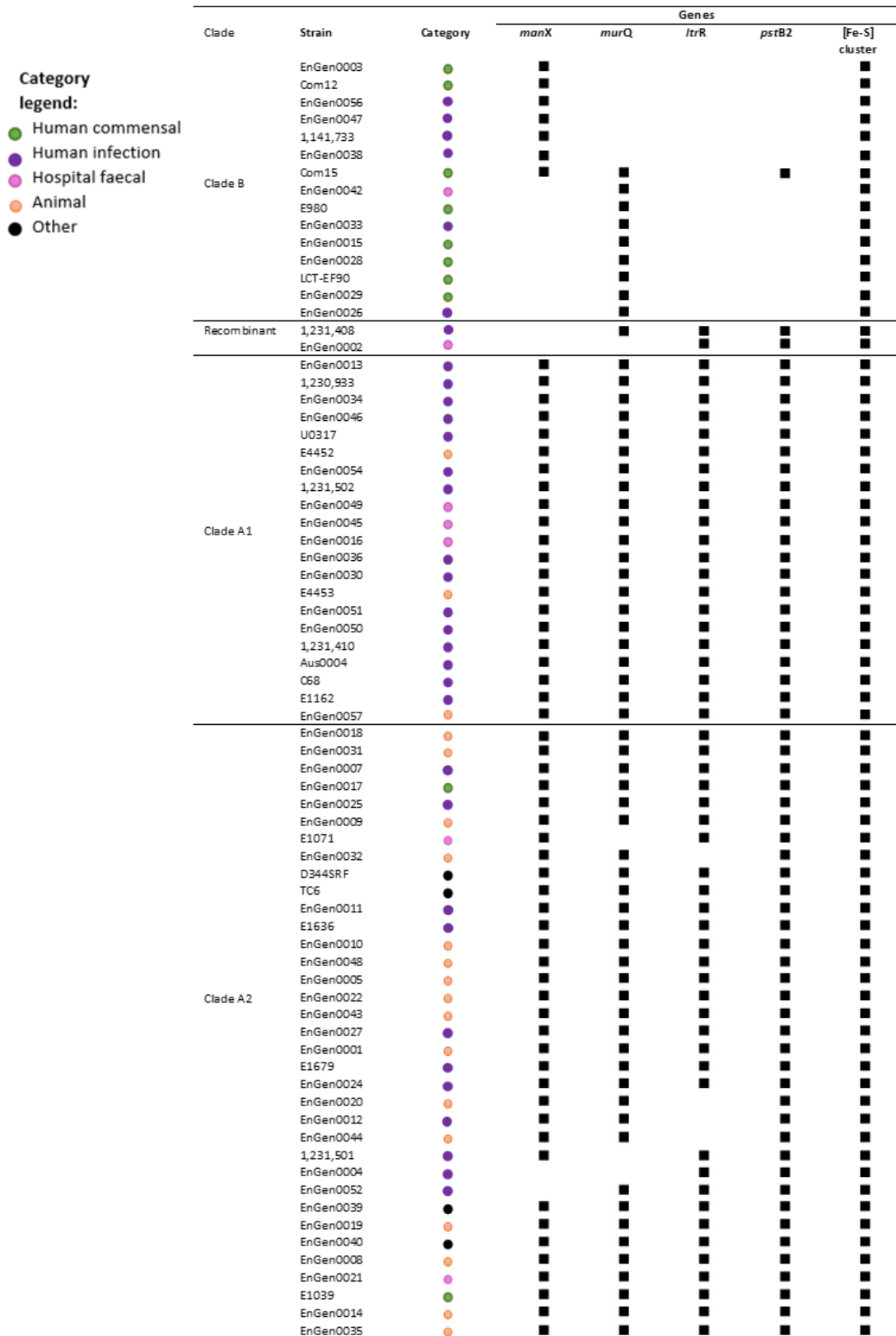
**Figure 70. Prevalence of the 5 studied genes involved in VRE gut colonization in different strains of *Enterococcus faecium*.** Each colour indicates strain origin: green indicates clade B; red: hybrid origin between clade A1 and B; purple: clade A1 and, orange: clade A2. Black squares indicate the presence or absence of the five selected candidate genes in each strain. On the left side is indicated the clade for which each strain belongs to.

As shown in Figure 70, some of the genes we have investigated are mainly present in clinical strains while others are present in both clinical and commensal strains. For those genes that were found in both clinical and commensal isolates (e.g. *man*X) a bootstrapped neighbour-joining tree was constructed in order to study if the sequences of the gene could differ depending on the origin of the strain (i.e. clinical vs commensal). Subsequently, we described the results obtained for each particular gene.

With respect to the genes *ltr*R (LacI family DNA-binding transcriptional regulator) and *pst*B2 (phosphate ABC transporter ATP-binding protein), we confirmed that these genes were absent in commensal strains, excepting *pst*B2 that was encoded by one commensal strain but with a sequence that significantly differ from the gene encoded by Aus0004 (92.57% of identity). The gene *lac*I was also absent in 4 strains of clade A, specifically from the clade A2 corresponding to the animal clade. However, the *pst*B2 gene was present in all the clade A strains studied. Thus these two genes are enriched in clade A strains.

In the case of the *man*X gene (PTS mannose transporter subunit EIIAB), we detected that this gene is present in all the strains from clade A (mainly clinical and animal strains) while it is present in less than 50% of the strains from clade B (7 out of the 15). This result suggested that these gene is enriched in strains from clinical and animal origin, a result that was verified statistically ($p < 0.0001$; Fischer test). As it can be checked in the phylogenetic tree (Figure 71A), the clade B strains containing the gene are present in a different branch of the tree compared to the strains from clade A. Nevertheless, the differences in the nucleotide sequence between the commensal and clinical strains are minimal since the nucleotide identity is higher than 98%, and at the amino acid level the proteins are identical. A similar result was obtained for the *mur*Q gene (N-acetylmuramic acid 6-phosphate etherase), being absent in in 6 out of the 15 clade B strains. But also absent in 1 recombinant strain and 3 clinical strains. However, in this particular case, when a tree was built with the sequences of murQ, we could not found any clear separation by clades (Figure 71B).

Finally, the [Fe-S] cluster was present in all strains, both from clade B and clade A. Interestinlgy, however, we detected certain degree of clustering by clade origin although not as clear as with *man*X (Figure 71C).

From these results we conclude that only *ltrR* and *pst*B2 are genes clearly enriched in *E. faecium* clade A strains (including those with clinical origin and animal origin). This result suggests that these genes contribute to intestinal colonization specifically in clinical strains and not human commensal strains, since the deletion of those genes impaired the VRE ability to colonize the murine gut. In addition, these genes will also probably contribute to the intestinal colonization of strains with animal origin. On the other hand, the fact that *man*X, *mur*Q and the Fe-S cluster are not specific of clinical strains suggest that these genes could also be relevant for the colonization process of both clinical and commensal strains.
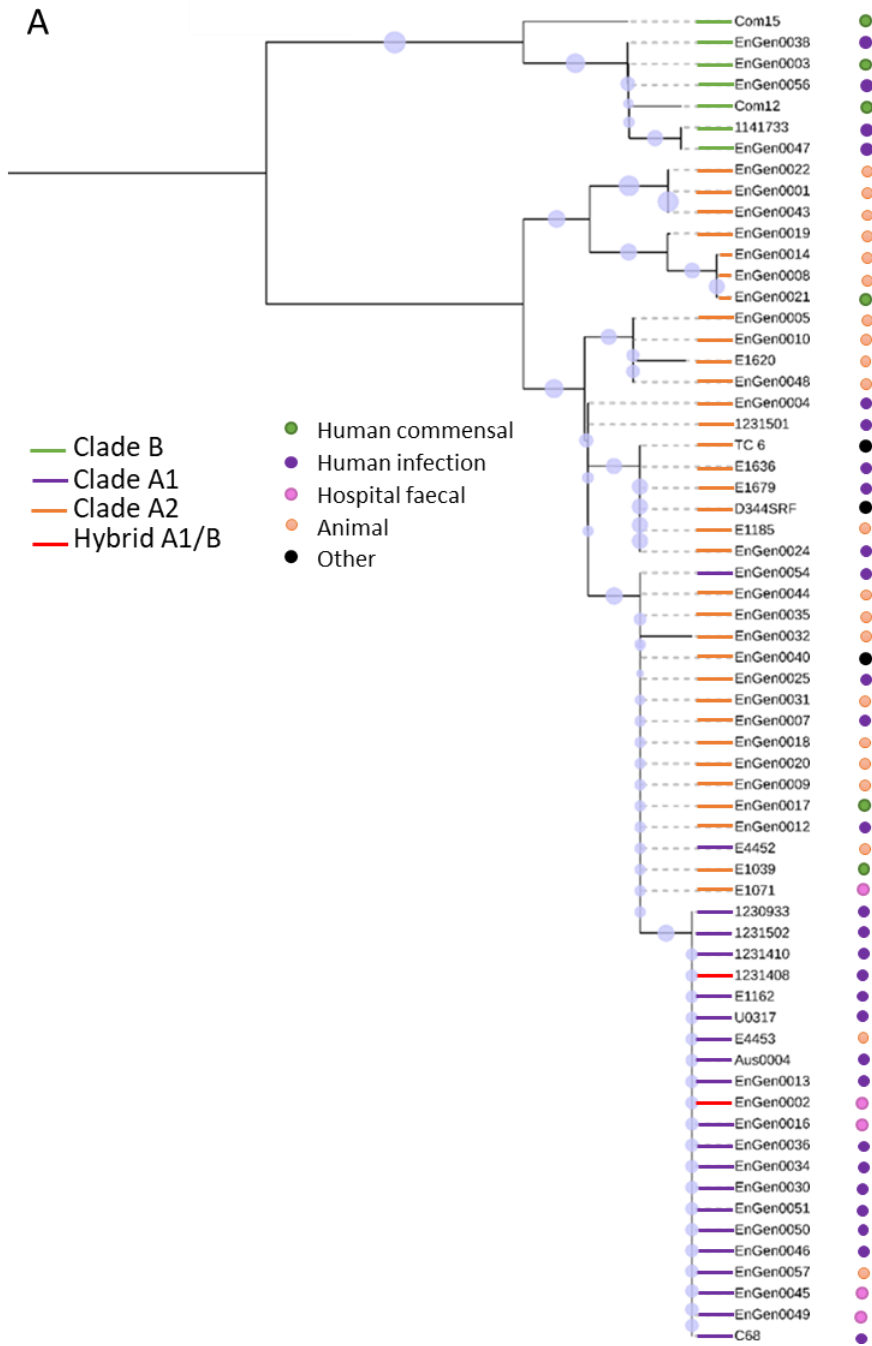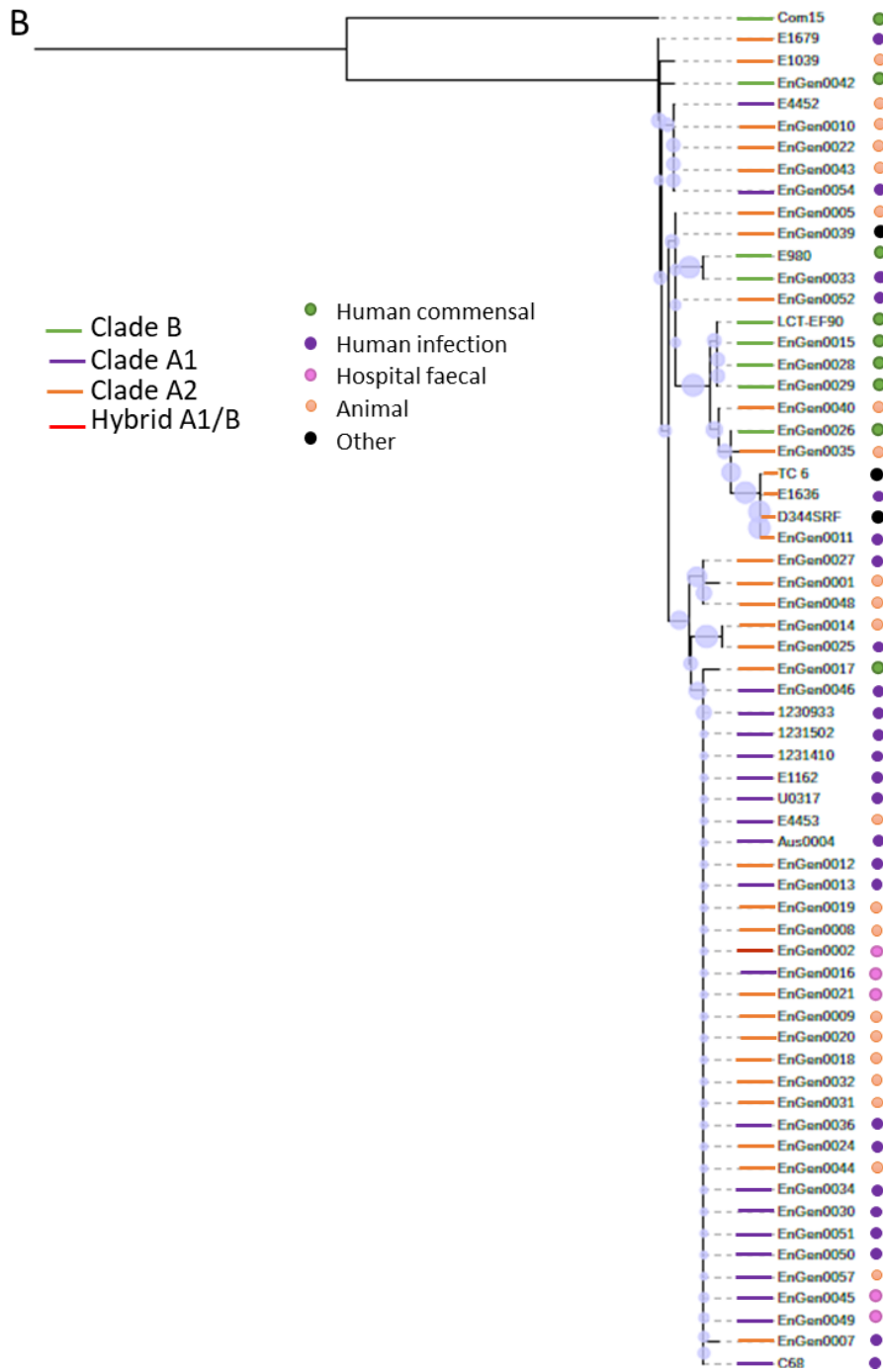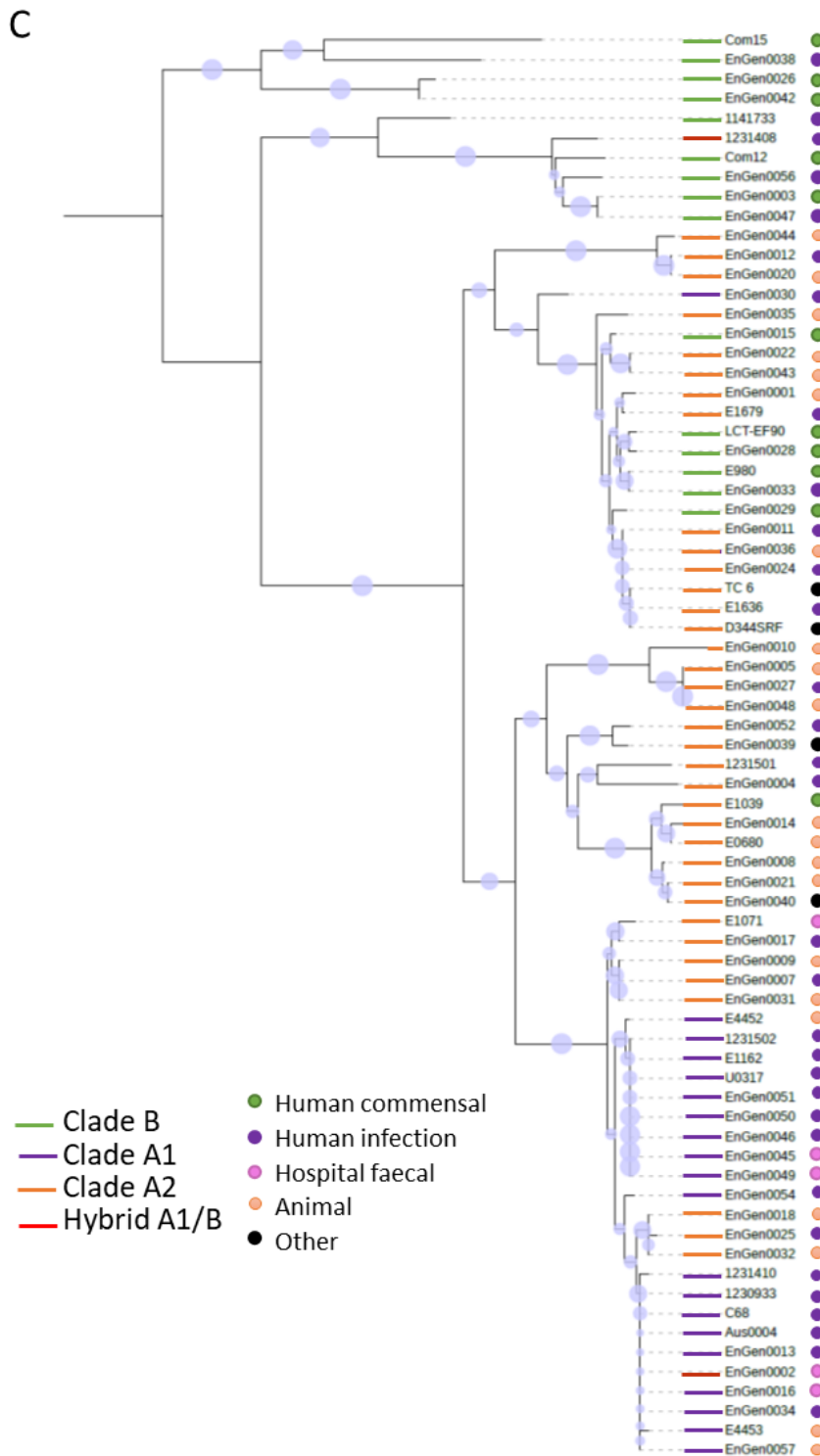
**Figure 71. Tree based on nucleotide similarity with the 73 *E. faecium* genomes.** Bootstrapping was performed with 500 replicates in MEGA X using statistical method of neighbour-joining. The nucleotide sequence of the three genes *man*X **(A)**, *mur*Q **(B)** and [Fe-S] cluster **(C)** from the Aus0004 strain were compared against the 73 *E. faecium* strains[57] to study the similarity of these genes in strains from different origin (clinical, human commensal, animal). The circles in purple display de bootstrap data.

## 1.11    Transcriptomic analysis comparing the gene expression from *in vivo* vs *in vitro* samples

Apart from using the transposon mutant library to study genes encoded by VRE to colonize the gut, we also performed a transcriptomic analysis in order to identify genes whose expression was upregulated *in vivo* as compare to *in vitro*. This could be an alternative approach to identify genes relevant for VRE gut colonization that our Tn-Seq approach could not identify (e.g. duplicate genes that would required a double mutation). We performed a preliminary analysis comparing the expression of the Aus0004 WT strain grown in BHI medium (N=3 replicates)*,* to the expression 3 days post-VRE inoculation of this strain in the colon content from mice that were treated with clindamycin (Material and Methods, section 23)*.* Initially, 3 mice were colonized, however for one of the mouse, the sequencing library did not work, so we could only analyse the VRE transcriptome *in vivo* of 2 mice. Nevertheless, we decided to include these results in the thesis although we are aware that the experiment should be repeated.

We compared the expression of each gene *in vivo* vs *in vitro* to identify the genes that are most expressed during intestinal colonization. Firstly, we performed a canonical correlation analysis in order to observe general differences between both groups of samples. Although the replicates per groups are limited, both groups were clearly separated (Figure 72).
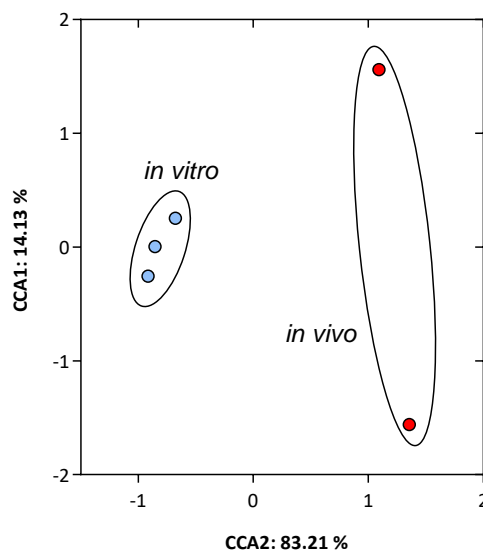


**Figure 72. Canonical correlation analysis (CCA) of *in vivo* and *in vitro* samples.** The samples are separated in two groups based on the expression of genes in each condition tested. CCA1 and CCA2 are coordinates 1 and 2 respectively, showing the percentage of the variability explained by each one.

Subsequently, we applied DeSeq2 test in order to detect genes whose expression differ *in vivo* vs *in vitro.* Among the significant genes identified, we found some of the genes that we had demonstrated through targeted mutagenesis to be involved in the gut

colonization process (2 from the 10 studied genes; p<0.05; FDR<0.1). Surprisingly, the expression of *suf*B (the gene of the Fe-S cluster) was higher *in vitro* than *in vivo* (p<0.05; Table 32). On the other hand, the bifunctional acetaldehyde-CoA/alcohol dehydrogenase was overexpressed in the colon samples (p<0.05). These results indicate that the level of expression not always correlate with the relevance of a particular gene for bacterial growth in a particular environment. Nevertheless, as mentioned, this analysis should be repeated with more mice.

**Table 32. Genes, from those previously studied through targeted-mutagenesis, whose expression significantly differ *in vivo* vs *in vitro*.**

| Protein | Product | Log2FC (a) |
|---|---|---|
| **EFAU004_RS02800** | Fe-S cluster assembly protein SufB | -2.03 |
| **EFAU004_RS01175** | bifunctional acetaldehyde-CoA/alcohol dehydrogenase | +3.56 |

[a]Log2FC obtained dividing the average gene expression (transcripts per million) detected *in vivo* vs *in vitro*. A positive value means a higher expression *in vivo* than *in vitro*.

On the other hand, we obtained 798 significant genes (p<0.05, FDR<0.1, log2 Fold Change (FC) > |2|), 681 of them that where more expressed *in vivo* than *in vitro* and 117 more expressed *in vitro* vs *in vivo* (Appendix, Table 36)*. We found several overexpressed genes *in vivo* related to energy adquisition. This is the case of ornithine carbamoyltransferase, which is an enzyme involved in the catabolic pathway of the arginine and agmatine degradation. Moreover, within those genes more expressed *in vivo* we detected several subunits of PTS for carbohydrates such as lactose (8 subunits), fructose (2), mannitol (2), mannose (5), or PTS for multiple sugars (mannose, fructose and sorbose, 4). However, we did not find any gene encoding a PTS for sugars overexpressed *in vitro*. These results reflect the importance of the PTS systems for VRE which facilitates the acquisition of different carbohydrates present in the gut.

## 2. STUDY OF COMMENSAL BACTERIA WITH PROTECTIVE CAPACITY AGAINST INTESTINAL COLONIZATION BY VRE

As indicated in the Introduction (section 2.2.3), results derived from the work of a previous PhD student from our group identified several commensal bacteria that were associated with protection against VRE colonization[168]. Moreover, strains from several of these commensal bacteria were isolated, including *Barnesiella, Olsenella, Alistipes, Oscillibacter* and unclassified_Ruminococcaceae. Notably, administration of a consortium containing these 5 bacterial isolates (ProBac – **pro**tective **bac**teria) to vancomycin treated mice was able to significantly restrict VRE intestinal colonization (Introduction, section 2.2.3, Figure 11).

The results described by the previous PhD student suggested that ProBac was key for suppressing VRE intestinal colonization. However, this result did not rule out the possibility that in general the administration of bacteria to antibiotic treated mice could have a protective effect against VRE. To evaluate this, a strain from the murine gut belonging to the genus *Bacteroides* was isolated. This genus was not associated with protection against VRE, as oppose to the other taxa included in ProBac (Introduction, section 2.2.3, Figure 9). To test if, as expected, *Bacteroides* would not have any effect against VRE colonization, we made use of a previously set up mouse model that was utilized to test the protective effect of ProBac. Briefly, mice were treated for one week with vancomycin in the drinking water and subsequently treatment was replaced by water. After antibiotic withdrawal, a group of mice was inoculated with the bacterial consortium ProBac, another group of mice received the control media suspension used to resuspend the bacteria (PBS with glycerol 20% and cysteine 0.1 %) and a third group of mice received the *Bacteroides* isolate. Then, mice were let for two weeks to allow the stabilization of the microbiota (as indicated in the Introduction, section 2.2.3, Figure 8). Subsequently, mice were inoculated with VRE and stool samples were collected two days after VRE inoculation.

The results showed that levels of VRE colonization were reduced in the group inoculated with ProBac, confirming their role in protection (Figure 73A). However, VRE levels were not reduced in the group that received *Bacteroides,* indicating that the protective effect was specific of ProBac and that the administration of other bacterium did not had any effect against VRE colonization. As shown in Figure 73B, *Bacteroides* could indeed colonize the intestine of mice to high levels, confirming the ineffective role of *Bacteroides* in protection against VRE.
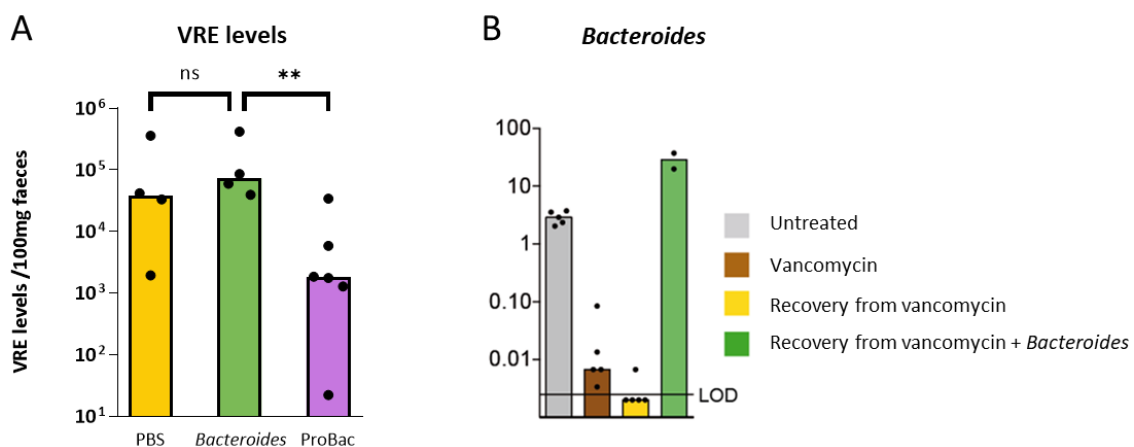


**Figure 73. VRE levels after the administration of ProBac or *Bacteroides*. (A)** Vancomycin treated mice were inoculated, after antibiotic withdrawal, with ProBac, *Bacteriodes* (bacteria not correlated with protection against VRE) or with a suspension control media (PBS with 20 % of Glycerol and 0.1 % of cysteine). After two weeks of recovery period mice were inoculated with VRE and two days after faeces were plated in selective media to evaluate the levels of colonization by VRE. **(B)** Bacteroides abundance of fecal samples of untreated mice; mice treated with vancomycin; mice treated with vancomycin that were allowed to recover for two weeks and that received *Bacteroides* (+*Bacteroides*) or not, immediately after stopping antibiotic treatment. ** $p < 0.01$, ns:not-significat $>0.05$, Wilcoxon test. Bars indicate the median.

## 2.1 Transcriptomic analysis demonstrated that ProBac colonization restores the expression of transporters for the internalization of sugars

Results previously obtained from our group have suggested that commensal bacteria, including ProBac, could inhibit VRE growth directly, in the absence of the host[28,168]. Moreover, as our group has previously shown, commensal bacteria can confer protection to VRE colonization in the absence of major components of the immune system (i.e. Toll like receptors, B cells and T cells). Altogether these results were suggesting that ProBac may be conferring protection to VRE colonization in a direct manner, without inducing the response of the immune system.

In order to decipher mechanisms of ProBac protection, we compared the bacterial functions expressed in the microbiomes of untreated mice or mice treated with vancomycin that received either ProBac or the bacterial control media suspension (PBS with glycerol 20 % and cysteine 0.1 %). In addition, we sequenced the genomes of ProBac isolates, which allow us to identify those bacterial functions specifically encoded and expressed by ProBac isolates.

We used the mouse model with two weeks of recovery period (Material and Methods, section 28.5, Figure 27). A group of mice did not receive antibiotics, another group was treated with vancomycin and inoculated with PBS-glycerol-cysteine (the suspension control media) and the last group was inoculated with ProBac. Mice were cohoused in pairs. Two weeks after antibiotic cessation, one of the mice of each pair from each cage was sacrificed. Part of the caecum was resuspended in RNA later for further RNA extraction (Material and Methods, section 23) while the other part of the caecum was frozen for subsequent metabolomics analysis (Material and Methods, section 26). The other mouse from each cage was inoculated with VRE and the capacity of VRE to colonize the intestine was monitor in faeces, as previously described, two days after VRE inoculation. We have previously shown that mice that are co-housed in the same cage recover a very similar microbiota and get colonize with very similar levels of VRE[168]. Thus, this mouse could be used to associate changes detected in the cecum with the capacity of VRE to colonize the intestinal tract.

To analyse the transcriptome of the different group of mice, we extracted the RNA from the caecal samples using RNeasy Power Microbiome Kit (Material and Methods, section 23). The library was constructed for sequencing on a Nextseq platform (150 bp, pair-end).

After checking that these bacteria were able to colonize the intestine (task performed by a previous PhD student, Introduction, section 2.2.3, Figure 12) we checked that the administered bacteria were functionally active. The expression of specific genes encoded by *Barnesiella*, *Alistipes* and *Olsenella* could be detected in treated-mice that received ProBac as well as in untreated mice (Figure 74). However, expression for the majority of these genes was undetectable in treated mice that did not received ProBac. This result indicates that these three isolates were able to colonize the gut and restore the expression of specific bacterial genes that was lost upon antibiotic treatment. In

167

contrast, expression of genes encoded by *Oscillibacter* and unclassified_Ruminococcaceae were identified in all three groups of mice. This is consistent with the previous colonization data (Figure 12), which showed that treated mice spontaneously recover similar bacterial isolates from these two taxa and strongly suggest that these two taxa are not sufficient to prevent VRE colonization.
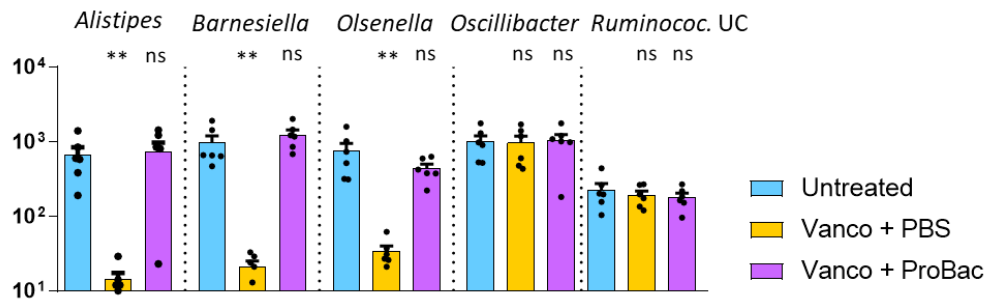


**Figure 74. Genes expressed by individual commensal bacteria from ProBac in each group of mice.** Number of genes expressed detected by the isolates that constitute ProBac in untreated mice and mice that were treated with vancomycin and received either ProBac (Vanco+ProBac) or PBS glycerol cysteine (Vanco+PBS). *$p<0.05$, **$p<0.01$, ns ns:not-significat >0.05, Wilcoxon test, compare to the untreated group. Bars represent the mean and whiskers represent the SEM.

Next, we evaluated to what extent the administration of Probac could modify the functional capacity of the microbiome. To this end, KEGG orthologs (KO) were assigned to non-redundant genes identified in the murine caecal transcriptome (Material and Methods, section 25).

Vancomycin treatment significantly diminished the diversity and richness of the expressed KOs (Figure 75). However, this diversity and richness was partially recovered in those mice that received ProBac. This last result suggests that the administration of ProBac was able to restore the expression of specific bacterial functions (KOs), whose expression was lost upon vancomycin administration and that therefore could be relevant for suppressing VRE colonization.
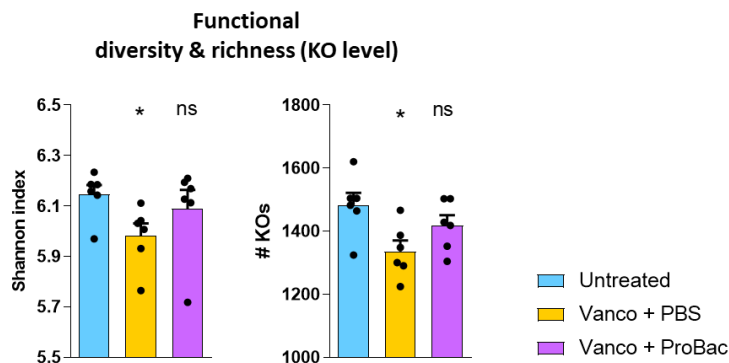


**Figure 75. Diversity and richness of the expressed KOs.** Diversity and richness of the KEGG orthologs expressed in the 3 groups of mice: untreated mice and mice that were treated with vancomycin and received either ProBac (Vanco+ProBac) or PBS glycerol cysteine (Vanco+PBS). *$p<0.05$, ns ns:not-significat >0.05, Wilcoxon test. Bars represent the mean and whiskers represent the SEM.

Subsequently, DeSeq2 statistical analysis was applied in order to identify those KOs whose expression significantly increased (>2 fold; q<0.05) upon Probac administration and that reach similar levels as the expression detected in untreated mice. After applying DeSeq2, we identified a total of 6 KOs that matched the defined criteria (Table 33).

**Table 33. KOs more expressed (>1log$_2$; q<0.05) in the group of mice receiving ProBac.**

| KOs | Function |
|---|---|
| K01295-cpg | peptidase |
| K00174-korA | oxidoreductase |
| K02770-fruA | fructose PTS system EIIBC or EIIC component |
| K01425-glsA | glutaminase |
| K16213-cbe | cellobiose epimerase |
| K03969-pspA | phage shock protein A |

Most of these KO were assigned to enzymes which have not been described to contribute to colonization resistance against pathogens, including K01295 (peptidase), K00174 (oxidoreductase), K01425 (glutaminase). In addition, K03969, putatively encoding a phage shock protein, was also overexpressed in the group of mice that received ProBac. Enterococcal phages have been shown to diminish intestinal colonization by VRE[179]. We found that K03969 was expressed by one of the commensals (*Olsenella*) in most of the detected mice (Figure 76A). *Olsenella* is an Actinobacteria and therefore phylogenetically unrelated to *Enterococcus* (Firmicutes). Taking into account the very narrow spectrum of phages, we thought that it was unlikely that a phage present in *Olsenella*, encoding this KO, could be contributing to resistance to VRE colonization.

Noteworthy, two additional KOs: K02770 and K16213, that were overexpressed in the group of mice that received ProBac, encoded for proteins required for the uptake and catabolism of specific sugars. Theoretically, the expression of these KOs could contribute to VRE inhibition through sugar depletion, a mechanism that has been characterized for other pathogens such as *Enterobacteriaceae* but not for VRE. K16213 encodes a cellobiose epimerase while K02770 encodes for *fru*A, a subunit of the phosphotransferase system, specific for the transport of fructose. K16213 was not found to be expressed by ProBac isolates in only 50% of the analysed mice (Figure 76A). However, the KO encoding for the fructose transporter was expressed by one of the administered isolates (i.e. *Olsenella*) in all analysed mice. *Olsenella* genome contains 4 genes that encode for *fru*A (KO2770). However, expression of *fru*A by *Olsenella* was mainly driven by one single gene: OLS _9050, in all the six analysed mice (Figure 76B). As expected, this gene could not be detected in vancomycin treated mice that did not received ProBac (Figure 76B), while its expression was detected in all analysed untreated mice. In addition to *fru*A, in most of the analysed mice that received ProBac (66-100%), it was possible to detect transcripts expressed by *Olsenella* that encoded for the other subunit of the fructose transporter (*fru*B:K02768) and the other genes required for the

incorporation of fructose into the glycolytic pathway: *fru*K (K00882) and fructose-biphosphate-aldolase (FBA, K01624) (Figure 76C). These results suggested that ProBac and more specifically *Olsenella* express *in vivo* bacterial functions for the uptake and utilization of fructose.
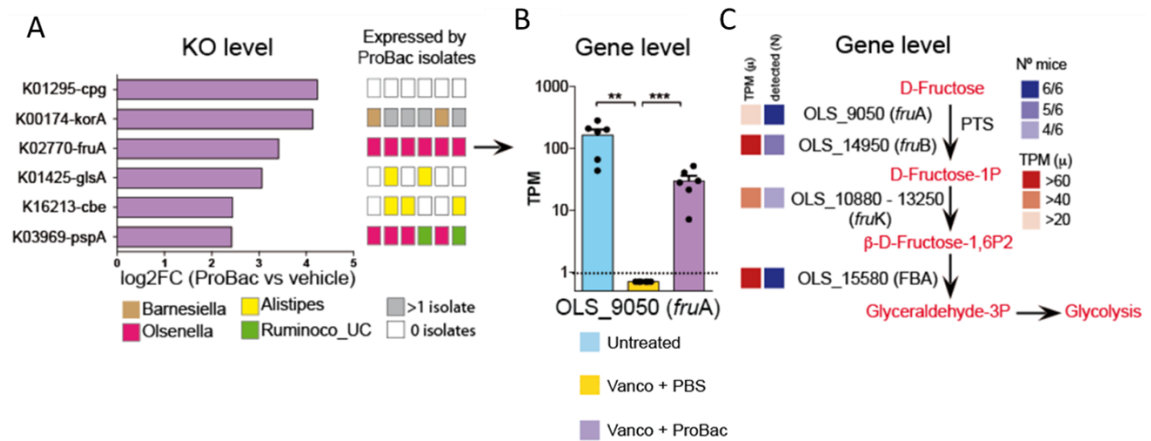


**Figure 76. ProBac restores the expression of specific bacterial functions, including a fructose transporter, which were lost after vancomycin treatment. (A)** KOs whose expression is significantly higher (>1log2; q<0.05) in the group of mice Vanco+ProBac as compare to the group Vanco+PBS but does not differ between Vanco+Probac and Untreated. Colours of the panel on the right indicate if any of the isolates of ProBac express that particular KO in each one of the mice from the group Vanco+Probac. **(B)** Expression in the different groups of mice of the gene encoded by *Olsenella* that was assigned to the KO2770, which encodes a subunit of a PTS transporter for fructose. **(C)** Expression of genes encoded by *Olsenella* that are required for the internalization and metabolization of fructose. The heatmap on the left indicates the TPM for each gene (red colours) and the number of mice in which that particular gene was detected in the group (Vanco+Probac). **p<0.01, ***p<0.001, ns:not-significat >0.05, two-sided t-test. Bars in B represent the mean. Whiskers represent the SEM.

## 2.2  ProBac diminish the levels of fructose in the large intestine

The transcriptomic analysis suggested a lower capacity of the cecum microbiota of vancomycin treated mice to internalize specific sugars, as compare to the microbiota of untreated mice or mice that received ProBac. If this was the case, we would expect to detect higher levels of those specific sugars in the cecum of mice treated with vancomycin. To this end, we performed targeted metabolomics analysis (Material and Methods, section 26). As expected, the levels of fructose were significantly higher in mice treated with vancomycin as compare to untreated mice (Figure 77). Interestingly, the levels of fructose were significantly reduced in those mice that received Probac. Altogether, these results suggest that during vancomycin administration, commensal bacteria that express gene orthologs required for sugar transport (i.e. fructose) are depleted, or the expression of these KOs is reduced, which leads to the higher availability of sugars (i.e. fructose). Conversely, Probac administration restores the expression of these KOs, which led to a decrease of fructose.
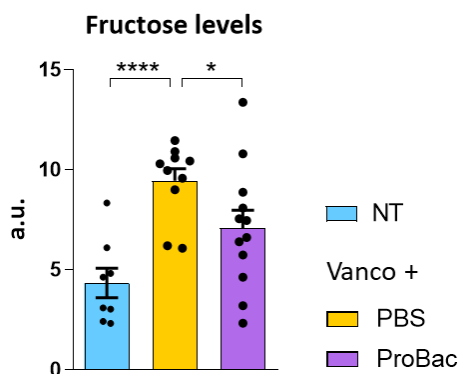
**Figure 77. Metabolomics analyses identify reduced levels of fructose in mice receiving ProBac.** Levels of fructose found in the different groups of mice: untreated mice (NT), vancomycin treatment (Vanco+PBS) and treated mice receiving ProBac (Vanco+Probac), showing that intestinal fructose is significantly reduced in mice that received ProBac. *p<0.05, ****p<0.0001, Unpaired t-test. Bars represent the mean. Whiskers represent the SEM.

## 2.3  Detection of main carbon sources required for VRE ATCC 700221 to grow

The results described above suggested that an increase availability of specific sugars (i.e. fructose) may promote the growth of VRE after vancomycin administration. On the other hand, Probac, by diminishing the level of these sugars, specifically fructose, may restrict VRE colonization. To test this hypothesis, we first determine the key carbohydrates used by VRE to grow *in vitro*. We repeated the Biolog methodology, the 96 well-plate containing a unique carbon source in each well, because in this case the clinical strain used was the ATCC 700221, whereas in the first part the construction of the transposon mutant library was performed with the clinical strain Aus0004.

The Figure 78A shows those sugars in which VRE growth was at least a 10 %, as compare to the sugar that allowed a highest growth. This result confirmed that fructose was one of the sugars that allow a highest VRE growth. Indeed, VRE was able to grow in fructose to a similar extend as in glucose (Figure 78B), one of the central sugars of bacterial metabolism. Besides fructose, we detected that VRE was able to grow in several of the sugars in which the strain Aus0004 was able to grow, including mannose.

A

**Carbon sources used by VRE**



B

**Growth curves**
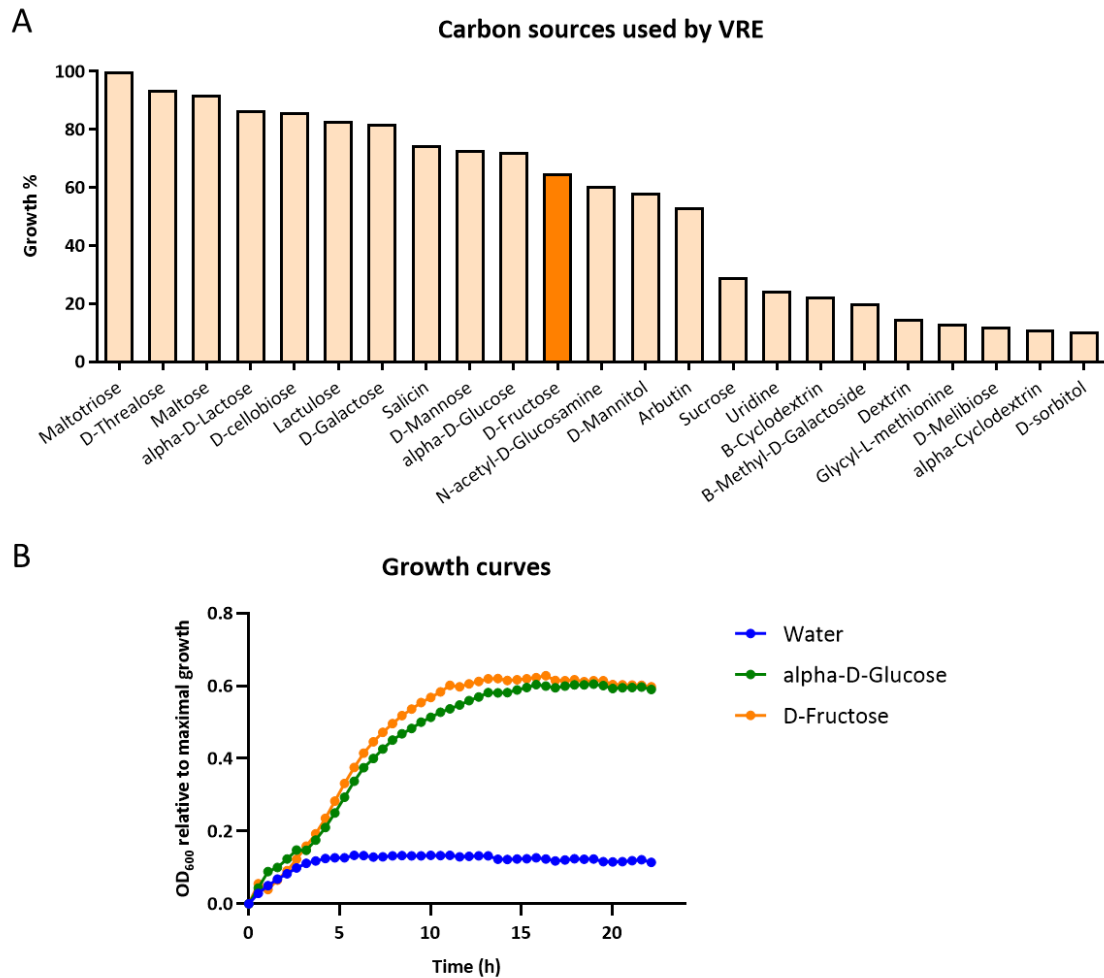


**Figure 78. Main carbohydrates used by VRE strain ATCC 700221 to grow.** VRE strain ATCC 700221 was added to Biolog plate containing 96 different carbon sources. **(A)** Growth is represented as the AUC referred to the sugar in which VRE reached the maximum growth (100%). Sugars with more than 10% of growth are shown. Other carbon sources that are not carbohydrates are not indicated in the figure. **(B)** Growth curve referred to the maximal growth (maltotriose) of the growh of VRE using fructose or glucose as carbon sources. Bars represent the AUC.

## 2.4 Fructose promotes VRE intestinal colonization

The previous results have shown that VRE efficiently utilize fructose as a nutrient carbon source *in vitro*. We next evaluate if fructose levels could be relevant for VRE to colonize the intestinal tract. To this end, we first performed a correlation analysis between the levels of fructose detected in mice treated with vancomycin that had received or not ProBac, and the levels of VRE colonization in the pair co-housed mice (Material and Methods, section 28.5). We used this indirect approach since (i) the sugar content of the cecum likely represents better the amount of sugars available for bacterial growth than the levels of sugar excreted in feces and (ii) analyzing the caecal levels of fructose post-VRE inoculation and the VRE colonization levels in the same mouse could be misleading since VRE could have an impact on fructose levels.

As shown in Figure 79, we found a significant and positive correlation between the levels of fructose and the capacity of VRE to colonize the intestinal tract of the co-housed mice. Although this is just an association, this result was suggesting that the availability of fructose in the gut was key for VRE to be able to colonize the intestinal tract.
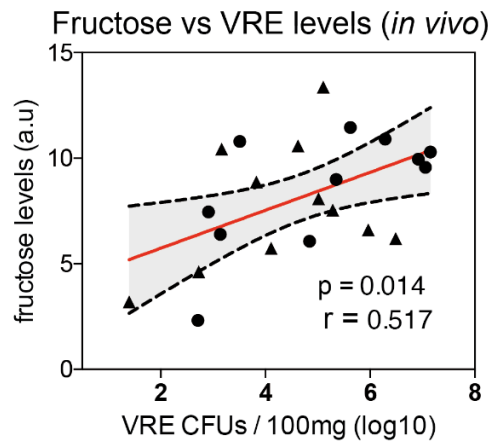


**Figure 79. Correlation between the levels of fructose present in the intestine and the ability of VRE to colonize the intestine.** Pearson correlation analysis positively correlates the ability of VRE to colonize the intestine (log10 CFUs/100mg) with the levels of fructose (a.u. area units) present in the caecum of co-housed mice. The P and rho values for the Pearson correlation analysis are shown in the Figure.

In order to test this hypothesis, we modified the mouse model experiment adding fructose to the drinking water of mice that had been previously treated with vancomycin and colonized with VRE after antibiotic cessation (Material and Methods, section 28.6, Figure 28). Faecal samples were collected immediately before fructose administration and two days after fructose addition to study the dynamics of VRE intestinal colonization in the presence or absence of fructose in the different groups: control (PBS) and *ProBac*. Notably, the levels of VRE significantly increased in those mice that received fructose, as compare to mice that were maintained in regular water (Figure 80). This result suggested that fructose promotes the growth of VRE *in vivo*. Moreover, the increase of VRE levels in the presence of fructose was abolished if mice had been previously colonized with ProBac suggesting that ProBac abolished fructose-dependent VRE expansion.

Altogether these results suggest that fructose enhance VRE intestinal colonization and that ProBac prevents VRE growth through fructose depletion.

**VRE levels comparison**
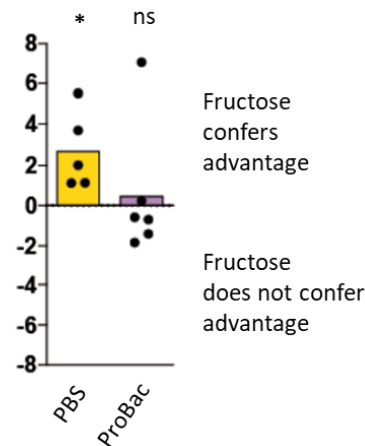


**Figure 80. Fructose in drinking water significantly increases VRE levels**. VRE growth advantage in mice that received fructose vs mice that did not received fructose 2 days after initiating fructose administration. See method section for the calculation of the growth advantage index and the experimental design. This experiment was performed in mice that were treated with vancomycin and allowed to recover for two weeks before VRE inoculation. Mice received during three consecutive days either PBS or ProBac, starting the day after antibiotic cessation. * $p < 0.05$, ns:not-significat $>0.05$. One-sample t-test from the value "0", a significant result indicates that fructose promotes VRE growth. Bars represent the mean.

## 2.5 *Ex vivo* experiments demonstrated protection by nutrient competition conferred by ProBac consortium against VRE

The results described above indicate that fructose is relevant for VRE to colonize the intestinal tract and that ProBac may be conferring protection against VRE colonization through nutrient competition by depleting the levels of fructose. In order to study if ProBac protects against VRE by nutrient competition, an *ex vivo* experiment was designed, using the caecal content of mice treated with vancomycin and allowed to recover from antibiotic treatment for two weeks. Briefly, the intestinal contents were collected, filtered and reduced in an anaerobic chamber (Material and Methods, section 28.8, Figure 30). Subsequently, the filtered contents were inoculated with $10^8$ CFUs of each bacterium of the ProBac mix and were incubated for 24 h in anaerobic conditions. Subsequently, VRE was added to the culture and aliquots were plated after 6 h of incubation to quantify VRE levels.

As shown in Figure 81, VRE was not able to grow in the presence of ProBac. As a control, we perform the same experiment but instead of ProBac we used *Bacteroides*, which did not confer resistance to VRE *in vivo* to verify that inhibition of VRE growth is confer by ProBac but not by other commensal bacteria. The results showed that, as expected, *Bacteroides* did not diminish VRE growth *ex vivo*.
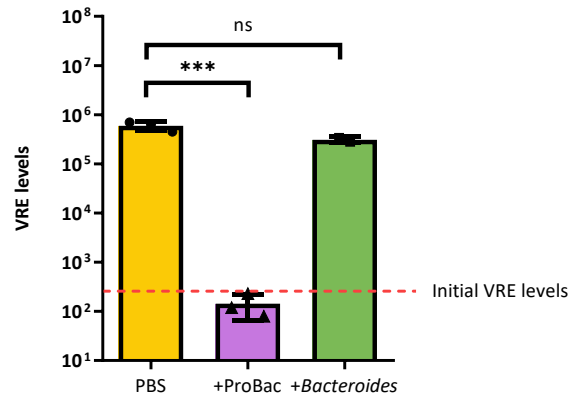
**Figure 81. ProBac diminish the growth of VRE *ex vivo*.** Filtered caecum samples were incubated for 24 h under anaerobic conditions with ProBac (N=3), *Bacteroides* (N=2) or without bacteria (control, N=3), adding subsequently VRE. Samples were plated in selective media after 6 h of incubation with VRE. The addition of ProBac clearly reduces the growth capability of VRE, observing significant differences. *** $p < 0.001$, ns: not significant $> 0.05$, Unpaired t-test. Bars represent the mean and whiskers represent the SEM.

Subsequently, we studied if the mechanism of inhibition of the pathogen was due to nutrient competition by carbohydrates. We repeated the same experiment but adding an excess of carbohydrates. We added a mix of sugars that are frequently found in the gut (Material and Methods, section 27, Table 8) or an individual carbohydrate (i.e. fructose) after the incubation of the caecal contents with the bacterial consortium, at the same time that we added VRE to the filtered caecum contents. As it is shown in Figure 82, VRE growth was partially restored when an excess of sugars was administered. Furthermore, a similar effect was detected when an excess of only fructose was added. This result suggests that VRE growth inhibition by ProBac is due to competition for nutrients and that this effect can be restored by the addition of an excess of sugars frequently present in the gut, including fructose. Interestingly, the excess of sugars did not enhance VRE growth in the absence of ProBac. This last result suggested that the filtered caecal contents of mice treated with vancomycin contain sufficient nutrient sources for VRE to reach maximal growth.
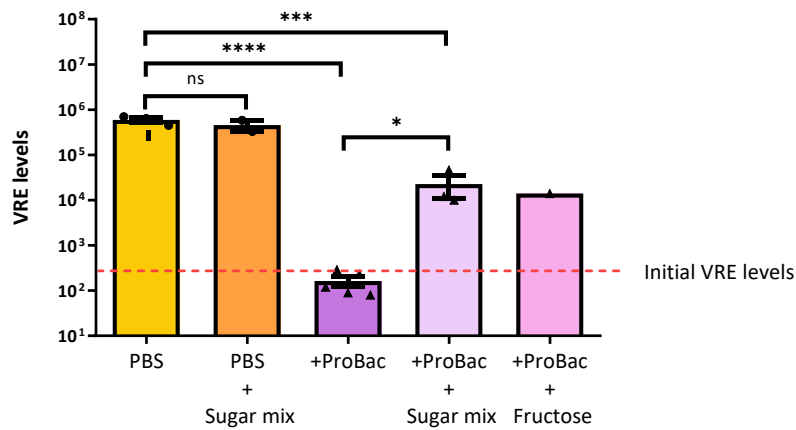
**Figure 82. The addition of a mixture of sugars or just fructose partially restores the growth of VRE in the presence of ProBac.** Sugars and/or VRE were added to filtered cecal contents incubated with ProBac or without ProBac (control) for 24 h. Subsequently the culture was incubated again 6h under anaerobic conditions and aliquots were plated in selective media to determine VRE levels. The same experiment was performed in filtered caecal contents without ProBac (Control). Between N=2 and N=5 samples per conditions. In this case we only introduced one sample with fructose but this result was repeated, as shown in Figure 85. * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$, ns:not significant $> 0.05$, Unpaired t-test. Bars represent the mean and whiskers represent the SEM.

Besides confirming the reduction of the VRE growth by the bacterial consortium ProBac, we tested the same type of experiment but with individual bacterium to study if any of the ProBac bacteria was sufficient to inhibit VRE growth. Figure 83 shows that *Olsenella* was able to inhibit VRE growth, obtaining similar results than when we add the five isolates together.
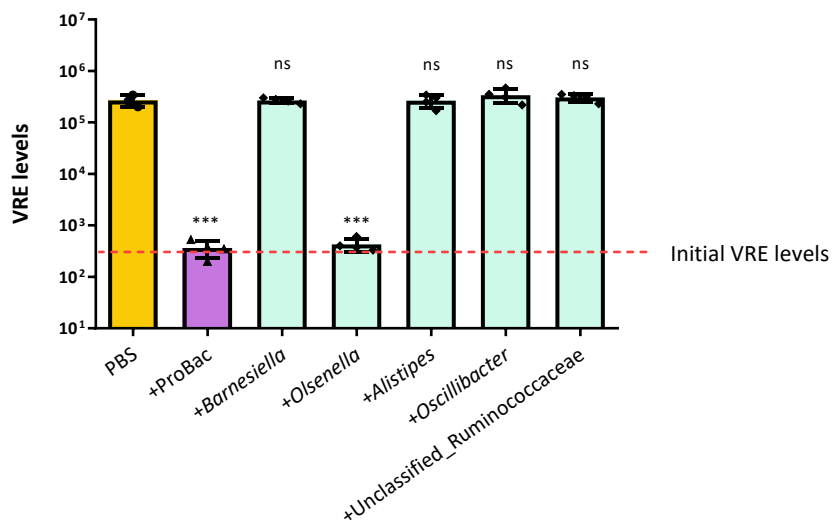


**Figure 83. *Olsenella* inhibits VRE growth *ex vivo* to a similar extend as Probac.** Filtered caecum was incubated both with ProBac and each individual bacterium from the consortium (N=4). After 24 h of incubation, VRE was added and the culture was incubated under anaerobic conditions at 37 ºC. Subsequently, aliquots were plated in selective media to quantify the levels of VRE after 6 h of incubation. *Olsenella* was able to inhibit VRE growth *ex vivo* similar to the ProBac consortium. *** $p < 0.001$, ns:not significant $> 0.05$, Unpaired t-test compare to the control group. Bars represent the mean and whiskers represent the SEM.

176

Notably, the rest of the bacteria were not able to inhibit VRE growth, despite similar growth of each commensal bacteria previous to VRE inoculation (Figure 84).
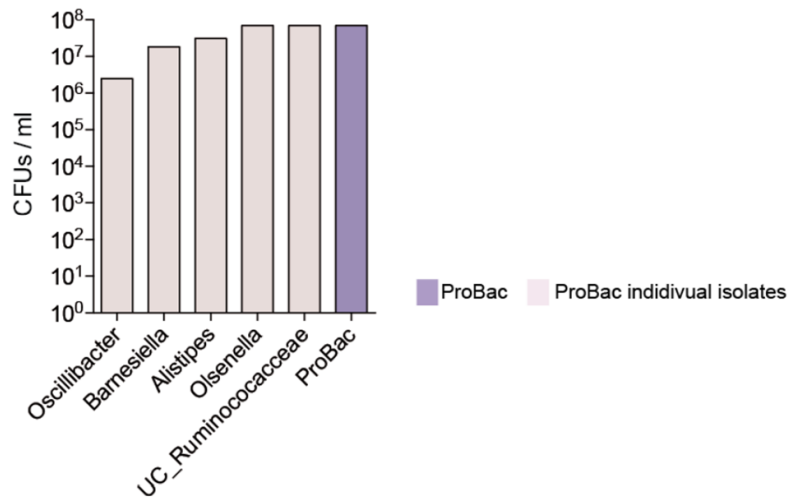


**Figure 84. Growth of ProBac strains *ex vivo*.** ProBac strains were frown individually or in combination in filtered supernatants of murine caecal contents at 37 ºC in anaerobic conditions. After 24 h, dilutions were plated in CBA plates and grown colonies were quantified after 3 days, except *Oscillibacter* and UC_Ruminococacceae that were count at 6 days due to their slower growth.

After checking that *Olsenella* individually was able to conferred protection *ex vivo*, we repeated the previous experiment adding fructose only with this bacterium incubated in the filtered caecum. Similar results were obtained adding *Olsenella* compare to previous results obtained with ProBac, as it can be observed in Figure 85. The results showed the decreased levels of VRE in the presence of *Olsenella* and the restoration of VRE growth after adding an excess of fructose.
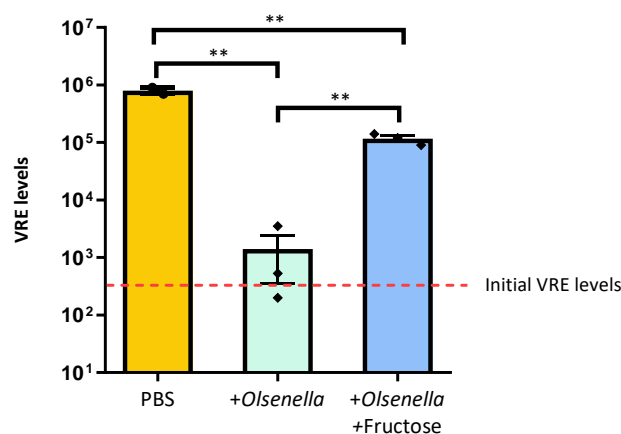


**Figure 85. Fructose supplementation restores VRE growth *ex vivo* in the presence of *Olsenella*.** *Olsenella* was incubated in anaerobic conditions at 37 ºC in the filtered caecum for 24 h, adding subsequently VRE and fructose. Subsequently, aliquots were plated in selective media to quantify the levels of VRE after 6 h of incubation. The addition of fructose partially restores the growth of VRE in the presence of *Olsenella*. ** $p < 0.01$, Unpaired t-test.

Altogether, these results suggest that ProBac is able to suppress VRE growth through nutrient competition for carbohydrates, mainly by depletion of fructose. In addition, the last result and the transcriptomic results suggest that *Olsenella* seems to be key in the protection conferred by ProBac since is the only bacterium that inhibit VRE growth by itself *ex vivo* and is the bacterium that express *in vivo* the genes required for the metabolism of fructose.

### 2.6 *Olsenella* consumes fructose as one of the main carbohydrate sources

Once we checked that *Olsenella* inhibit VRE growth *ex vivo* through a possible mechanism of competition by carbohydrates and that this growth was restored by the addition of an excess of fructose, we decided to check that indeed *Olsenella* is able to use fructose as a carbon source. For this purpose, we used again the Biolog methodology and grew the bacterium in the 96 well-plate containing different sugars.

Figure 86 shows the main sugars used by *Olsenella* for its growth. It is represented as the AUC, referred to the sugar in which the bacterium reached its maximum growth, mannose in this case. It is shown those carbohydrates with more than 20% of growth. The results showed that fructose is one of the sugars that the bacterium uses more efficiently for its growth. This result suggests that *Olsenella* could be contributing to the *in vivo* depletion of fructose detected after the administration of ProBac. Nevertheless, we do not discard that the other bacterium present in ProBac could be also contributing to fructose depletion *in vivo.* Besides fructose, we also detected that *Olsenella* utilize efficiently mannose and glucose as carbon sources, carbohydrates also used by VRE, and being mannose one of the most relevant sugars during the colonization process during antibiotic treatment, as shown in the previous section 1.9.4. Therefore, we do not discard the possibility that *Olsenella* may be inhibiting VRE growth, not only by depleting fructose but also by depleting other sugars such as mannose. This should be investigated in a future study that we did not had time to develop during the present thesis.
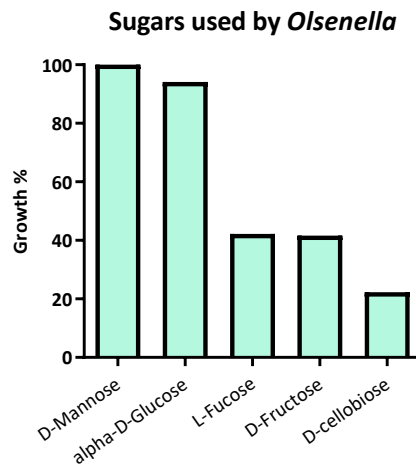
**Figure 86. Fructose is one of the main sugars used by *Olsenella* to grow.** *Olsenella* was added to Biolog 96 well-plate as described in the methods. The growth is represented as percentage referred to the AUC of the sugar that allow a higher growth of *Olsenella* (mannose, 100 %). Other carbon sources that are not carbohydrates are not indicated in the figure. Only those carbohydrates with more than 20% of growth, as compare to the growth in the presence of mannose, are represented in the graph. Bars represent the AUC.

## 2.7 *Olsenella* is sufficient to confer resistance against VRE intestinal colonization

As *Olsenella* was sufficient to suppress VRE growth *ex vivo,* we decided to investigate if this was also true *in vivo.* We repeated the *in vivo* experiments using the mouse model with two weeks of recovery period but administrating only *Olsenella* instead of the ProBac consortium after one week of vancomycin treatment. The results (Figura 87A) showed that *Olsenella* was sufficient to significantly restrict VRE intestinal colonization in mice that had been treated with vancomycin.

On the other hand, we tested a different mouse model (Material and Methods, section 28.7, Figure 29) to verify if *Olsenella* was able to eliminate VRE after being previously stablished in the gut. Briefly, mice were treated one week with vancomycin, inoculating VRE at day four of the treatment. Then, vancomycin was substituted by water and the next three consecutive days *Olsenella* was administered in order to analyze if *Olsenella* was able to decrease the levels of VRE. The results in Figure 87B show that after two weeks of antibiotic withdrawal, the levels of VRE were significantly reduced in the group of mice that received *Olsenella* as compared to the control group that received the suspension control media (PBS with 20 % of Glycerol and 0.1 % of cysteine).
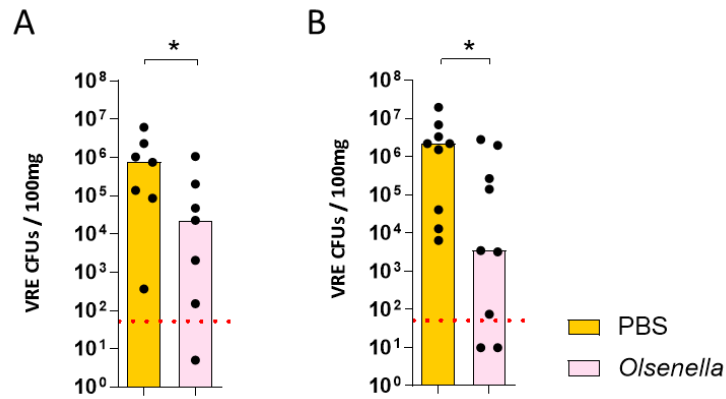
**Figure 87. VRE intestinal growth is suppressed by *Olsenella* inoculated to mice either before or after VRE colonization. (A)** Mice were treated with vancomycin and one day after stopping antibiotic treatment received either *Olsenella* or the suspension control media during three consecutive days (PBS with 20 % of Glycerol and 0.1 % of cysteine). Two weeks after antibiotic withdrawal, mice were inoculated through oral gavage with VRE and VRE levels were quantified 2 days after VRE inoculation. **(B)** Mice were treated with vancomycin during one week and inoculated with VRE at day four. One day after antibiotic withdrawal, *Olsenella* or the suspension control media were administered during three consecutive days. Two weeks after antibiotic withdrawal, faeces samples were collected to quantify the levels of VRE in the intestine. *$p<0.05$, **$p<0.01$, Wilcoxon-test. Bars represent the media.

All these results corroborate the inhibitory effect of *Olsenella in vivo* and indicate that this bacterium is able to displace VRE even if the pathogen is inoculated before the presence of *Olsenella* in the gut.

In summary, the previously results obtained by another PhD student identified a bacterial consortium composed of five protective bacteria, known as ProBac, that decreased the levels of VRE in the gut of treated mice. In my thesis we have identify that one of them, *Olsenella, seems to be* the key bacterium restricting VRE intestinal colonization. Moreover, through *in vivo* transcriptomic and metabolomics analysis besides Biolog assays and *ex vivo* experiments we propose a mechanism of protection by which ProBac and mainly *Olsenella*, deplete sugars (i.e. fructose) that are required for VRE gut colonization.

# V.   DISCUSSION

1.  **Efficiency of different *E. faecium* clinical strains to construct the transposon mutant library**

Although there has been a lack of genetic tools to efficiently manipulate *E. faecium* in the past, novel methodologies have been developed to facilitate the genomic study of *Enterococcus*.

One of these novel methodologies, based on the construction of a transposon mutant library, has been used to study the genes required by VRE to colonize the intestine through the use of an appropriate transposon delivery vector. This methodology, extensively used in other bacterial species, allows functional genomic studies in *E. faecium*. This procedure randomly inserts a transposon in one site of the genome of the bacteria (due to the low transposition frequency (around 1 %), double insertion is rare). Then the pool of mutants are grown under different conditions to further detect and compare the relative abundance of the mutants between the conditions tested, identifying the genes affected to grow in a particular condition.

This protocol was previously developed and tested by Zhang et al.[153] to identify genes necessary for VRE to grow in the presence of ampicillin. The strain selected to carry out this study was the clinical strain E1162. Thus, we reasoned to construct our mutant library using the same strain. A priori, we should have obtained a mutant library with the transposon inserted randomly in the genome of the bacteria, and with less than 10 % of the bacteria in the library containing the plasmid used for the mutagenesis, according to the previous results obtained by the authors of the study. However, our multiples mutant libraries constructed revealed higher presence of chloramphenicol resistant colonies (between 30-40 % to 100 % of bacterial cells resistant to chloramphenicol), suggesting that these colonies maintained the plasmid that should have been lost by growing the bacteria at the restrictive temperature for plasmid replication (i.e. 37 ºC). Although we tried to modify some steps of the protocol and increased to 42 ºC the temperature in order to favour the elimination of the plasmid, similar numbers of chloramphenicol resistant colonies were obtained.

Taking into account that chloramphenicol resistance can also arise spontaneously in *E. faecium* caused by gene mutations in the 23S rRNA gene[180], we checked individual chloramphenicol colonies from the library to discard this option. PCRs were performed, corroborating the presence of the plasmid in the library as the most likely cause of maintaining the chloramphenicol resistance.

As an exception, we obtained a transposon mutant library with less than 10 % of the plasmid but the transposon was inserted in the same point in all the clones tested, meaning no aleatory insertion (Results, section 1.2.2, Figure 33). This could be explained if this transposon integration occurred in the first event giving an advantage and bigger expansion to this particular clone. On the other hand, it is possible that this particular point in the genome represents a hot spot for transposon integration. The transposon used for this study is a gentamicin resistance *mariner* transposon constructed specifically for the study of ampicillin resistance determinants in *Enterococcus*

*faecium*[153]. It was constructed amplifying Himar1-mariner from pMMOrf and the gentamicin-resistance cassette from pAT392[153]. The Himar1-mariner transposon has been extensively and eficiently used in several bacteria such as *Pseudomonas aerugiona*, *Porphyromonas gingivalis*, streptococci and *Aggregatibacter actinomycetemcomitans*[181]. This transposon is preferentially inserted in sites with TA dinucleotides, which are abundant and distributed along the genome of the majority or microorganisms, being inserted randomly due to the distributed TA dinucleotides[182]. Nevertheless, it has been observed that the transposon seems to have preference TA site insertion affected by the DNA topology, being preferentially inserted in regions of bent or bendable DNA[182]. Thus, although the exact reason of our obtained result is unknown, it is possible that in the library with the transposon inserted in the same point of the genome, the transposon might have been inserted in one of these preferential places and that insertion in such preferential place was beneficial for bacterial growth. Both effects could have led to the expansion of this particular clone.

Finally, the library was constructed in the laboratory where the technique was developed. Same problems were obtained when the E1162 strain was tested again to obtain the mutant library, which prompt us to use new clinical *E. faecium* strains, Aus0004 and E1133. Moreover, as in the last attempts we had a higher presence of chloramphenicol resistant colonies than expected, we reasoned that adding an extra passage step at 37 ºC would increase the plasmid clearance rate.

Finally, two optimum mutant libraries were obtained adding the extra culture step at 37 ºC. One mutant library was obtained with Aus0004 strain with 0.0001 % of chloramphenicol resistant colonies and the other with E1162 strain with 10 % of chloramphenicol resistant colonies (Results, section 1.3, Figure 35). Moreover, we confirmed the random insertion of the transposon by sequencing. Although both libraries could be used, we decided to continue the study with the transposon mutant library obtained with Aus0004, as the plasmid presence was much lower.

The reasons that drive these differences in plasmid clearance with the different strains of *Enterococcus* are unknown and due to time limitation and since it was not the objective of my thesis, we did not study in depth the cause of plasmid stability in the E1162 strain as compare to the other strains tested. It is possible that the plasmid containing the transposon integrated into the genome of VRE through recombination. Therefore, differences in sequences of the genome of the different strains could drive a different frequency of recombination events. We search into the genome of E1162 and Aus0004 but we did not find a clear difference regarding the presence of chromosomal sequences that match the plasmid. In addition, a different expression of genes required for the recombination in the different strains could also influence the frequency of recombination. Another possibility that would allow plasmid maintenance at the restrictive temperature is that certain mutations would allow plasmid replication at the restrictive temperature. However, if this was the case, a similar rate of colonies containing the plasmid at the restrictive temperature would have been expected in all the strains tested.

Overall, our results demonstrate that, although the methodology was previously tested, it doesn't work with the same efficiency in the different strains checked. Nevertheless, the reasons for this difference are unknown.

## 2. Study of the effect of different antibiotics on the capacity of VRE to colonize the intestinal tract

As previously demonstrated, depletion of the commensal microbiota results in a dramatic increase of VRE growth in the intestine[28]. Thus, to study the genes that affect VRE colonization in the gut we first needed to set up a mouse model that allows VRE to reach high levels to further analyse the mutant library recovered from the intestine.

As it has been previously described that different antibiotic spectrum causes different effects and alteration of the microbiota[25] we decided to test different antibiotics (vancomycin, clindamycin, metronidazole, ampicillin, amoxicillin, amoxicillin with clavulanic acid and a cocktail of antibiotics including vancomycin, ampicillin and streptomycin) to select the antibiotic that contributes to higher levels of VRE in the intestine.

The VRE highest and homogeneous levels of gut colonization were reached after vancomycin and clindamycin treatments (Results, section 1.4.1, Figure 36), the two antibiotics that have been previously confirmed by our research group to cause more alterations to the microbiota, affecting the decrease in microbial richness, biomass and diversity to a greater extend[168] (results previously mentioned in the Introduction, section 2.2.3, Figure 8). This could be the reason of the higher colonization capacity of VRE and the mutant library with clindamycin and vancomycin among the different antibiotics tested. Surprisingly, the cocktail of three antibiotics supplied to mice did not allow the colonization of VRE in the gut. Theoretically, the mix of the three antibiotics with different spectrum should cause a higher depletion of the microbiota and therefore should promote higher or equal levels of colonization than the administration of vancomycin or clindamycin separately. We thought about two possibilities for explaining this phenotype. Firstly, expansion of specific bacteria after the administration of the mix but not after the therapy with vancomycin or clindamycin could be inhibiting gut colonization by VRE. Secondly, the effect of the three antibiotics could be directly affecting the viability of VRE, not being able to grow in the presence of all three antibiotics. In order to test the second hypothesis, we grew VRE in plates containing one or several of the antibiotics administered. Notably, when VRE was grown in plates containing vancomycin, ampicillin and streptomycin, VRE was not able to grow. This result supports the hypothesis that although the VRE strain used is resistant to each antibiotic separately, it is not able to grow in the presence of the 3 antibiotics. It has been previously demonstrated that the combination of different antibiotics can have different effects on bacterial survival than if they are added individually. The addition of two or more antibiotics can have an additive effect, meaning the sum of each antibiotic treatment or a cocktail of antibiotics can produce a synergistic effect, causing a superior

effect[183]. The antibiotic synergism has been studied in enterococci, demonstrating that the combination of antibiotics such as ampicillin with vancomycin, or vancomycin with streptomycin in 31 clinical isolates have synergic effect on enterococci survival[184].

Both clindamycin and vancomycin treatments were chosen to treat mice for further inoculation of the mutant transposon library and subsequent collection of samples from different parts of the intestinal tract. As we expected, according to the bibliography, *E. faecium* was found both in small and large intestine[18], reaching the highest levels of colonization in faeces and in the large intestine (Results, section 1.4.2, Figure 37). This is consistent with the overall load found in the intestinal tract in conditions where there is an intact microbiota. Higher intestinal loads are found in the large than in the small intestine[18], and VRE is not an exception, behaving like commensal *Enterococcus* colonizing the intestinal tract[110]. Indeed, higher intestinal loads of *Enterococcus* are found in the large intestine[185], probably due to their ability to obtain carbon and energy by fermentation of undigested substrates and fibers coming from upper intestinal tract.

Regarding the mucosa versus the lumen, significantly higher loads of VRE were found in the lumen than in the mucosa. The mucus is formed by complex molecules know as mucins, which are proteins decorated with carbohydrates[62]. Although strains of *E. faecium* are able to bound efficiently to mucus[186], mucins chains are protected from bacterial degradation by poly-O-acetylation and O-sulfation[62]. We reasoned that *Enterococcus* does not have the enzymes to efficiently breakdown the sugars from mucins but contain transporters that can internalize free available sugars in the lumen. Indeed, it has been shown that VRE requires other commensal bacteria that liberates the mucin sugars in order to be able to grow in the mucus[187]. According to our results, in a previously published study, the authors observed that the highest levels of VRE using clindamycin treatment in C57BL/6 mice were reached in the lumen rather than in the intestinal wall[188].

The experiment of VRE intestinal colonization after vancomycin or clindamycin treatment was done twice to confirm the reproducibility of the results. Indeed, similar results were obtained in both experiments confirming the reproducibility of the results. This is what we expected since for both experiments we used mouse derived from a colony of mice that has been bred in our animal facilities so that we can control not only the genotype, that it is exactly the same between different mice, but also the food and the environmental settings (i.e. cage, bedding, water). We decided to use our own mouse colony because our group has shown in the past that different colonies of mice that has been bred separately will derive in having a distinct microbiota[189]. Moreover, the differences detected among different colonies were even greater after vancomycin administration. In addition, recent studies have shown that mouse from different colonies, since they contain a slightly different microbiota, will have different susceptibility to colonization by intestinal pathogens (including *Citrobacter rondentium*[190] or *Salmonella*[191]). On the other hand, recent studies have shown that diet can also impact intestinal colonization by pathogens such as *Clostridium difficile*[192] or *Salmonella*[193]. Thus, it is very important for obtaining reproducible results, when

investigating colonization resistance, to use, if possible, mice that are derived from the same colony and bred in the same conditions and animal facility.

Afterwards, significant differences between both treatments were analysed and we detected that levels of VRE present during clindamycin treatment were significant higher and homogeneous, especially in the content of the intestine and faeces (Results, section 1.4.2, Table 20). We compared the variance of the samples from the same group from both antibiotic treatments and we obtained significant differences in the faeces colon and caecum samples of clindamycin treated mice as compare to mice treated with vancomycin. These differences in the variance between the antibiotic treatments could be explained by the administering procedure of the antibiotic to mice. It is also remarkable that the procedure to administer the antibiotic treatments differs from vancomycin and clindamycin. Vancomycin is administered in the drinking water whereas clindamycin is subcutaneously punctured twice a day. Thus, the vancomycin that reach the intestine depends on the amount of water that the mouse has drunk. Regarding the higher levels of VRE during clindamycin treatment, it is difficult with the data that we have to give an explanation. It is possible that the different microbiota remaining during vancomycin and clindamycin treatment differs and this could impact VRE intestinal colonization. In fact, we know from previous studies in our laboratory[168] that the microbiota of mice treated with clindamycin or vancomycin is distinct. On the other hand, it is possible that vancomycin could directly impact VRE growth. Indeed, the strain that we used for these experiments is resistant to vancomycin but its level of resistance is lower (containing the *van*B vancomycin resistance element[61], levels of resistance of 12 µg/ml[194]), as compare to other VRE strains, such as the VRE strain ATCC 700221 used in the second part of the study (containing the *van*A locus[151], whose MIC is 256 µg/ml[195]). Thus, it is possible that although both antibiotics, by depleting the microbiota, allow VRE intestinal colonization, the maximum levels are reached with clindamycin since oral administration of vancomycin leads to very high levels of this antibiotic in the large intestine[196], which may inhibit to a certain extent Aus0004 growth. Indeed, the colonies produce by Aus0004 in the presence of vancomycin (8 µg/ml) are smaller than in the absence of this antibiotic.

### 3. Transposition mutagenesis sequencing and analysis

To analyse those genes where the transposon was inserted, colon and caecum samples were prepared for sequencing by Illumina following the methodology described by Zhang et al[99].

We decided to use a high-throughput transposon mutant library sequencing approach (Tn-seq) methodology instead of the Microarray-based Transposon Mapping (M-TraM), the methodology used in the paper where the construction of the library was described[153]. The detection of the transposon insertion by M-Tram was based on the previously methodology of Genomic Array Footprinting (GAF), developed to identify essential genes in *S. pneumoniae*[197]. This method combines the transposon mutagenesis

and the technology of microarrays. The site of the transposon insertion is detected by labelling the DNA adjacent to the insertion point to be hybridized to a microarray[197]. This approach has some technical limitations, for example, the difficulty of detection of low signals on the microarray for those low abundant mutants in the library or on the contrary probe saturation, reaching signal saturation with those mutants highly present, being more difficult to identify the insertion sites with high resolution[198,199]. Moreover, these microarrays need to be previously designed with the probes offered by the company or the required probes, but they cannot be used for discovery of new regions of the genome if they are not previously described[200]. Tn-seq approaches solve the problems found with the use of microarrays. This methodology allows to obtantion of millions of sequences without the requirement to be previously described or known, detecting both low and high abundant mutants of the library, permitting higher resolution[199]. Furthermore, sequencing techniques are currently less costly than microarray-based approaches due to their increased development and the progressive reduction of its cost[198].

Once we prepared the samples for sequencing, we realized that *in vivo* samples had variable fragment sizes compare to *in vitro* samples, which had exactly the expected size (Results, section 1.5, Figure 39). To prepare samples for Illumina sequencing, firstly, we extracted the DNA using the phenol-chloroform following the protocol of Zhang et al[99]. This DNA was digested by MmeI restriction enzyme (located at the end of the transposon which cuts 18 nucleotides apart from the site of recognition) and phosphorylated adapters are ligated using T4 DNA ligase. Finally, a PCR is performed to add adapters for Illumina sequencing and to amplify the fragments that contain the DNA of interest. We reasoned that this variability *in vivo* came from the first step of DNA extraction (Results, section 1.5, Figure 38). We observed that DNA obtained from mice samples presented a smear whereas samples *in vitro* contained a single DNA band. The smear from caecum and colon samples could be derived from DNAses present in the intestinal lumen that could degrade the DNA[201]. We speculated that these unspecific fragments of different sizes could be affecting to the entire protocol, because after DNA fragmentation, adaptors will be added to these unspecific fragments. Thus, in the subsequent step of amplification, both the correct fragments containing the flanking region where the transposon was inserted and the unspecific fragments will be amplified, which will produce amplicons of different sizes. To solve this problem, we decided to first purify the DNA obtained after the first extraction. Effectively, eliminating the smear of different size fragments from the DNA extraction improved the obtained results, since after the amplification step, we were able to obtain fragments of the desired size (Results, section 1.5, Figure 40). This problem of unspecific fragments detected after the amplification process was not observed by the group that developed the methodology since in their study the authors exclusively analysed *in vitro* samples, derived from pure cultures of the libraries, grown in different conditions (i.e. absence, presence of ampicillin).

Subsequently, to screen for the transposon mutants involved in gut colonization, sequences were analysed using different bioinformatics tools. After filtering the

sequences by quality and size, they were matched to the coding regions of the genome of Aus0004 in order to identify the genes were the transposon has been inserted. On average, 73.25 % of the sequences from the *in vitro* samples aligned with VRE encoded genes. The sequences from *in vitro* samples did not match in higher percentage the VRE genome because we were discarding those sequences that aligned with an intergenic region or with more than one gene. Indeed, if these sequences we included, the number of sequences that aligned with the VRE genome increase significantly (i.e. 95.81 %). Still, a low number of sequences from the library did not match the VRE genome, which could be due to the fact that we were very stringent in the aligning step (maximum of mismatches accepted = 1). In the *in vivo* samples, the percentage of sequences that aligned with VRE encoded genes was lower (on average 45.5 %) (Results, section 1.6, Figure 41). This number also increased when the alignments were performed against the genome of VRE instead of only the CDSs (i.e. 54.4 %). We reasoned that in the case of the *in vivo* samples, the higher percentage of sequences not matching VRE encoded genes was due to the fact that mouse samples contain other bacteria or even tissue from mice, besides VRE. Some DNA fragments of other bacteria or of mouse origin could have been amplified and sequenced, even if they do not contain the transposon. Indeed, there were an important percentage of sequences from *in vivo* samples (an average of 87.15 % of sequences that not match VRE genome) that aligned against the mouse genome database, verifying that besides collecting colon and caecum samples we collected some mouse tissues and this DNA was extracted, amplified and sequenced. We reasoned that part of mouse genome and other bacteria was similar to the sequence of the transposon allowing the amplification of unexpected fragments, obtaining a percentage of sequences that does not correspond to VRE. However, this possibility remains to be demonstrated.

Subsequently, we analysed the coverage of the transposon in the entire genome of VRE. Insertion of the transposon was detected in the majority of the genes encoded by VRE, however, there were 111 genes (3.8 %) in which we could not detect any transposition event. We speculate that we were not able to detect the transposon in these genes probably because these genes were essential for both *in vitro* and *in vivo* growth of VRE. The transposon mutant libraries and Tn-seq methodologies have been extensively used to discover essential genes in numerous type of bacteria, referring as essential those genes required to bacteria survival under specific conditions, identifying the regions of the genome with no or low presence of transposon inserted[199]. For example it has been previously used to detect essential genes of *Helicobacter pylori*[202], identifying that 23 % of genes had zero transposon insertions, although in this case microarray hybridization methodology was used to identify the essential genes instead of Tn-seq. On the other hand, it has been applied to discover the colonization factors required by *Klebsiella pneumoniae* to colonize the gut[203]. However in this case, besides taking into account those mutants with zero transposon insertions, the mutants with minimal read count in the library were analysed with EL-ARTIST pipeline[176] to determine essentiality, applying Hidden Markow model (HMM), obtaining that approximately 14 % of genes were classified as indispensable. It has also been used a similar strategy with *Micobacterium*

*avium*, in which mutants with zero insertions or very few insertions were analysed with different filter procedures obtaining 5.5 % of gene essentiality[177]. Therefore, gene essentiality depends on the type of microorganism, the conditions tested and the different analysis performed taking into account genes with minimal read counts or only those that have zero transposon insertions. Besides genes that could be essential for the growth of VRE, some of the undetected genes could represent genes that are no longer present in the genome. In fact, several of the undetected genes theoretically encode for transposases, which could be included in mobile genetic elements. To evaluate this possibility, it will be important to re-sequence the genome of the strain, which could change overtime from the continuous growth of the strain in different laboratories as has been previously shown[204].

In addition, other undetected genes are hypothetical proteins with no previous function assigned. It will be therefore interesting, in future studies, to investigate the role of these genes in VRE growth. In addition, the transposon was never detected in genes encoding transcriptional regulators or ribosomal proteins (30 S ribosomal protein S4). Ribosomal proteins are essential for the translation of protein, being a structural constituent of ribosomes. Thus, it is not unexpected that the insertion of the transposon on genes encoding ribosomal proteins abolished VRE growth. Similarly, transcriptional regulators may regulate the expression of genes whose over or under expression could be key for VRE growth.

On the other hand, there is a list of genes, 688 (23.6 %), with low abundance of transposon mutants (less than five read counts in each sample) in all the conditions tested. These genes are candidates to be essential or nearly essential genes due to their low presence. Among them, we have detected several genes theoretically encoding for proteins of the different subunits of the ribosome that are essential for the translation of proteins; membrane proteins that are part of the cell structure; cell division proteins that control cell morphology and cell division; or DNA-dependent RNA polymerase that catalyses the transcription of DNA into RNA. Thus, some of these genes with low abundance of Tn insertions encode for basic functions in all type of bacteria. In addition, other genes detected in very low frequencies include putative transcriptional regulators such as PadR and MerR family transcriptional regulators that respond to environmental factors or stimuli[205,206]. This list of genes should be examined carefully and we could use different type of analysis as previously mentioned in other studies, which will help us to predict if actually these gene play an essential role due to their low abundance in the conditions tested. One option could be to apply EL-ARTIST pipeline[176], previously used in the study of colonization factors of *K. pneumoniae*[203], a methodology designed to identify essential regions using Tn-seq technology, detecting the genetic elements required for the bacteria to grow in specific conditions.

Subsequently, we performed a global analysis of the transposon insertions, comparing the samples collected *in vivo* and *in vitro*. To perform this analysis, we used PCoA, an unsupervised analysis, and CCA, a supervised analysis in which the information of the group that the sample belongs to is used to detect global differences between groups of

samples. In both types of analysis, we detected a clear separation between the samples collected *in vivo* and *in vitro* (Results, section 1.6, Figures 45 and 46). This result indicated major differences in the genes required for *in vitro* and *in vivo* growth. This is not unexpected given that it has been previously demonstrated the plasticity of the *E. faecium* genome, analysing the genomes of 161 *E. faecium* strains isolated from diverse environments. This plasticity contributes to the versatility and the ability to be adapted to different conditions[54]. In addition, we observed that while the *in vitro samples* clustered tightly in the PCoA, the *in vivo* samples were more disperse. One explanation to the higher variability detected *in vivo* could be that, as we and others have shown[207], the same antibiotic treatment can induce slightly different changes in the microbiota. These differences in the commensal bacteria remaining could impact the availability of nutrients, as we have also shown in this thesis, which may influence the expansion of different mutants. Since *in vitro*, the nutrients available are always the same, a more homogenous abundance of mutants would be expected to be detected *in vitro* than *in vivo*. Despite this more heterogeneity in the *in vivo* samples, no overall differences could be detected between *in vivo* samples collected from different locations of the large intestine (i.e. cecum vs colon) (Results, section 1.6, Table 24). We reasoned that we did not observe significant differences between both parts of the large intestine due to the fact that colon and caecum share a similar anaerobic environment and nutritional conditions. The majority of nutrients are absorbed in the small intestine but the indigestible food material is processed in the large intestine (both in the cecum and colon), allowing to obtain key nutrients that are not able to be transformed to usable nutrients in the small intestine[208]. In addition, both the cecum and the colon contain a similar microbiota, as compare to the small intestine, thus it was not unexpected that *E. faecium* mutants behave in a similar manner in both locations. It would have been interesting to have studied how the mutants behave in the small intestine. However, the loads in the small intestine were a priori too low for applying the Tn-Seq methodology.

To identify VRE genes involved in the gut colonization process, we analysed differences in the abundance of the Tn-mutants comparing the *in vitro* and *in vivo* samples, first combining all samples from both experiments and then analysing samples from each experiment separately. In addition, we compare the abundances of Tn-mutants between the two different parts of the intestine vs the *in vitro* samples. All these analyses allowed us to identify those genes that repeatedly affected gut colonization and therefore identify those candidate genes more relevant for VRE gut colonization. Curiously, after selecting those Tn-mutants whose abundance was significantly different between *in vivo* (both in colon and caecum) and *in vitro* samples, a higher number of Tn-mutants were found to be more abundant *in vivo* than *in vitro* (Results, section 1.6, Figure 49). One possible explanation for this type of mutations is that the gene mutated could be favouring the growth in the intestine, its mutation is favouring the fitness of the bacterium under these specific conditions. A second explanation could be that those Tn-mutants reduce fitness both *in vitro* and *in vivo*. However, *in vivo,* the competition between mutants could be less pronounce since a particular mutant could find a niche within the whole intestinal tract that would allow partial growth of the mutant. Finally,

as a third explanation, in the *in vitro* conditions we detected some Tn-mutants to be over-represented, with a very high abundance. It could be that a few clones are very efficient growing in the *in vitro* conditions, which could reduce the relative abundance of other mutants *in vitro*. *In vivo*, the decrease in the growth of the very efficient *in vitro* growing clones would cause the increase of other Tn-mutants in proportion, compared to those over-represented mutants. Nevertheless, as the objective of the thesis is to identify genes encoded by VRE that are required for gut colonization, we focused on those Tn-mutants more abundant *in vitro,* whose abundance was reduced in the intestinal tract (Results, section 1.6.1, Figure 51).

## 4.    Implication of identified genes in gut colonization

Within all the genes that could affect gut colonization by VRE, we focus on those 10 with the highest log2FC difference between *in vivo* and *in vitro*, that were significantly higher in both intestinal sites, as compare to *in vitro* samples, and that the phenotype was detected in both performed experiments. To confirm their role in gut colonization, we generated targeted mutants for each of the listed genes, ten in total, (Results, section 1.6.1, Table 28). The candidate genes were replaced by a gentamicin marker by homologous recombination using the strategy from Zhang et al[153]. To confirm that the gentamicin introduction in the genome of VRE did not affect its ability to grow both *in vitro* and *in vivo*, we inserted the gentamicin marker in the WT strain, in an intergenic region flanking a pathogenicity island. We reasoned that the introduction of a gene in this region of the genome would not affect VRE growth due to the plasticity of the genome and the facility to acquire mobile elements, especially in parts of the genome where there are already mobile genetic elements[60]. The introduction of the gentamicin marker in the selected intergenic region did not affect VRE growth, neither *in vivo* nor *in vitro* (Results, section 1.7.1). This first control was relevant before continuing with this strategy because it has been shown that some antibiotic resistant genes decrease the fitness of the bacterium when the antibiotic is not present in the environment. It has been previously demonstrated that the insertion of various mobile genetic elements such as transposons encoding for the gene that confers resistance to tetracycline resistance reduced the relative fitness of both *E. faecalis* and *E. faecium*[209]. After checking that the insertion of gentamycin resistance marker does not affect the growth of VRE, a targeted deletion mutant was obtained for each of the genes of interest and competition assays both *in vivo* and *in vitro* with the WT were tested for each generated mutant.

## 4.1  Study the phenotypic differences of mutant strains compared to WT *in vivo* and *in vitro*

Before analysing the *in vivo* effect of each generated mutant, we checked that the deletion of the mutant was not affecting *in vitro* growth (Results, section 1.7.2). For this purpose, we set up a competition experiment with the WT strain. For one of the deleted genes, the protein *Suf*B for the Fe-S cluster assembly, the competition assay showed a

clear disadvantage of this mutant *in vitro* compared to the WT strain (Results, section 1.7.2.1, Figure 56). We reasoned that the elimination of a single gene from the entire operon could be generating a [Fe-S] cluster that could be toxic for the bacteria due to the incorrect processing of the entire cluster. Indeed, the deletion of single genes from a similar iron cluster (Isc) induces severe growth defect in the case of *E. coli*[210,211].

Subsequently, we decided to obtain a complete deletion of the four genes within the operon. Interestingly, in the new mutant generated, we did not detect differences *in vitro* between the growth of the WT and the mutant strain (Results, section 1.7.2.1, Figure 57). Thus, deletion of a single gene of the operon (*suf*B) was somehow toxic for the bacteria probably due to the partial Fe-S cluster generated, rather than *suf*B gene being essential for the bacterial growth.

Next, to confirm the role of the selected genes *in vivo*, an *in vivo* competition assay was performed with each mutant strain and the wild type strain. We anticipated, taking into account the Tn-Seq data, that we would detect a lower abundance of the mutant strain in the competition assays. This was the case for most of the mutants tested. However, an unexpected result was obtained with the mutant for the Vat family streptogramin A O-acetyltransferase (Δ*vat*::*gen)*. In this particular case, the mutant had a higher growth capability *in vivo* than the WT (Results, section 1.7.2.1, Figure 55). The protein that encodes this gene is described to inactivate the group A of the streptogramin antibiotics by acetylation, conferring resistance to these "last-resort" compounds to treat infections caused by Gram-positive bacteria that are resistant to many antibiotics[212]. Streptogramin antibiotics are produced by species from the genera *Streptomyces*, from the phylum Actinobacteria. Analysis of the microbiota of mice treated with clindamycin did not reveal the presence of any bacterial species from the genera *Streptomyces*[168,213]. Nevertheless, it is possible that streptogramin could be synthetized by other commensal bacteria remaining after clindamycin treatment. If this was the case, the gene that confers resistance to streptogramin would benefit the growth of VRE *in vivo* only if the commensal bacteria producer of streptogramin was present in the gut. In the conditions in which no streptogramin producer was present in the gut, the gene could be detrimental for VRE growth, taking into account, as previously described, that antibiotic resistance genes can result in a fitness cost if the antibiotic is not present. Thus, it will be very interesting to study in subsequent studies: (i) if there is any bacterial species from the microbiota that survive the clindamycin treatment and is capable of producing streptogramin, (ii) if differences in the remaining microbiota after clindamycin treatment, specifically in streptogramin producers, influence the intestinal colonization of the mutant for the *vat* gene, (iii) if in the absence of streptogramin, the mutant for the *vat* gene confers a fitness advantage.

For the rest of the genes, we were able to confirm through targeted mutagenesis that they do not have any deffect on the *in vitro* growth but that their mutations specifically impact VRE intestinal colonization. Some of these genes (*man*X, *mur*Q, *pst*B2, *lac*I, [Fe-S] cluster), which are related to nutrient acquisition in the gut environment, were characterized in detail. For the rest of the genes (*yyc*I, *pgt*, *adh*E, *dsb*A), due to time

constraints, we were not able to characterize them in detail, however we do discuss in the next section their putative role in gut colonization.

### 4.1.1 Putative role of *yyc*I in VRE gut colonization

One of the genes that affected gut colonization by VRE encodes a hypothetical protein (assignation first given in the VRE Aus0004 genome deposited in NCBI) that was further classified by protein homology as YycI. We performed additional *in silico* analysis of this gene against YycI protein from *Bacillus subtilis* using blastp and we detected that this gene has 26 % of identity with 98 % of coverage to the YycI protein from *B. subtilis*. YycI is part of the YycFG two-component system, which has been described in *B. subtilis*. YycFG is essential for cell viability and regulates essential processes for the cell such as homeostasis, division and the maintenance of the cell membrane[214]. YycFG is formed, like other two component systems by an histidine kinase (YycG) that sense a stimulus from the environment, and the response regulator (YycF), that mediates the cellular response[214]. YycI is localized outside the cytoplasm and attached to the membrane by an N-terminal transmembrane sequence. In combination with YycH, it controls the autophosphorylation of the kinase YycG and therefore the activation of the two component system. In the strain that we studied, besides the ortholog of YycI, we found a gene with sequence similarity to YycH (Figure 88), and that was previously assigned in the genome of Aus0004 as a hypothetical protein. This gene is located downstream *yyc*I as occurs in the *Bacillus subtilis* genome, and shares 20 % of identify with 94 % of coverage. Since we found homology for these two proteins of the two-component system, we performed new alignments using blastp and the YycF and YycG protein sequences. We found again homology with the two following genes from Aus0004 genome, with 43 % of identity and 97 % of coverage to the YycF response regulator and 79 % of identity and 99% of coverage to the YycG, the histidine kinase (Figure 88). These putative proteins were assigned as cell wall metabolism sensor histidine kinase (*walk*) and response regulator transcription factor respectively in the Aus0004 genome. Moreover, we checked the abundance of the mutants for all these genes in the mutant transposon library and we found low abundance (both *in vitro* and *in vivo*) of transposon insertions for these last two genes that were similar to YycF and YycG, which would indicate that they are almost essential, as in the case of *B. subtilus* but also of *E. faecalis*[215]. We hypothesized that these proteins may be essential for VRE Aus0004 growth, but that probably the phenotype was even more pronounce *in vivo*, since the single deletion of the *yyc*I homologue affected more the *in vivo* than the *in vitro* growth.
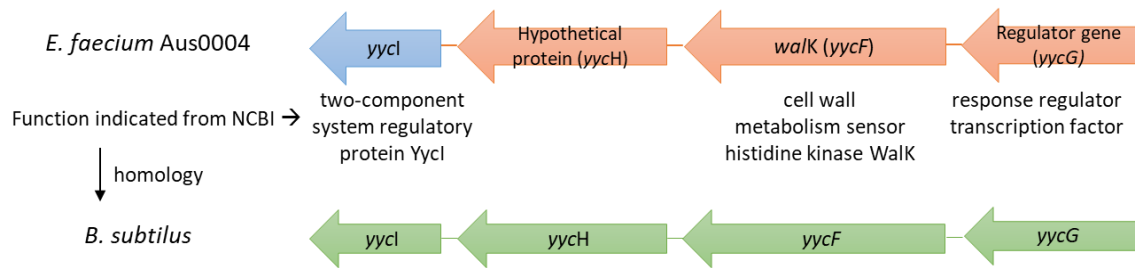
**Figure 88. Schematic representation of the two-component system YycFG in *E. faecium* strain Aus0004.** The four putative proteins found in Aus0004 are homologous to proteins encoded by the YycFG quorum sensing system previously described as essential in *B. subtilus* (operon coroulated in green). This system might be involved in the regulation of basic cellular functions such as cell viability, homeostasis and division of the cell membrane. In blue is coloured the gene deleted from the VRE genome that has been check to be involved in gut colonization. In orange are indicated the other proteins that are homologous to the proteins present in *B. subtilus*. Inside the arrow (in parentheses) is indicated the gene name of the homologue in the *B. Subtilis* genome. Below the orange arrows is indicated the predicted function of the gene that is indicated in the Aus0004 deposited genome in NCBI.

Nevertheless, at this point it is difficult to understand how the single deletion of YycI could affect the colonization of the intestinal tract. This will require extensive work involving the characterization of the role of the two component system in *E. faecium* biology, including the identification of target genes of the response regulator, their role in intestinal colonization or *E. faecium* viability and the impact that *yyc*I deletion could have on the expression of these genes. For this reason, we did not continue with the functional characterization of YycI, which will be part of a future study.

### 4.1.2 Putative role of *pgt* in VRE gut colonization

*pgt* encodes for a putative glycosyltransferase. This type of enzymes catalyses the transfer of saccharide moiety and the result of this transfer can be the formation of a carbohydrate, oligosacharide, polysaccharide or a glycoprotein[216]. Therefore, there are multiple activities of the bacterial cell that could be affected by the deletion of a glycosyltransferese and should be studied in order to identify the role of this particular glycosyltransferase on VRE intestinal colonization.

Interestingly, a study performed in *E. faecalis* using a transposon library to identify genes associated to the formation of biofilms, identified two glycosyltransferases, whose deletion affected the production of biofilm under subinhibitory concentrations of antibiotics. In addition, the authors found that deletion of one of these glycosyltransferases (*epa*OX) decreased cell envelop integrity and bacterial growth in the presence of detergent and bile salts[217]. Both biofilm formation and resistance to bile salts could play a role in gut colonization. Indeed, the loss of the ability to form biofilm impairs the overgrowth of enterococci in the gut[217]. In addition, bacteria have to be resistant to bile salts in order to colonize the intestinal tract[91]. Moreover, one particular study has demonstrated the importance of a different glycosyltransferase in *E. faecalis* gut colonization[218]. EpaX is involved in sugar composition of the cell membrane and its integrity, specifically in the composition of the rhamnopolysaccharid, being also

involved in resistance to bile salts and biofilm formation. Moreover this gene was enriched in clinical isolates and its deletion of gene impaired the ability of VRE to colonize the intestine of mice, demonstrating the importance of these proteins[218]. Therefore, it will be important to perform future studies to elucidate if the deletion of the Aus0004 *pgt* gene, encoding the glycosyltransferase, has any effect on both biofilm formation and growth in the presence of bile salts.

### 4.1.3 Putative role of *adhE* in VRE gut colonization

AdhE is a multifunctional protein that has been well characterized in *E. coli*[219]. It is used under anaerobic conditions in order to obtain energy and concordantly its overexpressed under anaerobic conditions[220]. It functions as an acetaldehyde-CoA/alcohol dehydrogenase that is involved in ethanol formation through the reduction of Acetyl-CoA to acetaldehyde and then this acetaldehyde is reduced to ethanol. Although it is required by *E. coli* to grow under anaerobic conditions, its activity depends on the electron acceptors available. For example, in the presence of fumarate, an electron acceptor utilized for anaerobic respiration by *E. coli*, AdhE activity is decreased. A similar gene to AdhE has been described in both *E. faecium* and *E. faecalis* genomes[221]. However their role in anaerobic respiration is unknown. It is known, however, that under nutrient-limited conditions, *Enterococcus* metabolism of mannitol or glucose produces ethanol as a fermentation end product and it has been suggested that *adh*E plays a role in ethanol production[222,223]. We have seen that *adh*E is relevant for the growth in the intestinal tract, an anaerobic environment. However, *adh*E was not required for growing under anaerobic conditions *in vitro*. It is possible that under the nutrient-limited conditions frequently found in the intestinal tract, VRE requires *adh*E to ferment glucose or even mannitol, two sugars frequently found in the intestinal tract. *In vitro*, however, the nutrients are not limited, and VRE may obtain energy using other metabolic routes. In addition, electron acceptors may differ *in vivo* and *in vitro* and like it has been shown in *Enterobacteriaceae*, this could have an impact on the activity of *adh*E and its requirement for VRE growth. Future studies will be required to test all these possibilities.

### 4.1.4 Putative role of *dsb*A in VRE gut colonization

DsbA is a soluble periplasmic oxidizing enzyme that catalyse directly and efficiently the formation of disulphide bonds in substrate proteins[224]. This protein is part of the oxidative pathway of the DSB machinery, which has been well described in *E. coli*[225], being the model for disulphide bond formation in bacteria and archaea, although is not universally conserved in all microorganisms[224]. The other protein involved in the oxidative pathway is DsbB, an inner membrane protein, which reoxidized and keeps DsbA in the active form, allowing to play its function to follow a catalytic cascade pathway[224]. Nevertheless, no significant similarity was found for DsbB protein in the genome of Aus0004. Although there is no homologous, another protein might play a similar role to DsbB protein, which oxide DsbA to maintain it in the active form.

Mutation or deletion of the DsbA protein of the DSB machinery has been associated with a decrease ability to colonize the gut of some pathogens such as *E. coli*[226], *V. cholerae*[227] or *V. parahaemolyticus*[228]. In the case of *E. coli*, the deletion of this gene caused a

reduction in the attachment to epithelial cells and also lower capability to biofilm formation[226]. Moreover, the deletion of the gene reduced the virulence of the pathogen decreasing the ability to colonize the gut of *C. elegans*. Although the authors did not find the protein or proteins affected by the deletion of the *dsb*A gene, they hypothesized that it might be a virulent factor. They reasoned that a non-identified protein involved in virulence was not able to play its function due to the inactivation of dsbA, which did not allow to catalyse the disulphide bond formation for the correct folding and function of the protein[226]. In the case of *Vibrio*, *dsb*A is known to activate the virulence gene regulator *tox*T[227]. ToxT is required for the activation of the transcription of the cholera toxin and the toxin-coregulated pilus, key in the *V. cholera* disease[229]. Deletion of *dsb*A reduces de activity of ToxT which diminish the capacity of *Vibrio* to colonize the intestine of mice[227].

We have checked that *dsb*A is relevant for the gut colonization in our mouse model, although we hypothesize that the deletion of this gene affects the proper folding of other proteins that acts as virulence factors and should be affecting to the pathogenesis of *Enterococcus*. Due to lack of time we have not elucidate with proteins are affected by DsbA and future studies will be necessary to elucidate the mechanism by which these proteins reduce the levels of VRE in the gut.

### 4.1.5 Characterization of *pst*B2, a gene related to phosphate uptake

One of the genes that we found using the transposon mutant library to affect most to the colonization of the intestine by VRE was *pst*B2, a subunit of a putative ABC transporter for phosphate uptake.

Phosphate is essential for bacterial growth. It is a component of nucleotides that are required both for energy storage (i.e. ATP) but also for the synthesis of nucleic acids. In addition, phosphate is required as a signalling molecule by two-component systems, essential for bacterial sensing of environmental signals[69].

Phosphate transporters have been extensively studied in *E. coli*. This bacterial species has two major transporters for phosphate. The Pit system is a low affinity system of acquisition of phosphate that does not require energy for phosphate uptake. This system is used by *E. coli* when the concentrations of phosphate are high. In addition, it encodes the Pst system, a high affinity ABC transporter that binds phosphate when the concentrations are low and requires energy for its uptake[69].

*E. coli* Pst system is formed by PstS, a periplasmic protein that binds phosphate. Then phosphate diffuses across the gram-negative outer membrane; PstA and PstC that form a membrane channel in the inner membrane of *E. coli* and PstB that interacts with PstA and PstC and provides energy for the process of internalization of phosphate across the inner membrane[69]. Interesting, the VRE Aus0004 strain, used in our study, contains a putative operon that besides *pst*B2, it encodes for *pst*B1, *pst*A, *pst*C and *pst*S (Figure 89). Furthermore, we have verified if the deletion of the other genes of the operon was essential for *E. faecium* checking their abundance in the transposon mutant library. In the case of the proteins *pst*B1, *pst*A and *pst*C we have found zero or few insertions (less

than twelve mutants per condition) that would indicate that these proteins are essential for VRE growth and they might interact in a similar way of the Pst system described in *E. coli*. Nevertheless, we found high number of Tn-mutants present in the library for the *pst*S gene, showing that this gene is not essential for the pathogen, which is able to grow both *in vitro* and *in vivo* conditions. It may not be essential for VRE since *Enterococcus* does not contain the gram negative external membrane and the system is able to work perfectly without the present of this single periplasmic protein.
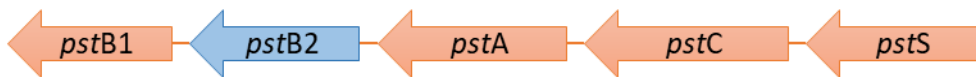


**Figure 89. Schematic representation of the Pst system found in *E. faecium* strain Aus0004.** The five proteins found in the Aus0004 genome to be similar to the previous Pst system described in *E. coli*. In blue is indicated the gene deleted from the genome of the pathogen that affects gut colonization and in orange are coloured those genes that are part of the system. These genes might be involved in the internalization of phosphate through the inner membrane due to their high similarity to the Pst system in *E. coli*.

Inorganic phosphate is absorbed through the small intestine[230]. Therefore, the concentrations in the large intestine are limited. In this conditions is where the Pst high affinity phosphate system should be playing a role in order to actively uptake the phosphate available. Therefore, a priori, our hypothesis about the role of PstB on gut colonization was its effect in the active uptake of phosphate.

To better understand the role of the Pst transporter (more specifically the *pst*B2 subunit of the transporter) in the gut colonization by *E. faecium,* we first checked that the elimination of this specific gene was enough to reduce the ability of VRE to expand in the gut and its deletion was not affecting its phenotype *in vitro*, where theoretically the concentrations of phosphate are not that limited. Indeed, this was the case. *pst*B deletion affected exclusively gut colonization but not *in vitro* growth when VRE was grown in a rich media. Subsequently, we decided to test if *pst*B was required for VRE growth under conditions in which phosphate is in low concentration. For this purpose, we utilized a minimal medium in which we were able to remove all phosphate compounds, adding all the other nutritional requirements needed for *E. faecium.* Therefore, we were able to add different concentrations of phosphate in order to check the growth differences between the WT and Δ*pst*B2. We tested two different minimal media previously used in *Bifidobacterium* to characterize the role of a similar phosphate transporter[152,161]. However, the wild type strain was not able to grow efficiently under any concentration of phosphate. We only detected growth of the WT strain with the highest concentrations of phosphate (>50mM). But even in this case, the growth was impaired since the lag phase was longer than 15 hours. A similar result was obtained with the mutant strain. We tested an alternative medium, also used for the study of *Bifidobacterium*, but VRE was not able to grow in that medium. Thus the results indicate that the medium used to study phosphate transport in *Bifidobacterium* cannot be used for studying VRE growth. This could be explained because both species have different nutrient requirements. The reasons that led us testing the *Bifidobacterium* media were

that no study of phosphate uptake has been published so far for VRE and the most close related species in which phosphate studies has been done was *Bifidobacterium*. Indeed, these experiments were performed in the lab where these studies were done and were this media was proposed. Thus future studies will require the development of a new defined media for *Enterococcus* in which the concentration of phosphate can be modified accurately.

In addition to phosphate uptake, the Pst system plays another important role in *E. coli*, that is the regulation of the so called Pho regulon[231]. This is a group of genes, most of them related to phosphate uptake whose expression is regulated by the Pst transporter interaction with a two component system formed by PhoR (the histidine kinase) and PhoB, the response regulator. Both, the two component system and the Pst interact through the PhoU protein[69]. Interestingly, the system is ON and the regulon is expressed when the concentrations of phosphate are low. While high concentrations of phosphate repress the expression of genes within the Pho regulon. This regulation has been also described in other bacterial species including *Vibrio cholerae*[232] and *Pseudomonas aeruginosa*[233]. Interestingly, in these species, the Pho regulon also controls de expression of virulence genes. In the case of *V. cholerae* it controls the expression of cholera toxins[232], while in the case of *P. aeruginosa* it controls the expression of the PQS+Fe3+ complex which produces gut epithelial disruption and death in mice[233]. Interestingly, the VRE Aus0004 genome encodes a KEGG ortholog of PhoU, and KEGG orthologs for PhoR (the histidine kinase of the two component system) and PhoP (the response regulator that has been found in *S. aureus* to have a similar role as PhoB – regulates the growth of *S. aureus* under phosphate starvation conditions[234]). It will be interesting to test if Pst, in combination with the two component system PhoP/PhoR controls not only the expression of genes related to phosphate uptake but also other genes that are essential to intestinal colonization. Future studies involving the deletion of the PhoP/PhoR two component system, transcription analysis under different concentrations of phosphate and deletion of genes controlled by the PhoP/PhoR two component system will be required in order to completely elucidate the role of the Pst system on VRE intestinal colonization.

### 4.1.6 Characterization of the Fe-S cluster

[Fe-S] cluster are prosthetic groups containing iron and sulphur at different molecular ratios, which contain either ferrous ($Fe^{2+}$) or ferric ($Fe^{3+}$) iron and sulfde ($S^{2-}$)[235]. It is a versatile group involved in electron transfer and redox potentials, iron/sulphur storage and gene regulation[236].

It has been identified three different Fe-S cluster pathways: the Nif (nitrogen fixation), the Isc (iron sulfur cluster) system and the Suf (sulfur formation)[211,235]. The cluster found in our study to be important for VRE intestinal colonization is part of the Suf system. SUF system plays its role in the Fe-S cluster biosynthesis and assembly under oxidative stress and limiting iron concentrations[235]. To produce the synthesis of this Fe-S cluster it is required the donation of iron and sulphide as ligand for iron ions. These two components are combine in a scaffold protein for cluster assembly, for which SufB is the

essential protein[235]. This system is also composed by other proteins, containing from two (SufB, SufC) to more than six (SufA, SufB, SufC, SufD, SufS, SufE, SufU), depending on the system found in each bacterium[237].

In *E. faecalis* strain V583, it has been detected a SUF system that contains subunits SufB, SufC, SufD, SufS and SufU[68]. We have compared these proteins by blastp to the genes detected from Aus0004 strain used in our study and we have found high percentage of identify and coverage (Table 34).

**Table 34. Percentages of identity and coverage of the different subunits of the [Fe-S] cluster between *E. faecium* Aus0004 and *E. faecalis* V583.**

| *E. faecalis* V583 | *E. faecium* Aus0004 | Identity | Coverage |
|---|---|---|---|
| SufB | SufB | 94.79 % | 99 % |
| SufU | NifU | 72.26 % | 99 % |
| SufS | Cysteine desulfurase (SufS) | 78.29 % | 99 % |
| SufD | SufD | 76.17 % | 100 % |

Iron is an essential nutrient and it is involved in different biological process, also acting as a co-factor in diverse metabolic pathways. Iron absorption occurs in the small intestine, by the epithelial cells[238]. Nevertheless, not all dietary iron is absorbed in the small intestine and the unabsorbed iron reach the large intestine, where plays an important role in the growth of intestinal pathogens[238].

This SUF system, which is the only Fe-S cluster present in Gram-positive bacteria belonging to Firmicutes, has been related to virulence in different microorganisms such as *Mycobacterium tuberculosis* and *Erwinia chrysanthemi*[68]. Specifically, an orthologous of the SufB protein in *M. tuberculosis* (Pps1 protein) has been checked to play an essential role under iron limitation and oxidative stress[239]. Its deletion caused a reduction of the survival of *M. tuberculosis* since this gene was essential for iron acquisition.

Regarding the use of iron by *E. faecium* and the contribution of the Fe-S cluster to its utilization and the role in gut colonization, we decided to use chelators *in vitro* with both strains, WT and [Fe-S] cluster mutant separately to verify the importance of these genes on the utilization of different iron states ($Fe^{2+}/Fe^{3+}$) for the growth of the bacteria. We used different chelators for the diverse forms of iron states ($Fe^{2+}$, $Fe^{3+}$ and chelator for both)[178]. We observed that the chelator that affected more the growth of the mutant was the chelator for both states of iron, indicating that the cluster could be involved in the transport of both ions (Figure 60). In the case of the chelator for the $Fe^{3+}$ we observed smaller but significant differences between the WT and the mutant strain. No differences were observed when $Fe^{2+}$ chelator was added to the media. In the intestine, $Fe^{3+}$ iron is the more common state[238], so this could be the reason that explain that *E. faecium* is more affected by the reduction of $Fe^{3+}$, because it may be the iron state preferentially used in the gut. Even though we detected some differences, the decrease

of the growth of the mutant was not so pronounced. Nevertheless it is possible that when the mutant has to grow in the presence of a WT competitor strain, the differences will be more pronounce. Therefore, this experiment should be repeated but doing a competition assay in order to verify if under these conditions the iron present in the medium is preferentially used by the WT strain.

### 4.1.7 Characterization of genes related to the use of carbon sources present in the gut

### 4.1.7.1 Characterization of *man*X

Sugars present in the gut are important nutrients for enterococci growth being able to metabolize diverse carbohydrates[62,64], presenting a core-genome to carbohydrate metabolism[6]. Moreover, one of the main differences between the *E. faecium* clade A (hospital associated strains) and clade B (commensal strains) is the enrichment for genes related to carbohydrate uptake. More specifically, strains of clade A have acquired more phosphotransferase systems (PTS) system inserted in mobile genetic elements and obsolete PTS have been removed from clade B strains, being the main genomic driver that separates both clades[57]. One of the genes that we found to be most relevant for gut colonization encodes for the PTS mannose transporter subunit IIAB.

Commonly, the PTS are classified in 2 types: (I) containing an histidine phosphocarrier protein (HPr) and (II) which are specific sugar membrane permeases[71]. ManX belongs to the second type, which is formed by two outer cytoplasmic subunits (IIA and IIB). These subunits are hydrophilic proteins that are specific for the sugar. In addition, the subunit IIC, a hydrophobic membrane subunit, that functions as a sugar permease with or without the IID subunit, a second integral membrane protein[71]. Specifically the PTS transporter identified in Aus0004 is theoretically composed by four subunits (Figure 90): (I) PTS sugar transporter subunit IIB, (II) PTS mannose transporter subunit IIAB (*man*X), (III) PTS mannose/fructose/sorbose transporter subunit IIC, (IV) PTS mannose/fructose/sorbose IID.



**Figure 90. Schematic representation of the phosphotransferase transporter system (PTS) found in Aus0004 *E. faecium* strain.** ManX (the PTS mannose transporter subunit IIAB detected to be involved in gut colonization) is a subunit of the PTS system that consists of a total of four subunits. In brown is indicated the gene removed from Aus0004 genome that has been checked to diminish VRE intestinal colonization and in orange are coloured the other subunits that form the PTS system.

We have checked the presence of the mutants for the different subunits in the transposon mutant library to verify if these proteins were essential for the VRE growth. We have found that the presence of the subunits IIB and IIC are too low or almost no detectable in any of the conditions tested *in vitro* or *in vivo*. It is possible that these two subunits are required for the internalization of other sugars indicated in the function assigned to this particular genes (e.g. fructose, sorbose) and therefore the phenotype will be more drastic that the one detected for the IIAB subunit, which only affects *in vivo*

growth. In the case of the case of the subunit IID, it was identified in the Tn-Seq library in all samples with an average of 40 counts per sample. This particular subunit, although it could be involved theoretically in the acquisition of fructose, sorbose and mannose, it has been detected only in PTS require for the acquisition of mannose. Thus, its function may be more specific of mannose and may not be required for the utilization of the other sugars. Nevertheless, this should be experimentally confirmed.

Indeed, we could confirm, using a Biolog plate that *man*X was mainly involved in the acquisition and utilization of mannose. Indeed, the growth of the mutant was not affected when other sugars were the main nutrient source for VRE. There was just an exception: glucose. Interestingly, in this sugar, the mutant was able to grow better than the WT. Moreover, as we indicate below, this enhanced growth was also detected in the mutant of *mur*Q (specific for n-acetyl muramic acid) (Results, section1.9.4, Figure 65). It is well know that PTS systems for sugar uptake can regulate the catabolism of the cell[240]. This is a phenomenon called "the inducer exclusion" in which the utilization of a sugar that is transported by a PTS system is going to repress the utilization of other sugars. It is therefore possible that the deletion of a PTS system in *Enterococcus* decreases the repression to utilize glucose as an energy source. If this was the case, in the absence of a PTS, genes related to the utilization of glucose will be overexpressed and therefore the capacity of *Enterococcus* to grow on the presence of glucose will be enhanced. However, we do not have an explanation of why glucose is the only sugar affected.

In order to corroborate the role of the PTS required for mannose uptake, we complemented the mutant strain with a plasmid encoding the *man*X gene whose expression was controlled by a constitutive promoter. The growth using mannose as a carbon source was restored in the complemented strain, which demonstrated that this gene is required for mannose uptake. Also, the growth of the mutant was restored *in vivo,* and did not differ from the growth of the WT strain*.* As a control of this *in vivo* complementation experiment, we included two mice that were colonized with the WT and the mutant strain (not being complemented). Surprisingly, the effect of the mutant in VRE gut colonization in this experiment was not as obvious as in previous experiments. One possible explanation for this differential effect of the mutant in different experiments in mice could be the different availability of mannose. Mannose can be found in the gut linked to the mucin (heavily glycosylated proteins that formed the mucus). But also it could be supplied by the diet: small amounts of mannose can be found in fruits and complex carbohydrates can also contain mannose. In this last case, complex carbohydrates could be fermented by commensal bacteria, mainly anaerobes, that will led to free available mannose that could be used by VRE but also by other bacteria present in the gut[187]. Therefore, the amount of free mannose available could differ depending on (I) the diet, (II) presence of commensal bacteria that liberates mannose, (III) presence of bacteria that consume mannose. Although we used the same mouse colonies, it is true that the microbiota could have evolve during the three years in which the experiments were performed and this could have an impact on the levels of mannose. Therefore, it will be important to check how different levels of mannose in the gut impact the colonization capacity of the *man*X mutant. In addition, it will be

important to check how the remaining microbiota could influence mannose availability. In addition, although the diet that we used is the same in terms of overall nutritional requirements for mouse breeding (i.e. % of carbohydrates, % proteins, % fat), the specific sources of the carbohydrates could have change. Thus this could have also impact the levels of mannose in the gut.

In a first attempt to check if increasing the amount of mannose in the gut could increase the VRE levels, we administrated mannose in the drinking water after antibiotic treatment. We observed a slightly increase in VRE levels in the group of mice that received mannose. However, this increase was not significant. One problem with this particular experiment is that the levels of VRE were already very high (close to the levels detected in germ-free mice[146]). Thus, in this conditions in which there is plenty of nutrients available it is possible that increasing mannose will not have a drastic effect. On the other hand, mannose can be uptaken by epithelial cells in the small intestine. Thus we don´t know to what extend the mannose supplied in the drinking water reach the large intestine, where we have been measured during the whole thesis the levels of VRE.   Thus, it will be relevant to check through mass spectrometry the amount of mannose in the large intestine of these mice. An analysis that could not be done for time constraints.

Comparing with previous literature available, other PTS systems have been related to increase the enterococci virulence. For example a PTS system known as *bep*A, that is highly express during the infective endocarditis process[78]. Deletion of this PTS causes a reduction in the thickness of the biofilm and a growth defect in the presence of β-methyl-D-glucoside, a glycosaminoglycan exposed on injured heart valves[78]. On the other hand, three subunits of a different PTS of *E. faecium* have been related to resistance to grow in human serum and their deletion reduced *E. faecium* growth in an *in vivo* zebrafish model[99]. In addition, in one particular study, the authors found a *Enterococcus* encoded PTS involved in gut colonization[72]. This PTS system is present in hospital-associated strains, but is absent in human commensals. Notably, the authors found that its deletion reduced the ability of VRE to colonize the intestinal tract. However, the specific sugar uptake by this PTS was not identified in that study.

### 4.1.7.2 Characterization of the *mur*Q gene

On the other hand, we have identified another gene, *mur*Q, related to the use of N-acetylmuramic acid sugar, to be involved in the gut colonization by VRE. As we did previously with the *man*X gene, we first checked the specificity of this gene for the described function assigned (the utilization of the N- acetylmuramic acid). For this purpose, we grew the Δ*mur*Q mutant in the M1 medium adding exclusively N-acetyl muramic acid (MurNAc). We demonstrated that the mutant was not able to grow in the presence of N-acetyl muramic acid. We also grew the Δ*mur*Q strain in Biolog plates. This allows us to verify the specificity of the gene. Growth of the Δ*mur*Q mutant strain was not affected in the presence of other sugars. Except for glucose, in which the growth of the mutant was increased, as previously described.

According to the bibliography, *mur*Q has been previously describe in *E. coli* as a N-acetylmuramic acid 6-phosphatase etherase, an enzyme required for the metabolism of MurNAc, derived from the cell wall or from the environment[241]. *mur*Q is located adjacent to *mur*P, a phosphotransferase system (PTS), which transports the MurNAc through the cytoplasmic membrane and introduce it in the cell as N-acetylmuramate 6-phosphate (MurNAc-P)[241,242]. *Mur*Q, however, is responsible for converting the MurNAc-P to N-acetyl-D-Glucosamine 6 phosphate[242]. The second step for using MurNAc as a carbon source.

By protein homology using blastp, we have compared the amino acid sequence of the described *mur*P encoded by *E. coli* with the Aus0004 genome. The best match with a coverage of 98 % and identify of 35.76 % belongs to the gene downstream *mur*Q. This gene encodes a protein that was claimed by protein homology to be a PTS glucose transporter subunit IIA. Nevertheless, this gene is located adjacent to *mur*Q, as it occurs in *E. coli*, suggesting that the function of this gene will be to internalize the external MurNAc. That said, this putative *mur*P was detected through Tn-Seq in high abundance (average of 2724 counts) both *in vivo* and *in vitro,* suggesting that its deletion is not required for VRE growth. It is possible that some redundancy exists in the genome of Aus0004 and that other transporters encoded by this strain could internalize this sugar. To discern this possibility, one could delete first *mur*P to verify the absence of phenotype. Subsequently, a Tn-Seq approach could be applied in the *mur*P mutant background to identify other genes involved in MurNAc acquisition.

MurNAc is one of the major components of the peptidoglycan, which is formed by alternating residues of beta-(1-4) linked N-acetylglucosamine and MurNAc. Lysozyme, produced by the host as a defensive mechanism, lyse bacterial cells by catalysing the hydrolysis of the 1-4-beta linkages of both sugars. It will be interesting to test if lysozyme, indirectly, could be promoting the expansion of VRE through liberation of MurNAc. In addition, phages encode lysins that could also promote the liberation of this sugar if a phage lysogen is induced, starts the lytic cycle and express the lysin. Certain antibiotics, such as ciprofloxacin or beta-lactams can induce the lytic cycle of the phage by inducing the SOS response[243,244]. It will be interesting to test if administration of antibiotics promotes the lysis of bacteria through SOS induction, increase the levels of MurNAC and consequently promotes the expansion of VRE in the intestinal tract.

### 4.1.8 Putative role of *ltrR* in VRE gut colonization

Another gene that was involved in gut colonization by VRE was a LacI family DNA-binding transcriptional regulator (*ltrR*). These type of transcriptional regulators are present in different types of bacteria, being broadly distributed[245]. A study to identify the function and evolution of these type of regulators from over 270 bacterial genomes identified that 90% of them were regulators of sugar pathways[245] and the other 10 % control different set of metabolic genes. A previous study performed in *E. faecium* identified a Lac-I family regulator involved in the metabolism of maltodextrin[104]. A blastn was performed to identify if the candidate gene that we selected was this previous identified gene, but not similarities were found. Therefore, the LacI family DNA-binding

transcriptional regulator from our study might be involucrate in a different metabolic process.

Metatranscriptomic analysis performed from the intestinal contents of mice inoculated with the mutant and WT strains have allowed us to elucidate that the significant genes affected by the deletion of this transcriptional regulator were mainly overexpressed in the mutant strain compared to the WT strain. Thus, we hypothesize that this LacI family DNA-binding transcriptional regulator is a repressor, whose elimination caused that some genes were overexpressed. Moreover, some of the genes most significantly overexpressed in the mutant strain are located in the chromosome next to the transcriptional regulator (Figure 91), but are transcribed in the opposite direction. Altogether, these results strongly suggest that the LacI family DNA-binding transcriptional regulator repress the expression of the genes localized next to it. Curiously, as happens with the majority of the LacI transcriptional regulators, the regulated genes seem to be involved (by protein homology) in the uptake of carbohydrates. Specifically, they seem to be part of an ABC (ATP-binding cassete) transporter. ABC transporters, as in the identified operon, are usually composed of four proteins: two membrane proteins (permeases), the ATP-binding protein and the substrate-binding protein[246,247].
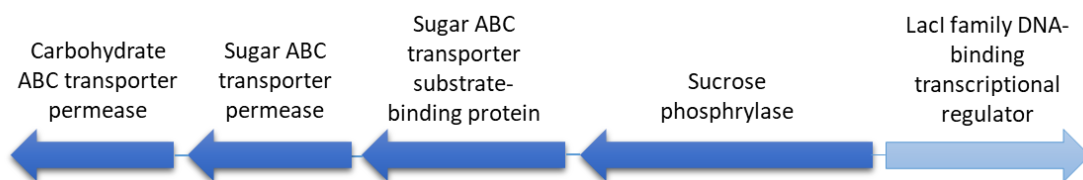


**Figure 91. Genes whose expression was affected by the deletion of the LacI family DNA-binding transcriptional regulator. Only those genes next to the transcriptional regulator in the chromosome are shown.** In light blue is represented the gene deleted from VRE genome and in dark blue the four genes whose expression in significantly increased with the deletion of the transcriptional regulator. Over the arrows is indicated the predicted function of the gene that is indicated in the Aus0004 deposited genome in NCBI.

The expression of this type of ABC transporters has been shown also in other organism to be controlled by LacI transcription regulators, such as in *Lactobacillus acidophilus*[248]. In this particular bacterium, a locus composed of a transcriptional regulator LacI along with the four proteins of the ABC transporter family was detected. Curiously, the expression was reduced in the presence of glucose[248]. However, the authors did not identify the sugars that activate the expression of the operon. Nevertheless, as it has been shown for other transporters of sugars such as PTS, its expression seems to be regulated by the sugars available in the environment.

Results obtained from the competition experiments between the WT and the LacI family DNA-binding transcriptional regulator mutant suggests that the overexpression of the sugar ABC transporter proteins is detrimental for gut colonization. Although we do not know the sugar that is transported by this ABC transporter, this result highlights how

important is a tight regulation of gene expression in bacteria for colonizing a given environment. Increasing the expression of certain genes that are not required for gut colonization could suppose a fitness cost for the mutant bacteria that will be outcompeted by a WT strain expressing lower levels of genes that may not be required for the colonization of the gut. In addition, as has been previously shown, the utilization of specific sugars could inhibit the usage of other type of sugars[240]. Thus, it is possible that utilization of sugars acquired by this particular overexpressed ABC transporter could be preventing the usage of sugars that promote more efficiently the growth of VRE. Future studies will be required to demonstrate this hypothesis.

### 5.  Implication of selected candidate genes in gut colonization under different antibiotic treatment conditions

As it was previously mentioned, different antibiotics affect microbiota in a different manner and we wanted to check if the genes that affected VRE gut colonization were relevant under different conditions caused by the disruption of the microbiota through different antibiotic treatments. Indeed, it has been previously shown for *C. difficile* that can this pathogen adapt the expression of metabolic related genes depending on the nutrients available in the different intestinal environments created by different antibiotic therapies[249].

We selected vancomycin because the levels of infection reached by VRE were similar to the levels reached with clindamycin, independently of the different changes produced to the microbiota by each antibiotic. The results obtained demonstrated that the key genes are relevant under different intestinal conditions created by different antibiotics. Nevertheless, a follow up study should be performed in order to evaluate how different are the intestinal environments produced by vancomycin or clindamycin. It is possible therefore that the environments created are not that different and that other antibiotics will impact in a different manner the growth of the mutants.

### 6.  Presence of the studied genes in clinical *E. faecium* strains

Previous studies have found differences in the presence of specific genes in clinical vs commensal enterococcal strains. Some examples of these genes have been previously mentioned, such as the *pst*D subunit of the PTS associated to increase gut colonization in mice, which is present mainly in clinical strains[72]. Another example is the *bep*A gene, which encodes a PTS, it enhances biofilm formation and is basically absent from commensal strains and enriched in *E. faecium* clinical isolates[98]. In addition, it has been also demonstrated that the virulence genes $hyl_{Efm}$ and *esp* were mainly found in vancomycin-resistant strains from hospitalized patients[250]. Thus, we wondered if those genes that we have better characterized in this study (*man*X, *mur*Q, *ltr*R, *pst*B2, proteins from [Fe-S] cluster) were also mainly present in clinical strains.

After aligning each gene with the 73 *E. faecium* strains we verified that two of the genes studied to be involved in gut colonization, *ltrR* and *pst*B2 were only present in strains from clade A (clinical and animal strains) and absent in strains from clade B (human commensal strains). This is relevant because if one would like to design therapies to prevent VRE intestinal colonization, ideally the targeted genes should be specific for clinical strains.

In the case of *man*X and *mur*Q genes, their presence predominates in strains from clade A but also they are found in some commensal strains, meaning that these genes are not specific traits of clinical strains. In the case of the [Fe-S] cluster, it was found both in commensal and clinical strains, suggesting that this cluster is essential for all strains from the species *Enterococus faecium* to survive in environments with iron limitation. Nevertheless, independently if the gene is or not specific of clinical strains, understanding the mechanisms used by *E. faecium* to colonize the gut will help us to design novel therapies to prevent infections by pathogenic strains of these species.

## 7.   Transcriptomic analysis from VRE grown *in vivo* vs *in vitro*

Transcriptomic analysis allowed us to detect those genes more expressed *in vivo* than *in vitro* revealing the importance of the PTS system during intestinal expansion since we found that several PTS for different carbohydrates such as lactose, fructose, mannose, mannitol or sorbose were overexpressed *in vivo* vs *in vitro*.

These results suggest that the capacity for acquiring sugars in the gut is crucial for VRE expansion and persistence. In concordance with results obtained in the second part of the thesis, showing that a main mechanism by which the microbiota confers protection is through fructose depletion. However, the two genes related to sugar utilization that were found to be relevant for gut colonization through targeted mutagenesis (i.e. *man*X and *mur*Q) were not overexpressed *in vivo*. One possibility for explaining this result is that these genes are constitutively expressed, thus their expression is not regulated by the concentration of substrate in the environment. Nevertheless, the results obtained in this transcriptomic analysis have to be taken with caution since the transcriptome of VRE *in vivo* was only analysed in two mice. Taking into account the possible variability between mice, additional experiments should be performed in order to verify if there are or not differences in the expression of the specific genes that in this thesis were found to be involved in gut colonization.

On the other hand, among the *in vivo* overexpressed genes, we detected a higher expression of an ornithine carbamoyltransferase, a gene related to the energy production. This is one of the three enzymes of the catabolic pathway involved in the degradation of arginine[251], one of the 20 essential amino acids to form proteins. The pathway to degrade this amino acid could be an essential mechanism for *E. faecium* since, as it was previously mentioned in the Introduction, section 1.7., arginine is decarboxylated to obtain energy[62,64]. Moreover, we have checked that Tn-mutants from this gene were significantly more abundant *in vitro* vs *in vivo* (log2FC >2). And

207

although it wasn't among the top 10 most significant Tn-mutants found differences, it was significantly less abundant both in colon and caecum conditions. Therefore, in this case, it could be that the gene is required for the colonization of the gut by VRE and also its expression needs to be induced *in vivo* in order to use arginine.

**8. Protection mechanism against VRE conferred by the administration of a bacteria consortium**

In a previous study from our group, a PhD student isolated five commensal bacteria (i.e. *Olsenella*, *Barnesiella, Alistiples, Oscillibacter and* unclassified_Ruminococcaceae), known as **Pro**tective **Bac**teria (ProBac), that can restrict VRE gut colonization in mice.

In this thesis, we have identified a novel mechanism by which these commensal bacteria, and more specifically *Olsenella,* can prevent VRE gut colonization: diminishing the availability of nutrients, specifically fructose, sugar that boosts VRE growth.

To determine the mechanisms by which ProBac was able to decrease VRE levels in the intestine, we performed transcriptomic and metabolomics analysis. Obtaining the information about the functions expressed by the commensal bacteria and the metabolites present in the gut. Firstly, transcriptomic analysis allowed us to assigned KEGG orthologs (KO) to non-redundant genes identified in the murine caecal transcriptome of untreated mice and vancomycin-treated mice that received or not ProBac. We detected that the administration of the five commensal bacteria was able to restore the expression of specific bacterial functions. Among those KO whose expression was significantly increased after the administration of ProBac, we identified *fru*A, a subunit of a phosphotransferase system, specific for the transport of fructose. Moreover, the fructose transporter was expressed by one of the five protective bacteria, *Olsenella,* which also expressed other genes required for the metabolism of fructose. Altogether these results pinpointed to a possible mechanism of protection by which ProBac, and more specifically *Olsenella*, could be diminishing the fructose availability in the gut which could be diminishing VRE colonization through nutrient competition. Consistent with this hypothesis, using metabolomics, we showed that the levels of fructose decreased after ProBac administration (Results, section 2.2, Figure 77). Moreover, we showed that the levels of fructose correlated with the capacity of VRE to colonize the gut. In addition, we found that fructose was one of the carbon sources that promoted higher VRE growth *in vitro*, besides glucose and mannose (Results, section 2.3, Figure 78).

The nutrient sources required for *Enterococcus* gut colonization have been not extensively studied and are largely unknown. A recent study shed some light into this question and identified lactose as a sugar that promotes *Enterococcus* growth in the murine intestine[102]. We attempt to analyse the levels of lactose in the caecal content from mice in order to investigate if ProBac could also be contributing to VRE inhibition through lactose depletion. However, we were not able to detect lactose in the caecal contents of the mice from our study (not shown). This result suggests that either lactose

is depleted by commensals that recovered after vancomycin treatment (independent of ProBac administration) or that most lactose was digested and absorbed by epithelial cells in the small intestine. Thus, at least in our mouse model, lactose may not be a key nutrient source for VRE growth in the large intestine.

Taking into account the transcriptomic and metabolomics data, we reasoned that the consumption of fructose by ProBac in the intestine (i.e. nutrient competition) could be the mechanism of the inhibition of VRE. To prove that, *ex vivo* experiments were set up with caecal filtered contents adding first the consortium of bacteria followed by VRE or VRE with an excess of carbohydrates. Effectively, ProBac was able to reduce the growth of VRE and the addition of mix of sugars to the caecal content was able to restore it, including the addition of only fructose (Results, section 2.3, Figure 82).

Sugars consumed by intestinal bacteria have two major origins: mucins (the major component of the intestinal mucus) and the diet[252,253]. Fructose is not frequently incorporated into mucins[253], but it is widely found in the diet as free monosaccharides, the disaccharide sucrose and as polymerized form (fructans). One of the major components of the mouse diet used in our study is wheat, which contains free fructose and fructans[254]. Although the majority of free fructose is absorbed in the small intestine[255], fructans cannot be absorbed[256] and serve as a nutrient source to bacteria in the intestine. Interestingly, fructans can be hydrolyzed into fructose by extracellular bacterial enzymes and the liberated fructose is then internalized through specific transporters[257]. Thus fructose sources in the large intestine for VRE growth may originate from monomers liberated from fructans by the action of the microbiota or from free fructose that was not absorbed in the small intestine. In the human population, the capacity of the small intestine to uptake fructose is limited and is frequently exceed due to the consumption of foods or beverage rich in fructose. For example, soft drinks that contain high fructose corn syrups as sweeteners and has been introduced in the consumer habits of modern societies[252]. The excess of fructose reaches the large intestine which leads to several symptoms, such as bloating, abdominal pain, derived from its fermentation by commensal bacteria[258]. Our study pinpoints an additional potential negative effect of high-fructose intake: increasing the gut levels of multidrug resistant pathogens such as VRE that can expand using this sugar as a carbon source.

Subsequently, we used the previously indicated *ex vivo* assay to investigate if any of the bacterium from the mix was sufficient to inhibit VRE growth (Figure 83). As expected, taking into account the transcriptomic data, we observed that the protection effect was caused by *Olsenella* but not by the other bacteria from ProBac. Subsequently, we performed *in vivo* experiments, inoculating *Olsenella* instead of ProBac and we confirmed that *Olsenella* was sufficient to diminish VRE intestinal colonization (Figure 87).

Since intestinal colonization is key for subsequent infections to develop in hospitalized patients[259], novel approaches, based on the microbiome, have been proposed in order to restrict VRE gut colonization. One of such strategies is the restoration of the

microbiota of patients by administering the fecal microbiota of a healthy donor (i.e. fecal microbiota transplant - FMT) [260]. However, such approach could be problematic[136], due to the complexity an incompletely defined composition of the FMT and the subsequent risk of introducing potentially pathogenic bacteria or virus. This will be even more risky in VRE infected patients, which are usually immunocompromised. For this reason, alternative strategies, involving the introduction of defined bacterial consortia or single strains have been proposed[261,262]. *Olsenella* therefore represents a novel candidate that could be used for restricting VRE gut colonization. Although our isolate has a murine origin, several species from the same genus have been isolated from humans, both in the oral cavity and faeces[263,264]. Interestingly, one of these species (*Olsenella profusa*) is able to ferment multiple intestinal carbohydrates, including fructose. Future studies that test the impact of this human isolate on VRE intestinal colonization are warranted. In addition, as an alternative, other bacteria with similar nutrient preferences as *Olsenella* could a priori be used to restrict VRE intestinal colonization. However, it is important to indicate that *Olsenella* was very efficient at colonizing the gastrointestinal tract upon antibiotic withdrawal and, as the transcriptome analysis pinpointed, it was functionally active *in vivo*. In addition, different bacteria may have differential preferences for sugars. The fact that one bacterium is able to consume a particular sugar *in vitro* does not necessarily mean that will consume it *in vivo* if other sugars are available[149,240]. All these considerations should be taken into account before testing alternative probiotics providing similar functions.

Although our study suggests that fructose is key for intestinal colonization of VRE, other sugars may also be relevant for VRE gut colonization and its depletion may be required for a complete eradication of VRE. Indeed, although we detected a significant correlation between the levels of fructose and the capacity of VRE to colonize the intestinal tract, a few mice could be efficiently colonized with VRE despite the fructose levels were close to those found in untreated mice. Our *in vitro* assays have pinpointed other sugars that can be found in the large intestine and that promote to a similar extend VRE growth *in vitro*. For example, galactose or mannose, sugars that are included in mucins or in other mucosal glycoproteins were efficiently used *in vitro* by the VRE strain used in this work and by other VRE strains[265]. In fact, mannose seems to be a relevant sugar for VRE growth in the intestinal tract as indicated by the study described in the first part of the thesis.

Future studies will be required to completely decipher the nutritional requirements of VRE in the intestinal tract and the identification of the members of the microbiota that prevent their utilization. Ultimately, all this knowledge could allow the design of personalized microbiome-based therapies to prevent VRE infections. Through metabolomics one could identify the nutrients available for VRE in the gut of a patient, then a proper cocktail of commensal bacteria that efficiently deplete those nutrients could be administered to complete eradicate VRE from the intestinal tract.

# VI. CONCLUSIONS

- The methodology of mutant transposon library is useful to study the genes of *Enterococcus* in different environments. However, the choice of strain is essential because the efficiency of library generation depends on the strain of *Enterococcus faecium* used.

- Tn-Seq analysis identified 71 candidate genes whose disruption by the transposon potentially decrease VRE gut colonization both in colon and caecum.

- Targeted mutagenesis of the genes with a potential higher effect on VRE intestinal colonization: *man*X, *mur*Q, *ltr*R, *pst*B2, proteins from [Fe-S] cluster, *yyc*I, *dsb*A, *adh*E, *pgt* confirmed their role in the VRE gut colonization process: their deletion diminished the capability of VRE to colonize the intestinal tract of mice.

- The selected candidate genes affected VRE gut colonization under different antibiotic treatments (clindamycin and vancomycin).

- *In vitro* assays demonstrated that Fe-S cluster is involved in the transport of iron ions, mainly $Fe^{3+}$.

- *In vitro* assays demonstrated that the *man*X and *mur*Q genes are required specifically for utilization of mannose and N-acetyl-muramic acid respectively.

- Transcriptomic analysis suggests that the *ltr*R gene is a transcriptional regulator of an operon involved in sugar acquisition.

- *ltr*R and *pst*B2 genes are encoded by *E. faecium* strains from clade A, containing hospital-associated strains and strains from animal origin, while are absent from strains from clade B (human commensal strains). *man*X and *mur*Q are enriched in strains from clade A, while the Fe-S cluster is present in all type of strains.

- The administration of the five commensal bacteria (ProBac), consisting in isolates from the taxa *Olsenella*, *Alistipes*, *Barnesiella*, *Oscillibacter* and unclassified_Ruminococacceae reduces the levels of VRE in the murine intestine as compare to the administration of bacteria not associated with protection against VRE (i.e. an isolate from the *Bacteroides* genus).

- The administration of ProBac to antibiotic treated mice restores the expression of specific bacterial functions, including genes related to the metabolism of fructose. Expression of these genes is mainly driven by one isolate belonging to the genus *Olsenella*.

- Vancomycin treatment increases the levels of fructose available in the murine large intestine, while the administration of ProBac reduces significantly the levels of fructose.

- A positive correlation was found between the levels of fructose in the gut and the capacity of VRE to colonize the intestinal tract of mice.

- *In vitro* assays showed that fructose is one of the sugars that allow a higher VRE and *Olsenella* growth.

- Fructose supplementation boosts VRE levels in the large intestine in mice. However, fructose-dependent VRE expansion is abolished if ProBac isolates are present in the murine gut.

- *Ex vivo* experiments demonstrated that ProBac inhibits VRE growth through nutrient depletion.

- *Olsenella* is sufficient to inhibit VRE growth *ex vivo* and *in vivo*.

# VII. RESUMEN

## INTRODUCCIÓN

Las bacterias resistentes a antibióticos son un problema cada vez más recurrente en pacientes hospitalizados[47]. *Enterococcus faecium* es un microorganismo resistente a múltiples antibióticos, entre ellos la vancomicina (Enterococo resistente a vancomicina, ERV) lo que dificulta su tratamiento, causando graves infecciones y siendo una causa importante de enfermedad nosocomial[32]. Se han descubierto nuevos antibióticos para tratar las infecciones por ERV, sin embargo, las cepas resistentes a estos nuevos antimicrobianos emergen rápidamente[47], lo que fomenta la implementación de otro tipo de estrategias que no sean antibióticos para prevenir infecciones causadas por este patógeno[47].

Bajo condiciones normales, el enterococo representa un porcentaje muy pequeño de la microbiota adulta, menor de un 1 %[49] y su presencia varía en función de la localización, estando presente tanto en el intestino delgado como en el grueso[110]. Distintas especies de enterococo están presentes en el intestino delgado, que se compone de comunidades microbianas más simples de anaerobios facultativos y aerobios[18]. También, están presentes en el ciego y colon, cuya microbiota está compuesta por una mayor variedad de anaerobios[18].

La presencia de *E. faecium* como parte del tracto gastrointestinal, junto a su tolerancia y capacidad para adaptarse y sobrevivir en entornos extremos hace que estos organismos se propaguen de forma muy rápida y fácil en ambientes hospitalarios[1,47]. Actualmente, se considera la segunda infección más frecuentemente asociada a hospitales en Estados Unidos[1,29] y la tercera causa más prevalente en Europa[30], siendo conocida como una de las enfermedades nosocomiales de mayor relevancia. La tasa de mortalidad de pacientes infectados con *E. faecium* depende del tipo de infección, el factor de riesgo, la duración de la exposición y el tipo de antibióticos, pudiendo variar entre el 20 y 60 %[26,31].

Las infecciones causadas por este patógeno ocurren tras un tratamiento con antibióticos y suelen comenzar donde se localiza el enterococo, en el intestino. Los antibióticos disminuyen la capa protectora del mucus del intestino[24,25] y además alteran la composición de la microbiota[24], facilitando su expansión en el intestino. En condiciones normales, la microbiota se compone de cientos de especies bacterianas comensales que juegan un papel fundamental en prevenir la colonización y expansión de microorganismos, lo que se conoce como "resistencia a la colonización" (CR)[114]. Sin embargo, dicha resistencia se pierde con la acción de los antibióticos. Tras el tratamiento antibiótico y disrupción de la microbiota, ERV coloniza el intestino alcanzando densidades muy altas, lo que favorece su propagación al torrente sanguíneo y a otros órganos[28]. A pesar de la relevancia clínica, se conoce muy poco acerca de las bacterias comensales necesarias para la protección frente a *E. faecium* y los mecanismos por los cuales la microbiota confiere protección. Recientemente, un estudio ha demostrado que *Blautia producta*, una bacteria comensal anaerobia inhibe la colonización intestinal de ERV en ratones mediante la producción de un lantibiótico[261,262]. De la misma manera, nuestro grupo ha identificado un consorcio de 5

217

bacterias (ProBac – <u>Bac</u>teria <u>Pro</u>tectoras) capaces de disminuir la colonización intestinal de ERV en un modelo de ratones tratados con antibióticos. Sin embargo, desconocemos el mecanismo por el cual ProBac disminuye la colonización intestinal de ERV, uno de los objetivos planteados en esta tesis.

Por otro lado, a pesar de la alta incidencia de este tipo de infecciones, se conoce muy poco sobre los genes expresados por ERV que son necesarios para colonizar el intestino o los nutrientes que necesita para poder sobrevivir en el tracto intestinal, siendo este otro de los objetivos principales de la tesis. En esta tesis por tanto, pretendemos conocer con más detalle cómo ERV es capaz de colonizar el intestino y cómo la microbiota confiere protección frente a dicho patógeno. Dichos conocimientos son imprescindibles si queremos prevenir las infecciones causadas por este tipo de patógenos en pacientes hospitalizados.

**OBJETIVOS**

El objetivo general del estudio es caracterizar los requerimientos tanto genéticos como nutricionales que el enterococo resistente a vancomicina requiere para colonizar el intestino, junto con la caracterización del mecanismo de protección de bacterias comensales que protegen frente a ERV.

Los objetivos específicos que se engloban en el objetivo principal para alcanzar ambas metas son:

1. La construcción de una librería de mutantes por transposición en ERV para identificar los genes necesarios para la colonización intestinal mediante secuenciación masiva y su posterior verificación por mutagénesis directa.

2. Caracterización funcional de los genes que afectan a la colonización intestinal, con especial atención a los relacionados con la adquisición de nutrientes.

3. Análisis transcriptómico para identificar los genes más expresados por la ERV durante el proceso de colonización intestinal.

4. Caracterización mediante transcriptómica y metabolómica de las funciones expresadas y cambios en los metabolitos producidos por la acción ProBac y que podrían ser relevantes para conferir resistencia frente a ERV.

5. Experimentos *in vivo* y *ex vivo* para demostrar los mecanismos específicos (sugeridos por los estudios ómicos) mediante los cuales ProBac confiere protección contra ERV.

Todos estos objetivos nos permitirán identificar los mecanismos moleculares por los que ERV es capaz de colonizar el intestino, así como los nutrientes necesarios que promueven su crecimiento y el mecanismo por el cual bacterias comensales específicas confieren protección contra ERV.

## METODOLOGÍA Y RESULTADOS

### 1. Identificación y caracterización de los genes necesarios para la colonización intestinal de ERV

#### 1.1 Construcción de la librería de mutantes por transposición

Para identificar qué genes son necesarios para la colonización intestinal de ERV hemos utilizado una estrategia previamente publicada por Zhang et al[99], a través de la cual se identifican genes necesarios para el crecimiento de dicho patógeno en suero humano. Como tuvimos dificultades para la correcta construcción de la librería, ya que teníamos problemas con la pérdida del plásmido y la inserción aleatoria del transposón, la librería se obtuvo en el laboratorio donde se publicó dicha técnica, en Utrecht, Holanda, bajo la supervisión del Dr. Willem van Schaik.

Finalmente y tras testar diversas cepas clínicas de *E. faecium,* obtuvimos una librería de mutantes por transposición. Para ello se utilizó un plásmido (pGPA) que contiene un replicón termosensible gram-positivo, un replicón ColE1 para la replicación de plásmidos en *E. coli,* una transposasa cuya expresión es inducida por nisina, un gen que confiere resistencia a cloranfenicol y un transposón que confiere resistencia a gentamicina. Después de la inducción de la transposasa por la nisina, el transposón resistente a gentamicina se inserta aleatoriamente en el genoma de la bacteria, consiguiendo así la librería de mutantes por transposición. Para confirmar que la librería era óptima para el desarrollo del estudio, se comprobó mediante PCR, tanto la pérdida del plásmido como la aleatoriedad de la inserción del transposón en el genoma de la bacteria.

La cepa utilizada para la construcción de la librería fue la cepa clínica Aus0004, aislada del torrente sanguíneo de un paciente hospitalizado en Australia, en 1998[61]. Esta cepa es resistente a múltiples antibióticos, como la ampicilina, los aminoglucósidos y la vancomicina.

Nuestros resultados demuestran que aunque la metodología había sido previamente publicada y testada, esta no funciona con la misma eficiencia en las distintas cepas que utilizamos.

#### 1.2 Estudio del efecto de los antibióticos en la colonización de ERV

Una vez obtuvimos las librería de mutantes se puso a punto el diseño de los experimentos *in vivo* con ratones (C57BL6). Primero se testaron distintos antibióticos con el fin de alterar la microbiota y seleccionar aquellos que promovieran una mayor colonización de ERV, y por tanto de la librería de mutantes. Los niveles de colonización más altos y homogéneos se alcanzaron con el tratamiento con vancomicina y clindamicina durante 4 días, seguido de la inoculación de ERV y la posterior cuantificación del mismo tres días tras la inoculación (séptimo día del tratamiento con antibiótico). Dichos antibióticos fueron por tanto seleccionados para inocular la librería de mutantes. A día siete post-tratamiento se recogieron muestras de distintas partes

del intestino incluyendo íleon, ciego y colon, junto a las paredes intestinales de cada localización. Se recogieron muestras tanto para evaluar los niveles de colonización en cada parte del intestino como para posteriormente analizar la librería de los mutantes de ERV por secuenciación masiva.

Comprobamos que los niveles de colonización del patógeno varían en función de la parte del intestino, siendo más elevados y similares en el intestino grueso (ciego y colon) y menores en el intestino delgado. Los niveles más bajos fueron detectados en las paredes intestinales. Cabe destacar que los niveles alcanzados en el modelo con clindamicina eran algo más elevados y homogéneos que con vancomicina. Teniendo en cuenta los resultados obtenidos, finalmente se seleccionaron las muestras de colon y ciego tras el tratamiento con clindamicina para la posterior secuenciación de la librería de mutantes inoculada.

### 1.3 Secuenciación de las muestras y análisis de las secuencias

Para secuenciar la librería crecida *in vitro* y la librería obtenida tras la inoculación de los ratones, se utilizó la técnica descrita por Zhang et al[99]. Dicha metodología consiste en extraer el DNA de las muestras, digerir y amplificar los fragmentos del genoma donde se ha insertado el transposón, amplificando la parte final del transposón junto a 16-18 nucleótidos del genoma de *E. faecium* Aus0004.

Tras la obtención de las secuencias, se puso a punto un programa para alinear dichas secuencias con el genoma de la bacteria y detectar los genes donde el transposón se había insertado. Dichas secuencias se alinearon con las regiones codificantes de Aus0004 utilizando blastn, con un mínimo de un 90 % de cobertura e identidad, es decir, permitiendo como máximo uno o dos nucleótidos incorrectos.

En primer lugar se realizaron análisis globales, un Análisis de Correspondencias Canónicas (CCA) y un Análisis de Coordinadas Principales (PCoA), para detectar diferencias en la abundancia relativa de los mutantes entre las muestras *in vitro* e *in vivo.* Ambos resultados mostraron una clara separación entre ambos tipos de muestras, *in vitro* e *in vivo*, aunque dentro de las muestras *in vivo* había más variabilidad. Además, no se detectó ninguna separación entre las muestras de colon y de ciego. Para confirmar estas diferencias entre grupos aplicamos el test PERMANOVA, siendo significativas las diferencias entre las muestras *in vitro* e *in vivo* ($p < 0.05$). Sin embargo, dichas diferencias no son significativas cuando comparamos ambas muestras *in vivo*, ciego vs colon. Los resultados obtenidos sugerían que ERV necesita un conjunto distinto de genes para crecer *in vitro* que para colonizar el intestino.

A continuación estudiamos diferencias significativas de manera concreta en la abundancia de cada uno de los genes del genoma de ERV en las muestras *in vitro* vs las muestras *in vivo*. En total detectamos 556 genes cuya abundancia (i.e. abundancia de los respectivos mutantes) difería significativamente (p<0.05; FDR<0.05; Log2 Fold Change >|1|; wilcoxon test) en las muestras *in vitro* vs las muestras *in vivo* (tanto en el

ciego como en el colon). Más concretamente, este análisis detectó un total de 71 genes cuya abundancia era mayor *in vitro* que *in vivo* y un total de 485 genes cuya abundancia era mayor *in vivo.* Resultados similares fueron obtenidos aplicando el test DeSeq2. Puesto que el objetivo de la tesis es el estudio de genes que afecten a la colonización intestinal, nos centramos en los genes que estaban más aumentados *in vitro* y cuya abundancia *in vivo* era muy baja. Dentro de estos genes, decidimos evaluar con más detalle los 10 cuya diferencia en la abundancia entre *in vivo* e *in vitro* era más significativa. Dichos genes codificaban, según un análisis in silico de su secuencia, para una subunidad EIIAB del transportador PTS – "phosphotransferase system" de manosa (*man*X), un regulador transcripcional de la familia LacI (*ltr*R), una ácido N-acetilmurámico 6-fosfato éterasa (*mur*Q), una subunidad de un transportador dependiente de ATP (ABC) (*pst*B2), SufB, una subunidad del complejo proteico [Fe-S] (*suf*B), una estreptogramina de la familia Vat acetiltransferasa (*vat)*, una proteína reguladora del sistema de dos componentes YycI (*yycI)*, una proteína de la familia DsbA (*dsb*A), la proteína acetaldehído bifuncional-CoA/alcohol deshidrogenasa (*adh*E) y la proteína glucosiltransferasa (*pgt)*.

### 1.4 Mutagénesis dirigida e implicación de los genes delecionados en la colonización intestinal

Tras la identificación de los 10 genes candidatos, se procedió a su deleción en el genoma de la cepa salvaje de ERV mediante un proceso de mutagénesis dirigida, sustituyendo el gen que queremos delecionar por el gen de gentamicina[153]. De esta manera podemos verificar la relevancia de genes específicos en la colonización intestinal de ERV.

Para tener un control con el marcador de gentamicina que nos permitiese comprobar que dicha inserción no afecta al crecimiento de la bacteria, se introdujo dicho gen en una región intergénica, al lado de una isla de patogenicidad. Efectivamente, la inserción de gentamicina no afectó al crecimiento de ERV ni *in vitro* ni *in vivo.*

Una vez aplicamos la metodología y obtuvimos los mutantes de los genes candidatos por deleción y sustitución por el gen que confiere resistencia a gentamicina, se llevaron a cabo experimentos de competición tanto *in vivo* como *in vitro*. Para los experimentos *in vivo,* utilizamos el modelo de ratón suministrando clindamicina, co-inoculando al cuarto día del tratamiento la cepa salvaje con cada uno de los mutantes y recogiendo las muestras a los tres días post-inoculación para plaquear en medio selectivo. Para los experimentos *in vitro,* se cultivó la cepa salvaje con cada una de las cepas mutantes, plaqueando en medio selectivo a las 24h de incubación. Tanto *in vivo* como *in vitro* se realizaron dos experimentos independientes testando todos los mutantes.

Los resultados *in vivo* mostraron que la cepa salvaje presenta una ventaja significativa de crecimiento sobre las cepas mutantes en todos los mutantes testados menos en uno de ellos (*Δvat*::*gen*), cuyo crecimiento *in vivo* era incluso algo mayor que la cepa salvaje. Dicho mutante fue descartado del estudio. Por el contrario, *in vitro* no se observó una ventaja de la cepa salvaje frente a las cepas mutantes, excepto frente al mutante para

el gen *suf*B. Este resultado indicaba que con excepción del gen *suf*B, todos los demás genes estaban implicados específicamente en la colonización intestinal de ERV. Respecto al resultado obtenido con el mutante para el gen *suf*B, razonamos que al haber eliminado solo uno de los genes del operón, el resto de proteínas derivadas del mismo podría generar un complejo de Fe-S tóxico para la bacteria, por lo que diseñamos la construcción de un nuevo mutante delecionando varias subunidades del complejo, además de SufB. Repetimos los experimentos con dicho mutante y continuamos observando la ventaja de la cepa salvaje *in vivo* en comparación con el mutante pero no observamos ninguna ventaja de crecimiento *in vitro,* ni en aerobiosis ni en anaerobiosis, de la cepa salvaje vs la cepa mutante. Este experimento confirmó también el papel específico en la colonización intestinal del operón que codifica para el complejo Fe-S.

Además de este modelo de ratón, también testamos la ventaja de la cepa salvaje respecto a las cepas mutantes utilizando el modelo de ratón en el que administrábamos vancomicina en agua de bebida. Con dicho experimento, quisimos comprobar si los mutantes se comportan de la misma manera con distintos tratamientos que causen un cambio diferente en la microbiota, pero que también permiten altos niveles de colonización de ERV. Por otro lado, la evolución en la proporción de ambas cepas en el intestino de los ratones se puede prolongar en el tiempo ya que la vancomicina se suministra en el agua de bebida, mientras que la clindamicina se administra vía subcutánea lo cual supone un stress para el ratón. Por tanto, en este modelo, además de recoger muestras a día 3 post-infección como en el caso del modelo con clindamicina, recogimos muestras a día 6 y 13 post-infección. Los resultados obtenidos fueron muy similares a clindamicina a día 3. Además, al continuar el experimento durante más días se pudo detectar que alguno de los mutantes se eliminaban por completo del intestino de algunos ratones a día 13, mientras que la cepa salvaje continuó permaneciendo en el intestino a niveles muy elevados (aproximadamente $1*10^9$/g heces). Estos resultados confirmaron que los genes delecionados son clave para la colonización intestinal de ERV bajo los efectos de distintos antibióticos.

### 1.5 Caracterización de los genes implicados en la colonización intestinal de ERV

Seguidamente, decidimos evaluar con más detalle la funcionalidad de algunos de los genes implicados en la colonización intestinal por ERV. De los genes descritos, decidimos caracterizar aquellos genes cuya potencial función en la colonización intestinal era más clara de inferir y estos son los genes relacionados con la adquisición de nutrientes: la subunidad EIIAB del transportador PTS de manosa (*man*X); el regulador transcripcional de la familia LacI (*ltr*R), frecuentemente relacionado con la regulación de genes necesarios para la captación de azúcares; el gen que codifica para una ácido N-acetilmurámico 6-fosfato éterasa (*mur*Q); la proteína transportadora dependiente de ATP (ABC) (*pst*B2) y el complejo proteico [Fe-S] (complejo [Fe-S]).

Sin embargo, los otros genes *yyc*I, *dsb*A, *adh*E y *pgt*, decidimos no caracterizarlos en detalle debido a que su papel en la colonización intestinal era (i) difícil de inferir según

la secuencia proteica, como la proteína con función hipotética (*yyc*I) y (ii) podrían estar implicados en distintos procesos, los cuales no podíamos estudiar en detalle en el periodo limitado de la tesis: el gen que codifica para una alcohol deshidrogenasa (*adh*E), el gen que codifica para una proteína de la familia (*dsb*A), y el gen que codifica para una glucosiltransferasa (*pgt*).

### 1.5.1 Genes involucrados en la utilización de iones presentes en el intestino

Comenzamos con el estudio de las proteínas asociadas al grupo Fe-S y al transportador ABC de fosfato, cuya caracterización fue el objetivo de la segunda estancia durante la tesis, en el laboratorio del Dr. Van Sinderen. El grupo receptor de la estancia ha publicado previamente varios artículos en los cuales han caracterizado genes requeridos por *Bifidobacterium* para crecer bajo condiciones limitantes tanto de hierro[178] como de fosfato[161].

En el caso del transportador de fosfato (*pst*B2) tratamos de poner a punto dos medios mínimos previamente utilizados para el crecimiento de *Bifidobacterium*[152,161], en los que se elimina por completo el fosfato. Desafortunadamente, no encontramos el medio mínimo óptimo para el crecimiento de ERV y del mutante de fosfato, por lo que tendremos que seguir buscando un medio que nos permita crecer la cepa salvaje y el mutante en un medio mínimo que no contenga fosfato y al cual podamos añadir concentraciones crecientes de fosfato para comprobar si realmente dicho transportador está involucrado en el transporte de fosfato en concentraciones limitantes.

En cuanto a la caracterización de las proteínas asociadas al grupo Fe-S, se utilizaron diferentes agentes quelantes para las distintas oxidaciones del hierro[178]. Como resultado observamos que se reduce el crecimiento del mutante comparado con la cepa salvaje al añadir el quelante para los dos estados de oxidación del hierro, $Fe^{2+}$ y $Fe^{3+}$ (fenantrolina). Esto podría indicar que dicho gen está involucrado en el transporte de iones de hierro en los dos estados de oxidación. También se observaron diferencias pero menos significativas agregando el quelante para el ion $Fe^{3+}$ (ciclopirox olamina). Sin embargo, no se observaron diferencias significativas entre ambas cepas cuando se añadió el quelante para $Fe^{2+}$. Estos resultados sugieren que el hierro es relevante en condiciones en las que ambas formas de hierro estaban presentes en concentraciones limitadas, siendo más importante el ion $Fe^{3+}$ que es el ion más comúnmente encontrado en el intestino.

### 1.5.2 Caracterización de la expresión génica del regulador transcripcional *ltr*R

Para caracterizar el regulador transcripcional LacI (*ltr*R) realizamos un experimento *in vivo* con el modelo de ratón de clindamicina, inoculando la cepa salvaje y la mutante (Δ*ltr*R::*gen*) a distintos ratones por separado.

Se realizaron análisis transcriptómicos y comparamos la expresión de cada gen codificado por ERV en los dos grupos de ratones, infectados con la cepa salvaje o la mutante. Con el fin de encontrar genes cuya expresión fuera significativa, aplicamos el test estadístico DeSeq2. De esta manera, observamos como la mayoría de los genes

significativos (26 de un total de 31) (p<0,05, FDR<0,1, log2 Fold Change (FC) > |1|) estaban sobreexpresados en el grupo de ratones inoculados con la cepa mutante. Por lo tanto, estos resultados sugerían que el regulador transcripcional LacI podría ser un represor y su deleción promueve la expresión de genes que estaban previamente reprimidos. Además, entre estos genes sobreexpresados se encontraron cuatro genes relacionados con el trasporte de azúcares. Más concretamente se trata de cuatro genes que codifican para proteínas de la familia de transportadores ABC que se localizan aguas abajo del regulador transcripcional *ltr*R. Este último resultado indicaba que el gen *ltr*R podría ser un regulador transcripcional de genes relacionados con la utilización de azúcares, tal y como se ha descrito para otros reguladores transcripcionales de la familia LacI.

### 1.5.3   Genes involucrados en la utilización de azúcares presentes en el intestino

Posteriormente nos centramos en otros dos genes relacionados con la utilización y adquisición de azúcares. Por un lado estudiamos el gen que codifica para la subunidad EIIAB del transportador PTS de manosa (*man*X). La predicción funcional para dicho gen es el transporte de manosa. Por otro lado, estudiamos el gen que codifica para una ácido N-acetilmurámico 6-fosfato eterasa (*mur*Q), enzima que teóricamente estaría involucrada en la utilización del N-acetilmurámico. Con el fin de corroborar la implicación de los genes en la utilización de azúcares, utilizamos las placas de Biolog con un medio mínimo (M1) optimizado para crecer ERV en presencia de fuentes de carbono únicas. Desafortunadamente, dichas placas no contenían N-acetilmurámico. Por tanto este azúcar fue testado de forma individual en posteriores experimentos. Los resultados obtenidos verificaron que el gen *man*X era relevante para la utilización de manosa, ya que el crecimiento de este mutante disminuía significativamente cuando la manosa era la única fuente de carbono disponible. Dicha metodología también nos permitió confirmar que los genes delecionados no afectan al crecimiento en presencia de otras fuentes de carbono.

A continuación, repetimos el experimento con los azúcares de interés, agregando al medio M1 ácido N-acetilmurámico, que no estaba presente en la placa de Biolog, o manosa, con el fin de comprobar los resultados obtenidos anteriormente. Los resultados obtenidos verificaron que estos dos genes (*man*X y *mur*Q) son relevantes para el crecimiento de ERV en presencia de estos dos azúcares específicos, manosa y N-acetilmurámico, respectivamente.

Para confirmar la función de ambos genes, introdujimos copias del gen a cada uno de los mutantes utilizando el plásmido pNZ44, que contiene un promotor de amplio espectro, sugerido y proporcionado por el Dr. Van Sinderen. La restauración de la expresión de dichos genes (*man*X y *mur*Q) restauró el crecimiento *in vitro* de los mutantes en presencia de manosa y N-acetilmurámico, respectivamente.

Seguidamente, verificamos que la función del gen se restablece *in vivo* mediante la complementación con los plásmidos que expresan los respectivos genes. Para ello utilizamos primero el modelo de ratón con clindamicina y seguidamente el de

vancomicina. En ambos casos observamos que la expresión de dichos genes favorece el crecimiento de la cepa complementada, de manera que ya no se observaban diferencias en el crecimiento *in vivo* entre la cepa complementada y la cepa salvaje. Sin embargo, sería necesario repetir dichos experimentos, ya que en el caso del grupo control inoculado con la cepa salvaje y el mutante para *man*X solo observamos un ligero defecto en el crecimiento del mutante respecto a la cepa salvaje, mientras que en anteriores experimentos detectamos diferencias más significativas.

Por último, añadimos manosa en el agua de bebida de los ratones para ver si un exceso de este azúcar aumentaba la concentración de ERV en el intestino, recogiendo muestras de heces un día y dos días tras suministrar el azúcar. Un día después de añadir manosa no vimos un aumento significativo de ERV. Sin embargo, después de dos días de adición de manosa, a pesar de no ser un resultado significativo, sí observamos una tendencia al aumento en los niveles de ERV comparado con aquellos ratones que no recibieron manosa, sugiriendo que la presencia de este azúcar *in vivo* favorece el crecimiento de VRE. No se pudo realizar un experimento similar con N-acetilmurámico por la poca cantidad disponible de azúcar para realizar experimentos *in vivo*.

## 1.6 Prevalencia de los cinco genes candidatos seleccionados en otras cepas de *E. faecium*

Una vez caracterizados con más detalle cinco genes (*man*X, *mur*Q, *ltr*R, *pst*B2, las proteínas del cluster [Fe-S]) que disminuían los niveles de colonización de ERV en el intestino del ratón, quisimos estudiar si la presencia de estos genes está favorecida en cepas de origen clínico (clado A del árbol filogenético de cepas de *E. faecium*), en comparación con cepas comensales humanas (clado B). Para ello, realizamos alineamientos de la secuencia de cada gen con los genes codificados por las 73 cepas descritas en el estudio de Lebreton et. al[57], donde se definen los dos clados en los que se pueden separar las cepas de *E. faecium*.

El gen *ltr*R (regulador transcripcional) estaba presente en todas las cepas clínicas del clado A y ausente en las cepas del clado B (cepas comensales). El gen *pst*B2 (transportador de fosfato) estaba presente en las cepas clínicas (clado A) y en tan solo una de las cepas del clado B. Sin embargo, en esta cepa comensal, la secuencia del gen difiere significativamente del gen codificado por Aus0004 (92.57 % de identidad). Por lo tanto, estos dos genes son específicos de las cepas del clado A (cepas clínicas).

En el caso del gen *man*X, detectamos su presencia en todas las cepas del clado A (cepas clínicas) mientras que está presente en menos del 50 % de las cepas del clado B (7 de las 15). Este resultado sugirió que este gen está enriquecido en las cepas de origen clínico. Un resultado similar se obtuvo para el gen *mur*Q.

Finalmente, el grupo de proteínas del grupo [Fe-S] estaba presente en todas las cepas, tanto en el clado B como en el clado A con un porcentaje similar de identidad de nucleótidos.

**1.7 Estudio de la expresión de los genes que utiliza ERV durante la colonización intestinal**

También se realizaron análisis trancriptómicos del patógeno *in vitro* e *in vivo* para identificar los genes que están más expresados en el intestino y que por tanto pudieran favorecer la colonización intestinal. Tras aplicar el test estadístico DeSeq2 identificamos una serie de genes significativos, 798, ($p<0,05$, FDR$<0,1$, log2 Fold Change (FC) $> |2|$), de los cuales, la mayoría (681) se expresaron más *in vivo* que *in vitro.* Centrándonos en genes que podrían tener un papel relevante durante la colonización de ERV, observamos una mayor expresión *in vivo* de diversas subunidades de diferentes PTS para la captación de lactosa (8 subunidades), fructosa (6), manitol (2), manosa (5), ascorbato (5) o PTS para la captación de más de un carbohidrato (manosa, fructosa y sorbosa, 4). Estos resultados sugieren que los sistemas PTS son relevantes para la colonización intestinal de ERV, los cuales facilitarían la adquisición de los diferentes hidratos de carbono presentes en el intestino.

Por otro lado, no se identificaron los 10 genes caracterizados de la librería de mutantes entre los genes más expresados, excepto el gen *adh*E. Este último resultado sugiere que la relevancia de un gen en el proceso de colonización intestinal no siempre va ligada a su mayor expresión.

**2. Estudio de bacterias comensales con capacidad protectora frente a la colonización por ERV**

Resultados obtenidos por una estudiante de doctorado anterior de nuestro grupo permitieron identificar varias bacterias comensales que se asociaron con la protección contra la colonización intestinal por parte de ERV. Además, se aislaron cepas de varias de estas bacterias comensales, incluyendo *Barnesiella*, Olsenella, *Alistipes*, *Oscillibacter* y Ruminococcaceae. Tal y como describió la anterior estudiante de doctorado, la administración conjunta de estos 5 aislados, nombrados como ProBac (Bacterias Protectoras), permitió reducir los niveles de ERV en el intestino. En esta tesis confirmamos que la administración de otras bacterias no asociadas a la protección frente a ERV (i.e. *Bacteroides*) no disminuyó la colonización de ERV. Este resultado confirmaba que el efecto protector era específico de ProBac y que otras bacterias no tenían dicho efecto protector.

**2.1 Estudio del mecanismo de protección de las bacterias comensales frente a ERV**

Resultados previos del grupo sugirieron que dichas bacterias comensales podrían inhibir el crecimiento de ERV de forma directa, en ausencia del sistema inmunitario. Por tanto, decidimos evaluar las funciones expresadas *in vivo* por este grupo de bacterias con el fin de detectar mecanismos por los que ProBac pudiera inhibir el crecimiento intestinal de

manera directa. Se realizaron análisis transcriptómicos y metabolómicos a distintos grupos de ratones: no tratados (resistentes a la colonización por ERV), tratados con vancomicina (por tanto susceptibles a la colonización por ERV) y ratones que habían sido tratados con vancomicina y posteriormente inoculados con ProBac (parcialmente resistentes a la colonización por ERV). Para ello, se estabularon 2 ratones por jaula, dos semanas después del tratamiento con vancomicina se sacrificó uno de los dos ratones de la jaula para extraer el ciego y analizarlo posteriormente tanto por transcriptómica como metabómica (a continuación). El otro ratón de la jaula se inoculó con ERV para evaluar su capacidad de colonización en cada uno de los grupos, ya que los ratones que se estabulan en la misma jaula recuperan una microbiota muy similar[168], colonizándose también con niveles similares del patógeno. Por tanto, este otro ratón se utilizó posteriormente para asociar los cambios transcriptómicos/metabolómicos detectados en el ciego con la capacidad de colonización de ERV.

El análisis de secuencias permitió estudiar las diferencias en la expresión de genes bacterianos en los distintos grupos de ratones. Cuando fue posible, se asignó una función a los genes identificados mediante la base de datos de la Enciclopedia de Genes y Genomas de Kyoto (KEGG). A continuación, se aplicó el análisis estadístico DeSeq2 para identificar aquellos ortólogos KEGG (KO) cuya expresión aumentó significativamente (> 2 veces; q <0.05) con la administración de Probac y que alcanzan niveles similares a la expresión detectada en ratones no tratados. Después de aplicar DeSeq2, identificamos un total de 6 KO que cumplían dichos criterios. Entre dichas KO encontramos dos de ellas que estaban sobre expresadas en el grupo de ratones que recibió el consorcio de bacterias y que codificaban para subunidades de transportadores de azúcares. K16213 codifica una celobiosa epimerasa, mientras que K02770 codifica para *fru*A, una subunidad del sistema de fosfotransferasa, específica para el transporte de fructosa. K16213 tan solo estaba expresado por los aislados de ProBac en el 50% de los ratones analizados. Sin embargo, el transportador de fructosa estaba expresado por un gen concreto de uno de los aislados administrados (*Olsenella*) en todos los ratones analizados. Como esperábamos, este gen no pudo detectarse en ratones tratados con vancomicina que no recibieron ProBac. Sin embargo, su expresión sí se detectó en todos los ratones no tratados. Además de *fru*A, en la mayoría de los ratones analizados que recibieron ProBac (66-100%), fue posible detectar la otra subunidad del transportador (*fru*B) expresada por *Olsenella,* así como otros genes necesarios para el catabolismo de la fructosa en la vía glicolítica: *fru*K (K00882) y fructosa-bifosfato-aldolasa (FBA, K01624). Estos resultados sugirieron que ProBac y más específicamente *Olsenella* restauraba ciertas funciones, eliminadas tras el tratamiento antibiótico, necesarias para la utilización de la fructosa presente en el intestino grueso. Por tanto, planteaban la hipótesis de que ProBac y más concretamente *Olsenella* estuvieran inhibiendo la colonización intestinal de ERV mediante la competición por nutrientes (i.e. fructosa) que pudiera necesitar ERV para colonizar el intestino. Para valorar esta hipótesis se realizaron distintos ensayos y análisis como se describen a continuación.

En primer lugar, con las mismas muestras utilizadas para los análisis transcriptómicos se realizaron análisis de metabolómica. Como esperábamos, los niveles de fructosa fueron

significativamente mayores en los ratones tratados con vancomicina en comparación con los ratones no tratados, siendo dichos niveles también reducidos significativamente en los ratones que recibieron ProBac.

Mediante el uso de la técnica Biolog detectamos los azúcares más importantes para el crecimiento de la cepa de ERV utilizada en este estudio, ATCC 700221 (cepa diferente a la utilizada en el anterior objetivo, siendo la cepa con la que se inició este estudio por la anterior estudiante de tesis doctoral). Los resultados de este análisis confirmaron que la fructosa y también la manosa son algunos de los azúcares en los que ERV es capaz de crecer de manera más eficiente.

A continuación, evaluamos si los niveles de fructosa podrían ser relevantes para que ERV colonice el tracto intestinal. Con este fin, primero realizamos un análisis de correlación entre los niveles de fructosa detectados en ratones tratados con antibióticos que habían recibido o no ProBac, y la capacidad de ERV de colonizar el intestino de ratones estabulados en las mismas jaulas. Encontramos una correlación altamente significativa y positiva entre los niveles de fructosa y la capacidad de ERV para colonizar el tracto intestinal de los ratones estabulados juntos. Este resultado sugiere que la disponibilidad de fructosa en el intestino es clave para que ERV pueda colonizar el tracto intestinal.

Para comprobar dicha hipótesis, realizamos un nuevo experimento de ratones con el modelo anterior pero ahora agregando fructosa a uno de los dos ratones estabulados juntos, para así comprobar si un exceso de fructosa podría aumentar de forma significativa ERV en el intestino. Como pudimos comprobar, los ratones que recibieron fructosa tenían unos niveles significativamente más elevados de patógeno en el intestino. También, es importante destacar que dicho aumento de ERV en presencia de fructosa no ocurre en los ratones que habían sido colonizados con ProBac antes de la colonización intestinal con ERV.

Todos estos resultados en conjunto sugerían que la fructosa favorece la colonización intestinal de ERV y que la reducción de los niveles de fructosa por ProBac disminuía la colonización intestinal de ERV.

## 2.2 Experimentos *ex vivo* demuestran protección por competición de nutrientes conferida por el consorcio ProBac y *Olsenella* frente a ERV

A partir de estos resultados y para estudiar si ProBac protege frente a ERV mediante competición por nutrientes (i.e. fructosa) se realizaron experimentos *ex vivo.* Para ello, utilizamos el contenido intestinal de ratones que se habían recuperado del tratamiento con vancomicina durante dos semanas. Los contenidos intestinales se filtraron para eliminar cualquier bacteria y se inoculó el mix de bacterias ProBac para permitir la depleción de nutrientes. A las 24 h se añadió ERV. Los resultados se analizaron tras 6 h de incubación con ERV y mostraron como el patógeno no es capaz de crecer en presencia de ProBac. Sin embargo, ERV si fue capaz de crecer en presencia de una bacteria no asociada a la protección (i.e. *Bacteroides*).

Posteriormente repetimos los experimentos pero agregando un exceso de carbohidratos (bien una mezcla o solo fructosa) a la vez que se añadió ERV. El exceso de azúcares permitió restablecer el crecimiento parcial de ERV, incluso añadiendo un exceso de únicamente fructosa. Dicho resultado sugiere que la inhibición del crecimiento de ERV por ProBac se debe a la competencia por los nutrientes, pudiendo restaurar el crecimiento con un exceso de azúcares. Este tipo de experimentos se repitió pero cultivando de forma individual cada uno de los aislados que conformaban ProBac. Tan solo uno de ellos, *Olsenella*, fue capaz de inhibir el crecimiento de ERV. Además, repetimos el experimento con solo *Olsenella,* añadiendo un exceso de fructosa y observamos una recuperación parcial del crecimiento de ERV. En conjunto, estos resultados sugieren que ProBac puede suprimir el crecimiento de ERV a través de la competición por carbohidratos, principalmente por la utilización de la fructosa. Además, *Olsenella* parece ser clave en la protección conferida por ProBac ya que es la única bacteria que inhibe el crecimiento de ERV por sí misma *ex vivo*.

Por otro lado, mediante Biolog, detectamos que uno de los principales azúcares para el crecimiento de *Olsenella* es la fructosa. Este resultado, junto con los análisis transcriptómicos anteriores, sugería que *Olsenella* es capaz de utilizar la fructosa *in vivo,* lo que inhibe el crecimiento de ERV.

Por tanto, repetimos los experimentos realizados para estudiar la protección *in vivo* que confiere Probac pero dando tan solo *Olsenella* para verificar si *in vivo* observamos el mismo efecto en la reducción de ERV. Los resultados mostraron que *Olsenella* es suficiente para disminuir de forma significativa la colonización intestinal de ERV.

Por tanto, el consorcio bacteriano previamente identificado compuesto por cinco bacterias protectoras (ProBac), disminuyó los niveles de ERV en el intestino de los ratones tratados y, más específicamente, uno de ellos, *Olsenella* es clave para restringir la colonización. Mediante el análisis transcriptómico y metabolómico *in vivo*, además de la metodología Biolog y los experimentos *ex vivo*, hemos identificado un posible mecanismo de protección a través de la competencia por azúcares, principalmente fructosa, como el mecanismo mediante el cual ProBac y más concretamente *Olsenella* sería capaz de proteger frente a ERV.

**CONCLUSIONES**

- La metodología de la librería de mutantes por transposición es útil para estudiar los genes de *Enterococcus* en distintos ambientes. Sin embargo, la elección de la cepa es esencial, puesto que la eficiencia de la generación de la librería depende de la cepa de *Enterococcus faecium* utilizada.

- El análisis Tn-Seq identificó 71 genes candidatos cuya interrupción por el transposón disminuye la colonización intestinal del ERV, tanto en el colon como en el ciego.

- La mutagénesis dirigida de los genes con un efecto potencialmente mayor en la colonización intestinal por ERV: *man*X, *mur*Q, *ltr*R, *pst*B2, las proteínas del grupo [Fe-S], *yyc*I, *dsb*A, *adh*E, *pgt* confirmaron su papel en el proceso de colonización intestinal por ERV ya que su eliminación disminuyó la capacidad de la ERV para colonizar el tracto intestinal de los ratones.

- Los genes candidatos seleccionados afectan a la colonización intestinal de ERV con distintos antibióticos (clindamicina y vancomicina).

- Los ensayos *in vitro* demostraron que el grupo Fe-S está involucrado en el transporte de iones de hierro, principalmente $Fe^{3+.}$

- Los ensayos *in vitro* demostraron que los genes *man*X y *mur*Q se requieren específicamente para la utilización de la manosa y el ácido N-acetil-murámico respectivamente.

- El análisis transcriptómico sugiere que el gen *ltr*R es un regulador transcripcional de un operón implicado en la adquisición de azúcar.

- Los genes *ltr*R y *pst*B2 están codificados por cepas de *E. faecium* del clado A, que contienen principalmente cepas asociadas al ambiente hospitalario y cepas de origen animal, mientras que están ausentes en las cepas del clado B (cepas humanas comensales). *man*X y *mur*Q están enriquecidos en las cepas del clado A, mientras que el clúster Fe-S está presente en todo tipo de cepas.

- La administración de las cinco bacterias comensales (ProBac), que consisten en aislados de los taxones *Olsenella*, *Alistipes*, *Barnesiella*, *Oscillibacter* y Ruminococceae reduce los niveles de ERV en el intestino murino, en comparación con la administración de bacterias no asociadas a la protección contra la ERV (es decir, un aislado del género *Bacteroides*).

- La administración de ProBac a ratones tratados con antibióticos restablece la expresión de funciones bacterianas específicas, incluidos los genes relacionados con el metabolismo de la fructosa. La expresión de estos genes es impulsada principalmente por un aislado del género *Olsenella*.

- El tratamiento con vancomicina aumenta los niveles de fructosa disponibles en el intestino grueso de ratón, mientras que la administración de ProBac reduce significativamente los niveles de fructosa.

- Se encontró una correlación positiva entre los niveles de fructosa en el intestino y la capacidad del ERV para colonizar el tracto intestinal de los ratones.

- Los ensayos *in vitro* mostraron que la fructosa es uno de los azúcares que permiten un mayor crecimiento de VRE y de *Olsenella*.

- Un suplemento de fructosa aumenta los niveles de ERV en el intestino grueso en los ratones. Sin embargo, la expansión del ERV dependiente de la fructosa se suprime si los aislados de ProBac están presentes en el intestino grueso.
- Los experimentos *ex vivo* demostraron que el consorcio bacteriano ProBac inhibe el crecimiento de ERV mediante la disminución de la presencia de nutrientes.
- *Olsenella* es suficiente para inhibir el crecimiento del ERV *ex vivo* e *in vivo*.

# VIII. BIBLIOGRAPHY

1.      Arias, C. A. & Murray, B. E. The rise of the Enterococcus: Beyond vancomycin resistance. *Nat. Rev. Microbiol.* **10**, 266–278 (2012).

2.      García-Solache, M. & Rice, L. B. The enterococcus: A model of adaptability to its environment. *Clin. Microbiol. Rev.* **32**, 1–28 (2019).

3.      Lebreton, F., Willems, R. J. L. & Gilmore, M. S. Enterococcus Diversity, Origins in Nature, and Gut Colonization. *Enterococci From Commensals to Lead. Causes Drug Resist. Infect.* 1–59 (2014).

4.      Sherman, J. M. the Streptococci. *Bacteriol. Rev.* **1**, 3–97 (1937).

5.      Klein, G. Taxonomy, ecology and antibiotic resistance of enterococci from food and the gastro-intestinal tract. *Int. J. Food Microbiol.* **88**, 123–131 (2003).

6.      Zhong, Z. *et al.* Comparative genomic analysis of the genus Enterococcus. *Microbiol. Res.* **196**, 95–105 (2017).

7.      Lebreton, F. *et al.* Tracing the Enterococci from Paleozoic Origins to the Hospital. *Cell* **169**, 849-861.e13 (2017).

8.      Schleifer, K. H. & Kilpper-Balz, R. Transfer of Streptococcus faecalis and Streptococcus faecium to the genus Enterococcus nom. rev. as Enterococcus faecalis comb. nov. and Enterococcus faecium comb. nov. *Int. J. Syst. Bacteriol.* **34**, 31–34 (1984).

9.      Byappanahalli, M. N., Nevers, M. B., Korajkic, A., Staley, Z. R. & Harwood, V. J. Enterococci in the Environment. *Microbiol. Mol. Biol. Rev.* **76**, 685–706 (2012).

10.     Collins, M., Jones, D., Farrow, J. A., Kilpper-Balz, R. & Shleifer, K. H. Enterococcus avium nom. rev., comb. nov.; E. casselipavus norn. rev., comb. nov.; E. durans norn. rev., comb. nov.; E. gallinarum comb. nov.; and E. malodoratus. *Int. J. Syst. Bacteriol.* 220–223 (1984).

11.     Ludwing, W., K-H, S. & WB, W. *Systematic Bacteriology-Second Edition,Volume Three, The Firmicutes*. (2009).

12.     MUNDT, J. O. Occurrence of enterococci in animals in a wild environment. *Appl. Microbiol.* **11**, 136–140 (1963).

13.     Gilmore, M. S., Lebreton, F. & van Schaik, W. Genomic transition of enterococci from gut commensals to leading causes of multidrug-resistant hospital infection in the antibiotic era. *Curr. Opin. Microbiol.* **16**, 10–16 (2013).

14.     MUNDT, J. O. Occurrence of Enterococci on Plants in a Wild Environment'. *Appl. Microbiol.* **11**, 141–144 (1963).

15.     Martin, J. D. & Mundt, J. O. Enterococci in insects. *Appl. Microbiol.* **24**, 575–580 (1972).

16.     Sghir, A. *et al.* Quantification of bacterial groups within human fecal flora by oligonucleotide probe hybridization. *Appl. Environ. Microbiol.* **66**, 2263–2266 (2000).

17.     Blaimont, B., Charlier, J. & Wauters, G. Comparative distribution of enterococcus species in faeces and clinical samples. *Microb. Ecol. Health Dis.* **8**, 87–92 (1995).

18.     Hayashi, H., Takahashi, R., Nishi, T., Sakamoto, M. & Benno, Y. Molecular analysis of jejunal, ileal, caecal and rectosigmoidal human colonic microbiota using 16S rRNA gene libraries and terminal restriction fragment length polymorphism. *J. Med. Microbiol.* **54**, 1093–1101 (2005).

19.     Bik, E. M. *et al.* Molecular analysis of the bacterial microbiota in the human stomach. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 732–737 (2006).

20.     Delgado, S., Cabrera-Rubio, R., Mira, A., Suárez, A. & Mayo, B. Microbiological Survey of the Human Gastric Ecosystem Using Culturing and Pyrosequencing Methods. *Microb. Ecol.* **65**, 763–772 (2013).

21.     Komiyama, E. Y. *et al.* Enterococcus species in the oral cavity: Prevalence, virulence factors and antimicrobial susceptibility. *PLoS One* **11**, 1–11 (2016).

22.     Smyth, C. J., Matthews, H., Halpenny, M. K., Brandis, H. & Colman, G. Biotyping, serotyping and phage typing of Streptococcus faecalis isolated from dental plaque in the human mouth. *J. Med. Microbiol.* **23**, 45–54 (1987).

23.     Wang, Q. Q., Zhang, C. F., Chu, C. H. & Zhu, X. F. Prevalence of Enterococcus faecalis in saliva and filled root canals of teeth associated with apical periodontitis. *Int. J. Oral Sci.* **4**, 19–23 (2012).

24.     Ubeda, C. & Pamer, E. G. Antibiotics, microbiota, and immune defense. *Trends Immunol.* **33**, 459–466 (2012).

25.     Ubeda, C. *et al.* Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *J. Clin. Invest.* **120**, 4332–4341 (2010).

26.     Jabbari, S. M., Shiadeh, Pormohammad, A., Hashemi, A. & Lak, P. Global prevalence of antibiotic resistance in blood-isolated enterococcus faecalis and enterococcus faecium: A systematic review and meta-analysis. *Infect. Drug Resist.* **12**, 2713–2725 (2019).

27.     Taur, Y. *et al.* Intestinal domination and the risk of bacteremia in patients undergoing allogeneic hematopoietic stem cell transplantation. *Clin. Infect. Dis.* **55**, 905–914 (2012).

28.     Ubeda, C. *et al.* Vancomycin-resistant Enterococcus domination of intestinal microbiota is enabled by antibiotic treatment in mice and precedes bloodstream invasion in humans. *J. Clin. Invest.* **120**, 4332–4341 (2010).

29.     Weiner, L. M. *et al.* Antimicrobial-Resistant Pathogens Associated with Healthcare-Associated Infections: Summary of Data Reported to the National Healthcare Safety Network at the Centers for Disease Control and Prevention, 2011-2014. *Infect. Control Hosp. Epidemiol.* **37**, 1288–1301 (2016).

30.     Suetens, C., Hopkins, S., Kolman, J. & Högberg, L. D. *Point prevalence survey of healthcare-associated infections and antimicrobial use in European acute care hospitals 2011–2012*. *European Centre for Disease Prevention and Control* (2013). doi:10.2807/ese.17.46.20316-en.

31.     Zhang, Y., Du, M., Chang, Y., Chen, L. an & Zhang, Q. Incidence, clinical characteristics, and outcomes of nosocomial Enterococcus spp. bloodstream infections in a tertiary-care hospital in Beijing, China: A four-year retrospective study. *Antimicrob. Resist. Infect. Control* **6**, 1–11 (2017).

32.     Puchter, L. *et al.* Economic burden of nosocomial infections caused by vancomycin-resistant enterococci. *Antimicrob. Resist. Infect. Control* **7**, 1–7 (2018).

33.     O'Driscoll, T. & Crank, C. W. Vancomycin-resistant enterococcal infections: Epidemiology, clinical manifestations, and optimal management. *Infect. Drug Resist.* **8**, 217–230 (2015).

34. Donskey, C. J. *et al.* Effect of antibiotic therapy on the density of vancomycin-resistant enterococci in the stool of colonized patients. *N. Engl. J. Med.* **343**, 1925–1932 (2000).

35. Kramer, A., Schwebke, I. & Kampf, G. How long do nosocomial pathogens persist on inanimate surfaces? A systematic review. *BMC Infect. Dis.* **6**, 1–8 (2006).

36. Noskin, G. A., Stosor, V., Cooper, I. & Peterson, L. R. Recovery of Vancomycin-Resistant Enterococci on Fingertips and Environmental Surfaces Published by : Cambridge University Press on behalf of The Society for Healthcare Linked references are available on JSTOR for this article : RECOVERY OF VANCOMYCIN-RESI. *Infect. Control Hosp. Epidemiol.* **16**, 577–581 (1995).

37. Davis, E. *et al.* Epidemiology of vancomycin-resistant Enterococcus faecium and Enterococcus faecalis colonization in nursing facilities. *Open Forum Infect. Dis.* **7**, (2020).

38. Barbara E. Murray, M. D. Vancomycin-Resistant Enterococcal Infections. *New Engl. J. Med. Drug* **342**, 710–721 (2000).

39. Ayobami, O., Willrich, N., Reuss, A., Eckmanns, T. & Markwart, R. The ongoing challenge of vancomycin-resistant Enterococcus faecium and Enterococcus faecalis in Europe: an epidemiological analysis of bloodstream infections. *Emerg. Microbes Infect.* **9**, 1180–1193 (2020).

40. Koch, S., Hufnagel, M., Theilacker, C. & Huebner, J. Enterococcal infections: Host response, therapeutic, and prophylactic possibilities. *Vaccine* **22**, 822–830 (2004).

41. Agudelo Higuita, N. I. & Huycke, M. M. Enterococcal Disease, Epidemiology, and Implications for Treatment. *Enterococci From Commensals to Lead. Causes Drug Resist. Infect.* 1–35 (2014).

42. Dahl, A. & Bruun, N. E. Enterococcus faecalis infective endocarditis: focus on clinical aspects. *Expert Rev. Cardiovasc. Ther.* **11**, 1247–1257 (2013).

43. Moses, V. *et al.* Enterococcal bacteremia is associated with prolonged stay in the medical intensive care unit. *J. Glob. Infect. Dis.* **4**, 26–30 (2012).

44. Moellering, R. C. Emergence of Enterococcus as a Significant Pathogen. *Clin. Infect. Dis.* **14**, 1173–1176 (1992).

45. Gorrie, C., Higgs, C., Carter, G., Stinear, T. P. & Howden, B. Genomics of vancomycin-resistant enterococcus faecium. *Microb. Genomics* **5**, (2019).

46. Tacconelli, E. *et al.* Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *Lancet Infect. Dis.* **18**, 318–327 (2018).

47. Miller, W. R., Munita, J. M. & Arias, C. A. Mechanisms of antibiotic resistance in enterococci. *Expert Rev. Anti. Infect. Ther.* **12**, 1221–1236 (2014).

48. Hollenbeck, B. L. & Rice, L. B. Intrinsic and acquired resistance mechanisms in enterococcus. *Virulence* **3**, 421–569 (2012).

49. Ahmed, M. O. & Baptiste, K. E. Vancomycin-Resistant Enterococci: A Review of Antimicrobial Resistance Mechanisms and Perspectives of Human and Animal Health. *Microb. Drug Resist.* **24**, 590–606 (2018).

50. Goh, H. M. S., Yong, M. H. A., Chong, K. K. L. & Kline, K. A. Model systems for the study of Enterococcal colonization and infection. *Virulence* **8**, 1525–1562 (2017).

51.     Palmer, K. L., van Schaik, W., Willems, R. J. L. & Gilmore, M. S. Enterococcal Genomics. *Enterococci From Commensals to Lead. Causes Drug Resist. Infect.* 1–50 (2014).

52.     Kim, E. B. & Marco, M. L. Nonclinical and clinical enterococcus faecium strains, but not enterococcus faecalis strains, have distinct structural and functional genomic features. *Appl. Environ. Microbiol.* **80**, 154–165 (2014).

53.     Hung, W. W. *et al.* Using groEL as the target for identification of Enterococcus faecium clades and 7 clinically relevant Enterococcus species. *J. Microbiol. Immunol. Infect.* **52**, 255–264 (2019).

54.     Zhong, Z. *et al.* Comparative genomic analysis revealed great plasticity and environmental adaptation of the genomes of Enterococcus faecium. *BMC Neurosci.* **20**, 1–13 (2019).

55.     Palmer, K. L. *et al.* Comparative Genomics of Enterococci : Variation in Enterococcus faecalis , Clade Structure in E . faecium , and Defining. *MBio* **3**, e00318-11 (2012).

56.     Galloway-Peña, J., Roh, J. H., Latorre, M., Qin, X. & Murray, B. E. Genomic and SNP analyses demonstrate a distant separation of the hospital and community-associated clades of enterococcus faecium. *PLoS One* **7**, (2012).

57.     Lebreton, F., Schaik, W. Van & Manson, A. Emergence of Epidemic Multidrug-Resistant Enterococcus faecium. *Am. Soc. Microbiol.* **4**, 1–10 (2013).

58.     Gouliouris, T. *et al.* Genomic surveillance of enterococcus faecium reveals limited sharing of strains and resistance genes between livestock and humans in the United Kingdom. *MBio* **9**, 1–15 (2018).

59.     Palmer, K. L. & Gilmore, M. S. Multidrug-Resistant Enterococci Lack CRISPR-cas. *Am. Soc. Microbiol.* 1–10 (2010) doi:10.1128/mBio.00227-10.Editor.

60.     van Schaik, W. *et al.* Pyrosequencing-based comparative genome analysis of the nosocomial pathogen Enterococcus faecium and identification of a large transferable pathogenicity island. *BMC Genomics* **11**, (2010).

61.     Lam, M. M. C. *et al.* Comparative analysis of the first complete Enterococcus faecium genome. *J. Bacteriol.* **194**, 2334–2341 (2012).

62.     Ramsey, M., Hartke, A. & Huycke, M. The Physiology and Metabolism of Enterococci. *Enterococci From Commensals to Lead. Causes Drug Resist. Infect.* 1–55 (2014).

63.     Niven, C. F. & Sherman, J. M. Nutrition of the Enterococci. *J. Bacteriol.* **47**, 335–42 (1943).

64.     Leblan, D. J. *The Prokaryotes: Vol. 4: Bacteria: Firmicutes, Cyanobacteria*. (2006).

65.     Murray, B. E. *et al.* Generation of restriction map of Enterococcus faecalis OG1 and investigation of growth requirements and regions encoding biosynthetic function. *J. Bacteriol.* **175**, 5216–5223 (1993).

66.     Doi, Y. Glycerol metabolism and its regulation in lactic acid bacteria. *Appl. Microbiol. Biotechnol.* **103**, 5079–5093 (2019).

67.     Vebø, H. C., Snipen, L., Nes, I. F. & Brede, D. A. The transcriptome of the nosocomial pathogen Enterococcus faecalis V583 reveals adaptive responses to growth in blood. *PLoS One* **4**, 15–21 (2009).

68. Riboldi, G. P., Verli, H. & Frazzon, J. Structural studies of the Enterococcus faecalis SufU [Fe-S] cluster protein. *BMC Biochem.* **10**, 1–10 (2009).

69. Chekabab, S. M., Harel, J. & Dozois, C. M. Interplay between genetic regulation of phosphate homeostasis and bacterial virulence. *Virulence* **5**, 786–793 (2014).

70. Lamarche, M. G., Wanner, B. L., Crépin, S. & Harel, J. The phosphate regulon and bacterial virulence: A regulatory network connecting phosphate homeostasis and pathogenesis. *FEMS Microbiol. Rev.* **32**, 461–473 (2008).

71. Saier, M. H. The Bacterial Phosphotransferase System: New Frontiers 50 Years after Its Discovery. *J. Mol. Microbiol. Biotechnol.* **25**, 73–78 (2015).

72. Zhang, X. *et al.* Identification of a genetic determinant in clinical enterococcus faecium strains that contributes to intestinal colonization during antibiotic treatment. *J. Infect. Dis.* **207**, 1780–1786 (2013).

73. Garsin, D. A. *et al.* Pathogenesis and Models of Enterococcal Infection. *Enterococci From Commensals to Lead. Causes Drug Resist. Infect.* 1–73 (2014).

74. Nallapareddy, S. R., Singh, K. V., Sillanpää, J., Zhao, M. & Murray, B. E. Relative contributions of Ebp Pili and the collagen adhesin ace to host extracellular matrix protein adherence and experimental urinary tract infection by Enterococcus faecalis OG1RF. *Infect. Immun.* **79**, 2901–2910 (2011).

75. Nallapareddy, S. R., Qin, X., Weinstock, G. M., Hook, M. & Murray, B. E. Enterococcus faecalis adhesin, Ace, mediates attachment to extracellular matrix proteins collagen type IV and laminin as well as collagen type I. *Infect. Immun.* **68**, 5218–5224 (2000).

76. Nallapareddy, S. R., Weinstock, G. M. & Murray, B. E. Clinical isolates of Enterococcus faecium exhibit strain-specific collagen binding mediated by Acm, a new member of the MSCRAMM family. *Mol. Microbiol.* **47**, 1733–1747 (2003).

77. Nallapareddy, S. R., Singh, K. V., Okhuysen, P. C. & Murray, B. E. A functional collagen adhesin gene, acm, in clinical isolates of Enterococcus faecium correlates with the recent success of this emerging nosocomial pathogen. *Infect. Immun.* **76**, 4110–4119 (2008).

78. Paganelli, F. L. *et al.* Enterococcus faecium Biofilm Formation: Identification of Major Autolysin AtlAEfm , Associated Acm Surface Localization, and AtlAEfm - Independent Extracellular DNA Release. *Am. Soc. Microbiol.* **4**, 1–10 (2013).

79. Sillanpää, J. *et al.* Identification and phenotypic characterization of a second collagen adhesin, Scm, and genome-based identification and analysis of 13 other predicted MSCRAMMs, including four distinct pilus loci, in Enterococcus faecium. *Microbiology* **154**, 3199–3211 (2008).

80. Sillpanpää, J. *et al.* Characterization of the ebpfm pilus-encoding operon of Enterococcus faecium and its role in biofilm formation and virulence in a murine model of urinary tract infection. *Virulence* **1**, 236–246 (2010).

81. Willems, R. J. L. *et al.* Variant esp gene as a marker of a distinct genetic lineage of vancomycin- resistant Enterococcus faecium spreading in hospitals For personal use only . Reproduce with permission from The Lancet Publishing Group . *Lancet* **357**, 853–855 (2001).

82.    Leavis, H. *et al.* A Novel Putative Enterococcal Pathogenicity Island Linked to the esp Virulence Gene of Enterococcus faecium and Associated with Epidemicity. *J. Bacteriol.* **186**, 672–682 (2004).

83.    Heikens, E., Bonten, M. J. M. & Willems, R. J. L. Enterococcal surface protein esp is important for biofilm formation of Enterococcus faecium E1162. *J. Bacteriol.* **189**, 8233–8240 (2007).

84.    Van Wamel, W. J. B. *et al.* Growth condition-dependent Esp expression by Enterococcus faecium affects initial adherence and biofilm formation. *Infect. Immun.* **75**, 924–931 (2007).

85.    Leendertse, M. *et al.* Enterococcal Surface Protein Transiently Aggravates Enterococcus faecium –Induced Urinary Tract Infection in Mice . *J. Infect. Dis.* **200**, 1162–1165 (2009).

86.    Heikens, E. *et al.* Contribution of the enterococcal surface protein Esp to pathogenesis of Enterococcus faecium endocarditis. *Microbes Infect.* **13**, 1185–1190 (2011).

87.    Hendrickx, A. P. A. *et al.* SgrA, a nidogen-binding LPXTG surface adhesin implicated in biofilm formation, and EcbA, a collagen binding MSCRAMM, are two novel adhesins of hospital-acquired Enterococcus faecium. *Infect. Immun.* **77**, 5097–5106 (2009).

88.    Nallapareddy, S. R., Singh, K. V. & Murray, B. E. Contribution of the collagen adhesin Acm to pathogenesis of Enterococcus faecium in experimental endocarditis. *Infect. Immun.* **76**, 4120–4128 (2008).

89.    Rice, L. B. *et al.* A Potential Virulence Gene, hyl Efm , Predominates in Enterococcus faecium of Clinical Origin . *J. Infect. Dis.* **187**, 508–512 (2003).

90.    Rice, L. B. *et al.* Transferable Capacity for Gastrointestinal Colonization in Enterococcus faecium in a Mouse Model . *J. Infect. Dis.* **199**, 342–349 (2009).

91.    Zhang, X. *et al.* Functional genomic analysis of bile salt resistance in Enterococcus faecium. *BMC Genomics* **14**, (2013).

92.    Wijaya, A. *et al.* Cloning of the bile salt hydrolase (bsh) gene from Enterococcus faecium FAIR-E 345 and chromosomal location of bsh genes in food Enterococci. *J. Food Prot.* **67**, 2772–2778 (2004).

93.    Choudhury, T., Singh, K. V., Sillanpää, J., Nallapareddy, S. R. & Murray, B. E. Importance of two Enterococcus faecium loci encoding Gls-like proteins for in vitro bile salts stress response and virulence. *J. Infect. Dis.* **203**, 1147–1154 (2011).

94.    Eaton, T. J. & Gasson, M. J. Molecular Screening of Enterococcus Virulence Determinants and Potential for Genetic Exchange between Food and Medical Isolates. *Appl. Environ. Microbiol.* **67**, 1628–1635 (2001).

95.    Soares, R. O., Fedi, A. C., Reiter, K. C., Caierão, J. & D'Azevedo, P. A. Correlation between biofilm formation and gelE, esp, and agg genes in enterococcus spp. clinical isolates. *Virulence* **5**, 634–637 (2014).

96.    Heikens, E., Van Schaik, W., Leavis, H. L., Bonten, M. J. M. & Willems, R. J. L. Identification of a novel genomic island specific to hospital-acquired clonal complex 17 Enterococcus faecium isolates. *Appl. Environ. Microbiol.* **74**, 7094–7097 (2008).

97.    Cacaci, M. *et al.* Expression profiling in a mammalian host reveals the strong induction of genes encoding LysM domain-containing proteins in Enterococcus faecium. *Sci. Rep.* **8**, 1–11 (2018).

98. Paganelli, F. L. *et al.* Genome-wide Screening Identifies Phosphotransferase System Permease BepA to Be Involved in Enterococcus faecium Endocarditis and Biofilm Formation. *J. Infect. Dis.* **214**, 189–195 (2016).

99. Zhang, X. *et al.* RNA-seq and Tn-seq reveal fitness determinants of vancomycin-resistant Enterococcus faecium during growth in human serum. *BMC Genomics* **18**, 1–12 (2017).

100. Fan, T.-J., Goeser, L., Naziripour, A., Redinbo, M. R. & Hansen, J. J. Enterococcus faecalis Gluconate Phosphotransferase System Accelerates Experimental Colitis and Bacterial Killing by Macrophages. *Infect. Immun.* **87**, 1–12 (2019).

101. Sauvageot, N. *et al.* Enterococcus faecalis Uses a Phosphotransferase System Permease and a Host Colonization-Related ABC Transporter for Maltodextrin Uptake. *J. Bacteriol.* **199**, 1–17 (2017).

102. Stein-Thoeringer, C. K. *et al.* Lactose drives Enterococcus expansion to promote graft-versus-host disease. *Science (80-. ).* **366**, 1143–1149 (2019).

103. Ruiz-Cruz, S., Espinosa, M., Goldmann, O. & Bravo, A. Global regulation of gene expression by the MafR protein of Enterococcus faecalis. *Front. Microbiol.* **6**, 1–13 (2016).

104. Zhang, X. *et al.* A LacI-Family Regulator Activates Maltodextrin Metabolism of Enterococcus faecium. *PLoS One* **8**, 1–9 (2013).

105. Top, J. *et al.* The Enterococcus faecium Enterococcal Biofilm Regulator, EbrB, Regulates the esp Operon and Is Implicated in Biofilm Formation and Intestinal Colonization. *PLoS One* **8**, (2013).

106. Nicholson, J. K., Holmes, E. & Wilson, I. D. Gut microorganisms, mammalian metabolism and personalized health care. *Nat. Rev. Microbiol.* **3**, 431–438 (2005).

107. Jandhyala, S. M. *et al.* Role of the normal gut microbiota. *World J. Gastroenterol.* **21**, 8836–8847 (2015).

108. Huttenhower, C. *et al.* Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207–214 (2012).

109. Rajilić-Stojanović, M. Function of the microbiota. *Best Pract. Res. Clin. Gastroenterol.* **27**, 5–16 (2013).

110. Sekirov, I., Russell, S. L., Caetano M Antunes, L. & Finlay, B. B. Gut microbiota in health and disease. *Physiol. Rev.* **90**, 859–904 (2010).

111. O'Hara, A. M. & Shanahan, F. The gut flora as a forgotten organ. *EMBO Rep.* **7**, 688–693 (2006).

112. Rinninella, E. *et al.* What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. *Microorganisms* **7**, (2019).

113. Bäckhed, F., Ley, R. E., Sonnenburg, J. L., Peterson, D. A. & Gordon, J. I. Host-bacterial mutualism in the human intestine. *Science (80-. ).* **307**, 1915–1920 (2005).

114. Van Der Waaij, D., Berghuis-de Vries, J. M. & Lekkerkerk-Van Der Wees, J. E. C. Colonization resistance of the digestive tract in conventional and antibiotic-treated mice. *J. Hyg. (Lond).* **69**, 405–411 (1971).

115. Becattini, S., Taur, Y. & Pamer, E. G. Antibiotic-Induced Changes in the Intestinal Microbiota and Disease. *Trends Mol. Med.* **22**, 458–478 (2016).

116. Kim, S., Covington, A. & Pamer, E. G. The intestinal microbiota: Antibiotics, colonization resistance, and enteric pathogens. *Immunol. Rev.* **279**, 90–105 (2017).

117. Vollaard, E. J. & Clasener, H. A. L. Colonization resistance. *Antimicrob. Agents Chemother.* **38**, 409–414 (1994).

118. Dubin, K. & Pamer, E. G. Enterococci and Their Interactions with the Intestinal Microbiome. *Microbiol. Spectr.* **5**, 1–16 (2017).

119. Vaishnava, S., Behrendt, C. L., Ismail, A. S., Eckmann, L. & Hooper, L. V. Paneth cells directly sense gut commensals and maintain homeostasis at the intestinal host-microbial interface. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 20858–20863 (2008).

120. Mukherjee, S. *et al.* Antibacterial membrane attack by a pore-forming intestinal C-type lectin. *Nature* **505**, 103–107 (2014).

121. Brandl, K. *et al.* Vancomycin-resistant enterococci exploit antibiotic-induced innate immune deficits. *Nature* **455**, 804–807 (2008).

122. Johansson, E. V. M. *et al.* The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15064–15069 (2008).

123. Wlodarska, M. *et al.* Antibiotic treatment alters the colonic mucus layer and predisposes the host to exacerbated Citrobacter rodentium-induced colitis. *Infect. Immun.* **79**, 1536–1545 (2011).

124. Loonen, L. M. P. *et al.* REG3γ-deficient mice have altered mucus distribution and increased mucosal inflammatory responses to the microbiota and enteric pathogens in the ileum. *Mucosal Immunol.* **7**, 939–947 (2014).

125. Zheng, D., Liwinski, T. & Elinav, E. Interaction between microbiota and immunity in health and disease. *Cell Res.* **30**, 492–506 (2020).

126. Ubeda, C., Djukovic, A. & Isaac, S. Roles of the intestinal microbiota in pathogen protection. *Clin. Transl. Immunol.* **6**, e128-10 (2017).

127. Borrero, J., Chen, Y., Dunny, G. M. & Kaznessis, Y. N. Modified lactic acid bacteria detect and inhibit multiresistant enterococci. *ACS Synth. Biol.* **4**, 299–306 (2015).

128. Den Besten, G. *et al.* The role of short-chain fatty acids in the interplay between diet, gut microbiota, and host energy metabolism. *J. Lipid Res.* **54**, 2325–2340 (2013).

129. Shin, R., Suzuki, M. & Morishita, Y. Influence of intestinal anaerobes and organic acids on the growth of enterohaemorrhagic Escherichia coli O157:H7. *J. Med. Microbiol.* **51**, 201–206 (2002).

130. Hsiao, A. *et al.* Members of the human gut microbiota involved in recovery from Vibrio cholerae infection. *Nature* **515**, 423–426 (2014).

131. Maltby, R., Leatham-Jensen, M. P., Gibson, T., Cohen, P. S. & Conway, T. Nutritional Basis for Colonization Resistance by Human Commensal Escherichia coli Strains HS and Nissle 1917 against E. coli O157:H7 in the Mouse Intestine. *PLoS One* **8**, 1–10 (2013).

132. Pultz, N. J., Stiefel, U., Subramanyan, S., Helfand, M. S. & Donskey, C. J. Mechanisms by which anaerobic microbiota inhibit the establishment in mice of intestinal colonization by vancomycin-resistant Enterococcus. *J. Infect. Dis.* **191**, 949–956 (2005).

133. Aas, J., Gessert, C. E. & Bakken, J. S. Recurrent Clostridium difficile Colitis: Case Series Involving 18 Patients Treated with Donor Stool Administered via a Nasogastric Tube . *Clin. Infect. Dis.* **36**, 580–585 (2003).

134. Van Nood, E. *et al.* Duodenal infusion of donor feces for recurrent clostridium difficile. *N. Engl. J. Med.* **368**, 407–415 (2013).

135. Garborg, K., Waagsbø, B., Stallemo, A., Matre, J. & Sundøy, A. Results of faecal donor instillation therapy for recurrent Clostridium difficile-associated diarrhoea. *Scand. J. Infect. Dis.* **42**, 857–861 (2010).

136. Pamer, E. G. Fecal microbiota transplantation: Effectiveness, complexities, and lingering concerns. *Mucosal Immunol.* **7**, 210–214 (2014).

137. Dubberke, E. R. *et al.* Clearance of vancomycin-resistant Enterococcus concomitant with administration of a microbiota-based drug targeted at recurrent Clostridium difficile infection. *Open Forum Infect. Dis.* **3**, 1–6 (2016).

138. Ubeda, C. *et al.* Intestinal microbiota containing Barnesiella species cures vancomycin-resistant Enterococcus faecium colonization. *Infect. Immun.* **81**, 965–973 (2013).

139. Bilinski, J. *et al.* Fecal Microbiota Transplantation in Patients With Blood Disorders Inhibits Gut Colonization With Antibiotic-Resistant Bacteria: Results of a Prospective, Single-Center Study. *Clin. Infect. Dis.* **65**, 364–370 (2017).

140. Davido, B. *et al.* Is faecal microbiota transplantation an option to eradicate highly drug-resistant enteric bacteria carriage? *J. Hosp. Infect.* **95**, 433–437 (2017).

141. DeFilipp, Z. *et al.* Drug-Resistant E. coli Bacteremia Transmitted by Fecal Microbiota Transplant . *N. Engl. J. Med.* **381**, 2043–2050 (2019).

142. Vidal, M. *et al.* Probiotics and intestinal colonization by vancomycin-resistant enterococci in mice and humans. *J. Clin. Microbiol.* **48**, 2595–2598 (2010).

143. Crouzet, L. *et al.* Lactobacillus paracasei CNCM I-3689 reduces vancomycin-resistant Enterococcus persistence and promotes Bacteroidetes resilience in the gut following antibiotic challenge. *Sci. Rep.* **8**, 1–11 (2018).

144. Li, X. *et al.* Two strains of lactobacilli effectively decrease the colonization of VRE in a mouse model. *Front. Cell. Infect. Microbiol.* **9**, 1–12 (2019).

145. Caballero, S. *et al.* Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium. *Cell Host Microbe* **21**, 592-602.e4 (2017).

146. Kim, S. G. *et al.* Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. *Nature* **572**, 665–669 (2019).

147. Kommineni, S. *et al.* Bacteriocin production augments niche competition by enterococci in the mammalian gastrointestinal tract. *Nature* **526**, 719–722 (2015).

148. Fukuda, S. *et al.* Bifidobacteria can protect from enteropathogenic infection through production of acetate. *Nature* **469**, 543–549 (2011).

149. Kamada, N. *et al.* Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science (80-. ).* **336**, 1325–1329 (2012).

150. Willems, R. J. L. & van Schaik, W. Transition of Enterococcus faecium from commensal organism to nosocomial pathogen. *Future Microbiol.* **4**, 1125–1135 (2009).

151. McKenney, P. T., Ling, L., Wang, G., Mane, S. & Pamer, E. G. Complete genome sequence of Enterococcus faecium ATCC 700221. *Genome Announc.* **4**, 2015–2016 (2016).

152. Alvarez-Martin, P. *et al.* A conserved two-component signal transduction system controls the response to phosphate starvation in Bifidobacterium breve UCC2003. *Appl. Environ. Microbiol.* **78**, 5258–5269 (2012).

153. Zhang, X. *et al.* Genome-wide identification of ampicillin resistance determinants in enterococcus faecium. *PLoS Genet.* **8**, (2012).

154. Klein, B. A. *et al.* Identification of essential genes of the periodontal pathogen Porphyromonas gingivalis. *BMC Genomics* **13**, (2012).

155. O'Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–D745 (2016).

156. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).

157. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

158. Neuwirth, E. RColorBrewer: ColorBrewer Palettes. *R package version 1.1-2* https://cran.r-project.org/package=RColorBrewer (2014).

159. R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing,. *Vienna, Austria* https://www.r-project.org/.

160. Carver, T., Thomson, N., Bleasby, A., Berriman, M. & Parkhill, J. DNAPlotter: Circular and linear interactive genome visualization. *Bioinformatics* **25**, 119–120 (2009).

161. Otto, R., Brink, B. ten, Veldkamp, H. & Konings, W. N. The relation between growth rate and electrochemical proton gradient of Streptococcus cremoris. *FEMS Microbiol. Lett.* **16**, 69–74 (1983).

162. Zhang, X., Vrijenhoek, J. E. P., Bonten, M. J. M., Willems, R. J. L. & Van Schaik, W. A genetic element present on megaplasmids allows Enterococcus faecium to use raffinose as carbon source. *Environ. Microbiol.* **13**, 518–528 (2011).

163. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).

164. Magoč, T. & Salzberg, S. L. FLASH: Fast length adjustment of short reads to improve genome assemblies. *Bioinformatics* **27**, 2957–2963 (2011).

165. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

166. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

167. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLoS Comput. Biol.* **9**, 1–10 (2013).

168. Isaac, S. Role of the microbiota in the defense against infections by Enterococci, PhD thesis. *Univ. València - CSISP - FISABIO (Área Genómica y Salud Pública)* (2017).

169. Quast, C. *et al.* The SILVA ribosomal RNA gene database project: Improved data processing and web-based tools. *Nucleic Acids Res.* **41**, 590–596 (2013).

170. Consortium, M. G. S. *et al.* Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).

171. Zhu, W., Lomsadze, A. & Borodovsky, M. Ab initio gene identification in metagenomic sequences. *Nucleic Acids Res.* **38**, 1–15 (2010).

172. Mirdita, M. *et al.* Uniclust databases of clustered and deeply annotated protein sequences and alignments. *Nucleic Acids Res.* **45**, D170–D176 (2017).

173. Rognes, T., Flouri, T., Nichols, B., Quince, C. & Mahé, F. VSEARCH: A versatile open source tool for metagenomics. *PeerJ* **2016**, 1–22 (2016).

174. Medeiros, P. M. & Simoneit, B. R. T. Analysis of sugars in environmental samples by gas chromatography-mass spectrometry. *J. Chromatogr. A* **1141**, 271–278 (2007).

175. McGrath, S., Fitzgerald, G. F. & Van Sinderen, D. Improvement and optimization of two engineered phage resistance mechanisms in Lactococcus lactis. *Appl. Environ. Microbiol.* **67**, 608–616 (2001).

176. Pritchard, J. R. *et al.* ARTIST: High-Resolution Genome-Wide Assessment of Fitness Using Transposon-Insertion Sequencing. *PLoS Genet.* **10**, (2014).

177. Matern, W. M., Jenquin, R. L., Bader, J. S. & Karakousis, P. C. Identifying the essential genes of Mycobacterium avium subsp. hominissuis with Tn-Seq using a rank-based filter procedure. *Sci. Rep.* **10**, 1–8 (2020).

178. Lanigan, N., Bottacini, F., Casey, P. G., Motherway, M. O. C. & van Sinderen, D. Genome-wide search for genes required for bifidobacterial growth under iron-limitation. *Front. Microbiol.* **8**, 1–15 (2017).

179. Shlezinger, M., Coppenhagen-Glazer, S., Gelman, D., Beyth, N. & Hazan, R. Eradication of vancomycin-resistant enterococci by combining phage and vancomycin. *Viruses* **11**, (2019).

180. Wardenburg, K. E. *et al.* Phenotypic and genotypic characterization of linezolid-resistant Enterococcus faecium from the USA and Pakistan. *J. Antimicrob. Chemother.* **74**, 3445–3452 (2019).

181. Ding, Q. & Tan, K. S. Himar1 transposon for efficient random mutagenesis in Aggregatibacter actinomycetemcomitans. *Front. Microbiol.* **8**, 1–9 (2017).

182. Lampe, D. J., Grant, T. E. & Robertson, H. M. Factors affecting transposition of the Himar1 mariner transposon in vitro. *Genetics* **149**, 179–187 (1998).

183. Kohanski, M. A., Dwyer, D. J. & Collins, J. J. How antibiotics kill bacteria: From targets to networks. *Nat. Rev. Microbiol.* **8**, 423–435 (2010).

184. Harwick, H. J., Kalmanson, G. M. & Guze, L. B. In vitro activity of ampicillin or vancomycin combined with gentamicin or streptomycin against enterococci. *Antimicrob. Agents Chemother.* **4**, 383–387 (1973).

185. Murray, P. R., Rosenthal, K. S. & Pfaller, M. A. *Medical Microbiology, Eighth Edition*. *Elsevier* (2015). doi:10.1017/CBO9781107415324.004.

186. Tytgat, H. L. P. *et al.* Lactobacillus rhamnosus GG Outcompetes Enterococcus faecium via Mucus-Binding Pili: Evidence for a Novel and Heterospecific Probiotic Mechanism. **82**, 5756–5762 (2016).

187. Sharma, V., Ichikawa, M. & Freeze, H. H. Mannose metabolism: More than meets the eye. *Biochem. Biophys. Res. Commun.* **453**, 220–228 (2014).

188. Archambaud, C., Derré-Bobillot, A., Lapaque, N., Rigottier-Gois, L. & Serror, P. Intestinal translocation of enterococci requires a threshold level of enterococcal overgrowth in the lumen. *Sci. Rep.* **9**, 1–12 (2019).

189. Ubeda, C. *et al.* Familial transmission rather than defective innate immunity shapes the distinct intestinal microbiota of TLR-deficient mice. *J. Exp. Med.* **209**, 1445–1456 (2012).

190. Osbelt, L. *et al.* Variations in microbiota composition of laboratory mice influence Citrobacter rodentium infection via variable short-chain fatty acid production. *PLoS Pathog.* **16**, 1–27 (2020).

191. Velazquez, E. M. *et al.* Endogenous Enterobacteriaceae underlie variation in susceptibility to Salmonella infection. *Nat. Microbiol.* **4**, 1057–1064 (2019).

192. Hryckowian, A. J. *et al.* Microbiota-Accessible carbohydrates suppress Clostridium difficile infection in a murine model. *Nat. Microbiol.* **3**, 662–669 (2018).

193. Wotzka, S. Y. *et al.* Escherichia coli limits Salmonella Typhimurium infections after diet shifts and fat-mediated microbiota perturbation in mice. *Nat. Microbiol.* **4**, 2164–2174 (2019).

194. Lam, M. M. C. *et al.* Comparative analysis of the complete genome of an epidemic hospital sequence type 203 clone of vancomycin-resistant Enterococcus faecium. *BMC Genomics* **14**, 1 (2013).

195. Mohammad, H. *et al.* Phenylthiazole Antibacterial Agents Targeting Cell Wall Synthesis Exhibit Potent Activity In Vitro and In Vivo against Vancomycin-resistant Enterococci. *J. Med. Chem.* **60**, 2425–2438 (2017).

196. Gonzales, M. *et al.* Faecal pharmacokinetics of orally administered vancomycin in patients with suspected Clostridium difficile infection. *BMC Infect. Dis.* **10**, (2010).

197. Bijlsma, J. J. E. *et al.* Development of genomic array footprinting for identification of conditionally essential genes in Streptococcus pneumoniae. *Appl. Environ. Microbiol.* **73**, 1514–1524 (2007).

198. Barquist, L., Boinett, C. J. & Cain, A. K. Approaches to querying bacterial genomes with transposon-insertion sequencing. *RNA Biol.* 1161–1169 (2013).

199. Kwon, Y. M., Ricke, S. C. & Mandal, R. K. Transposon sequencing: methods and expanding applications. *Appl. Microbiol. Biotechnol.* **100**, 31–43 (2016).

200. Tachibana, C. Transcriptomics today: Microarrays, RNA-seq, and more. *Science (80-. ).* (2015) doi:10.1126/science.opms.p1500095.

201. Abdel-Gany, S. S., El-Badry, M. O., Fahmy, A. S. & Mohamed, S. A. Purification and characterization of deoxyribonuclease from small intestine of camel Camelus dromedarius. *J. Genet. Eng. Biotechnol.* **15**, 463–467 (2017).

202. Salama, N. R., Shepherd, B. & Falkow, S. Global transposon mutagenesis and essential gene analysis of Helicobacter pylori. *J. Bacteriol.* **186**, 7926–7935 (2004).

203. Hea-Jin Jung *et al.* crossm Genome-Wide Screening for Enteric Colonization Factors in. 1–15 (2019).

204. Bisanz, J. E. *et al.* A Genomic Toolkit for the Mechanistic Dissection of Intractable Human Gut Bacteria. *Cell Host Microbe* **27**, 1001-1013.e9 (2020).

205. Park, S. C., Kwak, Y. M., Song, W. S., Hong, M. & Yoon, S. Il. Structural basis of effector and operator recognition by the phenolic acid-responsive transcriptional regulator PadR. *Nucleic Acids Res.* **45**, 13080–13093 (2017).

206. Brown, N. L., Stoyanov, J. V., Kidd, S. P. & Hobman, J. L. The MerR family of transcriptional regulators. *FEMS Microbiol. Rev.* **27**, 145–163 (2003).

207. Isaac, S. *et al.* Short- and long-term effects of oral vancomycin on the human intestinal microbiota. *J. Antimicrob. Chemother.* **72**, 128–136 (2017).

208. Azzouz, L. L. & Sharma, S. Physiology, Large Intestine. *StatPearls [Internet]. Treasure Isl. StatPearls Publ. 2020 Jan-. Available from https//www.ncbi.nlm.nih.gov/books/NBK507857/* 1–3 (2020).

209. Starikova, I. *et al.* Fitness costs of various mobile genetic elements in enterococcus faecium and enterococcus faecalis. *J. Antimicrob. Chemother.* **68**, 2755–2765 (2013).

210. Tokumoto, U. & Takahashi, Y. Genetic analysis of the isc operon in Escherichia coli involved in the biogenesis of cellular iron-sulfur proteins. *J. Biochem.* **130**, 63–71 (2001).

211. Ayala-Castro, C., Saini, A. & Outten, F. W. Fe-S Cluster Assembly Pathways in Bacteria. *Microbiol. Mol. Biol. Rev.* **72**, 110–125 (2008).

212. Stogios, P. J. *et al.* Potential for reduction of streptogramin a resistance revealed by structural analysis of acetyltransferase VatA. *Antimicrob. Agents Chemother.* **58**, 7083–7092 (2014).

213. Buffie, C. G. *et al.* Profound alterations of intestinal microbiota following a single dose of clindamycin results in sustained susceptibility to Clostridium difficile-induced colitis. *Infect. Immun.* **80**, 62–73 (2012).

214. Szurmant, H., Mohan, M. A., Imus, P. M. & Hoch, J. A. YycH and YycI interact to regulate the essential YycFG two-component system in Bacillus subtilis. *J. Bacteriol.* **189**, 3280–3289 (2007).

215. Hancock, L. E. & Perego, M. Systematic inactivation and phenotypic characterization of two-component signal transduction systems of Enterococcus faecalis V583. *J. Bacteriol.* **186**, 7951–7958 (2004).

216. Lairson, L. L., Henrissat, B., Davies, G. J. & Withers, S. G. Glycosyltransferases: Structures, Functions, and Mechanisms. *Annu. Rev. Biochem.* **77**, 521–555 (2008).

217. Dale, J. L., Cagnazzo, J., Phan, C. Q., Barnes, A. M. T. & Dunny, G. M. Multiple roles for Enterococcus faecalis glycosyltransferases in biofilm-associated antibiotic resistance, cell envelope integrity, and conjugative transfer. *Antimicrob. Agents Chemother.* **59**, 4094–4105 (2015).

218. Rigottier-Gois, L. *et al.* The surface rhamnopolysaccharide epa of enterococcus faecalis is a key determinant of intestinal colonization. *J. Infect. Dis.* **211**, 62–71 (2015).

219. Shasmal, M., Dey, S., Shaikh, T. R., Bhakta, S. & Sengupta, J. E. coli metabolic protein aldehyde-alcohol dehydrogenase-E binds to the ribosome: A unique moonlighting action revealed. *Sci. Rep.* **6**, 1–12 (2016).

220. Chen, Y. M. & Lin, E. C. C. Regulation of the adhE gene, which encodes ethanol dehydrogenase in Escherichia coli. *J. Bacteriol.* **173**, 8009–8013 (1991).

221. Arnau, J., Jørgensen, F., Madsen, S. M., Vrang, A. & Israelsen, H. Cloning of the Lactococcus lactis adhE gene, encoding a multifunctional alcohol dehydrogenase, by complementation of a fermentative mutant of Escherichia coli. *J. Bacteriol.* **180**, 3049–3055 (1998).

222. Sarantinopoulos, P., Kalantzopoulos, G. & Tsakalidou, E. Citrate Metabolism by Enterococcus faecalis FAIR-E 229. *Appl. Environ. Microbiol.* **67**, 5482–5487 (2001).

223. Snoep, J. L., Joost, M., de Mattos, T. & Neijssel, O. M. Effect of the energy source on the NADH/NAD ratio and on pyruvate catabolism in anaerobic chemostat cultures of Enterococcus faecalis NCTC 775. *FEMS Microbiol. Lett.* **81**, 63–66 (1991).

224. Heras, B. *et al.* DSB proteins and bacterial pathogenicity. *Nat. Rev. Microbiol.* **7**, 215–225 (2009).

225. Akiyama, Y., Kamitani, S., Kusukawa, N. & Ito, K. In vitro catalysis of oxidative folding of disulfide-bonded proteins by the Escherichia coli dsbA (ppfA) gene product. *J. Biol. Chem.* **267**, 22440–22445 (1992).

226. Lee, Y. *et al.* The role of disulfide bond isomerase A (DsbA) of Escherichia coli O157:H7 in biofilm formation and virulence. *FEMS Microbiol. Lett.* **278**, 213–222 (2008).

227. Yang, M. *et al.* Bile salt-induced intermolecular disulfide bond formation activates Vibrio cholerae virulence. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 2348–2353 (2013).

228. Wu, C. Q. *et al.* Two DsbA Proteins Are Important for Vibrio parahaemolyticus Pathogenesis. *Front. Microbiol.* **10**, 1–14 (2019).

229. Weber, G. G. & Klose, K. E. The complexity of Tox T-dependent transcription in Vibrio cholerae. *Indian J. Med. Res.* **133**, 201–206 (2011).

230. Sabbagh, Y., Giral, H., Caldas, Y., Levi, M. & Schiavi, S. Intestinal Phosphate Transport. *Adv. Chronic Kidney Dis.* 1–7 (2011) doi:10.1038/jid.2014.371.

231. Hsieh, Y. J. & Wanner, B. L. Global regulation by the seven-component Pi signaling system. *Curr. Opin. Microbiol.* **13**, 198–203 (2010).

232. Pratt, J. T., Ismail, A. M. & Camilli, A. PhoB regulates both environmental and virulence gene expression in Vibrio cholerae. *Mol. Microbiol.* **77**, 1595–1605 (2010).

233. Zaborin, A. *et al.* Red death in Caenorhabditis elegans caused by Pseudomonas aeruginosa PAO1. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 6327–6332 (2009).

234. Starvation, P. & Environment-, P. crossm PhoPR Contributes to Staphylococcus aureus Growth during. **86**, 1–15 (2018).

235. Pérard, J. & Ollagnier de Choudens, S. Iron–sulfur clusters biogenesis by the SUF machinery: close to the molecular mechanism understanding. *J. Biol. Inorg. Chem.* **23**, 581–596 (2018).

236. Johnson, D. C., Dean, D. R., Smith, A. D. & Johnson, M. K. Structure, Function, and Formation of Biological Iron-Sulfur Clusters. *Annu. Rev. Biochem.* **74**, 247–281 (2005).

237. Riboldi, G. P., De Oliveira, J. S. & Frazzon, J. Enterococcus faecalis SufU scaffold protein enhances SufS desulfurase activity by acquiring sulfur from its cysteine-153. *Biochim. Biophys. Acta - Proteins Proteomics* **1814**, 1910–1918 (2011).

238. Parmanand, B. A. *et al.* A decrease in iron availability to human gut microbiome reduces the growth of potentially pathogenic gut bacteria; an in vitro colonic fermentation study. *J. Nutr. Biochem.* **67**, 20–27 (2019).

239. Huet, G., Daffé, M. & Saves, I. Identification of the Mycobacterium tuberculosis SUF machinery as the exclusive mycobacterial system of [Fe-S] cluster assembly: Evidence for its implication in the pathogen's survival. *J. Bacteriol.* **187**, 6137–6146 (2005).

240. Deutscher, J., Francke, C. & Postma, P. W. How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 939–1031 (2006).

241. Jaeger, T., Arsic, M. & Mayer, C. Scission of the lactyl ether bond of N-acetylmuramic acid by Escherichia coli 'etherase'. *J. Biol. Chem.* **280**, 30100–30106 (2005).

242. Uehara, T., Suefuji, K., Jaeger, T., Mayer, C. & Park, J. T. MurQ Etherase Is Required by Escherichia coli in Order To Metabolize Anhydro-N-Acetylmuramic Acid Obtained either from the Environment or from Its Own Cell Wall. *J. Bacteriol.* **188**, 1660–1662 (2006).

243. Beaber, J. W., Hochhut, B. & Waldor, M. K. SOS response promotes horizontal dissemination of antibiotic resistance genes. *Nature* **427**, 72–74 (2004).

244. Miller, C. *et al.* SOS response induction by β-lactams and bacterial defense against antibiotic lethality. *Science (80-. ).* **305**, 1629–1631 (2004).

245. Ravcheev, D. A. *et al.* Comparative genomics and evolution of regulons of the LacI-family transcription factors. *Front. Microbiol.* **5**, 1–1 (2014).

246. Cuthbertson, L., Kos, V. & Whitfield, C. ABC Transporters Involved in Export of Cell Surface Glycoconjugates. *Microbiol. Mol. Biol. Rev.* **74**, 341–362 (2010).

247. Tomii, K. & Kanehisa, M. A comparative analysis of ABC transporters in complete microbial genomes. *Genome Res.* **8**, 1048–1059 (1998).

248. Barrangou, R., Altermann, E., Hutkins, R., Cano, R. & Klaenhammer, T. R. Functional and comparative genomic analyses of an operon involved in fructooligosaccharide utilization by Lactobacillus acidophilus. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 8957–8962 (2003).

249. Jenior, M. L., Leslie, J. L., Young, V. B. & Schloss, P. D. Clostridium difficile Colonizes Alternative Nutrient Niches during Infection across Distinct Murine Gut Microbiomes. *mSystems* **2**, 1–19 (2017).

250. Rice, L. B. *et al.* A potential virulence gene, hylEfm, predominates in Enterococcus faecium of clinical origin. *J. Infect. Dis.* **187**, 508–512 (2003).

251. Shi, D. *et al.* Crystal structure and biochemical properties of putrescine carbamoyltransferase from Enterococcus faecalis: Assembly, active site, and allosteric regulation. *Proteins Struct. Funct. Bioinforma.* **80**, 1436–1447 (2012).

252. Gibson, P. R., Newnham, E., Barrett, J. S., Shepherd, S. J. & Muir, J. G. Review article: Fructose malabsorption and the bigger picture. *Aliment. Pharmacol. Ther.* **25**, 349–363 (2007).

253. Tailford, L. E., Crost, E. H., Kavanaugh, D. & Juge, N. Mucin glycan foraging in the human gut microbiome. *Front. Genet.* **5**, (2015).

254. Biesiekierski, J. R. *et al.* Quantification of fructans, galacto-oligosacharides and other short-chain carbohydrates in processed grains and cereals. *J. Hum. Nutr. Diet.* **24**, 154–176 (2011).

255. Pessin, J. Mammalian Facilitative Glucose Transporter Family: Structure And Molecular Regulation. *Annu. Rev. Physiol.* **54**, 911–930 (1992).

256. Molis, C. A., The, B. F., Marie-fr, O., Lartigue, S. & Flouri, J. Fructooligosaccharides and Energy Value of in Healthy Humans13. *Nutrition* 324–328 (1996).

257. Goh, Y. J. & Klaenhammer, T. R. Genetic Mechanisms of Prebiotic Oligosaccharide Metabolism in Probiotic Microbes. *Annu. Rev. Food Sci. Technol.* **6**, 137–156 (2015).

258. Putkonen, L., Yao, C. K. & Gibson, P. R. Fructose malabsorption syndrome. *Curr. Opin. Clin. Nutr. Metab. Care* **16**, 473–477 (2013).

259. Taur, Y. *et al.* Intestinal Domination and the Risk of Bacteremia in Patients Undergoing Allogeneic Hematopoietic Stem Cell Transplantation. *Clin. Infect. Dis.* (2012).

260. Santiago, M. *et al.* Microbiome predictors of dysbiosis and VRE decolonization in patients with recurrent C. difficile infections in a multi-center retrospective study. *AIMS Microbiol.* **5**, 1–18 (2019).

261. Caballero, S. *et al.* Cooperating Commensals Restore Colonization Resistance to Vancomycin-Resistant Enterococcus faecium. *Cell Host Microbe* **21**, 592-602.e4 (2017).

262. Kim, S. G. *et al.* Microbiota-derived lantibiotic restores resistance against vancomycin-resistant Enterococcus. *Nature* 1–23 (2019).

263. Dewhirst, F. E. *et al.* Characterization of novel human oral isolates and cloned 16S rDNA sequences that fall in the family Coriobacteriaceae: Description of Olsenella gen. nov., reclassification of Lactobacillus uli as Olsenella uli comb. nov. and description of Olsenella profu. *Int. J. Syst. Evol. Microbiol.* **51**, 1797–1804 (2001).

264. Han, K. Il *et al.* Olsenella faecalis sp. Nov., an anaerobic actinobacterium isolated from human faeces. *Int. J. Syst. Evol. Microbiol.* **69**, 2323–2328 (2019).

265. Pultz, N. J., Hoskins, L. C. & Donskey, C. J. Vancomycin-resistant enterococci may obtain nutritional support by scavenging carbohydrate fragments generated during mucin degradation by the anaerobic microbiota of the colon. *Microb. Drug Resist.* **12**, 63–67 (2006).

# IX.  APPENDIX

**Table 35. Genes not detected in Tn-seq analysis.**

| Accesion | Product |
| --- | --- |
| WP_002285758.1 | hypothetical protein |
| WP_002287659.1 | transposase |
| WP_002296127.1 | IS66 family transposase |
| WP_002316736.1 | transposase |
| WP_071974486.1 | hypothetical protein |
| WP_002293413.1 | 30S ribosomal protein S14 |
| WP_002295743.1 | IS982 family transposase |
| WP_002295755.1 | N-acetyltransferase |
| WP_044051741.1 | hypothetical protein |
| WP_000222572.1 | IS30 family transposase |
| WP_073466279.1 | DDE transposase |
| WP_014386990.1 | IS30 family transposase |
| WP_002286587.1 | integrase |
| WP_002296617.1 | hypothetical protein |
| WP_002296616.1 | hypothetical protein |
| WP_002296615.1 | phage protein |
| WP_002296613.1 | toxin |
| WP_002296612.1 | transcriptional regulator |
| WP_002290310.1 | phage antirepressor protein |
| WP_002296611.1 | XRE family transcriptional regulator |
| WP_002286573.1 | hypothetical protein |
| WP_002296610.1 | hypothetical protein |
| WP_002286568.1 | hypothetical protein |
| WP_002286562.1 | phage regulatory protein |
| WP_002286559.1 | hypothetical protein |
| WP_002286557.1 | hypothetical protein |
| WP_002286553.1 | hypothetical protein |
| WP_002286552.1 | helix-turn-helix domain-containing protein |
| WP_002286696.1 | hypothetical protein |
| WP_002286695.1 | hypothetical protein |
| WP_002296606.1 | antitoxin |
| WP_002286694.1 | hypothetical protein |
| WP_002286550.1 | hypothetical protein |
| WP_002296604.1 | hypothetical protein |
| WP_002286547.1 | hypothetical protein |
| WP_002286545.1 | hypothetical protein |
| WP_002286693.1 | hypothetical protein |
| WP_002286542.1 | autolysin |
| WP_002300143.1 | hypothetical protein |
| WP_002311723.1 | hypothetical protein |
| WP_002296600.1 | thymidylate synthase |
| WP_002286540.1 | HNH endonuclease |
| WP_002296599.1 | hypothetical protein |
| WP_002286538.1 | terminase |
| WP_002286533.1 | terminase |
| WP_002286530.1 | phage portal protein |
| WP_002286527.1 | peptidase S14 |
| WP_002286525.1 | major capsid protein |
| WP_002286524.1 | hypothetical protein |
| WP_002286523.1 | hypothetical protein |
| WP_002286522.1 | head-tail adaptor protein |
| WP_002296598.1 | hypothetical protein |
| WP_002286516.1 | hypothetical protein |
| WP_014386994.1 | hypothetical protein |
| WP_002299170.1 | hypothetical protein |
| WP_002305377.1 | hypothetical protein |
| WP_002286502.1 | phage tail tape measure protein |
| WP_002286500.1 | phage tail protein |

| | | |
|---|---|---|
| | **WP_014386995.1** | phage minor structural protein%2C region |
| | **WP_014386996.1** | hypothetical protein |
| | **WP_002286491.1** | hypothetical protein |
| | **WP_002296594.1** | XkdX family protein |
| | **WP_014386997.1** | hypothetical protein |
| | **WP_014386998.1** | holin |
| | **WP_014386999.1** | N-acetylmuramoyl-L-alanine amidase |
| | **WP_077974336.1** | hypothetical protein |
| | **WP_002286474.1** | hypothetical protein |
| | **WP_002296902.1** | hypothetical protein |
| | **WP_002286473.1** | hypothetical protein |
| | **WP_002286680.1** | hypothetical protein |
| | **WP_002296623.1** | ISL3 family transposase |
| | **WP_000222572.1** | IS30 family transposase |
| | **WP_014387011.1** | ISL3 family transposase |
| | **WP_000997695.1** | IS256 family transposase ISEf1 |
| | **WP_002287107.1** | IS110 family transposase |
| | **WP_000997695.1** | IS256 family transposase ISEf1 |
| | **WP_002322034.1** | hypothetical protein |
| | **WP_002287656.1** | ISL3 family transposase |
| | **WP_002287760.1** | IS30 family transposase |
| | **WP_002296623.1** | ISL3 family transposase |
| | **WP_002287656.1** | ISL3 family transposase |
| | **WP_002305377.1** | hypothetical protein |
| | **WP_014386994.1** | hypothetical protein |
| | **WP_002286516.1** | hypothetical protein |
| | **WP_002296598.1** | hypothetical protein |
| | **WP_002286522.1** | head-tail adaptor protein |
| | **WP_002286523.1** | hypothetical protein |
| | **WP_002286524.1** | hypothetical protein |
| | **WP_002286525.1** | major capsid protein |
| | **WP_002286527.1** | peptidase S14 |
| | **WP_002286530.1** | phage portal protein |
| | **WP_002286533.1** | terminase |
| | **WP_002286538.1** | terminase |
| | **WP_002296599.1** | hypothetical protein |
| | **WP_002286696.1** | hypothetical protein |
| | **WP_002286552.1** | helix-turn-helix domain-containing protein |
| | **WP_002296611.1** | XRE family transcriptional regulator |
| | **WP_002287107.1** | IS110 family transposase |
| | **WP_014387082.1** | transposase |
| | **WP_002327793.1** | IS30 family transposase |
| | **WP_002288457.1** | DUF3042 domain-containing protein |
| | **WP_002296127.1** | IS66 family transposase |
| | **WP_002287659.1** | transposase |
| | **WP_002285758.1** | hypothetical protein |
| | **WP_002286940.1** | group II intron reverse transcriptase/maturase |
| | **WP_002296840.1** | IS256 family transposase IS16 |
| | **WP_002285758.1** | hypothetical protein |
| | **WP_000222572.1** | IS30 family transposase |
| | **WP_002322034.1** | hypothetical protein |
| | **WP_002296840.1** | IS256 family transposase IS16 |
| | **WP_002322034.1** | hypothetical protein |
| | **WP_002322034.1** | hypothetical protein |

**Table 36. Significant genes (log2FC > |2|) from transcriptomic analysis comparing *in vivo* (colon) and *in vitro* samples.**

| Accesion[a] | Definition[b] | FDR[c] | Log2FC[d] *In vivo* (colon) vs *in vitro* | Description[e] |
|---|---|---|---|---|
| WP_002286858.1 | complete genome | 2,71E-45 | 6,99 | galactokinase |
| WP_002304196.1 | complete genome | 1,47E-40 | 6,67 | galactose-1-epimerase |
| WP_002288830.1 | complete genome | 8,41E-29 | 6,39 | PTS ascorbate transporter subunit IIB |
| WP_002289048.1 | complete genome | 2,01E-37 | 6,23 | L-xylulose 5-phosphate 3-epimerase |
| WP_002288832.1 | complete genome | 1,42E-33 | 5,93 | 3-keto-L-gulonate-6-phosphate decarboxylase |
| WP_002295327.1 | complete genome | 1,28E-31 | 5,85 | ribulose-5-phosphate 4-epimerase |
| WP_002288405.1 | complete genome | 1,37E-21 | 5,69 | glyoxalase |
| WP_002289415.1 | complete genome | 5,01E-23 | 5,45 | hsp20/alpha crystallin family protein |
| WP_002288827.1 | complete genome | 2,64E-17 | 5,30 | PTS ascorbate transporter subunit IIC |
| WP_002296672.1 | complete genome | 8,19E-11 | 4,87 | iron-sulfur cluster biosynthesis family protein |
| WP_002287118.1 | complete genome | 1,19E-14 | 4,85 | glucosamine--fructose-6-phosphate aminotransferase |
| NC_017032.1-gene55 | plasmid Aus0004_p1 | 4,50E-14 | 4,77 | hypothetical protein |
| WP_002288825.1 | complete genome | 3,83E-16 | 4,74 | L-ascorbate-6-phosphate lactonase |
| WP_002295331.1 | complete genome | 2,45E-14 | 4,74 | PTS ascorbate transporter subunit IIA |
| NC_017032.1-gene54 | plasmid AUS0004_p1 | 4,29E-14 | 4,61 | hypothetical protein |
| WP_002296674.1 | complete genome | 9,18E-14 | 4,55 | magnesium transporter |
| WP_002287119.1 | complete genome | 1,32E-15 | 4,53 | glucosamine--fructose-6-phosphate aminotransferase |
| WP_002292560.1 | complete genome | 7,99E-18 | 4,53 | aquaporin family protein |
| WP_002296670.1 | complete genome | 4,70E-11 | 4,46 | glycosyl transferase family 8 |
| WP_002287115.1 | complete genome | 1,27E-11 | 4,38 | PTS sugar transporter subunit IIC |
| WP_002296624.1 | complete genome | 9,75E-12 | 4,37 | PTS sorbitol transporter subunit IIA |
| WP_002288437.1 | complete genome | 1,93E-26 | 4,32 | ornithine carbamoyltransferase |
| WP_002288318.1 | complete genome | 3,64E-16 | 4,28 | hypothetical protein |
| WP_002288434.1 | complete genome | 2,58E-24 | 4,25 | carbamate kinase |
| WP_002295768.1 | complete genome | 2,53E-11 | 4,22 | hypothetical protein |
| WP_002292856.1 | complete genome | 2,90E-16 | 4,13 | chitin-binding protein |
| WP_002286789.1 | complete genome | 7,30E-19 | 4,13 | manganese catalase |
| WP_002288285.1 | complete genome | 7,14E-15 | 4,12 | PTS mannose transporter subunit IIA |
| WP_002289418.1 | complete genome | 1,28E-13 | 4,08 | DUF378 domain-containing protein |
| WP_002296675.1 | complete genome | 3,16E-12 | 4,08 | oxidoreductase |
| WP_002288064.1 | complete genome | 6,16E-17 | 4,08 | hypothetical protein |
| WP_002297001.1 | complete genome | 6,24E-12 | 4,07 | hypothetical protein |
| WP_002297290.1 | complete genome | 9,75E-16 | 4,02 | DNA starvation/stationary phase protection protein |
| WP_002295944.1 | complete genome | 2,87E-12 | 3,98 | Asp23/Gls24 family envelope stress response protein |
| WP_002287116.1 | complete genome | 3,43E-09 | 3,98 | PTS mannose transporter subunit IID |
| WP_002287246.1 | complete genome | 4,98E-11 | 3,94 | mannitol-1-phosphate 5-dehydrogenase |
| WP_002286047.1 | complete genome | 1,59E-10 | 3,93 | PTS lactose transporter subunit IIB |
| WP_002287114.1 | complete genome | 9,22E-09 | 3,93 | PTS mannose/fructose/sorbose transporter subunit IIB |
| WP_002288281.1 | complete genome | 2,98E-17 | 3,93 | PTS galactitol transporter subunit IIC |
| WP_002311093.1 | complete genome | 1,75E-09 | 3,92 | transaldolase |
| WP_002297122.1 | complete genome | 1,13E-19 | 3,92 | flotillin |
| WP_002294666.1 | complete genome | 2,93E-11 | 3,88 | glycerol-3-phosphate dehydrogenase/oxidase |
| WP_002289734.1 | complete genome | 7,01E-10 | 3,88 | type I glyceraldehyde-3-phosphate dehydrogenase |
| WP_002296425.1 | complete genome | 7,21E-11 | 3,85 | universal stress protein |
| WP_002288278.1 | complete genome | 4,98E-15 | 3,84 | aldolase |
| WP_002286065.1 | complete genome | 3,53E-11 | 3,78 | hypothetical protein |
| WP_002289293.1 | complete genome | 7,62E-11 | 3,75 | dihydroxyacetone kinase subunit L |

| | | | | |
|---|---|---|---|---|
| **WP_002287112.1** | complete genome | 7,78E-07 | 3,75 | PTS mannose transporter subunit IIA |
| **WP_002326789.1** | complete genome | 2,57E-08 | 3,72 | sodium:solute symporter |
| **WP_002286273.1** | complete genome | 2,64E-11 | 3,72 | accessory gene regulator B |
| **WP_002296671.1** | complete genome | 4,90E-09 | 3,71 | universal stress protein A |
| **WP_002294668.1** | complete genome | 2,29E-17 | 3,70 | glycerol kinase |
| **WP_002289467.1** | complete genome | 8,16E-09 | 3,68 | universal stress protein |
| **WP_002294638.1** | complete genome | 3,93E-12 | 3,64 | dihydroxyacetone kinase subunit DhaK |
| **WP_002289423.1** | complete genome | 3,33E-06 | 3,62 | PTS sorbitol transporter subunit IIC |
| **WP_002296424.1** | complete genome | 3,58E-13 | 3,61 | hypothetical protein |
| **WP_002296473.1** | complete genome | 6,96E-11 | 3,59 | L-lactate oxidase |
| **WP_002296567.1** | complete genome | 1,94E-13 | 3,56 | bifunctional acetaldehyde-CoA/alcohol dehydrogenase |
| **WP_002296654.1** | complete genome | 6,09E-08 | 3,53 | mannonate dehydratase |
| **WP_002290825.1** | complete genome | 6,52E-16 | 3,53 | hypothetical protein |
| **WP_002296653.1** | complete genome | 3,09E-05 | 3,52 | PTS fructose transporter subunit IIA |
| **WP_002295336.1** | complete genome | 7,17E-04 | 3,51 | diacylglycerol kinase |
| **WP_002296656.1** | complete genome | 4,46E-07 | 3,51 | HAD family phosphatase |
| **WP_002287756.1** | complete genome | 6,73E-04 | 3,50 | hypothetical protein |
| **WP_002294422.1** | complete genome | 7,59E-05 | 3,50 | hypothetical protein |
| **WP_002289259.1** | complete genome | 8,30E-10 | 3,49 | hypothetical protein |
| **WP_002288252.1** | complete genome | 1,06E-07 | 3,48 | AraC family transcriptional regulator |
| **WP_002295945.1** | complete genome | 8,31E-08 | 3,46 | DUF2273 domain-containing protein |
| **WP_002288964.1** | complete genome | 1,26E-02 | 3,40 | helix-turn-helix domain-containing protein |
| **WP_002297333.1** | complete genome | 3,20E-03 | 3,37 | sigma-70 family RNA polymerase sigma factor |
| **WP_002320809.1** | complete genome | 1,01E-02 | 3,36 | helix-turn-helix domain-containing protein |
| **WP_002287755.1** | complete genome | 4,31E-05 | 3,35 | sugar ABC transporter ATP-binding protein |
| **WP_002289045.1** | complete genome | 2,63E-04 | 3,34 | hypothetical protein |
| **WP_002286467.1** | complete genome | 3,56E-08 | 3,33 | TetR family transcriptional regulator |
| **WP_002297125.1** | complete genome | 1,62E-05 | 3,30 | histidine kinase |
| **WP_002295447.1** | complete genome | 3,58E-09 | 3,29 | hypothetical protein |
| **WP_002295947.1** | complete genome | 1,15E-07 | 3,28 | hypothetical protein |
| **WP_002287647.1** | complete genome | 3,04E-08 | 3,27 | heavy metal translocating P-type ATPase |
| **WP_002287954.1** | complete genome | 1,69E-06 | 3,27 | redox-sensing transcriptional repressor Rex |
| **WP_002340434.1** | complete genome | 2,39E-04 | 3,26 | hypothetical protein |
| **WP_002287583.1** | complete genome | 4,17E-10 | 3,25 | beta-glucosidase |
| **WP_002286530.1** | complete genome | 7,93E-06 | 3,25 | phage portal protein |
| **WP_002289226.1** | complete genome | 1,05E-07 | 3,24 | ketose-bisphosphate aldolase |
| **WP_014387054.1** | complete genome | 1,28E-02 | 3,23 | hypothetical protein |
| **WP_002296408.1** | complete genome | 3,43E-06 | 3,22 | DUF1538 domain-containing protein |
| **WP_002296645.1** | complete genome | 1,34E-05 | 3,22 | oxidoreductase |
| **WP_002287147.1** | complete genome | 4,01E-02 | 3,22 | hypothetical protein |
| **WP_002291790.1** | complete genome | 2,64E-05 | 3,21 | PRD domain-containing protein |
| **WP_002296760.1** | complete genome | 6,48E-06 | 3,21 | PTS sorbitol transporter subunit IIB |
| **WP_002297127.1** | complete genome | 1,24E-05 | 3,21 | histidine kinase |
| **WP_002289218.1** | complete genome | 7,44E-06 | 3,20 | RNA polymerase subunit sigma-54 |
| **WP_002289225.1** | complete genome | 4,68E-07 | 3,19 | PTS fructose transporter subunit IIBC |
| **WP_002288002.1** | complete genome | 2,96E-06 | 3,19 | sensor histidine kinase |
| **WP_002287247.1** | complete genome | 9,66E-06 | 3,17 | PTS mannitol transporter subunit IIA |
| **WP_002286527.1** | complete genome | 2,68E-05 | 3,17 | peptidase S14 |
| **WP_002289187.1** | complete genome | 2,63E-05 | 3,16 | L-threonine 3-dehydrogenase |
| **WP_002288811.1** | complete genome | 3,54E-04 | 3,15 | carbohydrate deacetylase |
| **WP_002302669.1** | complete genome | 1,72E-06 | 3,14 | AP endonuclease |
| **WP_002285948.1** | complete genome | 1,35E-05 | 3,13 | PTS cellbiose transporter subunit IIC |
| **WP_002287569.1** | complete genome | 9,96E-06 | 3,13 | hypothetical protein |
| **WP_002289535.1** | complete genome | 6,71E-07 | 3,12 | glutamine amidotransferase |
| **WP_002286048.1** | complete genome | 2,65E-08 | 3,12 | transcription antiterminator BglG |
| **WP_002287584.1** | complete genome | 6,04E-08 | 3,12 | PTS lactose transporter subunit IIC |
| **WP_002296610.1** | complete genome | 1,77E-02 | 3,12 | hypothetical protein |
| **WP_002347051.1** | complete genome | 1,36E-03 | 3,11 | YopX protein |
| **WP_002286500.1** | complete genome | 5,46E-04 | 3,10 | phage tail protein |

| | | | | |
|---|---|---|---|---|
| **WP_002287997.1** | complete genome | 7,60E-05 | 3,10 | LacI family transcriptional regulator |
| **WP_002285960.1** | complete genome | 1,85E-04 | 3,10 | hypothetical protein |
| **WP_002288635.1** | complete genome | 4,13E-04 | 3,08 | PTS lactose/cellobiose transporter subunit IIA |
| **WP_002286562.1** | complete genome | 2,12E-04 | 3,08 | phage regulatory protein |
| **WP_002296611.1** | complete genome | 1,73E-03 | 3,07 | XRE family transcriptional regulator |
| **WP_002287570.1** | complete genome | 2,53E-06 | 3,07 | oleate hydratase |
| **WP_002293648.1** | complete genome | 8,20E-04 | 3,06 | hypothetical protein |
| **WP_002295200.1** | complete genome | 3,30E-04 | 3,03 | P-II family nitrogen regulator |
| **WP_002287250.1** | complete genome | 5,12E-06 | 3,03 | PTS mannitol transporter subunit IICB |
| **WP_044031074.1** | complete genome | 1,66E-02 | 3,03 | hypothetical protein |
| **WP_002288004.1** | complete genome | 1,04E-04 | 3,01 | DNA-binding response regulator |
| **WP_002286530.1** | complete genome | 1,33E-05 | 3,00 | phage portal protein |
| **WP_002286041.1** | complete genome | 6,58E-05 | 3,00 | PTS ascorbate transporter subunit IIC |
| **WP_002321678.1** | complete genome | 3,12E-04 | 2,99 | hypothetical protein |
| **WP_002289421.1** | complete genome | 3,83E-05 | 2,99 | transcription antiterminator BglG |
| **WP_002287587.1** | complete genome | 6,03E-09 | 2,98 | glyceraldehyde 3-phosphate reductase |
| **WP_002321266.1** | complete genome | 2,86E-03 | 2,97 | hypothetical protein |
| **WP_002295938.1** | complete genome | 6,83E-03 | 2,97 | glutaminase A |
| **WP_002289507.1** | complete genome | 6,47E-08 | 2,96 | sugar ABC transporter ATP-binding protein |
| **WP_002296648.1** | complete genome | 1,29E-03 | 2,95 | PTS mannose/fructose/sorbose transporter subunit IIB |
| **WP_002288009.1** | complete genome | 3,51E-06 | 2,95 | bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/inosine monophosphate cyclohydrolase |
| **WP_002319777.1** | complete genome | 1,12E-04 | 2,92 | antirepressor |
| **WP_002296225.1** | complete genome | 4,94E-04 | 2,91 | PTS sugar transporter subunit IIA |
| **WP_002288636.1** | complete genome | 7,48E-05 | 2,90 | 6-phospho-beta-galactosidase |
| **WP_002288633.1** | complete genome | 6,15E-06 | 2,90 | arabinogalactan endo-1%2C4-beta-galactosidase |
| **WP_002287534.1** | complete genome | 5,28E-03 | 2,90 | PTS ascorbate transporter subunit IIB |
| **WP_002296640.1** | complete genome | 1,06E-03 | 2,90 | 6-phosphogluconate dehydrogenase |
| **WP_002288487.1** | complete genome | 6,18E-05 | 2,90 | hypothetical protein |
| **WP_002286527.1** | complete genome | 1,29E-04 | 2,89 | peptidase S14 |
| **WP_002286523.1** | complete genome | 2,33E-03 | 2,89 | hypothetical protein |
| **WP_002289280.1** | complete genome | 7,40E-05 | 2,88 | GlsB/YeaQ/YmgE family stress response membrane protein |
| **WP_002304144.1** | complete genome | 5,56E-06 | 2,88 | PTS beta-glucoside transporter subunit EIIBCA |
| **WP_002290310.1** | complete genome | 1,68E-04 | 2,88 | phage antirepressor protein |
| **WP_074394427.1** | complete genome | 5,65E-02 | 2,88 | hypothetical protein |
| **WP_002289130.1** | complete genome | 1,93E-03 | 2,87 | carbohydrate ABC transporter permease |
| **WP_002296488.1** | complete genome | 2,55E-03 | 2,87 | terminase |
| **WP_002296647.1** | complete genome | 6,56E-04 | 2,87 | PTS mannose transporter subunit IID |
| **WP_002286523.1** | complete genome | 6,64E-03 | 2,87 | hypothetical protein |
| **WP_014387040.1** | complete genome | 1,49E-02 | 2,87 | phage tail protein |
| **WP_002286780.1** | complete genome | 8,62E-04 | 2,86 | peptidyl-prolyl cis-trans isomerase |
| **WP_002288249.1** | complete genome | 5,35E-05 | 2,86 | L-rhamnose isomerase |
| **WP_002288013.1** | complete genome | 8,45E-05 | 2,86 | amidophosphoribosyltransferase |
| **WP_002320997.1** | complete genome | 2,09E-03 | 2,86 | hypothetical protein |
| **WP_002294699.1** | complete genome | 7,40E-05 | 2,86 | iron ABC transporter substrate-binding protein |
| **WP_002288809.1** | complete genome | 3,74E-04 | 2,86 | transcriptional antiterminator |
| **WP_002318517.1** | complete genome | 4,17E-04 | 2,86 | hypothetical protein |
| **NC_017032.1-gene31** | plasmid Aus0004_p1 | 1,37E-04 | 2,86 | hypothetical protein |
| **NC_017032.1-gene29** | plasmid Aus0004_p1 | 3,18E-04 | 2,85 | hypothetical protein |
| **WP_002289135.1** | complete genome | 1,39E-03 | 2,85 | AraC family transcriptional regulator |
| **WP_002289800.1** | complete genome | 8,15E-04 | 2,85 | lactate dehydrogenase |
| **WP_002287999.1** | complete genome | 1,58E-04 | 2,85 | LacI family transcriptional regulator |
| **WP_002286524.1** | complete genome | 4,99E-04 | 2,84 | hypothetical protein |
| **WP_002287995.1** | complete genome | 2,48E-04 | 2,83 | sugar ABC transporter ATP-binding protein |

| | | | | |
|---|---|---|---|---|
| **WP_049788865.1** | complete genome | 5,89E-04 | 2,83 | topoisomerase |
| **WP_071974494.1** | complete genome | 4,38E-02 | 2,83 | hypothetical protein |
| **WP_071974454.1** | complete genome | 3,56E-02 | 2,82 | hypothetical protein |
| **WP_002285974.1** | complete genome | 5,16E-06 | 2,81 | ABC transporter substrate-binding protein |
| **WP_002288525.1** | complete genome | 3,42E-02 | 2,81 | maltose ABC transporter permease |
| **WP_002287295.1** | complete genome | 1,07E-04 | 2,80 | ABC transporter substrate-binding protein |
| **WP_002286530.1** | complete genome | 4,07E-04 | 2,80 | phage portal protein |
| **WP_002285949.1** | complete genome | 2,74E-05 | 2,80 | transcription antiterminator BglG |
| **WP_002299170.1** | complete genome | 3,59E-02 | 2,80 | hypothetical protein |
| **WP_002286523.1** | complete genome | 6,59E-03 | 2,79 | hypothetical protein |
| **WP_002296094.1** | complete genome | 2,73E-07 | 2,79 | hypothetical protein |
| **WP_014387053.1** | complete genome | 1,14E-02 | 2,78 | hypothetical protein |
| **WP_002296646.1** | complete genome | 2,12E-03 | 2,78 | PTS sugar transporter subunit IIC |
| **WP_002289543.1** | complete genome | 9,31E-06 | 2,78 | NADH dehydrogenase |
| **WP_002286547.1** | complete genome | 1,59E-02 | 2,77 | hypothetical protein |
| **WP_002295474.1** | complete genome | 2,89E-03 | 2,77 | phosphoribosylformylglycinamidine synthase subunit PurS |
| **WP_002287812.1** | complete genome | 1,59E-03 | 2,76 | hypothetical protein |
| **WP_002294399.1** | complete genome | 6,88E-08 | 2,76 | hypothetical protein |
| **WP_002299170.1** | complete genome | 1,72E-02 | 2,76 | hypothetical protein |
| **WP_014386997.1** | complete genome | 1,97E-02 | 2,76 | hypothetical protein |
| **WP_002286522.1** | complete genome | 6,32E-03 | 2,75 | head-tail adaptor protein |
| **WP_002321731.1** | complete genome | 4,24E-04 | 2,74 | phosphoribosylformylglycinamidine cyclo-ligase |
| **WP_002296491.1** | complete genome | 3,61E-03 | 2,74 | hypothetical protein |
| **WP_002291274.1** | complete genome | 1,62E-04 | 2,74 | DUF2273 domain-containing protein |
| **WP_002288970.1** | complete genome | 6,07E-05 | 2,74 | transcriptional regulator |
| **WP_002288823.1** | complete genome | 2,63E-03 | 2,73 | DeoR/GlpR transcriptional regulator |
| **WP_002289643.1** | complete genome | 1,35E-03 | 2,73 | carbohydrate ABC transporter substrate-binding protein |
| **WP_044031073.1** | complete genome | 6,82E-03 | 2,73 | hypothetical protein |
| **WP_002322021.1** | complete genome | 1,50E-02 | 2,73 | hypothetical protein |
| **WP_002288812.1** | complete genome | 8,76E-04 | 2,73 | PTS acetylglucosamine transporter subunit IIB |
| **WP_002286545.1** | complete genome | 5,98E-04 | 2,72 | hypothetical protein |
| **WP_002289420.1** | complete genome | 1,74E-04 | 2,72 | NAD(P)-dependent oxidoreductase |
| **WP_002296650.1** | complete genome | 8,83E-03 | 2,72 | hypothetical protein |
| **WP_002296259.1** | complete genome | 8,23E-03 | 2,72 | hypothetical protein |
| **WP_002285971.1** | complete genome | 1,23E-03 | 2,72 | sugar ABC transporter permease |
| **WP_002289220.1** | complete genome | 1,61E-02 | 2,72 | sugar permease |
| **WP_002287050.1** | complete genome | 1,96E-05 | 2,71 | flavocytochrome c |
| **WP_002305377.1** | complete genome | 4,14E-03 | 2,71 | hypothetical protein |
| **WP_002288010.1** | complete genome | 2,03E-03 | 2,71 | phosphoribosylglycinamide formyltransferase |
| **WP_002286550.1** | complete genome | 5,68E-03 | 2,71 | hypothetical protein |
| **WP_002293677.1** | complete genome | 2,59E-03 | 2,71 | hypothetical protein |
| **WP_014386994.1** | complete genome | 1,30E-03 | 2,71 | hypothetical protein |
| **WP_002321938.1** | complete genome | 1,72E-04 | 2,70 | hypothetical protein |
| **WP_014387037.1** | complete genome | 2,38E-03 | 2,70 | glycosyl hydrolase family 18 |
| **WP_002294874.1** | complete genome | 2,00E-06 | 2,70 | hypothetical protein |
| **WP_002287817.1** | complete genome | 3,90E-04 | 2,70 | hypothetical protein |
| **WP_002296598.1** | complete genome | 2,17E-02 | 2,69 | hypothetical protein |
| **WP_002288162.1** | complete genome | 3,43E-06 | 2,69 | MerR family transcriptional regulator |
| **WP_002289120.1** | complete genome | 9,07E-06 | 2,68 | histidine phosphatase family protein |
| **WP_002303716.1** | complete genome | 3,07E-04 | 2,68 | maltose-6'-phosphate glucosidase |
| **WP_014387073.1** | complete genome | 1,48E-02 | 2,68 | sigma-70 family RNA polymerase sigma factor |
| **WP_002296628.1** | complete genome | 8,51E-07 | 2,68 | CPBP family intramembrane metalloprotease |
| **WP_002286693.1** | complete genome | 5,80E-03 | 2,68 | hypothetical protein |
| **NC_017032.1-gene48** | plasmid Aus0004_p1 | 6,49E-04 | 2,68 | transposase |
| **WP_002286784.1** | complete genome | 3,55E-03 | 2,68 | membrane protein |
| **NC_017032.1-gene33** | plasmid Aus0004_p1 | 1,04E-03 | 2,67 | hypothetical protein |

| | | | | |
|---|---|---|---|---|
| **WP_002296641.1** | complete genome | 6,34E-04 | 2,67 | 2-hydroxyacid dehydrogenase |
| **WP_002297115.1** | complete genome | 1,21E-04 | 2,67 | hypothetical protein |
| **WP_002286552.1** | complete genome | 4,65E-03 | 2,67 | helix-turn-helix domain-containing protein |
| **WP_010776523.1** | complete genome | 1,09E-02 | 2,66 | hypothetical protein |
| **WP_002296604.1** | complete genome | 1,42E-02 | 2,66 | hypothetical protein |
| **WP_014748731.1** | complete genome | 3,96E-02 | 2,66 | hypothetical protein |
| **WP_002286500.1** | complete genome | 2,06E-02 | 2,66 | phage tail protein |
| **NC_017032.1-gene3** | plasmid Aus0004_p1 | 5,98E-04 | 2,65 | hypothetical protein |
| **WP_002296471.1** | complete genome | 2,22E-03 | 2,65 | cell surface protein |
| **WP_002288021.1** | complete genome | 1,68E-03 | 2,65 | phosphoribosylaminoimidazolesuccinocarboxamide synthase |
| **WP_002286296.1** | complete genome | 7,67E-04 | 2,65 | PTS sucrose transporter subunit IIBC |
| **WP_002287742.1** | complete genome | 8,25E-04 | 2,65 | DNA-binding response regulator |
| **WP_002286562.1** | complete genome | 8,53E-03 | 2,65 | phage regulatory protein |
| **WP_002285954.1** | complete genome | 2,06E-03 | 2,64 | PTS cellobiose transporter subunit IIA |
| **WP_002289186.1** | complete genome | 7,89E-05 | 2,64 | glycine C-acetyltransferase |
| **WP_002297337.1** | complete genome | 2,49E-03 | 2,64 | pyridine nucleotide-disulfide oxidoreductase |
| **WP_002288261.1** | complete genome | 5,05E-04 | 2,64 | glycosyl hydrolase |
| **WP_002294402.1** | complete genome | 5,30E-04 | 2,64 | hypothetical protein |
| **WP_002311723.1** | complete genome | 2,78E-02 | 2,64 | hypothetical protein |
| **WP_002296335.1** | complete genome | 9,20E-03 | 2,63 | hypothetical protein |
| **WP_002287057.1** | complete genome | 1,07E-03 | 2,63 | hypothetical protein |
| **WP_002317192.1** | complete genome | 4,01E-04 | 2,63 | 1-phosphofructokinase |
| **WP_044031436.1** | complete genome | 1,45E-03 | 2,63 | phage/plasmid primase P4 family domain-containing protein |
| **WP_002305377.1** | complete genome | 2,45E-02 | 2,63 | hypothetical protein |
| **WP_014387051.1** | complete genome | 6,00E-03 | 2,63 | gamma-glutamylcyclotransferase |
| **WP_002311095.1** | complete genome | 2,29E-04 | 2,63 | peptidase |
| **WP_002295465.1** | complete genome | 2,86E-03 | 2,63 | sugar ABC transporter permease |
| **WP_002296514.1** | complete genome | 5,31E-02 | 2,63 | hypothetical protein |
| **WP_002286772.1** | complete genome | 7,83E-04 | 2,62 | class C sortase |
| **WP_002286502.1** | complete genome | 2,86E-04 | 2,62 | phage tail tape measure protein |
| **WP_002286525.1** | complete genome | 6,04E-04 | 2,62 | major capsid protein |
| **WP_002311272.1** | complete genome | 4,29E-05 | 2,62 | DNA helicase |
| **WP_002347046.1** | complete genome | 1,36E-02 | 2,62 | hypothetical protein |
| **WP_000713874.1** | complete genome | 2,63E-03 | 2,62 | lactococcin 972 family bacteriocin |
| **WP_002296611.1** | complete genome | 1,12E-02 | 2,61 | XRE family transcriptional regulator |
| **WP_002321528.1** | complete genome | 3,66E-03 | 2,61 | hypothetical protein |
| **WP_002294171.1** | complete genome | 6,50E-04 | 2,61 | lactaldehyde reductase |
| **WP_002286928.1** | complete genome | 7,98E-03 | 2,61 | hypothetical protein |
| **WP_002286527.1** | complete genome | 1,01E-03 | 2,61 | peptidase S14 |
| **WP_002287282.1** | complete genome | 5,05E-04 | 2,61 | L-arabinose isomerase |
| **WP_002286559.1** | complete genome | 4,47E-03 | 2,60 | hypothetical protein |
| **WP_077495106.1** | plasmid | 5,30E-02 | 2,60 | hypothetical protein |
| **WP_002289136.1** | complete genome | 1,18E-03 | 2,60 | sensor histidine kinase |
| **WP_014387104.1** | complete genome | 2,92E-03 | 2,60 | PTS mannose/fructose/sorbose IIA component |
| **WP_002287140.1** | complete genome | 1,68E-04 | 2,59 | hypothetical protein |
| **WP_044031099.1** | complete genome | 2,25E-02 | 2,59 | hypothetical protein |
| **WP_002290310.1** | complete genome | 2,36E-03 | 2,59 | phage antirepressor protein |
| **WP_002286696.1** | complete genome | 4,18E-02 | 2,59 | hypothetical protein |
| **WP_002287284.1** | complete genome | 9,97E-04 | 2,59 | L-ribulose-5-phosphate 4-epimerase |
| **WP_002286816.1** | complete genome | 6,95E-05 | 2,58 | cysteine hydrolase |
| **WP_014386995.1** | complete genome | 4,98E-04 | 2,58 | phage minor structural protein%2C region |
| **WP_002296108.1** | complete genome | 2,59E-03 | 2,58 | amino acid permease |
| **WP_002295229.1** | complete genome | 2,06E-02 | 2,58 | 30S ribosomal protein S9 |
| **WP_002340386.1** | complete genome | 2,17E-04 | 2,57 | beta-galactosidase large subunit |
| **WP_002286522.1** | complete genome | 2,15E-02 | 2,57 | head-tail adaptor protein |
| **WP_044031091.1** | complete genome | 7,61E-02 | 2,57 | HNH endonuclease |
| **WP_002289648.1** | complete genome | 3,90E-04 | 2,57 | gfo/Idh/MocA family oxidoreductase |
| **WP_002304527.1** | complete genome | 1,41E-03 | 2,57 | endonuclease |

| | | | | |
|---|---|---|---|---|
| WP_002286538.1 | complete genome | 1,72E-02 | 2,56 | terminase |
| WP_002288007.1 | complete genome | 6,81E-04 | 2,56 | phosphoribosylamine--glycine ligase |
| WP_014386997.1 | complete genome | 5,49E-02 | 2,56 | hypothetical protein |
| WP_002317191.1 | complete genome | 3,30E-04 | 2,56 | PTS fructose transporter subunit IIC |
| WP_002288258.1 | complete genome | 3,54E-03 | 2,56 | TRAP transporter permease DctQ |
| WP_002296783.1 | complete genome | 4,02E-06 | 2,56 | transcriptional regulator |
| WP_002296412.1 | complete genome | 5,75E-04 | 2,56 | alpha-galactosidase |
| WP_002321369.1 | complete genome | 1,42E-02 | 2,56 | glucose uptake protein |
| WP_002286596.1 | complete genome | 6,24E-02 | 2,56 | DUF3188 domain-containing protein |
| WP_002288430.1 | complete genome | 7,00E-03 | 2,55 | phosphonoacetaldehyde hydrolase |
| WP_002289059.1 | complete genome | 6,87E-04 | 2,55 | FUSC family protein |
| WP_002288982.1 | complete genome | 1,88E-03 | 2,54 | class C sortase |
| WP_002286557.1 | complete genome | 3,72E-03 | 2,54 | hypothetical protein |
| WP_002296192.1 | complete genome | 6,71E-02 | 2,54 | phenylalanyl-tRNA synthetase subunit beta |
| WP_002286516.1 | complete genome | 3,57E-02 | 2,54 | hypothetical protein |
| WP_002347042.1 | complete genome | 3,74E-03 | 2,54 | ArpU family phage transcriptional regulator |
| WP_002347047.1 | complete genome | 1,40E-02 | 2,54 | hypothetical protein |
| WP_002296669.1 | complete genome | 2,06E-04 | 2,54 | HAD family phosphatase |
| WP_002296611.1 | complete genome | 1,88E-02 | 2,54 | XRE family transcriptional regulator |
| WP_001791010.1 | complete genome | 1,66E-02 | 2,54 | tetracycline resistance protein |
| NC_017032.1-gene22 | plasmid Aus0004_p1 | 6,78E-03 | 2,54 | hypothetical protein |
| WP_033657400.1 | complete genome | 2,41E-02 | 2,53 | hypothetical protein |
| WP_002287656.1 | complete genome | 1,31E-03 | 2,53 | ISL3 family transposase |
| WP_002298631.1 | complete genome | 1,25E-03 | 2,53 | group II intron reverse transcriptase/maturase |
| WP_002297304.1 | complete genome | 1,48E-02 | 2,53 | cobalt transporter |
| WP_002294701.1 | complete genome | 2,68E-03 | 2,53 | hypothetical protein |
| NC_017032.1-gene45 | plasmid Aus0004_p1 | 1,29E-02 | 2,53 | hypothetical protein |
| WP_002296600.1 | complete genome | 2,82E-02 | 2,53 | thymidylate synthase |
| WP_002289266.1 | complete genome | 1,64E-02 | 2,53 | hypothetical protein |
| WP_002304700.1 | complete genome | 3,19E-03 | 2,52 | MoxR family ATPase |
| WP_002311951.1 | complete genome | 5,68E-02 | 2,52 | hypothetical protein |
| WP_002286818.1 | complete genome | 1,96E-05 | 2,52 | pyridine nucleotide-disulfide oxidoreductase |
| WP_002286516.1 | complete genome | 6,32E-02 | 2,52 | hypothetical protein |
| WP_002303719.1 | complete genome | 1,99E-02 | 2,52 | hypothetical protein |
| WP_002296480.1 | complete genome | 2,41E-04 | 2,52 | anion permease |
| WP_002288817.1 | complete genome | 1,14E-02 | 2,51 | PTS lactose transporter subunit IIC |
| WP_002294179.1 | complete genome | 1,47E-03 | 2,51 | hypothetical protein |
| WP_002304820.1 | complete genome | 3,65E-02 | 2,51 | hypothetical protein |
| WP_002311314.1 | complete genome | 1,90E-03 | 2,51 | tagatose-6-phosphate kinase |
| WP_002291912.1 | complete genome | 3,24E-05 | 2,51 | PTS sugar transporter subunit IIC |
| WP_002289133.1 | complete genome | 7,52E-04 | 2,50 | carbohydrate ABC transporter substrate-binding protein |
| WP_002286470.1 | complete genome | 7,14E-02 | 2,50 | hypothetical protein |
| WP_002286502.1 | complete genome | 3,21E-04 | 2,50 | phage tail tape measure protein |
| NC_017032.1-gene23 | plasmid Aus0004_p1 | 5,37E-03 | 2,50 | hypothetical protein |
| WP_014387039.1 | complete genome | 2,59E-03 | 2,50 | endopeptidase |
| WP_002320982.1 | complete genome | 7,86E-02 | 2,49 | hypothetical protein |
| WP_002286180.1 | complete genome | 1,96E-03 | 2,49 | competence protein ComGB |
| WP_002289228.1 | complete genome | 2,45E-05 | 2,49 | hypothetical protein |
| WP_002321251.1 | complete genome | 1,86E-02 | 2,49 | hypothetical protein |
| WP_002286524.1 | complete genome | 5,59E-03 | 2,49 | hypothetical protein |
| WP_002286694.1 | complete genome | 6,70E-03 | 2,49 | hypothetical protein |
| WP_014387130.1 | complete genome | 2,05E-03 | 2,49 | hypothetical protein |
| WP_002304138.1 | complete genome | 1,41E-03 | 2,49 | PTS beta-glucoside transporter subunit IIBC |
| WP_002286056.1 | complete genome | 1,29E-03 | 2,48 | hypothetical protein |
| WP_002286552.1 | complete genome | 2,44E-02 | 2,48 | helix-turn-helix domain-containing protein |
| WP_002288439.1 | complete genome | 1,11E-06 | 2,48 | arginine deiminase |
| WP_002289647.1 | complete genome | 4,06E-03 | 2,48 | carbohydrate ABC transporter permease |

| | | | | |
|---|---|---|---|---|
| WP_002296429.1 | complete genome | 1,13E-05 | 2,48 | hypothetical protein |
| WP_002296637.1 | complete genome | 4,89E-03 | 2,48 | alpha/beta hydrolase |
| WP_002296487.1 | complete genome | 3,96E-03 | 2,48 | terminase |
| WP_002295800.1 | complete genome | 2,56E-05 | 2,48 | DUF871 domain-containing protein |
| WP_002317278.1 | complete genome | 1,29E-02 | 2,47 | plasmid mobilization relaxosome protein MobC |
| WP_002286559.1 | complete genome | 1,21E-02 | 2,47 | hypothetical protein |
| WP_002286601.1 | complete genome | 1,29E-03 | 2,47 | PTS cellobiose transporter subunit IIC |
| WP_002288934.1 | complete genome | 2,02E-02 | 2,47 | antirestriction protein ArdA |
| WP_002286592.1 | complete genome | 4,77E-02 | 2,47 | hypothetical protein |
| WP_002296495.1 | complete genome | 2,25E-02 | 2,47 | hypothetical protein |
| WP_002321736.1 | complete genome | 1,05E-02 | 2,47 | hypothetical protein |
| WP_002296275.1 | complete genome | 1,88E-03 | 2,47 | mannonate dehydratase |
| WP_002286693.1 | complete genome | 1,47E-02 | 2,46 | hypothetical protein |
| WP_002317138.1 | complete genome | 1,93E-04 | 2,46 | glyoxalase |
| WP_014387033.1 | complete genome | 4,37E-03 | 2,46 | recombinase family protein |
| WP_002289215.1 | complete genome | 2,17E-02 | 2,46 | PTS mannose/fructose/sorbose transporter subunit IIB |
| WP_002294677.1 | complete genome | 6,47E-04 | 2,46 | 6-phospho-beta-glucosidase |
| WP_002304825.1 | complete genome | 7,27E-02 | 2,46 | helix-turn-helix domain-containing protein |
| WP_002293786.1 | complete genome | 7,72E-03 | 2,45 | ABC transporter permease |
| WP_002296598.1 | complete genome | 2,78E-02 | 2,44 | hypothetical protein |
| WP_002289649.1 | complete genome | 7,05E-03 | 2,44 | hypothetical protein |
| WP_002296606.1 | complete genome | 1,93E-02 | 2,44 | antitoxin |
| WP_014387056.1 | complete genome | 5,32E-03 | 2,44 | DNA cytosine methyltransferase |
| WP_002322093.1 | complete genome | 1,94E-02 | 2,44 | HNH endonuclease |
| WP_014386999.1 | complete genome | 6,78E-03 | 2,44 | N-acetylmuramoyl-L-alanine amidase |
| WP_002288981.1 | complete genome | 3,35E-03 | 2,44 | peptidase |
| WP_002288015.1 | complete genome | 8,17E-04 | 2,44 | phosphoribosylformylglycinamidine synthase II |
| WP_002288977.1 | complete genome | 7,31E-03 | 2,44 | hypothetical protein |
| WP_002296810.1 | complete genome | 1,31E-03 | 2,44 | beta-fructofuranosidase |
| WP_002347165.1 | complete genome | 1,49E-03 | 2,43 | hypothetical protein |
| WP_014387093.1 | complete genome | 6,20E-03 | 2,43 | PTS beta-glucoside IIBC component |
| WP_002311313.1 | complete genome | 2,89E-03 | 2,43 | tagatose-bisphosphate aldolase |
| WP_002347298.1 | complete genome | 1,38E-03 | 2,43 | GntR family transcriptional regulator |
| WP_002347048.1 | complete genome | 2,61E-02 | 2,43 | hypothetical protein |
| WP_002286711.1 | complete genome | 2,09E-03 | 2,42 | aminotransferase |
| WP_002289122.1_1117 | complete genome | 8,64E-05 | 2,42 | heavy metal translocating P-type ATPase |
| WP_002296683.1 | complete genome | 9,09E-03 | 2,42 | ammonium transporter |
| WP_002286525.1 | complete genome | 2,76E-03 | 2,42 | major capsid protein |
| WP_002296499.1 | complete genome | 2,50E-02 | 2,42 | hypothetical protein |
| WP_002287544.1 | complete genome | 1,78E-02 | 2,42 | malonate decarboxylase holo-[acyl-carrier-protein] synthase |
| WP_002296485.1 | complete genome | 1,02E-03 | 2,42 | phage major capsid protein |
| WP_002311723.1 | complete genome | 3,53E-02 | 2,42 | hypothetical protein |
| WP_002294839.1 | complete genome | 5,01E-03 | 2,42 | hypothetical protein |
| WP_014386994.1 | complete genome | 1,54E-02 | 2,42 | hypothetical protein |
| WP_002294365.1 | complete genome | 6,76E-03 | 2,42 | PTS system%2C cellobiose-specific IIC component |
| WP_002286012.1 | complete genome | 5,49E-03 | 2,41 | membrane protein |
| WP_002288159.1 | complete genome | 4,93E-05 | 2,41 | sugar ABC transporter substrate-binding protein |
| NC_017032.1-gene18 | plasmid Aus0004_p1 | 1,07E-02 | 2,41 | hypothetical protein |
| WP_002296679.1 | complete genome | 1,85E-04 | 2,41 | Asp23/Gls24 family envelope stress response protein |
| WP_002286788.1 | complete genome | 2,04E-04 | 2,41 | ABC transporter substrate-binding protein |
| WP_002304525.1 | complete genome | 5,31E-02 | 2,41 | hypothetical protein |
| WP_002286808.1 | complete genome | 1,19E-04 | 2,41 | DUF159 family protein |
| WP_002288275.1 | complete genome | 4,37E-03 | 2,41 | ring-cleaving dioxygenase |

| | | | | |
|---|---|---|---|---|
| **WP_002321172.1** | complete genome | 4,48E-02 | 2,41 | two-component system response regulator DcuR |
| **WP_002296606.1** | complete genome | 1,47E-02 | 2,41 | antitoxin |
| **WP_014387062.1** | complete genome | 1,47E-02 | 2,41 | ATP-dependent helicase |
| **WP_002296598.1** | complete genome | 6,13E-02 | 2,40 | hypothetical protein |
| **WP_002296720.1** | complete genome | 1,69E-02 | 2,40 | hypothetical protein |
| **WP_002286545.1** | complete genome | 1,22E-02 | 2,40 | hypothetical protein |
| **WP_002285961.1** | complete genome | 3,98E-03 | 2,40 | PTS cellobiose transporter subunit IIC |
| **WP_002288976.1** | complete genome | 1,34E-02 | 2,40 | cell wall anchor |
| **WP_010731016.1** | complete genome | 2,45E-02 | 2,40 | carbohydrate ABC transporter permease |
| **WP_002312920.1** | complete genome | 2,20E-02 | 2,39 | hypothetical protein |
| **WP_002286573.1** | complete genome | 5,68E-03 | 2,39 | hypothetical protein |
| **WP_002291914.1** | complete genome | 1,49E-05 | 2,39 | 2-dehydropantoate 2-reductase |
| **WP_002286524.1** | complete genome | 1,23E-02 | 2,39 | hypothetical protein |
| **WP_002289334.1** | complete genome | 9,00E-03 | 2,39 | ABC transporter permease |
| **WP_014387032.1** | complete genome | 3,42E-02 | 2,39 | recombinase |
| **WP_002297364.1** | complete genome | 2,41E-02 | 2,39 | hypothetical protein |
| **WP_002321468.1** | complete genome | 4,88E-03 | 2,39 | hypothetical protein |
| **WP_002288634.1** | complete genome | 1,18E-03 | 2,38 | PTS lactose transporter subunit IIBC |
| **WP_002301705.1** | complete genome | 3,28E-03 | 2,38 | transposase |
| **WP_002288247.1** | complete genome | 1,28E-02 | 2,38 | L-rhamnose mutarotase |
| **WP_002296096.1** | complete genome | 2,10E-02 | 2,38 | tetracycline resistance MFS efflux pump |
| **WP_002286924.1** | complete genome | 5,06E-03 | 2,38 | cell wall surface anchor protein |
| **WP_002286926.1** | complete genome | 2,38E-02 | 2,38 | DUF961 domain-containing protein |
| **WP_002345019.1** | complete genome | 1,36E-04 | 2,37 | tetracycline resistance protein tetM |
| **WP_002294496.1** | complete genome | 1,20E-03 | 2,37 | glycoside hydrolase family 28 |
| **WP_002288175.1** | complete genome | 8,07E-03 | 2,37 | mannitol dehydrogenase |
| **WP_002333679.1** | complete genome | 6,39E-02 | 2,37 | LPXTG-domain-containing protein cell wall anchor domain |
| **WP_002286053.1** | complete genome | 3,02E-03 | 2,37 | cell wall surface anchor protein |
| **WP_002320995.1** | complete genome | 4,03E-04 | 2,37 | relaxase |
| **WP_044031089.1** | complete genome | 9,81E-03 | 2,37 | S-adenosylmethionine synthetase |
| **WP_014387052.1** | complete genome | 8,36E-03 | 2,37 | amidoligase |
| **WP_002287965.1** | complete genome | 2,62E-02 | 2,37 | acyl-CoA thioester hydrolase |
| **WP_002297365.1** | complete genome | 1,04E-02 | 2,37 | transposase |
| **WP_002296500.1** | complete genome | 2,59E-02 | 2,37 | hypothetical protein |
| **WP_002286061.1** | complete genome | 6,18E-04 | 2,37 | ABC transporter permease |
| **WP_014387010.1** | complete genome | 2,44E-02 | 2,36 | sugar ABC transporter permease |
| **WP_002287256.1** | complete genome | 7,16E-04 | 2,36 | mannitol operon transcriptional antiterminator |
| **WP_002296189.1** | complete genome | 3,07E-03 | 2,36 | transcriptional regulator |
| **WP_002296258.1** | complete genome | 4,14E-02 | 2,36 | hypothetical protein |
| **WP_002289516.1** | complete genome | 5,67E-03 | 2,35 | beta-galactosidase |
| **WP_077495183.1** | complete genome | 4,90E-02 | 2,35 | hypothetical protein |
| **WP_002286925.1** | complete genome | 2,44E-02 | 2,34 | hypothetical protein |
| **WP_002287754.1** | complete genome | 2,55E-02 | 2,34 | sugar ABC transporter ATP-binding protein |
| **WP_002287269.1** | complete genome | 3,88E-03 | 2,34 | TetR family transcriptional regulator |
| **NC_017032.1-gene26** | plasmid Aus0004_p1 | 5,82E-03 | 2,34 | peptidase M23 |
| **WP_002294167.1** | complete genome | 8,16E-04 | 2,34 | gluconate 5-dehydrogenase |
| **WP_002340435.1** | complete genome | 7,93E-02 | 2,34 | hypothetical protein |
| **WP_002288858.1** | complete genome | 1,10E-03 | 2,34 | PTS cellobiose transporter subunit IIC |
| **WP_002287546.1** | complete genome | 9,69E-03 | 2,34 | biotin-independent malonate decarboxylase subunit beta |
| **WP_014387055.1** | complete genome | 2,61E-02 | 2,34 | hypothetical protein |
| **WP_002287297.1** | complete genome | 3,42E-03 | 2,34 | alpha-N-arabinofuranosidase |
| **WP_002296199.1** | complete genome | 8,82E-04 | 2,33 | hypothetical protein |
| **WP_002311665.1** | complete genome | 3,54E-02 | 2,33 | transposase |
| **WP_014386999.1** | complete genome | 6,83E-03 | 2,33 | N-acetylmuramoyl-L-alanine amidase |
| **WP_002288248.1** | complete genome | 1,21E-02 | 2,33 | rhamnulose-1-phosphate aldolase |
| **WP_002300930.1** | complete genome | 7,41E-02 | 2,33 | hypothetical protein |
| **WP_002293650.1** | complete genome | 1,55E-02 | 2,33 | portal protein |

| | | | | |
|---|---|---|---|---|
| **WP_002296629.1** | complete genome | 5,98E-03 | 2,33 | peptidase |
| **WP_002286542.1** | complete genome | 1,23E-02 | 2,33 | autolysin |
| **WP_002289315.1** | complete genome | 1,80E-02 | 2,32 | hypothetical protein |
| **WP_002287579.1** | complete genome | 2,27E-02 | 2,32 | PTS lactose/cellobiose transporter subunit IIA |
| **WP_002296295.1** | complete genome | 2,87E-02 | 2,32 | hypothetical protein |
| **WP_002286598.1** | complete genome | 1,96E-02 | 2,32 | hypothetical protein |
| **WP_002311595.1** | complete genome | 2,14E-02 | 2,32 | hypothetical protein |
| **WP_002286533.1** | complete genome | 3,12E-03 | 2,32 | terminase |
| **WP_002286533.1** | complete genome | 2,57E-03 | 2,32 | terminase |
| **WP_014386998.1** | complete genome | 6,72E-02 | 2,32 | holin |
| **WP_002288017.1** | complete genome | 7,98E-03 | 2,32 | phosphoribosylformylglycinamidine synthase I |
| **WP_002286787.1** | complete genome | 7,65E-03 | 2,32 | sugar ABC transporter permease |
| **WP_002286695.1** | complete genome | 4,48E-02 | 2,32 | hypothetical protein |
| **WP_002287577.1** | complete genome | 1,18E-02 | 2,31 | PTS lactose transporter subunit IIC |
| **WP_002296526.1** | complete genome | 3,85E-02 | 2,31 | MerR family transcriptional regulator |
| **WP_002303476.1** | complete genome | 1,72E-02 | 2,31 | hypothetical protein |
| **WP_002294811.1** | complete genome | 4,53E-04 | 2,31 | FAD:protein FMN transferase |
| **WP_014386996.1** | complete genome | 8,60E-04 | 2,31 | hypothetical protein |
| **WP_002286776.1** | complete genome | 5,69E-03 | 2,31 | peptidase |
| **WP_002286525.1** | complete genome | 5,20E-03 | 2,31 | major capsid protein |
| **WP_002311319.1** | complete genome | 8,69E-03 | 2,30 | PTS fructose transporter subunit IID |
| **WP_002289214.1** | complete genome | 7,53E-02 | 2,30 | PTS fructose transporter subunit IIA |
| **WP_002286604.1** | complete genome | 1,58E-02 | 2,30 | PTS lactose/cellobiose transporter subunit IIA |
| **WP_044031432.1** | complete genome | 4,99E-02 | 2,29 | toxin secretion/phage lysis holin |
| **WP_002286016.1** | complete genome | 4,53E-03 | 2,29 | hypothetical protein |
| **WP_002286694.1** | complete genome | 3,39E-02 | 2,29 | hypothetical protein |
| **WP_002286553.1** | complete genome | 6,77E-02 | 2,29 | hypothetical protein |
| **NC_017032.1-gene28** | plasmid Aus0004_p1 | 3,72E-03 | 2,29 | hypothetical protein |
| **WP_002286375.1** | complete genome | 1,23E-02 | 2,29 | beta-phosphoglucomutase |
| **WP_002293788.1** | complete genome | 3,83E-03 | 2,28 | amidohydrolase |
| **WP_002296185.1** | complete genome | 2,02E-03 | 2,28 | methionine ABC transporter substrate-binding protein |
| **WP_002288256.1** | complete genome | 8,04E-03 | 2,28 | membrane protein |
| **WP_002289801.1** | complete genome | 5,40E-02 | 2,28 | malate permease |
| **WP_002296167.1** | complete genome | 1,34E-02 | 2,27 | ABC transporter permease |
| **WP_014387092.1** | complete genome | 2,85E-02 | 2,27 | PTS sugar transporter subunit IIC |
| **WP_002286695.1** | complete genome | 1,55E-02 | 2,27 | hypothetical protein |
| **WP_002301068.1** | complete genome | 5,38E-02 | 2,27 | hypothetical protein |
| **WP_002311663.1** | complete genome | 7,64E-03 | 2,27 | ABC transporter ATP-binding protein |
| **WP_002288979.1** | complete genome | 4,00E-02 | 2,27 | cell wall anchor |
| **WP_002287830.1** | complete genome | 4,37E-02 | 2,26 | hypothetical protein |
| **WP_002296147.1** | complete genome | 3,46E-03 | 2,26 | 6-phospho-beta-glucosidase |
| **WP_002311310.1** | complete genome | 6,70E-03 | 2,26 | tagatose-bisphosphate aldolase |
| **WP_033700366.1** | complete genome | 6,36E-02 | 2,26 | cell wall anchor protein |
| **WP_002321810.1** | complete genome | 5,14E-02 | 2,26 | DNA methyltransferase |
| **WP_002286557.1** | complete genome | 2,45E-02 | 2,26 | hypothetical protein |
| **WP_002320996.1** | complete genome | 2,86E-02 | 2,26 | hypothetical protein |
| **WP_002302892.1** | complete genome | 1,80E-03 | 2,26 | transcription antiterminator BglG |
| **NC_017032.1-gene43** | plasmid Aus0004_p1 | 1,00E-02 | 2,26 | hypothetical protein |
| **WP_002296631.1** | complete genome | 2,59E-02 | 2,25 | hypothetical protein |
| **WP_002286361.1** | complete genome | 3,35E-02 | 2,25 | acetyltransferase |
| **WP_002285972.1** | complete genome | 6,76E-03 | 2,25 | carbohydrate ABC transporter permease |
| **WP_002289025.1** | complete genome | 1,89E-04 | 2,25 | FMN-binding domain-containing protein |
| **NC_017032.1-gene44** | plasmid Aus0004_p1 | 2,98E-03 | 2,25 | hypothetical protein |
| **WP_002340433.1** | complete genome | 3,96E-03 | 2,25 | hypothetical protein |
| **WP_002286054.1** | complete genome | 8,90E-03 | 2,25 | peptidase |
| **NC_017032.1-gene21** | plasmid Aus0004_p1 | 1,38E-03 | 2,25 | hypothetical protein |

| | | | | |
|---|---|---|---|---|
| WP_002289422.1 | complete genome | 6,19E-02 | 2,25 | DeoR family transcriptional regulator |
| WP_002322120.1 | complete genome | 2,70E-03 | 2,24 | DNA internalization-related competence protein ComEC/Rec2 |
| WP_002321407.1 | complete genome | 2,55E-02 | 2,24 | transcriptional regulator |
| WP_002321406.1 | complete genome | 4,83E-02 | 2,24 | DNA-binding protein |
| WP_014386996.1 | complete genome | 2,88E-03 | 2,24 | hypothetical protein |
| WP_014386995.1 | complete genome | 6,11E-03 | 2,24 | phage minor structural protein%2C region |
| WP_002319736.1 | complete genome | 5,89E-02 | 2,24 | transcriptional regulator |
| WP_002296472.1 | complete genome | 2,33E-02 | 2,24 | WxL domain-containing protein |
| WP_002293649.1 | complete genome | 1,21E-02 | 2,24 | universal stress protein |
| WP_002302663.1 | complete genome | 3,55E-03 | 2,23 | amino acid ABC transporter permease |
| WP_002286782.1 | complete genome | 2,92E-02 | 2,23 | phosphatase PAP2 family protein |
| WP_002304157.1 | complete genome | 8,41E-05 | 2,23 | PTS mannose transporter subunit IIA |
| WP_002288250.1 | complete genome | 9,51E-03 | 2,23 | rhamnulokinase |
| WP_002296334.1 | complete genome | 1,07E-02 | 2,23 | membrane protein |
| WP_002296814.1 | complete genome | 4,00E-03 | 2,22 | beta-fructofuranosidase |
| WP_002297368.1 | complete genome | 6,50E-03 | 2,22 | cell wall surface anchor protein |
| WP_002287661.1 | complete genome | 4,18E-02 | 2,22 | hypothetical protein |
| WP_002287286.1 | complete genome | 3,47E-03 | 2,22 | ATPase |
| WP_002296599.1 | complete genome | 4,96E-02 | 2,22 | hypothetical protein |
| WP_002297352.1 | complete genome | 5,66E-02 | 2,22 | DUF3789 domain-containing protein |
| WP_002286552.1 | complete genome | 3,97E-02 | 2,21 | helix-turn-helix domain-containing protein |
| WP_002289055.1 | complete genome | 2,11E-02 | 2,21 | conjugal transfer protein |
| WP_002286522.1 | complete genome | 5,79E-02 | 2,21 | head-tail adaptor protein |
| WP_002288160.1 | complete genome | 3,96E-03 | 2,21 | alpha-glycosidase |
| WP_014387124.1 | complete genome | 4,74E-03 | 2,21 | hypothetical protein |
| WP_002346967.1 | complete genome | 1,13E-02 | 2,21 | hypothetical protein |
| WP_002288384.1 | complete genome | 3,08E-04 | 2,21 | pyruvate oxidase |
| WP_002311317.1 | complete genome | 2,82E-02 | 2,21 | PTS sugar transporter subunit IIC |
| WP_002305377.1 | complete genome | 5,93E-02 | 2,20 | hypothetical protein |
| WP_002286553.1 | complete genome | 4,28E-02 | 2,20 | hypothetical protein |
| WP_002321044.1 | complete genome | 6,24E-02 | 2,20 | membrane protein |
| NC_017032.1-gene42 | plasmid Aus0004_p1 | 3,28E-02 | 2,20 | hypothetical protein |
| WP_002294831.1 | complete genome | 1,54E-02 | 2,20 | N-acetylmuramoyl-L-alanine amidase |
| WP_002286275.1 | complete genome | 1,44E-03 | 2,20 | ATP-binding protein |
| WP_002286015.1 | complete genome | 6,82E-03 | 2,20 | sugar ABC transporter substrate-binding protein |
| WP_002295879.1 | complete genome | 1,54E-02 | 2,20 | MFS transporter |
| WP_080019252.1 | complete genome | 2,59E-03 | 2,20 | IS5/IS1182 family transposase |
| WP_002303842.1 | complete genome | 6,17E-03 | 2,20 | hypothetical protein |
| WP_002317395.1 | complete genome | 1,86E-02 | 2,19 | competence protein ComYD |
| WP_002296338.1 | complete genome | 1,38E-02 | 2,19 | hydrolase |
| WP_014386994.1 | complete genome | 5,42E-02 | 2,19 | hypothetical protein |
| WP_002296173.1 | complete genome | 3,13E-03 | 2,19 | DUF871 domain-containing protein |
| WP_002289057.1 | complete genome | 1,42E-02 | 2,18 | lipoprotein |
| NC_017032.1-gene20 | plasmid Aus0004_p1 | 2,05E-03 | 2,18 | hypothetical protein |
| WP_002286055.1 | complete genome | 8,83E-03 | 2,18 | VWA domain-containing protein |
| WP_002289668.1 | complete genome | 5,76E-03 | 2,18 | melibiose carrier protein |
| WP_002286533.1 | complete genome | 1,23E-02 | 2,18 | terminase |
| NC_017032.1-gene9 | plasmid Aus0004_p1 | 4,11E-03 | 2,18 | cell wall surface anchor family protein |
| WP_002320964.1 | complete genome | 2,20E-02 | 2,18 | ISL3 family transposase |
| WP_044031087.1 | complete genome | 4,99E-02 | 2,18 | lactate dehydrogenase |
| WP_002296600.1 | complete genome | 8,13E-02 | 2,18 | thymidylate synthase |
| WP_002287753.1 | complete genome | 4,89E-03 | 2,18 | glycosyl hydrolase family 88 |
| WP_002321677.1 | complete genome | 4,33E-02 | 2,18 | alcohol dehydrogenase |
| WP_014387031.1 | complete genome | 2,46E-02 | 2,18 | recombinase family protein |
| WP_002287757.1 | complete genome | 1,03E-02 | 2,17 | ABC transporter substrate-binding protein |
| WP_002296194.1 | complete genome | 5,43E-03 | 2,17 | PTS beta-glucoside transporter subunit EIIBCA |

| | | | | |
|---|---|---|---|---|
| **WP_002286550.1** | complete genome | 2,92E-02 | 2,17 | hypothetical protein |
| **WP_014387125.1** | complete genome | 8,04E-03 | 2,17 | phage minor structural protein |
| **WP_002286367.1** | complete genome | 5,53E-03 | 2,17 | phosphatase |
| **WP_002287267.1** | complete genome | 2,27E-03 | 2,17 | ABC transporter ATP-binding protein |
| **NC_017032.1-gene46** | plasmid Aus0004_p1 | 6,92E-02 | 2,17 | hypothetical protein |
| **WP_002296476.1** | complete genome | 2,10E-02 | 2,16 | NAD-dependent malic enzyme |
| **WP_002288157.1** | complete genome | 1,07E-02 | 2,16 | sugar ABC transporter permease |
| **WP_002295743.1** | complete genome | 1,53E-03 | 2,16 | IS982 family transposase |
| **WP_002286538.1** | complete genome | 7,00E-02 | 2,16 | terminase |
| **WP_014387131.1** | complete genome | 7,18E-02 | 2,16 | hypothetical protein |
| **WP_014387042.1** | complete genome | 3,71E-02 | 2,16 | phage major tail protein |
| **WP_002305373.1** | complete genome | 5,09E-02 | 2,15 | phage tail protein |
| **WP_014387072.1** | complete genome | 1,42E-02 | 2,15 | DNA polymerase I |
| **NC_017032.1-gene30** | plasmid Aus0004_p1 | 4,86E-02 | 2,15 | hypothetical protein |
| **WP_002296825.1** | complete genome | 1,54E-03 | 2,15 | hypothetical protein |
| **WP_002296599.1** | complete genome | 5,94E-02 | 2,15 | hypothetical protein |
| **WP_002289216.1** | complete genome | 4,31E-02 | 2,15 | PTS trehalose transporter subunit IIBC |
| **WP_002293573.1** | complete genome | 3,17E-02 | 2,15 | ABC transporter permease |
| **WP_002296186.1** | complete genome | 1,74E-02 | 2,14 | methionine ABC transporter ATP-binding protein |
| **WP_002296677.1** | complete genome | 1,33E-03 | 2,14 | hypothetical protein |
| **WP_071974492.1** | complete genome | 7,48E-02 | 2,14 | hypothetical protein |
| **WP_002296486.1** | complete genome | 2,07E-02 | 2,14 | portal protein |
| **WP_014387041.1** | complete genome | 2,55E-02 | 2,14 | phage tail protein |
| **WP_002286007.1** | complete genome | 5,81E-03 | 2,14 | AraC family transcriptional regulator |
| **WP_002286021.1** | complete genome | 2,87E-02 | 2,13 | endopolygalacturonase |
| **WP_014387063.1** | complete genome | 7,68E-02 | 2,13 | VRR-NUC domain-containing protein |
| **WP_002297028.1** | complete genome | 7,04E-02 | 2,13 | hypothetical protein |
| **WP_002286774.1** | complete genome | 6,04E-03 | 2,13 | peptidase |
| **NC_017032.1-gene32** | plasmid AUS0004_p1 | 1,03E-02 | 2,13 | hypothetical protein |
| **WP_002288541.1** | complete genome | 7,04E-03 | 2,13 | hypothetical protein |
| **WP_002296157.1** | complete genome | 2,79E-02 | 2,13 | branched-chain amino acid transporter II carrier protein |
| **WP_002288357.1** | complete genome | 2,02E-02 | 2,13 | site-specific integrase |
| **WP_077828743.1** | complete genome | 7,82E-02 | 2,13 | hypothetical protein |
| **WP_002286940.1** | complete genome | 6,83E-03 | 2,13 | group II intron reverse transcriptase/maturase |
| **WP_002296489.1** | complete genome | 4,29E-02 | 2,13 | DUF2292 domain-containing protein |
| **NC_017032.1-gene27** | plasmid Aus0004_p1 | 1,29E-02 | 2,12 | hypothetical protein |
| **WP_002288643.1** | complete genome | 5,93E-02 | 2,12 | hypothetical protein |
| **WP_002293405.1** | complete genome | 4,50E-03 | 2,12 | choline ABC transporter permease |
| **WP_002289745.1** | complete genome | 3,27E-03 | 2,12 | 50S ribosomal protein L25 |
| **WP_002286683.1** | complete genome | 4,10E-02 | 2,12 | holin |
| **WP_002297218.1** | complete genome | 8,62E-03 | 2,12 | ISL3 family transposase |
| **NC_017032.1-gene34** | plasmid Aus0004_p1 | 1,48E-02 | 2,12 | hypothetical protein |
| **WP_002288815.1** | complete genome | 5,68E-02 | 2,12 | sugar permease |
| **WP_002347278.1** | complete genome | 3,06E-02 | 2,12 | 2-aminoethylphosphonate--pyruvate transaminase |
| **WP_002296149.1** | complete genome | 3,53E-03 | 2,12 | 6-phospho-beta-glucosidase |
| **WP_002289275.1** | complete genome | 9,43E-03 | 2,12 | hypothetical protein |
| **WP_002287549.1** | complete genome | 9,02E-03 | 2,11 | malonate decarboxylase subunit alpha |
| **WP_009269250.1** | complete genome | 6,14E-02 | 2,11 | hypothetical protein |
| **WP_002289471.1** | complete genome | 7,54E-02 | 2,11 | transcriptional regulator |
| **NC_017032.1-gene17** | plasmid Aus0004_p1 | 6,39E-02 | 2,11 | hypothetical protein |
| **WP_002286930.1** | complete genome | 2,10E-02 | 2,11 | cell division protein FtsK |
| **WP_002298781.1** | complete genome | 2,55E-02 | 2,11 | HAD family phosphatase |

| | | | | |
|---|---|---|---|---|
| **WP_010730282.1** | complete genome | 4,69E-02 | 2,11 | IS4/IS5 family transposase |
| **WP_002296967.1** | complete genome | 8,50E-03 | 2,10 | 6-phospho-beta-glucosidase |
| **WP_001009054.1** | complete genome | 8,34E-02 | 2,10 | hypothetical protein |
| **WP_002286542.1** | complete genome | 3,93E-02 | 2,10 | autolysin |
| **NC_017032.1-gene11** | plasmid Aus0004_p1 | 1,77E-02 | 2,10 | hypothetical protein |
| **WP_002311312.1** | complete genome | 1,00E-02 | 2,10 | tagatose-6-phosphate ketose |
| **WP_002296623.1** | complete genome | 2,62E-02 | 2,10 | ISL3 family transposase |
| **WP_073466279.1** | complete genome | 4,99E-03 | 2,10 | DDE transposase |
| **WP_002286060.1** | complete genome | 3,91E-02 | 2,10 | peptide ABC transporter ATP-binding protein |
| **WP_002294551.1** | complete genome | 1,27E-02 | 2,09 | ABC transporter permease |
| **WP_002287558.1** | complete genome | 2,09E-02 | 2,09 | amino acid permease |
| **WP_002288239.1** | complete genome | 9,75E-04 | 2,09 | 5-keto-4-deoxyuronate isomerase |
| **WP_002293681.1** | complete genome | 9,62E-03 | 2,09 | MFS transporter |
| **NC_017032.1-gene10** | plasmid Aus0004_p1 | 2,32E-02 | 2,09 | class C sortase |
| **WP_002296599.1** | complete genome | 7,79E-02 | 2,09 | hypothetical protein |
| **WP_002294687.1** | complete genome | 2,31E-03 | 2,09 | hypothetical protein |
| **WP_002299428.1** | complete genome | 2,17E-02 | 2,08 | amidophosphoribosyltransferase |
| **WP_002288135.1** | complete genome | 2,35E-02 | 2,08 | sugar ABC transporter substrate-binding protein |
| **WP_002297366.1** | complete genome | 3,70E-02 | 2,08 | transposase |
| **WP_002287536.1** | complete genome | 1,71E-02 | 2,08 | transcription antiterminator BglG |
| **WP_002296705.1** | complete genome | 8,01E-02 | 2,08 | AraC family transcriptional regulator |
| **WP_002295923.1** | complete genome | 5,42E-02 | 2,08 | maltodextrose utilization protein MalA |
| **WP_002286181.1** | complete genome | 2,35E-02 | 2,08 | secretion system protein E |
| **WP_002287128.1** | complete genome | 3,60E-02 | 2,08 | multidrug ABC transporter ATP-binding protein |
| **WP_002294893.1** | complete genome | 4,17E-02 | 2,08 | PTS fructose transporter subunit IIA |
| **WP_002303475.1** | complete genome | 7,41E-02 | 2,07 | XkdX family protein |
| **WP_002289037.1** | complete genome | 3,61E-02 | 2,07 | transcriptional regulator |
| **WP_002287287.1** | complete genome | 2,68E-02 | 2,07 | MFS transporter |
| **WP_002288240.1** | complete genome | 6,57E-03 | 2,07 | sugar kinase |
| **WP_002296776.1** | complete genome | 8,15E-03 | 2,07 | beta-glucosidase |
| **WP_002304701.1** | complete genome | 7,63E-02 | 2,07 | hypothetical protein |
| **WP_002286024.1** | complete genome | 1,85E-02 | 2,07 | pectinesterase |
| **WP_002287759.1** | complete genome | 4,93E-03 | 2,07 | CoA-disulfide reductase |
| **WP_002296141.1** | complete genome | 5,16E-03 | 2,07 | ABC transporter ATP-binding protein |
| **WP_002347297.1** | complete genome | 5,30E-02 | 2,06 | MurR/RpiR family transcriptional regulator |
| **WP_002286000.1** | complete genome | 1,98E-02 | 2,06 | uronate isomerase |
| **WP_002285995.1** | complete genome | 1,47E-02 | 2,06 | ABC transporter ATP-binding protein |
| **WP_002298394.1** | complete genome | 8,26E-02 | 2,06 | hypothetical protein |
| **NC_017032.1-gene13** | plasmid Aus0004_p1 | 1,38E-02 | 2,06 | isopeptide-forming domain-containing fimbrial protein |
| **WP_002288348.1** | complete genome | 1,35E-03 | 2,06 | ATP-dependent chaperone ClpB |
| **WP_002296449.1** | complete genome | 2,25E-02 | 2,06 | ornithine cyclodeaminase |
| **WP_002293448.1** | complete genome | 1,59E-02 | 2,05 | hypothetical protein |
| **WP_002304270.1** | complete genome | 3,49E-02 | 2,05 | fructokinase |
| **WP_002352859.1** | complete genome | 1,40E-02 | 2,05 | 6-phospho-beta-glucosidase |
| **WP_002287021.1** | complete genome | 4,92E-02 | 2,04 | glycoside hydrolase |
| **WP_002296509.1** | complete genome | 7,26E-03 | 2,04 | ABC transporter |
| **WP_002295743.1** | complete genome | 1,11E-02 | 2,04 | IS982 family transposase |
| **WP_002289134.1** | complete genome | 2,70E-02 | 2,03 | response regulator |
| **WP_010815683.1** | complete genome | 4,29E-02 | 2,03 | glycopeptide resistance accessory protein VanW-B |
| **WP_002286067.1** | complete genome | 1,73E-03 | 2,03 | acetate--CoA ligase |
| **WP_002288005.1** | complete genome | 8,73E-04 | 2,03 | aldo/keto reductase |
| **WP_080019247.1** | complete genome | 1,83E-02 | 2,03 | IS5/IS1182 family transposase |
| **WP_002296493.1** | complete genome | 2,50E-02 | 2,03 | helicase |
| **WP_002297356.1** | complete genome | 4,29E-02 | 2,03 | DNA methylase |
| **WP_002288137.1** | complete genome | 2,36E-02 | 2,02 | sucrose phosphorylase |
| **WP_014387129.1** | complete genome | 5,75E-02 | 2,02 | hypothetical protein |

| | | | | |
|---|---|---|---|---|
| **WP_002288559.1** | complete genome | 2,05E-02 | 2,02 | WxL domain-containing protein |
| **WP_002286721.1** | complete genome | 2,38E-02 | 2,02 | arginine-ornithine antiporter |
| **WP_002327793.1** | complete genome | 6,75E-02 | 2,02 | IS30 family transposase |
| **WP_002297227.1** | complete genome | 2,33E-02 | 2,02 | GCN5 family acetyltransferase |
| **WP_002286749.1** | complete genome | 1,36E-02 | 2,02 | peptidase |
| **WP_000584387.1** | complete genome | 4,27E-02 | 2,02 | conjugal transfer protein |
| **WP_002321715.1** | complete genome | 5,34E-02 | 2,02 | hypothetical protein |
| **WP_002288986.1** | complete genome | 3,98E-02 | 2,02 | MerR family transcriptional regulator |
| **WP_002287575.1** | complete genome | 5,33E-02 | 2,02 | N-acetyltransferase |
| **WP_002288586.1** | complete genome | 2,49E-02 | 2,02 | hypothetical protein |
| **WP_002286002.1** | complete genome | 3,88E-02 | 2,02 | altronate hydrolase |
| **WP_002289264.1** | complete genome | 2,94E-03 | 2,02 | pyruvate formate lyase-activating protein |
| **WP_002287550.1** | complete genome | 3,12E-02 | 2,02 | ACP S-malonyltransferase |
| **WP_002347052.1** | complete genome | 4,76E-02 | 2,02 | phage protein |
| **WP_002288127.1** | complete genome | 4,42E-02 | 2,01 | dehydrogenase |
| **WP_002286368.1** | complete genome | 5,36E-03 | 2,01 | PTS glucose transporter subunit IIABC |
| **WP_002288119.1** | complete genome | 1,24E-02 | 2,01 | peptide-methionine (S)-S-oxide reductase |
| **WP_002295114.1** | complete genome | 1,41E-02 | 2,01 | isomerase |
| **WP_002288393.1** | complete genome | 1,57E-02 | 2,00 | hemolysin III |
| **WP_002297010.1** | complete genome | 2,58E-02 | 2,00 | ABC transporter ATP-binding protein |
| **WP_002295260.1** | complete genome | 2,91E-02 | 2,00 | hypothetical protein |
| **WP_002296055.1** | complete genome | 1,10E-16 | -2,01 | lysine--tRNA ligase |
| **WP_002290449.1** | complete genome | 1,28E-12 | -2,01 | 50S ribosomal protein L23 |
| **WP_033582077.1** | complete genome | 2,12E-15 | -2,03 | ribosome silencing factor RsfS |
| **WP_002287686.1** | complete genome | 9,40E-16 | -2,04 | Fe-S cluster assembly protein SufB |
| **WP_002303833.1** | complete genome | 1,59E-18 | -2,04 | penicillin-binding protein 4 |
| **WP_002290446.1** | complete genome | 2,59E-16 | -2,05 | 50S ribosomal protein L4 |
| **WP_002288713.1** | complete genome | 2,47E-14 | -2,06 | 30S ribosomal protein S13 |
| **WP_002297292.1** | complete genome | 2,70E-11 | -2,08 | manganese ABC transporter ATP-binding protein |
| **WP_002288688.1** | complete genome | 6,65E-19 | -2,09 | preprotein translocase subunit SecY |
| **WP_002288659.1** | complete genome | 9,23E-18 | -2,10 | 30S ribosomal protein S3 |
| **WP_002289282.1** | complete genome | 1,93E-16 | -2,11 | 30S ribosomal protein S15 |
| **WP_002288679.1** | complete genome | 5,50E-16 | -2,11 | 50S ribosomal protein L18 |
| **WP_002296753.1** | complete genome | 1,29E-12 | -2,12 | bacteriocin |
| **WP_002290442.1** | complete genome | 1,12E-20 | -2,13 | translation elongation factor Tu |
| **WP_002288710.1** | complete genome | 1,63E-17 | -2,13 | 50S ribosomal protein L36 |
| **WP_002288661.1** | complete genome | 2,11E-14 | -2,14 | 50S ribosomal protein L16 |
| **WP_002290440.1** | complete genome | 3,57E-14 | -2,15 | 30S ribosomal protein S7 |
| **WP_002289452.1** | complete genome | 2,56E-13 | -2,16 | 50S ribosomal protein L27 |
| **WP_002286302.1** | complete genome | 3,65E-15 | -2,17 | ABC transporter permease |
| **WP_002317276.1** | complete genome | 2,75E-14 | -2,17 | DNA-binding protein |
| **WP_002288669.1** | complete genome | 1,56E-15 | -2,17 | 50S ribosomal protein L14 |
| **WP_002288687.1** | complete genome | 3,52E-19 | -2,19 | 50S ribosomal protein L15 |
| **WP_002288673.1** | complete genome | 4,16E-17 | -2,20 | 30S ribosomal protein S14 |
| **WP_002295352.1** | complete genome | 7,26E-20 | -2,21 | hypothetical protein |
| **WP_002295907.1** | complete genome | 2,43E-12 | -2,21 | hypothetical protein |
| **WP_002287463.1** | complete genome | 2,30E-18 | -2,22 | mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase |
| **WP_002296565.1** | complete genome | 1,62E-22 | -2,23 | tRNA preQ1(34) S-adenosylmethionine ribosyltransferase-isomerase QueA |
| **WP_002304026.1** | complete genome | 1,06E-16 | -2,23 | hypothetical protein |
| **WP_002290445.1** | complete genome | 4,28E-17 | -2,24 | 50S ribosomal protein L3 |
| **WP_002288972.1** | complete genome | 6,29E-02 | -2,24 | CAAX amino protease |
| **WP_002288533.1** | complete genome | 1,27E-16 | -2,25 | asparagine synthetase B |
| **WP_002296517.1** | complete genome | 1,37E-15 | -2,27 | aspartate racemase |
| **WP_002288491.1** | complete genome | 2,47E-14 | -2,27 | GTP-binding protein |
| **WP_079158009.1** | complete genome | 3,51E-10 | -2,30 | hypothetical protein |
| **WP_002289451.1** | complete genome | 3,47E-15 | -2,33 | hypothetical protein |
| **WP_002288368.1** | complete genome | 9,40E-16 | -2,34 | single-stranded DNA-binding protein |
| **WP_002287763.1** | complete genome | 2,84E-10 | -2,36 | hypothetical protein |

| | | | | |
|---|---|---|---|---|
| **WP_002288668.1** | complete genome | 1,84E-22 | -2,37 | 30S ribosomal protein S17 |
| **WP_002289439.1** | complete genome | 1,48E-15 | -2,37 | S-ribosylhomocysteine lyase |
| **WP_002298578.1** | complete genome | 5,46E-21 | -2,37 | GMP synthase (glutamine-hydrolyzing) |
| **WP_002286400.1** | complete genome | 1,32E-19 | -2,38 | tyrosine decarboxylase |
| **WP_002287684.1** | complete genome | 4,46E-17 | -2,39 | iron-sulfur cluster assembly scaffold protein |
| **WP_002288675.1** | complete genome | 2,01E-21 | -2,40 | 30S ribosomal protein S8 |
| **WP_002286075.1** | complete genome | 1,75E-14 | -2,40 | DegV family protein |
| **WP_002296444.1** | complete genome | 3,80E-18 | -2,41 | inosine-uridine preferring nucleoside hydrolase |
| **WP_002286769.1** | complete genome | 8,05E-20 | -2,41 | glucose transporter GlcU |
| **WP_002288681.1** | complete genome | 1,56E-18 | -2,41 | 30S ribosomal protein S5 |
| **WP_002287639.1** | complete genome | 9,83E-21 | -2,42 | SAM-dependent methyltransferase |
| **WP_002288686.1** | complete genome | 3,23E-17 | -2,43 | 50S ribosomal protein L30 |
| **WP_002290463.1** | complete genome | 4,87E-18 | -2,44 | 50S ribosomal protein L28 |
| **WP_002285906.1** | complete genome | 7,74E-17 | -2,45 | rRNA methyltransferase |
| **WP_002289717.1** | complete genome | 1,45E-17 | -2,46 | L-lactate dehydrogenase |
| **WP_002293607.1** | complete genome | 2,51E-14 | -2,46 | nucleoside deoxyribosyltransferase |
| **WP_002292188.1** | complete genome | 2,72E-14 | -2,47 | 30S ribosomal protein S20 |
| **WP_002287321.1** | complete genome | 1,19E-20 | -2,47 | 30S ribosomal protein S21 |
| **WP_002294265.1** | complete genome | 3,77E-16 | -2,50 | 30S ribosomal protein S4 |
| **WP_002289450.1** | complete genome | 1,30E-22 | -2,51 | 50S ribosomal protein L21 |
| **WP_002288251.1** | complete genome | 1,30E-14 | -2,51 | AraC family transcriptional regulator |
| **WP_002294320.1** | complete genome | 1,15E-18 | -2,53 | resolvase |
| **WP_002289509.1** | complete genome | 1,75E-15 | -2,53 | folate ECF transporter |
| **WP_002289309.1** | complete genome | 8,35E-15 | -2,54 | 50S ribosomal protein L35 |
| **WP_002320870.1** | complete genome | 1,50E-23 | -2,55 | 30S ribosomal protein S10 |
| **WP_002296138.1** | complete genome | 1,56E-18 | -2,55 | divalent metal cation transporter |
| **WP_002288670.1** | complete genome | 4,09E-22 | -2,57 | 50S ribosomal protein L24 |
| **WP_002294057.1** | complete genome | 7,42E-24 | -2,58 | fructose-1%2C6-bisphosphate aldolase%2C class II |
| **WP_002287318.1** | complete genome | 1,22E-21 | -2,59 | hypothetical protein |
| **WP_002289272.1** | complete genome | 1,48E-15 | -2,60 | hypothetical protein |
| **WP_002288370.1** | complete genome | 1,03E-21 | -2,60 | 30S ribosomal protein S18 |
| **WP_002287272.1** | complete genome | 3,88E-22 | -2,61 | FMN-dependent NADH-azoreductase |
| **WP_002288664.1** | complete genome | 1,08E-19 | -2,62 | 50S ribosomal protein L29 |
| **WP_002288714.1** | complete genome | 1,48E-25 | -2,64 | 30S ribosomal protein S11 |
| **WP_002290330.1** | complete genome | 1,87E-21 | -2,65 | 50S ribosomal protein L31 |
| **WP_002288657.1** | complete genome | 3,99E-24 | -2,68 | 50S ribosomal protein L22 |
| **WP_002294067.1** | complete genome | 1,48E-16 | -2,72 | cold-shock protein |
| **WP_002286049.1** | complete genome | 7,86E-22 | -2,74 | hypothetical protein |
| **WP_002290274.1** | complete genome | 4,88E-15 | -2,75 | 30S ribosomal protein S16 |
| **WP_002286306.1** | complete genome | 6,25E-22 | -2,76 | amino acid ABC transporter ATP-binding protein |
| **WP_002290277.1** | complete genome | 1,99E-07 | -2,76 | RNA-binding protein |
| **WP_002290970.1** | complete genome | 3,23E-17 | -2,77 | 50S ribosomal protein L32 |
| **WP_002294587.1** | complete genome | 9,08E-28 | -2,83 | 50S ribosomal protein L1 |
| **WP_002288092.1** | complete genome | 2,36E-19 | -2,84 | D-alanine--poly(phosphoribitol) ligase subunit 2 |
| **WP_002288113.1** | complete genome | 3,76E-18 | -2,86 | teichoic acid D-Ala incorporation-associated protein DltX |
| **WP_002291638.1** | complete genome | 1,84E-21 | -2,86 | 50S ribosomal protein L11 |
| **WP_002288548.1** | complete genome | 1,45E-27 | -2,88 | peptidyl-prolyl cis-trans isomerase |
| **WP_002286304.1** | complete genome | 8,19E-24 | -2,90 | bacitracin ABC transporter ATP-binding protein |
| **WP_002286913.1** | complete genome | 1,84E-22 | -2,94 | 50S ribosomal protein L19 |
| **WP_002289438.1** | complete genome | 1,80E-25 | -2,95 | cystathionine gamma-synthase |
| **WP_002288655.1** | complete genome | 2,36E-19 | -2,96 | 30S ribosomal protein S19 |
| **WP_002287624.1** | complete genome | 3,99E-24 | -2,97 | acetyl-CoA carboxylase carboxyl transferase subunit alpha |
| **WP_002291634.1** | complete genome | 1,91E-31 | -2,97 | 50S ribosomal protein L10 |
| **WP_002296521.1** | complete genome | 7,44E-24 | -2,99 | glutamine ABC transporter permease |
| **WP_002294295.1** | complete genome | 5,97E-23 | -3,03 | acyl carrier protein |
| **WP_002289310.1** | complete genome | 7,05E-24 | -3,07 | 50S ribosomal protein L20 |

| | | | | |
|---|---|---|---|---|
| **WP_002293880.1** | complete genome | 2,38E-20 | -3,10 | 30S ribosomal protein S2 |
| **WP_002287621.1** | complete genome | 6,81E-22 | -3,11 | beta-hydroxyacyl-ACP dehydratase |
| **WP_002292069.1** | complete genome | 5,51E-27 | -3,11 | trigger factor |
| **WP_002287622.1** | complete genome | 9,31E-32 | -3,13 | acetyl-CoA carboxylase biotin carboxylase subunit |
| **WP_002288716.1** | complete genome | 5,74E-34 | -3,15 | 50S ribosomal protein L17 |
| **WP_002296950.1** | complete genome | 6,96E-30 | -3,15 | malonyl CoA-acyl carrier protein transacylase |
| **WP_002294588.1** | complete genome | 1,16E-26 | -3,24 | 50S ribosomal protein L7/L12 |
| **WP_002295975.1** | complete genome | 1,59E-36 | -3,29 | 50S ribosomal protein L13 |
| **WP_002287623.1** | complete genome | 8,15E-28 | -3,31 | acetyl-CoA carboxylase carboxyl transferase subunit beta |
| **WP_002287619.1** | complete genome | 7,25E-34 | -3,34 | beta-ketoacyl-[acyl-carrier-protein] synthase II |
| **WP_002289436.1** | complete genome | 1,28E-31 | -3,42 | cysteine synthase |
| **WP_002287617.1** | complete genome | 8,13E-36 | -3,46 | beta-ketoacyl-ACP reductase |
| **WP_002289156.1** | complete genome | 3,14E-32 | -3,53 | transcription elongation factor GreA |
| **WP_002287505.1** | complete genome | 8,28E-30 | -3,66 | 30S ribosomal protein S9 |
| **WP_002295795.1** | complete genome | 3,68E-31 | -3,78 | enoyl-[acyl-carrier-protein] reductase |
| **WP_002287620.1** | complete genome | 1,26E-26 | -3,97 | acetyl-CoA carboxylase%2C biotin carboxyl carrier protein |
| **WP_002287001.1** | complete genome | 1,08E-37 | -3,99 | type I glutamate--ammonia ligase |
| **WP_002287002.1** | complete genome | 8,96E-39 | -4,02 | MerR family transcriptional regulator |
| **WP_002292113.1** | complete genome | 2,32E-35 | -4,07 | amino acid ABC transporter ATP-binding protein |
| **WP_002295797.1** | complete genome | 2,61E-29 | -4,37 | beta-hydroxyacyl-ACP dehydratase |
| **WP_002294321.1** | complete genome | 5,23E-46 | -4,42 | phosphoglyceromutase |
| **WP_002287610.1** | complete genome | 2,83E-50 | -4,94 | MarR family transcriptional regulator |
| **WP_002287612.1** | complete genome | 2,25E-66 | -5,32 | ketoacyl-ACP synthase III |
| **WP_002287614.1** | complete genome | 5,95E-68 | -5,89 | acyl carrier protein |

[a] Indicates the protein ID number in the Aus0004 genome. [b] Indicates the location of the protein in the Aus0004 genome. [c] Indicates the adjusted p-value (FDR) obtained applying DeSeq2. [d] Log2 Fold-change (FC) derived from the ratio of the abundance of the gene present in *in vivo* (colon) as compare to *in vitro* samples. A positive FC indicates that the gene was found to be in higher levels *in vivo*. [e] Indicates the predicted function.

**Table 37. Tn-mutants that have a significantly different abundance *in vitro* vs *in vivo* (colon and caecum) (p<0.05; FDR<0.05; log2FC > |1|).**

| Accesion[a] | Description[b] | FDR[c] *In vitro* vs colon | Log2FC[d] *In vitro* vs colon | FDR[e] *In vitro* vs caecum | Log2FC[f] *In vitro* vs caecum |
|---|---|---|---|---|---|
| WP_002289456.1 | PTS mannose transporter subunit EIIAB | 5,24E-04 | 5,88 | 3,17E-04 | 5,13 |
| WP_002287582.1 | PTS sugar transporter subunit IIB | 5,24E-04 | 4,71 | 1,88E-03 | 5,81 |
| WP_014387028.1 | Vat family streptogramin A O-acetyltransferase | 6,19E-05 | 4,71 | 3,17E-04 | 4,84 |
| WP_002287686.1 | Fe-S cluster assembly protein SufB | 5,24E-04 | 4,00 | 1,88E-03 | 5,75 |
| WP_002287094.1 | N-acetylmuramic acid 6-phosphate etherase | 6,19E-05 | 3,43 | 3,17E-04 | 2,83 |
| WP_002288139.1 | LacI family transcriptional regulator | 6,19E-05 | 3,25 | 3,17E-04 | 3,61 |
| WP_002288853.1 | hypothetical protein | 6,19E-05 | 3,20 | 3,17E-04 | 3,80 |
| WP_002287653.1 | MurR/RpiR family transcriptional regulator | 5,24E-04 | 3,20 | 3,06E-03 | 2,93 |
| WP_002287026.1 | PTS mannose/fructose/sorbose transporter subunit IIB | 6,60E-04 | 2,95 | 1,88E-03 | 5,30 |
| WP_002296567.1 | bifunctional acetaldehyde-CoA/alcohol dehydrogenase | 6,19E-05 | 2,67 | 3,17E-04 | 2,81 |
| WP_002289486.1 | phosphate ABC transporter ATP-binding protein | 6,19E-05 | 2,63 | 3,17E-04 | 5,69 |
| WP_002289850.1 | DsbA family protein | 6,19E-05 | 2,60 | 3,17E-04 | 3,39 |
| WP_002289860.1 | glycosyl transferase | 6,19E-05 | 2,54 | 3,17E-04 | 2,55 |
| WP_002288437.1 | ornithine carbamoyltransferase | 6,19E-05 | 2,53 | 3,17E-04 | 3,26 |
| WP_002295336.1 | diacylglycerol kinase | 4,58E-03 | 2,45 | 1,45E-02 | 1,57 |
| WP_002289265.1 | formate C-acetyltransferase | 3,20E-04 | 2,42 | 3,17E-04 | 2,67 |
| WP_002289453.1 | transcription antiterminator BglG | 6,19E-05 | 2,41 | 3,17E-04 | 3,72 |
| WP_002288878.1 | carbamoyl phosphate synthase large subunit | 6,19E-05 | 2,31 | 3,17E-04 | 2,58 |
| WP_002296447.1 | DUF3298 domain-containing protein | 6,19E-05 | 2,24 | 3,17E-04 | 3,79 |
| WP_002286858.1 | galactokinase | 6,19E-05 | 2,20 | 3,17E-04 | 2,09 |
| WP_002288015.1 | phosphoribosylformylglycinamide synthase II | 6,19E-05 | 2,17 | 3,17E-04 | 2,56 |
| WP_002317234.1 | sugar transferase | 3,20E-04 | 2,03 | 6,98E-03 | 1,65 |
| WP_002289154.1 | two-component sensor histidine kinase | 6,19E-05 | 1,90 | 3,17E-04 | 1,81 |
| WP_002286455.1 | glycosyl transferase family 2 | 9,17E-04 | 1,89 | 3,17E-04 | 4,83 |
| WP_002288811.1 | carbohydrate deacetylase | 1,21E-02 | 1,88 | 1,88E-03 | 3,26 |
| WP_002296213.1 | tyrosine protein kinase | 6,19E-05 | 1,87 | 3,17E-04 | 1,78 |
| WP_008144451.1 | transposase | 1,77E-03 | 1,83 | 5,31E-04 | 2,48 |
| WP_002295101.1 | ATP synthase subunit A | 6,19E-05 | 1,83 | 3,17E-04 | 2,31 |
| WP_002287536.1 | transcription antiterminator BglG | 9,17E-04 | 1,81 | 1,88E-03 | 3,94 |
| WP_002296453.1 | V-type ATP synthase subunit I | 7,38E-03 | 1,78 | 3,62E-03 | 2,58 |
| WP_002348700.1 | hypothetical protein | 6,19E-05 | 1,77 | 3,17E-04 | 1,77 |
| WP_002333679.1 | LPXTG-domain-containing protein cell wall anchor domain | 2,46E-03 | 1,75 | 6,28E-03 | 1,57 |
| WP_014387065.1 | hypothetical protein | 1,77E-03 | 1,73 | 1,27E-02 | 1,00 |
| WP_002291634.1 | 50S ribosomal protein L10 | 4,55E-02 | 1,69 | 4,75E-03 | 5,46 |
| WP_002288955.1 | exonuclease sbcCD subunit D | 6,60E-04 | 1,61 | 5,31E-04 | 1,87 |
| WP_002286596.1 | DUF3188 domain-containing protein | 3,87E-02 | 1,58 | 3,51E-02 | 3,91 |
| WP_002286614.1 | sugar fermentation stimulation protein SfsA | 6,60E-04 | 1,57 | 3,17E-04 | 1,76 |
| WP_002287297.1 | alpha-N-arabinofuranosidase | 6,19E-05 | 1,56 | 3,17E-04 | 1,79 |
| WP_002287681.1 | Fe-S cluster assembly protein SufD | 1,12E-04 | 1,52 | 2,51E-03 | 1,33 |
| WP_002297171.1 | glycosyl transferase family 1 | 9,17E-04 | 1,51 | 5,31E-04 | 1,72 |
| WP_002296211.1 | tyrosine protein kinase | 5,24E-04 | 1,51 | 1,88E-03 | 1,55 |
| WP_002287001.1 | type I glutamate--ammonia ligase | 6,60E-04 | 1,49 | 3,17E-04 | 3,41 |
| NC_017032.1-gene8 | class A sortase | 4,61E-04 | 1,45 | 9,55E-04 | 1,47 |
| WP_002286285.1 | alpha%2Calpha-phosphotrehalase | 9,17E-04 | 1,41 | 9,55E-04 | 2,33 |
| WP_002296810.1 | beta-fructofuranosidase | 5,24E-04 | 1,39 | 3,17E-04 | 1,86 |
| WP_002287683.1 | cysteine desulfurase | 2,96E-02 | 1,38 | 9,55E-04 | 3,67 |

| | | | | | |
|---|---|---|---|---|---|
| WP_002296473.1 | L-lactate oxidase | 1,77E-03 | 1,36 | 1,62E-02 | 1,25 |
| WP_002294670.1 | pyridine nucleotide-disulfide oxidoreductase | 6,19E-05 | 1,35 | 3,17E-04 | 1,33 |
| WP_002296458.1 | endonuclease MutS2 | 4,61E-04 | 1,34 | 3,17E-04 | 1,52 |
| WP_002297160.1 | 3-hydroxy-3-methylglutaryl-CoA lyase | 6,19E-05 | 1,32 | 3,17E-04 | 1,44 |
| WP_002301912.1 | transcriptional regulator | 6,19E-05 | 1,31 | 3,62E-03 | 1,13 |
| WP_002296542.1 | membrane protein | 3,60E-02 | 1,31 | 2,51E-03 | 2,22 |
| WP_002353096.1 | hypothetical protein | 6,60E-04 | 1,27 | 3,17E-04 | 1,19 |
| WP_002289203.1 | BMP family ABC transporter substrate-binding protein | 3,20E-04 | 1,25 | 2,51E-03 | 1,23 |
| WP_002289507.1 | sugar ABC transporter ATP-binding protein | 2,02E-04 | 1,24 | 3,17E-04 | 1,58 |
| WP_002289202.1 | heme ABC transporter ATP-binding protein | 1,28E-03 | 1,22 | 3,17E-04 | 1,58 |
| WP_002289334.1 | ABC transporter permease | 6,19E-05 | 1,19 | 3,17E-04 | 1,21 |
| WP_002288872.1 | aspartate carbamoyltransferase | 2,02E-04 | 1,19 | 3,17E-04 | 1,25 |
| WP_002287403.1 | transcription-repair coupling factor | 6,19E-05 | 1,18 | 3,17E-04 | 1,44 |
| WP_002288948.1 | adaptor protein MecA | 6,19E-05 | 1,17 | 3,17E-04 | 1,09 |
| WP_002295400.1 | LytR family transcriptional regulator | 3,60E-02 | 1,17 | 9,59E-03 | 2,17 |
| WP_002297196.1 | ATPase P | 4,61E-04 | 1,15 | 3,17E-04 | 1,63 |
| WP_002305249.1 | relaxase | 6,19E-05 | 1,15 | 3,17E-04 | 1,01 |
| WP_002285913.1 | muramidase | 6,19E-05 | 1,11 | 3,17E-04 | 1,09 |
| WP_002287134.1 | hypothetical protein | 1,28E-03 | 1,11 | 3,17E-04 | 1,48 |
| WP_002289201.1 | ABC transporter permease | 5,66E-03 | 1,10 | 9,55E-04 | 1,41 |
| WP_002294157.1 | dipeptidase PepV | 4,61E-04 | 1,08 | 3,17E-04 | 1,01 |
| WP_002286317.1 | DUF554 domain-containing protein | 4,31E-03 | 1,08 | 2,63E-02 | 1,01 |
| WP_002287799.1 | ferrous iron transport protein B | 6,19E-05 | 1,06 | 3,17E-04 | 1,08 |
| WP_002288021.1 | phosphoribosylaminoimidazolesuccinocarbox amide synthase | 6,60E-04 | 1,04 | 9,55E-04 | 1,29 |
| WP_002287889.1 | ABC transporter ATP-binding protein | 1,21E-02 | 1,02 | 1,62E-02 | 1,15 |
| WP_002289239.1 | prolyl-tRNA synthetase associated domain-containing protein | 5,24E-04 | -2,34 | 2,09E-02 | -1,09 |
| WP_002289315.1 | hypothetical protein | 2,45E-03 | -1,46 | 9,59E-03 | -1,11 |
| WP_002317283.1 | hypothetical protein | 5,24E-04 | -1,56 | 3,17E-04 | -1,14 |
| WP_002296417.1 | DNA helicase PcrA | 1,12E-04 | -2,20 | 4,97E-03 | -1,14 |
| WP_002294466.1 | prolyl-tRNA editing protein | 4,61E-04 | -2,84 | 4,21E-02 | -1,16 |
| WP_002286818.1 | pyridine nucleotide-disulfide oxidoreductase | 6,19E-05 | -2,19 | 2,63E-02 | -1,20 |
| WP_002286631.1 | cell division protein FtsA | 1,28E-03 | -2,84 | 4,21E-02 | -1,22 |
| WP_002288825.1 | L-ascorbate-6-phosphate lactonase | 2,02E-04 | -2,81 | 4,21E-02 | -1,23 |
| WP_014387137.1 | hypothetical protein | 6,19E-05 | -3,04 | 2,09E-02 | -1,25 |
| WP_002288348.1 | ATP-dependent chaperone ClpB | 6,19E-05 | -3,13 | 1,62E-02 | -1,25 |
| WP_002287112.1 | PTS mannose transporter subunit IIA | 3,27E-03 | -2,63 | 1,27E-02 | -1,29 |
| WP_002288009.1 | bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/inosine monophosphate cyclohydrolase | 1,12E-04 | -3,67 | 4,21E-02 | -1,31 |
| WP_002287956.1 | cysteine desulfurase | 1,21E-02 | -3,25 | 4,21E-02 | -1,31 |
| WP_002288890.1 | fibronectin/fibrinogen-binding protein | 1,21E-02 | -1,20 | 4,21E-02 | -1,31 |
| WP_002294252.1 | peptide-methionine (R)-S-oxide reductase | 3,96E-03 | -2,52 | 4,14E-02 | -1,34 |
| WP_002288958.1 | competence protein CoiA | 6,19E-05 | -3,18 | 1,62E-02 | -1,34 |
| WP_002289003.1 | UDP-N-acetylglucosamine 1-carboxyvinyltransferase | 1,01E-03 | -2,63 | 4,98E-02 | -1,36 |
| WP_002293880.1 | 30S ribosomal protein S2 | 2,39E-02 | -1,41 | 3,35E-02 | -1,38 |
| WP_002286621.1 | isoleucine--tRNA ligase | 1,12E-04 | -3,11 | 2,51E-03 | -1,42 |
| WP_002294735.1 | DUF4767 domain-containing protein | 6,19E-05 | -3,00 | 2,09E-02 | -1,43 |
| WP_002289036.1 | hypothetical protein | 6,19E-05 | -2,92 | 2,09E-02 | -1,46 |
| WP_002289328.1 | transposase | 6,19E-05 | -3,37 | 4,21E-02 | -1,51 |
| WP_002289260.1 | DUF2179 domain-containing protein | 4,61E-04 | -2,52 | 3,62E-03 | -1,52 |
| NC_017032.1-gene9 | cell wall surface anchor family protein | 6,19E-05 | -2,65 | 5,31E-04 | -1,56 |
| WP_002294577.1 | hypothetical protein | 6,19E-05 | -3,23 | 6,98E-03 | -1,58 |
| WP_002296493.1 | helicase | 4,61E-04 | -2,62 | 2,09E-02 | -1,58 |
| WP_002287379.1 | ribosome small subunit-dependent GTPase A | 5,24E-04 | -1,85 | 2,51E-03 | -1,58 |
| WP_002296629.1 | peptidase | 9,17E-04 | -2,77 | 1,27E-02 | -1,59 |

| | | | | | |
|---|---|---|---|---|---|
| WP_002289388.1 | ABC transporter ATP-binding protein | 3,20E-04 | -2,99 | 2,63E-02 | -1,59 |
| WP_002287130.1 | RNase J family beta-CASP ribonuclease | 6,19E-05 | -3,13 | 2,63E-02 | -1,59 |
| WP_002287299.1 | transcriptional regulator | 6,19E-05 | -2,00 | 3,17E-04 | -1,60 |
| WP_002296982.1 | ribosome biogenesis GTPase Der | 3,20E-04 | -2,80 | 1,62E-02 | -1,67 |
| WP_002311272.1 | DNA helicase | 6,19E-05 | -3,41 | 5,31E-04 | -1,67 |
| WP_002296167.1 | ABC transporter permease | 6,19E-05 | -2,01 | 3,17E-04 | -1,68 |
| WP_002295483.1 | inositol monophosphatase | 2,52E-03 | -2,26 | 3,35E-02 | -1,72 |
| WP_002296158.1 | HDIG domain-containing protein | 4,61E-04 | -3,46 | 2,63E-02 | -1,72 |
| WP_002296488.1 | terminase | 1,12E-04 | -2,84 | 9,59E-03 | -1,73 |
| WP_002287797.1 | ferrous iron transporter A | 4,61E-04 | -2,83 | 2,09E-02 | -1,73 |
| NC_017032.1-gene51 | ATP-binding protein | 6,19E-05 | -2,20 | 1,27E-02 | -1,73 |
| WP_002290168.1 | YfcC family protein | 6,19E-05 | -3,25 | 1,51E-03 | -1,74 |
| WP_002296571.1 | single-stranded-DNA-specific exonuclease RecJ | 6,19E-05 | -3,39 | 2,51E-03 | -1,74 |
| WP_002296814.1 | beta-fructofuranosidase | 9,17E-04 | -3,45 | 1,62E-02 | -1,77 |
| WP_002287605.1 | glycosyl transferase | 6,19E-05 | -3,27 | 2,09E-02 | -1,77 |
| WP_002295906.1 | hypothetical protein | 1,28E-03 | -2,61 | 4,95E-02 | -1,78 |
| WP_002286048.1 | transcription antiterminator BglG | 6,19E-05 | -3,03 | 5,31E-04 | -1,78 |
| WP_002289647.1 | carbohydrate ABC transporter permease | 6,60E-04 | -3,86 | 7,89E-03 | -1,79 |
| WP_002289477.1 | alpha/beta hydrolase | 2,02E-04 | -2,95 | 4,21E-02 | -1,79 |
| WP_002286002.1 | altronate hydrolase | 6,19E-05 | -3,45 | 1,51E-03 | -1,79 |
| WP_002296588.1 | preprotein translocase subunit SecA | 3,20E-04 | -3,35 | 1,51E-03 | -1,79 |
| WP_002287015.1 | site-specific integrase | 6,19E-05 | -3,68 | 3,35E-02 | -1,81 |
| WP_002294496.1 | glycoside hydrolase family 28 | 6,19E-05 | -3,89 | 1,27E-02 | -1,81 |
| WP_002321266.1 | hypothetical protein | 8,12E-04 | -3,17 | 2,70E-02 | -1,81 |
| WP_002353032.1 | MATE domain protein | 6,19E-05 | -3,43 | 4,21E-02 | -1,82 |
| WP_002296865.1 | ribose-5-phosphate isomerase | 2,02E-04 | -3,66 | 2,63E-02 | -1,85 |
| WP_002303716.1 | maltose-6'-phosphate glucosidase | 2,45E-03 | -1,68 | 3,35E-02 | -1,86 |
| WP_002287958.1 | haloacid dehalogenase | 1,03E-03 | -3,82 | 1,00E-02 | -1,87 |
| WP_002303746.1 | DNA topoisomerase IV subunit B | 1,12E-04 | -3,30 | 6,98E-03 | -1,88 |
| WP_002287896.1 | ribonuclease III | 4,61E-04 | -2,64 | 2,63E-02 | -1,88 |
| WP_002296647.1 | PTS mannose transporter subunit IID | 9,17E-04 | -2,40 | 4,97E-03 | -1,90 |
| WP_014386982.1 | AraC family transcriptional regulator | 4,61E-04 | -1,62 | 3,17E-04 | -1,91 |
| WP_002287560.1 | amino acid permease | 1,12E-04 | -3,04 | 4,21E-02 | -1,91 |
| WP_002296457.1 | excinuclease ABC subunit C | 2,02E-04 | -3,76 | 2,63E-02 | -1,91 |
| WP_002288231.1 | mevalonate kinase | 1,00E-03 | -2,72 | 1,86E-02 | -1,91 |
| WP_002296412.1 | alpha-galactosidase | 6,19E-05 | -3,60 | 3,62E-03 | -1,95 |
| WP_002287639.1 | SAM-dependent methyltransferase | 6,19E-05 | -3,11 | 3,17E-04 | -1,96 |
| WP_002287480.1 | hypothetical protein | 6,19E-05 | -3,42 | 4,97E-03 | -1,97 |
| WP_002285962.1 | UDP-N-acetylmuramoyl-L-alanyl-D-glutamate--2%2C6-diaminopimelate ligase | 3,20E-04 | -2,22 | 3,35E-02 | -1,97 |
| WP_002289731.1 | methionine adenosyltransferase | 2,02E-04 | -3,36 | 4,97E-03 | -1,97 |
| WP_002286373.1 | maltose phosphorylase | 6,19E-05 | -3,61 | 1,88E-03 | -1,97 |
| WP_002311095.1 | peptidase | 6,19E-05 | -3,67 | 2,63E-02 | -1,98 |
| WP_002317189.1 | cytochrome oxidase biogenesis protein OxaA | 5,24E-04 | -2,78 | 1,62E-02 | -1,98 |
| WP_002287257.1 | PTS sugar transporter subunit IIC | 6,19E-05 | -3,69 | 1,51E-03 | -1,99 |
| WP_002287959.1 | haloacid dehalogenase | 5,24E-04 | -3,18 | 3,35E-02 | -2,02 |
| WP_002296633.1 | transcriptional regulator | 8,09E-04 | -3,60 | 1,54E-02 | -2,05 |
| WP_002296311.1 | RDD family protein | 1,00E-03 | -3,71 | 2,25E-02 | -2,05 |
| WP_002287107.1 | IS110 family transposase | 3,20E-04 | -3,06 | 9,59E-03 | -2,05 |
| WP_002347315.1 | hypothetical protein | 6,19E-05 | -3,82 | 1,27E-02 | -2,06 |
| WP_014387020.1 | hypothetical protein | 2,02E-04 | -3,01 | 2,51E-03 | -2,08 |
| WP_002288368.1 | single-stranded DNA-binding protein | 6,19E-05 | -3,96 | 6,98E-03 | -2,08 |
| WP_002321715.1 | hypothetical protein | 1,61E-03 | -2,84 | 1,25E-02 | -2,08 |
| WP_002297134.1 | DNA mismatch repair protein MutL | 6,19E-05 | -3,36 | 2,51E-03 | -2,09 |
| WP_002295350.1 | phosphogluconate dehydrogenase (NADP(+)-dependent%2C decarboxylating) | 3,20E-04 | -3,66 | 1,62E-02 | -2,10 |
| WP_080019250.1 | hypothetical protein | 6,19E-05 | -3,17 | 1,51E-03 | -2,11 |
| WP_002291653.1 | catabolite control protein A | 3,20E-04 | -3,12 | 1,51E-03 | -2,11 |
| WP_002289227.1 | DeoR family transcriptional regulator | 8,12E-04 | -3,13 | 2,26E-02 | -2,12 |

| | | | | | |
|---|---|---|---|---|---|
| **WP_002286776.1** | peptidase | 6,19E-05 | -3,68 | 5,31E-04 | -2,12 |
| **WP_002286201.1** | HIT family protein | 6,60E-04 | -3,42 | 6,98E-03 | -2,12 |
| **WP_002288390.1** | ABC transporter ATP-binding protein | 6,19E-05 | -3,83 | 1,88E-03 | -2,14 |
| **WP_002296189.1** | transcriptional regulator | 2,11E-03 | -3,11 | 3,58E-02 | -2,14 |
| **WP_002289593.1** | histidine phosphatase family protein | 3,87E-02 | -2,08 | 1,65E-02 | -2,15 |
| **WP_002295964.1** | DNA-directed RNA polymerase subunit beta | 6,19E-05 | -3,57 | 3,17E-04 | -2,16 |
| **WP_002287128.1** | multidrug ABC transporter ATP-binding protein | 1,49E-03 | -2,73 | 2,90E-02 | -2,18 |
| **WP_002287415.1** | hypoxanthine phosphoribosyltransferase | 4,58E-02 | -2,57 | 3,22E-02 | -2,19 |
| **WP_002289366.1** | dehydrogenase | 8,12E-04 | -3,59 | 2,26E-02 | -2,19 |
| **WP_002287302.1** | NAD(+) synthetase | 7,61E-04 | -3,67 | 1,46E-02 | -2,20 |
| **WP_002289687.1** | dihydrolipoamide acetyltransferase | 6,60E-04 | -4,07 | 1,25E-02 | -2,23 |
| **WP_002289495.1** | hypothetical protein | 6,19E-05 | -4,40 | 1,51E-03 | -2,24 |
| **WP_002286311.1** | hypothetical protein | 1,28E-03 | -3,31 | 1,54E-02 | -2,26 |
| **NC_017032.1-gene1** | hypothetical protein | 6,19E-05 | -4,30 | 5,31E-04 | -2,28 |
| **WP_010731019.1** | polysaccharide deacetylase | 6,19E-05 | -4,48 | 3,62E-03 | -2,28 |
| **WP_002289402.1** | ATP-dependent Clp protease proteolytic subunit | 2,02E-04 | -3,97 | 5,31E-04 | -2,29 |
| **WP_002289695.1** | ATP-binding protein | 6,19E-05 | -3,03 | 1,27E-02 | -2,35 |
| **WP_002288823.1** | DeoR/GlpR transcriptional regulator | 6,50E-04 | -3,89 | 1,23E-02 | -2,36 |
| **WP_002297127.1** | histidine kinase | 2,02E-04 | -4,20 | 4,97E-03 | -2,38 |
| **WP_002296820.1** | two-component sensor histidine kinase | 1,12E-04 | -2,93 | 2,51E-03 | -2,39 |
| **WP_002297169.1** | polysaccharide biosynthesis protein | 6,60E-04 | -2,86 | 1,88E-03 | -2,39 |
| **WP_002289159.1** | aminodeoxychorismate lyase | 2,02E-04 | -3,01 | 3,17E-04 | -2,39 |
| **WP_002292842.1** | V-type ATP synthase subunit B | 6,19E-05 | -4,04 | 5,31E-04 | -2,40 |
| **WP_002320813.1** | iron ABC transporter permease | 9,17E-04 | -1,89 | 9,55E-04 | -2,40 |
| **WP_014387033.1** | recombinase family protein | 6,19E-05 | -4,77 | 2,09E-02 | -2,41 |
| **WP_002298781.1** | HAD family phosphatase | 6,19E-05 | -4,56 | 3,62E-03 | -2,42 |
| **WP_002288411.1** | DNA topoisomerase I | 6,19E-05 | -4,04 | 1,88E-03 | -2,43 |
| **WP_002285974.1** | ABC transporter substrate-binding protein | 6,19E-05 | -4,62 | 5,31E-04 | -2,43 |
| **WP_002296524.1** | hypothetical protein | 2,51E-03 | -3,37 | 3,03E-03 | -2,45 |
| **WP_002288017.1** | phosphoribosylformylglycinamidine synthase I | 1,75E-02 | -3,11 | 3,35E-02 | -2,45 |
| **WP_002295869.1** | hypothetical protein | 7,61E-04 | -4,37 | 1,86E-02 | -2,46 |
| **WP_002286786.1** | carbohydrate ABC transporter permease | 5,24E-04 | -3,58 | 4,98E-02 | -2,48 |
| **WP_002289137.1** | prephenate dehydratase | 6,19E-05 | -4,15 | 3,17E-04 | -2,49 |
| **WP_002296462.1** | peptide ABC transporter ATP-binding protein | 2,02E-04 | -4,26 | 1,62E-02 | -2,49 |
| **WP_002297289.1** | bifunctional biotin--[acetyl-CoA-carboxylase] synthetase/biotin operon repressor | 6,19E-05 | -4,43 | 3,62E-03 | -2,51 |
| **WP_002287369.1** | peptide deformylase | 3,10E-03 | -2,98 | 3,32E-02 | -2,51 |
| **WP_002287155.1** | multifunctional 2'%2C3'-cyclic-nucleotide 2'-phosphodiesterase/5'-nucleotidase/3'-nucleotidase | 6,19E-05 | -3,82 | 5,31E-04 | -2,51 |
| **WP_002289648.1** | gfo/Idh/MocA family oxidoreductase | 5,24E-04 | -3,65 | 4,94E-03 | -2,52 |
| **WP_002288361.1** | DNA polymerase III subunit beta | 9,12E-03 | -2,66 | 1,54E-02 | -2,53 |
| **WP_002295354.1** | LysR family transcriptional regulator | 6,19E-05 | -3,86 | 9,55E-04 | -2,55 |
| **WP_002287288.1** | GntR family transcriptional regulator | 1,62E-03 | -4,02 | 7,89E-03 | -2,56 |
| **WP_002296256.1** | DUF4767 domain-containing protein | 1,12E-04 | -4,73 | 2,51E-03 | -2,57 |
| **WP_002340436.1** | hypothetical protein | 5,24E-04 | -4,71 | 1,25E-02 | -2,59 |
| **WP_002287793.1** | ribonucleoside-diphosphate reductase subunit alpha | 6,19E-05 | -4,23 | 5,31E-04 | -2,60 |
| **WP_002288364.1** | DNA gyrase subunit B | 5,24E-04 | -3,54 | 3,06E-03 | -2,61 |
| **WP_014387138.1** | leucine--tRNA ligase | 6,19E-05 | -3,20 | 3,17E-04 | -2,62 |
| **WP_002293649.1** | universal stress protein | 1,12E-04 | -4,07 | 5,31E-04 | -2,62 |
| **WP_002294134.1** | isoprenyl transferase | 5,24E-04 | -3,52 | 3,03E-03 | -2,64 |
| **WP_002289340.1** | DUF536 domain-containing protein | 6,19E-05 | -4,76 | 5,31E-04 | -2,64 |
| **WP_002289763.1** | proline/glycine betaine ABC transporter ATP-binding protein | 6,19E-05 | -4,19 | 1,51E-03 | -2,64 |
| **WP_002317290.1** | hypothetical protein | 1,07E-02 | -2,23 | 3,22E-02 | -2,65 |
| **WP_002296465.1** | NAD(P)/FAD-dependent oxidoreductase | 6,19E-05 | -3,99 | 3,17E-04 | -2,67 |

XXI

| | | | | | |
|---|---|---|---|---|---|
| WP_002294889.1 | RNA methyltransferase | 5,24E-04 | -5,16 | 1,54E-02 | -2,67 |
| WP_002288722.1 | energy-coupling factor transporter ATPase | 6,19E-05 | -3,66 | 3,17E-04 | -2,67 |
| WP_002287997.1 | LacI family transcriptional regulator | 6,19E-05 | -4,36 | 1,51E-03 | -2,67 |
| WP_000237797.1 | site-specific integrase | 6,19E-05 | -4,06 | 1,51E-03 | -2,68 |
| WP_002289035.1 | ATP-dependent Clp protease ATP-binding subunit | 6,19E-05 | -4,70 | 3,17E-04 | -2,69 |
| WP_002289136.1 | sensor histidine kinase | 6,19E-05 | -3,80 | 3,17E-04 | -2,70 |
| WP_002289045.1 | hypothetical protein | 6,19E-05 | -4,51 | 5,31E-04 | -2,72 |
| WP_002287150.1 | Fis family transcriptional regulator | 1,12E-04 | -3,65 | 3,17E-04 | -2,73 |
| WP_049788871.1 | hypothetical protein | 5,24E-04 | -4,20 | 1,59E-02 | -2,75 |
| WP_002287746.1 | redox-sensing transcriptional repressor Rex | 5,24E-04 | -4,61 | 1,88E-03 | -2,77 |
| WP_002287279.1 | membrane protein | 6,19E-05 | -4,13 | 3,17E-04 | -2,77 |
| WP_002296645.1 | oxidoreductase | 7,92E-04 | -3,33 | 3,69E-03 | -2,78 |
| WP_002296648.1 | PTS mannose/fructose/sorbose transporter subunit IIB | 6,19E-05 | -4,82 | 3,17E-04 | -2,79 |
| WP_002286055.1 | VWA domain-containing protein | 6,19E-05 | -4,33 | 3,62E-03 | -2,80 |
| WP_002287463.1 | mannosyl-glycoprotein endo-beta-N-acetylglucosaminidase | 6,19E-05 | -4,74 | 5,31E-04 | -2,81 |
| WP_002289390.1 | ABC transporter permease | 7,61E-04 | -4,28 | 1,54E-02 | -2,82 |
| WP_002296409.1 | DUF1538 domain-containing protein | 5,80E-04 | -5,41 | 6,98E-03 | -2,83 |
| WP_002288447.1 | tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase complex dimerization subunit type 1 TsaB | 6,19E-05 | -4,41 | 1,27E-02 | -2,83 |
| WP_002288329.1 | Replication initiation and membrane attachment | 6,19E-05 | -4,19 | 5,31E-04 | -2,85 |
| WP_002289438.1 | cystathionine gamma-synthase | 5,24E-04 | -8,83 | 3,62E-03 | -2,86 |
| WP_002287494.1 | ABC transporter ATP-binding protein | 3,94E-03 | -4,10 | 3,35E-02 | -2,88 |
| WP_002311595.1 | hypothetical protein | 5,24E-04 | -4,45 | 2,87E-02 | -2,89 |
| WP_002296998.1 | DNA topoisomerase IV subunit A | 6,19E-05 | -4,24 | 3,17E-04 | -2,90 |
| WP_002286405.1 | tyrosine--tRNA ligase | 4,77E-03 | -3,54 | 2,70E-02 | -2,90 |
| WP_002289566.1 | phage holin family protein | 3,41E-03 | -4,57 | 1,88E-03 | -2,93 |
| WP_002296274.1 | peptidase | 5,24E-04 | -4,84 | 2,35E-03 | -2,93 |
| WP_002295450.1 | FUSC family protein | 4,31E-03 | -5,56 | 9,59E-03 | -2,93 |
| WP_002288688.1 | preprotein translocase subunit SecY | 5,24E-04 | -4,77 | 7,89E-03 | -2,93 |
| WP_002303894.1 | DNA ligase (NAD(+)) LigA | 6,19E-05 | -4,09 | 3,17E-04 | -2,95 |
| WP_002289599.1 | two-component sensor histidine kinase | 5,24E-04 | -4,87 | 1,88E-03 | -2,96 |
| WP_002287078.1 | nicotinamide mononucleotide transporter | 2,99E-03 | -3,84 | 2,20E-02 | -2,98 |
| WP_002295470.1 | xanthine permease | 6,60E-04 | -4,70 | 4,91E-03 | -2,98 |
| WP_002294001.1 | TatD family deoxyribonuclease | 6,19E-05 | -3,93 | 5,31E-04 | -2,99 |
| WP_002287068.1 | phosphate acetyltransferase | 5,24E-04 | -4,72 | 1,88E-03 | -2,99 |
| WP_002294579.1 | preprotein translocase subunit SecE | 1,28E-03 | -5,64 | 2,55E-02 | -3,00 |
| WP_002287478.1 | transcriptional regulator | 6,60E-04 | -4,42 | 9,83E-03 | -3,00 |
| WP_002287457.1 | DUF2798 domain-containing protein | 5,24E-04 | -3,85 | 4,37E-02 | -3,00 |
| WP_014387062.1 | ATP-dependent helicase | 3,96E-03 | -3,17 | 1,54E-02 | -3,01 |
| WP_002288118.1 | gluconokinase | 6,19E-05 | -4,54 | 3,17E-04 | -3,02 |
| WP_002286368.1 | PTS glucose transporter subunit IIABC | 6,19E-05 | -4,56 | 3,17E-04 | -3,02 |
| WP_002285918.1 | ABC transporter ATP-binding protein | 6,19E-05 | -4,77 | 3,17E-04 | -3,02 |
| WP_002295138.1 | DUF2975 domain-containing protein | 5,80E-04 | -4,49 | 2,36E-02 | -3,02 |
| WP_002287048.1 | alcohol dehydrogenase | 6,19E-05 | -4,64 | 3,17E-04 | -3,03 |
| WP_002288352.1 | nucleoside deoxyribosyltransferase | 5,24E-04 | -5,50 | 4,61E-03 | -3,06 |
| WP_007897440.1 | hypothetical protein | 1,28E-03 | -3,88 | 2,43E-03 | -3,06 |
| WP_002287617.1 | beta-ketoacyl-ACP reductase | 6,19E-05 | -5,17 | 3,17E-04 | -3,07 |
| WP_002286656.1 | farnesyl-diphosphate synthase | 7,61E-04 | -4,55 | 1,86E-02 | -3,07 |
| WP_002295857.1 | hypothetical protein | 5,24E-04 | -4,54 | 3,83E-03 | -3,07 |
| WP_002287041.1 | quinone oxidoreductase | 5,24E-04 | -5,34 | 9,48E-03 | -3,07 |
| WP_044031073.1 | hypothetical protein | 5,24E-04 | -4,42 | 2,63E-02 | -3,08 |
| WP_014387016.1 | NAD-dependent epimerase | 6,19E-05 | -4,19 | 3,17E-04 | -3,09 |
| WP_002294563.1 | HAD family hydrolase | 5,24E-04 | -4,58 | 2,95E-03 | -3,10 |
| WP_002289190.1 | ATP-dependent helicase | 6,19E-05 | -4,45 | 3,17E-04 | -3,11 |
| WP_014387075.1 | glyoxalase-like domain protein | 6,19E-05 | -4,74 | 3,17E-04 | -3,13 |

| | | | | | |
|---|---|---|---|---|---|
| WP_002286638.1 | UDP-N-acetylmuramoyl-L-alanine--D-glutamate ligase | 5,24E-04 | -3,53 | 2,13E-02 | -3,14 |
| WP_002287274.1 | PTS sorbitol transporter subunit IIA | 5,24E-04 | -4,27 | 1,61E-02 | -3,14 |
| WP_002295265.1 | serine--tRNA ligase | 6,19E-05 | -4,52 | 3,17E-04 | -3,14 |
| WP_002295917.1 | glutamyl-tRNA(Gln) amidotransferase | 6,19E-05 | -4,57 | 5,31E-04 | -3,16 |
| WP_002295087.1 | hypothetical protein | 1,03E-03 | -3,74 | 1,97E-02 | -3,16 |
| WP_002288392.1 | dihydrofolate reductase | 8,12E-04 | -5,32 | 3,83E-03 | -3,17 |
| WP_002352380.1 | hypothetical protein | 6,37E-03 | -3,15 | 1,97E-02 | -3,17 |
| WP_002296196.1 | class A sortase | 5,24E-04 | -4,40 | 2,36E-02 | -3,18 |
| WP_002288023.1 | adenylosuccinate lyase | 6,19E-05 | -4,40 | 3,17E-04 | -3,19 |
| WP_002292874.1 | N-acetylneuraminate synthase | 6,19E-05 | -4,98 | 3,17E-04 | -3,19 |
| NC_017032.1-gene30 | hypothetical protein | 6,19E-05 | -4,83 | 3,62E-03 | -3,20 |
| WP_002289241.1 | amino acid ABC transporter ATP-binding protein | 6,19E-05 | -4,15 | 3,17E-04 | -3,21 |
| WP_002286302.1 | ABC transporter permease | 6,19E-05 | -4,82 | 3,17E-04 | -3,21 |
| WP_002286213.1_2609 | hypothetical protein | 6,19E-05 | -6,47 | 3,17E-04 | -3,23 |
| WP_002286924.1 | cell wall surface anchor protein | 6,19E-05 | -4,41 | 3,17E-04 | -3,23 |
| WP_002286343.1 | MFS transporter | 6,19E-05 | -4,43 | 3,17E-04 | -3,23 |
| WP_002288397.1 | peptide-methionine (S)-S-oxide reductase | 2,51E-03 | -4,65 | 3,62E-03 | -3,23 |
| WP_002287792.1 | ribonucleoside-diphosphate reductase | 6,19E-05 | -4,93 | 3,62E-03 | -3,26 |
| WP_002295566.1 | MFS transporter | 6,19E-05 | -4,91 | 3,17E-04 | -3,27 |
| WP_002288852.1 | hypothetical protein | 6,19E-05 | -5,28 | 3,17E-04 | -3,29 |
| WP_014387148.1 | D-alanine--(R)-lactate ligase VanB | 6,19E-05 | -5,14 | 3,17E-04 | -3,30 |
| WP_002287752.1 | hypothetical protein | 6,19E-05 | -4,80 | 3,17E-04 | -3,32 |
| WP_002289219.1 | fructokinase | 5,24E-04 | -5,25 | 6,04E-03 | -3,32 |
| WP_002287490.1 | tRNA-dihydrouridine synthase | 5,24E-04 | -4,78 | 9,59E-03 | -3,32 |
| WP_002289118.1 | MBL fold metallo-hydrolase | 6,60E-04 | -4,85 | 1,61E-02 | -3,33 |
| WP_002289270.1 | hydroxymethylglutaryl-CoA synthase | 5,24E-04 | -5,55 | 1,88E-03 | -3,34 |
| WP_002289545.1 | MarR family transcriptional regulator | 8,09E-04 | -4,85 | 1,88E-03 | -3,34 |
| WP_002297304.1 | cobalt transporter | 5,24E-04 | -5,40 | 3,69E-03 | -3,35 |
| WP_002295818.1 | transketolase | 6,19E-05 | -5,08 | 3,17E-04 | -3,36 |
| WP_002286026.1 | xylan 1%2C4-beta-xylosidase | 6,19E-05 | -5,03 | 3,17E-04 | -3,36 |
| WP_002286009.1 | protein lplB | 6,19E-05 | -4,71 | 3,17E-04 | -3,37 |
| WP_002285981.1 | GntR family transcriptional regulator | 6,19E-05 | -4,58 | 5,31E-04 | -3,37 |
| WP_002288473.1 | Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase GatCAB subunit B | 5,24E-04 | -4,59 | 3,06E-03 | -3,38 |
| WP_002321467.1 | hypothetical protein | 4,61E-04 | -4,96 | 3,57E-02 | -3,38 |
| WP_002289207.1 | pyrimidine-nucleoside phosphorylase | 5,24E-04 | -4,41 | 3,22E-02 | -3,39 |
| WP_002287468.1_ | acetyltransferase | 3,41E-03 | -5,13 | 3,03E-03 | -3,41 |
| WP_002298578.1 | GMP synthase (glutamine-hydrolyzing) | 5,24E-04 | -5,67 | 3,62E-03 | -3,42 |
| WP_002287431.1 | large-conductance mechanosensitive channel MscL | 3,27E-03 | -4,52 | 4,90E-02 | -3,43 |
| WP_002288250.1 | rhamnulokinase | 6,19E-05 | -4,36 | 1,88E-03 | -3,44 |
| WP_002294562.1 | RNA pseudouridine synthase | 5,24E-04 | -4,36 | 4,94E-03 | -3,44 |
| WP_002287348.1 | thiamine transporter thia | 6,20E-04 | -4,69 | 1,86E-02 | -3,46 |
| WP_000248477.1 | hypothetical protein | 6,20E-04 | -4,34 | 4,61E-03 | -3,46 |
| WP_002294546.1 | PTS mannose/fructose/sorbose transporter subunit IIC | 2,51E-03 | -4,61 | 9,74E-03 | -3,49 |
| WP_002288633.1 | arabinogalactan endo-1%2C4-beta-galactosidase | 5,24E-04 | -4,82 | 4,94E-03 | -3,49 |
| WP_002296946.1 | cell division protein FtsI | 6,19E-05 | -4,67 | 5,31E-04 | -3,49 |
| WP_002297290.1 | DNA starvation/stationary phase protection protein | 5,99E-04 | -4,89 | 9,83E-03 | -3,54 |
| WP_002288377.1 | DegV family protein | 7,94E-03 | -4,03 | 4,37E-02 | -3,54 |
| WP_002294560.1 | copper homeostasis protein CutC | 9,36E-04 | -4,05 | 2,26E-02 | -3,55 |
| WP_002294975.1 | YibE/F family protein | 5,24E-04 | -5,00 | 1,88E-03 | -3,55 |
| WP_002296555.1 | alanine racemase | 6,19E-05 | -4,95 | 3,17E-04 | -3,58 |
| WP_002294021.1 | hypothetical protein | 1,28E-03 | -3,78 | 1,62E-02 | -3,59 |

| | | | | | |
|---|---|---|---|---|---|
| WP_002286008.1 | carbohydrate ABC transporter permease | 6,19E-05 | -5,30 | 9,55E-04 | -3,59 |
| WP_002320976.1 _2169 | hypothetical protein | 3,41E-03 | -5,87 | 1,88E-03 | -3,60 |
| WP_002296539.1 | class I SAM-dependent methyltransferase | 6,19E-05 | -5,21 | 3,17E-04 | -3,60 |
| WP_002289237.1 | bacteriocin immunity protein | 5,24E-04 | -5,08 | 1,34E-02 | -3,62 |
| WP_002305538.1 | transcriptional regulator | 5,80E-04 | -5,13 | 2,56E-03 | -3,62 |
| WP_002288323.1 _1607 | DNA-formamidopyrimidine glycosylase | 5,24E-04 | -5,08 | 3,79E-03 | -3,64 |
| WP_002324227.1 | ABC transporter ATP-binding protein | 5,24E-04 | -3,90 | 7,27E-03 | -3,64 |
| WP_002286589.1 | acetyltransferase | 6,19E-05 | -5,17 | 3,17E-04 | -3,65 |
| WP_002291698.1 | PhoH family protein | 6,19E-05 | -5,42 | 3,17E-04 | -3,65 |
| WP_002416070.1 | D-Ala-D-Ala carboxypeptidase VanY-B | 5,24E-04 | -5,32 | 2,95E-03 | -3,67 |
| WP_002287244.1 | UDP-N-acetylmuramyl peptide synthase | 6,19E-05 | -5,46 | 3,17E-04 | -3,69 |
| WP_002287630.1 | RNA-binding protein | 2,60E-03 | -3,74 | 2,07E-03 | -3,72 |
| WP_002288005.1 | aldo/keto reductase | 5,24E-04 | -5,13 | 1,88E-03 | -3,72 |
| WP_002296549.1 | hypothetical protein | 5,24E-04 | -5,49 | 2,36E-02 | -3,72 |
| WP_002288445.1 | ribosomal-protein-alanine N-acetyltransferase RimI | 5,24E-04 | -4,45 | 2,92E-02 | -3,73 |
| WP_002304106.1 | peptidase | 5,99E-04 | -4,60 | 1,58E-02 | -3,74 |
| WP_002286627.1 | cell division protein SepF | 1,67E-03 | -5,59 | 2,90E-02 | -3,74 |
| WP_002297350.1 | antirestriction protein ArdA | 5,24E-04 | -4,95 | 4,27E-02 | -3,74 |
| WP_002296282.1 | branched chain amino acid aminotransferase | 5,24E-04 | -5,32 | 1,88E-03 | -3,75 |
| WP_002289206.1 | 2-deoxyribose-5-phosphate aldolase | 5,24E-04 | -5,44 | 1,88E-03 | -3,75 |
| WP_002293906.1 | ferrichrome ABC transporter substrate-binding protein | 5,24E-04 | -5,16 | 9,83E-03 | -3,76 |
| WP_002294095.1 | uracil phosphoribosyltransferase | 5,24E-04 | -5,03 | 1,54E-02 | -3,79 |
| WP_002288135.1 | sugar ABC transporter substrate-binding protein | 5,24E-04 | -5,22 | 1,88E-03 | -3,81 |
| WP_002294955.1 | hypothetical protein | 1,54E-03 | -4,77 | 7,11E-03 | -3,83 |
| WP_002289619.1 | acetolactate decarboxylase | 6,19E-05 | -5,44 | 5,31E-04 | -3,83 |
| WP_002296996.1 | CoA-binding protein | 5,24E-04 | -5,39 | 4,27E-03 | -3,86 |
| WP_002288753.1 | GTPase Era | 5,24E-04 | -5,43 | 1,88E-03 | -3,87 |
| WP_002287743.1 | phosphate ABC transporter substrate-binding protein | 6,60E-04 | -4,51 | 1,88E-03 | -3,89 |
| WP_002294583.1 | L-lactate oxidase | 5,24E-04 | -5,52 | 1,88E-03 | -3,90 |
| WP_002288715.1 | DNA-directed RNA polymerase subunit alpha | 5,24E-04 | -5,67 | 2,43E-03 | -3,91 |
| WP_002287795.1 | NrdH-redoxin | 6,19E-05 | -5,31 | 3,17E-04 | -3,94 |
| WP_002295114.1 | isomerase | 6,19E-05 | -5,04 | 3,17E-04 | -3,96 |
| WP_002289138.1 | shikimate kinase | 6,37E-03 | -4,68 | 2,36E-02 | -3,97 |
| WP_002293709.1 | signal recognition particle protein | 2,02E-04 | -4,43 | 3,35E-02 | -3,98 |
| WP_002287754.1 | sugar ABC transporter ATP-binding protein | 5,24E-04 | -5,59 | 1,88E-03 | -4,01 |
| WP_002296582.1 | ROK family protein | 5,24E-04 | -4,49 | 1,54E-02 | -4,01 |
| WP_002287129.1 | ABC transporter permease | 6,19E-05 | -5,27 | 3,17E-04 | -4,02 |
| WP_002296055.1 | lysine--tRNA ligase | 6,19E-05 | -5,80 | 3,17E-04 | -4,03 |
| WP_002295465.1 | sugar ABC transporter permease | 5,24E-04 | -5,23 | 1,88E-03 | -4,04 |
| WP_002289121.1 | undecaprenyl-diphosphatase | 3,20E-04 | -5,19 | 3,17E-04 | -4,04 |
| NC_017032.1-gene39 | nucleotidyl transferase AbiEii/AbiGii toxin family protein | 6,19E-05 | -5,51 | 3,17E-04 | -4,04 |
| WP_002382867.1 | VanB-type vancomycin resistance histidine kinase VanS | 6,19E-05 | -5,36 | 3,17E-04 | -4,05 |
| WP_002286469.1 | aminopeptidase | 6,19E-05 | -5,46 | 3,17E-04 | -4,05 |
| WP_002289210.1 | lactoylglutathione lyase | 5,24E-04 | -6,04 | 4,94E-03 | -4,05 |
| WP_002296544.1 | dTDP-glucose 4%2C6-dehydratase | 6,19E-05 | -5,67 | 3,17E-04 | -4,06 |
| WP_002287439.1 | DNA-binding protein | 5,24E-04 | -6,04 | 1,88E-03 | -4,08 |
| WP_002288979.1 | cell wall anchor | 4,61E-04 | -5,40 | 2,09E-02 | -4,08 |
| WP_002286653.1 | ArgR family transcriptional regulator | 4,34E-03 | -5,41 | 1,88E-03 | -4,08 |
| WP_002285909.1 | TIGR01212 family radical SAM protein | 5,24E-04 | -5,55 | 1,88E-03 | -4,09 |
| WP_002287344.1 | metal ABC transporter permease | 6,19E-05 | -5,81 | 3,17E-04 | -4,10 |
| WP_002287953.1 | valine--tRNA ligase | 6,19E-05 | -6,24 | 3,17E-04 | -4,11 |
| WP_002288071.1 | DNA polymerase III subunit delta' | 6,19E-05 | -5,45 | 3,62E-03 | -4,12 |
| WP_002286006.1 | LacI family transcriptional regulator | 6,19E-05 | -5,84 | 3,17E-04 | -4,13 |

| WP_002296398.1 | dehydratase | 5,24E-04 | -4,79 | 6,30E-03 | -4,14 |
|---|---|---|---|---|---|
| WP_002286361.1 | acetyltransferase | 5,24E-04 | -5,30 | 6,98E-03 | -4,14 |
| WP_002288649.1 | FUSC family protein | 5,24E-04 | -6,02 | 1,88E-03 | -4,16 |
| WP_002288500.1 | transcription termination/antitermination protein NusA | 6,19E-05 | -5,60 | 3,17E-04 | -4,16 |
| WP_033700366.1 | cell wall anchor protein | 5,24E-04 | -5,23 | 1,00E-02 | -4,17 |
| WP_002296572.1 | repressor LexA | 1,64E-03 | -4,45 | 2,09E-02 | -4,18 |
| WP_014387013.1 | hypothetical protein | 5,24E-04 | -6,05 | 4,90E-02 | -4,18 |
| WP_002287161.1 | TIGR01906 family membrane protein | 5,24E-04 | -5,28 | 1,88E-03 | -4,19 |
| WP_002286616.1 | Cro/CI family transcriptional regulator | 5,24E-04 | -5,24 | 1,88E-03 | -4,21 |
| WP_002295194.1 | L-cystine transporter | 6,19E-05 | -5,50 | 3,17E-04 | -4,21 |
| WP_002289868.1 | threonine/serine exporter | 5,24E-04 | -5,46 | 7,99E-03 | -4,22 |
| WP_002288588.1 | two-component sensor histidine kinase | 6,19E-05 | -5,18 | 3,17E-04 | -4,22 |
| WP_002296474.1 | DUF4809 domain-containing protein | 1,19E-02 | -2,93 | 1,86E-02 | -4,25 |
| WP_002289473.1 | cardiolipin synthase | 5,24E-04 | -5,25 | 1,88E-03 | -4,25 |
| WP_002294533.1 | type I methionyl aminopeptidase | 6,19E-05 | -5,98 | 3,17E-04 | -4,28 |
| WP_002294107.1 | hypothetical protein | 6,37E-03 | -4,39 | 6,30E-03 | -4,29 |
| WP_002289091.1 | hypothetical protein | 5,24E-04 | -5,79 | 1,88E-03 | -4,29 |
| WP_014387131.1 | hypothetical protein | 7,61E-04 | -4,93 | 1,54E-02 | -4,31 |
| WP_002288089.1 | D-alanine--poly(phosphoribitol) ligase | 5,24E-04 | -5,80 | 1,88E-03 | -4,33 |
| WP_002288477.1 | Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase GatCAB subunit C | 9,33E-04 | -3,51 | 1,33E-02 | -4,34 |
| WP_002287080.1 | hypothetical protein | 5,24E-04 | -5,93 | 1,88E-03 | -4,35 |
| WP_002295123.1 | biotin/lipoate A/B protein ligase family protein | 5,24E-04 | -5,50 | 3,83E-03 | -4,38 |
| WP_044031087.1 | lactate dehydrogenase | 5,24E-04 | -6,14 | 1,79E-03 | -4,43 |
| WP_002287231.1 | RNA-binding protein | 3,27E-03 | -5,14 | 5,31E-04 | -4,45 |
| WP_002287824.1 | phosphoglucosamine mutase | 5,24E-04 | -5,42 | 1,88E-03 | -4,47 |
| WP_002288657.1 | 50S ribosomal protein L22 | 2,13E-02 | -3,34 | 4,27E-02 | -4,47 |
| WP_002286176.1 | competence protein ComGF | 6,19E-05 | -5,62 | 3,17E-04 | -4,49 |
| WP_002304032.1 | chromosome segregation protein SMC | 6,19E-05 | -6,08 | 3,17E-04 | -4,49 |
| WP_002288736.1 | DNA helicase RecQ | 6,19E-05 | -5,85 | 3,17E-04 | -4,50 |
| WP_002288107.1 | O-methyltransferase | 5,24E-04 | -5,64 | 2,39E-03 | -4,52 |
| WP_002317192.1 | 1-phosphofructokinase | 5,24E-04 | -6,15 | 1,88E-03 | -4,53 |
| WP_002289277.1 | MATE family efflux transporter | 6,19E-05 | -5,48 | 3,17E-04 | -4,54 |
| WP_002288962.1 | hypothetical protein | 1,28E-03 | -5,10 | 6,28E-03 | -4,54 |
| WP_002287619.1 | beta-ketoacyl-[acyl-carrier-protein] synthase II | 5,24E-04 | -6,66 | 1,88E-03 | -4,56 |
| WP_002297340.1 | hypothetical protein | 5,24E-04 | -5,84 | 1,88E-03 | -4,57 |
| WP_042959752.1 | ATPase | 5,24E-04 | -6,33 | 1,88E-03 | -4,58 |
| WP_002294192.1 | glycine/betaine ABC transporter ATP-binding protein | 5,24E-04 | -6,70 | 1,88E-03 | -4,60 |
| WP_002289337.1 | hypothetical protein | 5,24E-04 | -5,92 | 1,62E-02 | -4,60 |
| WP_002293681.1 | MFS transporter | 6,19E-05 | -5,38 | 3,17E-04 | -4,61 |
| WP_002296175.1 | Lsa family ABC-F type ribosomal protection protein | 6,19E-05 | -5,72 | 3,17E-04 | -4,64 |
| WP_014387074.1 | hypothetical protein | 5,11E-04 | -4,91 | 2,07E-03 | -4,64 |
| WP_002288966.1 | hypothetical protein | 5,24E-04 | -6,02 | 9,74E-03 | -4,65 |
| WP_002288415.1 | tyrosine recombinase XerC | 5,24E-04 | -4,90 | 7,99E-03 | -4,69 |
| WP_002295256.1 | membrane protein | 5,24E-04 | -6,20 | 1,88E-03 | -4,73 |
| WP_002297325.1 | IS30 family transposase | 5,24E-04 | -6,15 | 1,88E-03 | -4,76 |
| WP_002289620.1 | acetolactate synthase AlsS | 6,19E-05 | -4,17 | 3,17E-04 | -4,76 |
| WP_002325125.1 | NUDIX hydrolase | 5,24E-04 | -5,46 | 1,88E-03 | -4,76 |
| WP_002287961.1 | L-asparaginase | 5,24E-04 | -5,11 | 3,62E-03 | -4,76 |
| WP_002296589.1 | cardiolipin synthase | 6,19E-05 | -5,99 | 3,17E-04 | -4,79 |
| WP_002286211.1 | 3'-5' exoribonuclease YhaM | 6,19E-05 | -6,04 | 3,17E-04 | -4,80 |
| WP_002289069.1 | glycosyl transferase family 1 | 6,19E-05 | -6,28 | 3,17E-04 | -4,80 |
| WP_002292069.1 | trigger factor | 6,19E-05 | -5,90 | 3,17E-04 | -4,80 |
| WP_002293992.1 | hypothetical protein | 6,60E-04 | -4,98 | 5,76E-03 | -4,80 |
| WP_002286875.1 | 2%2C3%2C4%2C5-tetrahydropyridine-2%2C6-dicarboxylate N-acetyltransferase | 5,11E-04 | -5,95 | 4,75E-03 | -4,84 |

| | | | | | |
|---|---|---|---|---|---|
| WP_002294701.1 | hypothetical protein | 1,28E-03 | -5,52 | 4,75E-03 | -4,84 |
| WP_002287905.1 | iron ABC transporter permease | 5,24E-04 | -5,77 | 1,79E-03 | -4,84 |
| WP_002300928.1 | DNA-binding protein | 1,67E-03 | -5,38 | 2,39E-03 | -4,85 |
| WP_002292337.1 | nucleoid-associated protein%2C YbaB/EbfC family | 5,24E-04 | -5,80 | 1,88E-03 | -4,86 |
| WP_002288690.1 | adenylate kinase | 5,24E-04 | -6,59 | 1,88E-03 | -4,87 |
| WP_002289057.1 | lipoprotein | 6,19E-05 | -4,87 | 3,17E-04 | -4,88 |
| WP_002289686.1 | dihydrolipoyl dehydrogenase | 6,19E-05 | -6,50 | 3,17E-04 | -4,88 |
| WP_002296885.1 | helicase-exonuclease AddAB subunit AddA | 6,19E-05 | -6,66 | 3,17E-04 | -4,92 |
| WP_002287346.1 | bifunctional N-acetylglucosamine-1-phosphate uridyltransferase/glucosamine-1-phosphate acetyltransferase | 5,24E-04 | -6,58 | 1,88E-03 | -4,95 |
| WP_002288242.1 | bifunctional 2-keto-4-hydroxyglutarate aldolase/2-keto-3-deoxy-6-phosphogluconate aldolase | 5,24E-04 | -5,89 | 9,74E-03 | -4,96 |
| WP_014387054.1 | hypothetical protein | 1,69E-03 | -4,00 | 3,35E-02 | -4,97 |
| WP_002289734.1 | type I glyceraldehyde-3-phosphate dehydrogenase | 5,24E-04 | -5,93 | 2,22E-03 | -4,98 |
| WP_002294057.1 | fructose-1%2C6-bisphosphate aldolase%2C class II | 5,24E-04 | -6,70 | 1,88E-03 | -5,02 |
| WP_002298392.1 | hypothetical protein | 5,24E-04 | -6,30 | 2,39E-03 | -5,04 |
| WP_002289610.1 | DNA replication protein DnaD | 5,24E-04 | -6,14 | 1,88E-03 | -5,06 |
| WP_002288475.1 | Asp-tRNA(Asn)/Glu-tRNA(Gln) amidotransferase GatCAB subunit A | 6,19E-05 | -7,29 | 3,17E-04 | -5,07 |
| WP_002303833.1 | penicillin-binding protein 4 | 6,19E-05 | -6,75 | 3,17E-04 | -5,08 |
| WP_002288559.1 | WxL domain-containing protein | 5,24E-04 | -5,91 | 1,88E-03 | -5,08 |
| WP_002297345.1 | conjugal transfer protein | 5,24E-04 | -6,36 | 1,88E-03 | -5,08 |
| WP_002285994.1 | YlbF family regulator | 1,28E-03 | -5,37 | 6,08E-03 | -5,08 |
| WP_002321602.1 | hypothetical protein | 1,28E-03 | -5,02 | 4,75E-03 | -5,08 |
| WP_002287165.1 | NAD(P)/FAD-dependent oxidoreductase | 8,12E-04 | -4,98 | 1,54E-02 | -5,10 |
| WP_002294585.1 | L-serine dehydratase%2C iron-sulfur-dependent subunit beta | 6,19E-05 | -6,40 | 3,17E-04 | -5,12 |
| WP_002296444.1 | inosine-uridine preferring nucleoside hydrolase | 6,19E-05 | -6,93 | 3,17E-04 | -5,15 |
| WP_002286728.1 | metallophosphatase | 1,50E-02 | -4,26 | 9,59E-03 | -5,15 |
| WP_002289284.1 | peptide deformylase | 5,24E-04 | -6,64 | 1,88E-03 | -5,18 |
| WP_002287389.1 | membrane protein | 3,94E-03 | -4,43 | 9,83E-03 | -5,18 |
| WP_002304795.1 | hypothetical protein | 5,24E-04 | -5,82 | 2,39E-03 | -5,22 |
| WP_002295261.1 | tRNA uridine-5-carboxymethylaminomethyl(34) synthesis GTPase MnmE | 6,19E-05 | -4,58 | 2,63E-02 | -5,22 |
| WP_014387060.1 | hypothetical protein | 4,27E-03 | -4,66 | 1,47E-02 | -5,23 |
| WP_002317291.1 | hypothetical protein | 6,19E-05 | -5,67 | 3,17E-04 | -5,26 |
| WP_002289810.1 | hypothetical protein | 5,11E-04 | -5,77 | 1,52E-03 | -5,27 |
| WP_002289105.1 | aspartate--tRNA ligase | 5,24E-04 | -7,12 | 1,88E-03 | -5,27 |
| WP_002286933.1 | hypothetical protein | 9,33E-04 | -5,16 | 1,30E-03 | -5,28 |
| WP_002317292.1 | conjugative transfer protein | 5,11E-04 | -6,47 | 2,07E-03 | -5,30 |
| WP_002289198.1 | purine-nucleoside phosphorylase | 5,11E-04 | -5,97 | 1,52E-03 | -5,33 |
| WP_002322015.1 | PspC family transcriptional regulator | 4,61E-04 | -5,73 | 1,64E-03 | -5,35 |
| WP_002288258.1 | TRAP transporter permease DctQ | 4,31E-03 | -4,50 | 2,51E-03 | -5,35 |
| WP_002321528.1 | hypothetical protein | 5,24E-04 | -6,31 | 2,07E-03 | -5,35 |
| WP_002286655.1 | TlyA family rRNA (cytidine-2'-O)-methyltransferase | 5,11E-04 | -6,63 | 1,52E-03 | -5,35 |
| WP_002295461.1 | N-acetylmuramoyl-L-alanine amidase | 1,67E-03 | -4,39 | 1,01E-02 | -5,38 |
| WP_002287347.1 | ribose-phosphate pyrophosphokinase | 6,19E-05 | -7,08 | 3,17E-04 | -5,39 |
| WP_002304808.1 | heptaprenyl diphosphate synthase subunit I | 5,24E-04 | -6,63 | 1,88E-03 | -5,40 |
| WP_002287124.1 | sugar isomerase | 5,24E-04 | -6,27 | 4,77E-03 | -5,42 |
| WP_002296291.1 | hypothetical protein | 5,24E-04 | -7,04 | 1,79E-03 | -5,42 |
| WP_002321810.1 | DNA methyltransferase | 5,24E-04 | -5,77 | 2,70E-02 | -5,43 |
| WP_002288360.1 | chromosomal replication initiation protein DnaA | 5,24E-04 | -6,05 | 9,74E-03 | -5,49 |

| | | | | | |
|---|---|---|---|---|---|
| WP_002286660.1 | bifunctional methylenetetrahydrofolate dehydrogenase/methenyltetrahydrofolate cyclohydrolase | 5,24E-04 | -5,90 | 2,07E-03 | -5,50 |
| WP_002289325.1 | methionine--tRNA ligase | 6,19E-05 | -7,49 | 3,17E-04 | -5,54 |
| WP_002286668.1 | DUF956 domain-containing protein | 5,24E-04 | -6,51 | 2,56E-03 | -5,55 |
| WP_002295807.1 | hypothetical protein | 5,24E-04 | -6,50 | 2,39E-03 | -5,59 |
| WP_002286869.1 | aminotransferase A | 5,24E-04 | -7,10 | 1,88E-03 | -5,60 |
| WP_002289439.1 | S-ribosylhomocysteine lyase | 4,61E-04 | -6,29 | 3,37E-03 | -5,61 |
| WP_002296152.1 | phosphoglucomutase | 6,19E-05 | -7,05 | 3,17E-04 | -5,63 |
| WP_002295469.1 | xanthine phosphoribosyltransferase | 5,24E-04 | -6,45 | 2,56E-03 | -5,66 |
| WP_014387010.1 | sugar ABC transporter permease | 5,24E-04 | -6,69 | 1,88E-03 | -5,67 |
| WP_002288727.1 | tRNA pseudouridine(38-40) synthase TruA | 5,24E-04 | -7,08 | 1,88E-03 | -5,68 |
| WP_002286674.1 | hypothetical protein | 5,11E-04 | -6,02 | 4,75E-03 | -5,80 |
| WP_002288098.1 | hypothetical protein | 6,60E-04 | -5,38 | 1,88E-03 | -5,83 |
| WP_002289452.1 | 50S ribosomal protein L27 | 5,24E-04 | -5,39 | 5,54E-03 | -5,87 |
| WP_002287538.1 | hypothetical protein | 5,24E-04 | -6,60 | 1,88E-03 | -5,91 |
| WP_002286808.1 | DUF159 family protein | 4,61E-04 | -7,05 | 1,30E-03 | -5,93 |
| WP_002288637.1 | DeoR/GlpR transcriptional regulator | 3,20E-04 | -7,52 | 1,51E-03 | -6,00 |
| WP_002296179.1 | amino acid permease | 6,19E-05 | -7,54 | 3,17E-04 | -6,01 |
| WP_002296525.1 | nucleotide exchange factor GrpE | 6,19E-05 | -7,60 | 3,17E-04 | -6,09 |
| WP_002288976.1 | cell wall anchor | 5,24E-04 | -7,39 | 1,88E-03 | -6,10 |
| WP_010815683.1 | glycopeptide resistance accessory protein VanW-B | 5,24E-04 | -7,55 | 1,79E-03 | -6,18 |
| WP_002287350.1 | patatin | 6,19E-05 | -7,70 | 3,17E-04 | -6,18 |
| WP_002295088.1 | colicin V production protein CvpA | 6,60E-04 | -8,04 | 1,88E-03 | -6,24 |
| WP_002317262.1 | hypothetical protein | 5,11E-04 | -6,25 | 1,52E-03 | -6,25 |
| WP_002287554.1 | hypothetical protein | 5,24E-04 | -7,68 | 1,88E-03 | -6,27 |
| WP_002287328.1 | DNA recombination/repair protein RecA | 5,24E-04 | -8,35 | 1,88E-03 | -6,31 |
| WP_002297139.1 | metallophosphoesterase | 4,61E-04 | -5,92 | 3,37E-03 | -6,33 |
| WP_002295844.1 | membrane protein | 5,24E-04 | -7,54 | 1,88E-03 | -6,35 |
| WP_002289509.1 | folate ECF transporter | 1,70E-02 | -5,53 | 7,79E-03 | -6,36 |
| WP_071974492.1 | hypothetical protein | 3,77E-04 | -5,65 | 1,47E-02 | -6,37 |
| NC_017032.1-gene10 | class C sortase | 5,24E-04 | -7,02 | 1,88E-03 | -6,41 |
| WP_014387039.1 | endopeptidase | 5,24E-04 | -6,00 | 1,62E-02 | -6,42 |
| WP_002287900.1 | signal recognition particle-docking protein FtsY | 6,19E-05 | -7,59 | 3,17E-04 | -6,46 |
| WP_001009054.1 | hypothetical protein | 5,11E-04 | -7,60 | 1,52E-03 | -6,49 |
| WP_002321654.1 | bacteriocin-type signal sequence | 5,24E-04 | -7,91 | 1,79E-03 | -6,49 |
| WP_002288444.1 | tRNA (adenosine(37)-N6)-threonylcarbamoyltransferase complex transferase subunit TsaD | 6,19E-05 | -8,04 | 3,17E-04 | -6,52 |
| WP_002296499.1 | hypothetical protein | 5,11E-04 | -7,24 | 4,75E-03 | -6,55 |
| WP_002289010.1 | ATP synthase F0 subunit B | 4,61E-04 | -6,83 | 1,30E-03 | -6,57 |
| WP_002287607.1 | hypothetical protein | 4,61E-04 | -6,28 | 3,37E-03 | -6,58 |
| WP_002295502.1 | esterase | 3,77E-04 | -5,54 | 1,47E-02 | -6,61 |
| WP_002285954.1 | PTS cellobiose transporter subunit IIA | 5,24E-04 | -8,10 | 1,88E-03 | -6,62 |
| WP_002296902.1 | hypothetical protein | 4,61E-04 | -6,75 | 1,30E-03 | -6,67 |
| WP_002294584.1 | L-serine dehydratase%2C iron-sulfur-dependent subunit alpha | 5,24E-04 | -7,44 | 1,88E-03 | -6,71 |
| WP_002287621.1 | beta-hydroxyacyl-ACP dehydratase | 5,24E-04 | -7,43 | 1,88E-03 | -6,72 |
| WP_002293782.1 | hypothetical protein | 4,61E-04 | -6,90 | 1,30E-03 | -6,72 |
| WP_014387041.1 | phage tail protein | 1,77E-03 | -5,79 | 3,17E-04 | -6,74 |
| WP_002289134.1 | response regulator | 3,77E-04 | -5,75 | 2,17E-03 | -6,74 |
| WP_002296580.1 | N-acetyltransferase | 5,24E-04 | -8,24 | 1,88E-03 | -6,82 |
| WP_002287763.1 | hypothetical protein | 4,61E-04 | -6,94 | 1,30E-03 | -6,88 |
| NC_017024.1-gene2 | hypothetical protein | 5,24E-04 | -7,47 | 1,79E-03 | -6,90 |
| WP_002317278.1 | plasmid mobilization relaxosome protein MobC | 3,77E-04 | -5,91 | 5,76E-03 | -6,94 |
| WP_002297348.1 | ATP/GTP-binding protein | 1,12E-04 | -6,58 | 9,55E-04 | -6,98 |

| | | | | | |
|---|---|---|---|---|---|
| **WP_002289692.1** | pyruvate dehydrogenase (acetyl-transferring) E1 component subunit alpha | 4,61E-04 | -6,51 | 3,17E-04 | -7,03 |
| **WP_002296829.1** | hydrolase | 5,24E-04 | -6,17 | 3,79E-03 | -7,06 |
| **WP_002296402.1** | SprT family protein | 6,19E-05 | -8,67 | 3,17E-04 | -7,08 |
| **WP_002304099.1** | DUF998 domain-containing protein | 5,11E-04 | -7,79 | 1,52E-03 | -7,08 |
| **WP_002287484.1** | branched-chain amino acid ABC transporter | 3,77E-04 | -5,74 | 2,17E-03 | -7,12 |
| **WP_002296598.1** | hypothetical protein | 5,11E-04 | -7,90 | 1,52E-03 | -7,19 |
| **WP_002286774.1** | peptidase | 2,02E-04 | -6,50 | 9,55E-04 | -7,21 |
| **WP_014387052.1** | amidoligase | 5,11E-04 | -7,91 | 1,52E-03 | -7,22 |
| **WP_014387047.1** | phage major capsid protein | 3,77E-04 | -6,69 | 9,55E-04 | -7,22 |
| **WP_002289327.1** | NUDIX hydrolase | 5,24E-04 | -9,28 | 1,88E-03 | -7,30 |
| **WP_002287602.1** | phosphocarrier protein HPr | 5,24E-04 | -8,08 | 1,88E-03 | -7,37 |
| **WP_014387032.1** | recombinase | 5,24E-04 | -7,90 | 6,30E-03 | -7,38 |
| **WP_002317279.1** | endonuclease | 5,24E-04 | -8,71 | 1,88E-03 | -7,45 |
| **WP_002296503.1** | Holliday junction resolvase RuvX | 6,19E-05 | -9,18 | 3,17E-04 | -7,45 |
| **WP_002288635.1** | PTS lactose/cellobiose transporter subunit IIA | 5,11E-04 | -7,12 | 4,75E-03 | -7,46 |
| **WP_002298265.1** | two-component sensor histidine kinase | 6,19E-05 | -8,59 | 3,17E-04 | -7,48 |
| **WP_002288970.1** | transcriptional regulator | 5,11E-04 | -7,94 | 1,52E-03 | -7,48 |
| **WP_002287604.1** | glycosyl transferase | 6,19E-05 | -8,70 | 3,17E-04 | -7,61 |
| **WP_002287651.1** | threonine--tRNA ligase | 6,19E-05 | -8,79 | 3,17E-04 | -7,67 |
| **WP_002296185.1** | methionine ABC transporter substrate-binding protein | 5,24E-04 | -8,64 | 1,88E-03 | -7,68 |
| **WP_014387051.1** | gamma-glutamylcyclotransferase | 4,61E-04 | -7,58 | 1,30E-03 | -7,70 |
| **WP_044031076.1** | head-tail adaptor protein | 4,61E-04 | -7,40 | 1,30E-03 | -7,83 |
| **WP_002289292.1** | hypothetical protein | 5,24E-04 | -9,05 | 1,88E-03 | -7,87 |
| **WP_042959759.1** | tRNA-specific adenosine deaminase | 5,24E-04 | -9,45 | 1,88E-03 | -7,90 |
| **WP_044031432.1** | toxin secretion/phage lysis holin | 3,77E-04 | -5,99 | 2,17E-03 | -7,93 |
| **WP_002289040.1** | copper-binding protein | 5,11E-04 | -8,03 | 1,52E-03 | -7,94 |
| **WP_002320968.1** | hypothetical protein | 5,24E-04 | -8,87 | 1,79E-03 | -8,04 |
| **WP_002288471.1** | diacylglycerol kinase | 3,20E-04 | -7,41 | 3,17E-04 | -8,11 |
| **WP_002290045.1** | hypothetical protein | 1,69E-03 | -4,82 | 2,17E-03 | -9,08 |
| **WP_002288642.1** | hypothetical protein | 5,24E-04 | -10,43 | 1,88E-03 | -9,21 |
| **WP_002296613.1** | toxin | 3,77E-04 | -8,51 | 9,55E-04 | -9,23 |
| **WP_002287367.1** | primosomal protein N' | 2,02E-04 | -9,06 | 9,55E-04 | -9,57 |
| **WP_002292860.1** | cold-shock protein | 3,77E-04 | -8,93 | 9,55E-04 | -10,32 |
| **WP_014387005.1** | hypothetical protein | 3,77E-04 | -10,14 | 9,55E-04 | -11,08 |
| **WP_002294265.1** | 30S ribosomal protein S4 | 5,24E-04 | -12,49 | 1,88E-03 | -11,73 |
| **WP_009269250.1** | hypothetical protein | 3,77E-04 | -10,67 | 9,55E-04 | -11,86 |
| **WP_014387066.1** | hypothetical protein | 3,77E-04 | -10,63 | 9,55E-04 | -12,03 |
| **WP_002321678.1** | hypothetical protein | 5,24E-04 | -11,57 | 3,17E-04 | -12,60 |

[a] Indicates the protein ID number in the Aus0004 genome. [b] Indicates the predicted function. [c] Indicates the adjusted p-value (FDR) obtained applying Wilcoxon test comparing *in vitro* vs colon samples. [d] Log2 Fold-change (FC) derived from the ratio of the relative abundance of the Tn-mutant present in *in vitro* as compare to *in vivo* (colon) samples. A positive FC indicates that the gene was found to be in higher levels *in vitro*. [e] Indicates the adjusted p-value (FDR) obtained applying Wilcoxon test comparing *in vitro* vs caecum samples. [f] Log2 Fold-change (FC) derived from the ratio of the relative abundance of the Tn-mutant present in *in vitro* as compare to *in vivo* (caecum) samples. A positive FC indicates that the gene was found to be in higher levels *in vitro*.