



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Escola Tècnica Superior  
d'Enginyeria Agronòmica i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Agronómica  
y del Medio Natural

Optimización de un procedimiento de secuenciación de  
nueva generación de amplicones cortos para la  
identificación de alelos de riesgo de enfermedad celíaca e  
intolerancia a la lactosa.

Trabajo Fin de Grado

Grado en Biotecnología

AUTOR/A: Millán Castillo, Miriam

Tutor/a: Gadea Vacas, José

Cotutor/a externo: CHAVES MARTINEZ, FELIPE JAVIER

Director/a Experimental: GARCIA GARCIA, ANA BARBARA

CURSO ACADÉMICO: 2022/2023

# UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA  
AGRONÒMICA I DEL MEDI NATURAL (ETSIAMN)



TRABAJO FIN DE GRADO EN BIOTECNOLOGÍA

## **Optimización de un procedimiento de Secuenciación de Nueva Generación de amplicones cortos para la identificación de alelos de riesgo de enfermedad celíaca e intolerancia a la lactosa**

**AUTORA:**

Miriam Millán Castillo

**TUTOR:**

José Gadea Vacas

**COTUTOR EXTERNO:**

Felipe Javier Chaves Martínez

**DIRECTORA EXPERIMENTAL:**

Ana Bárbara García García

Curso académico 2022/2023

Valencia, 3 de julio de 2023

## TÍTULO

Optimización de un procedimiento de Secuenciación de Nueva Generación de amplicones cortos para la identificación de alelos de riesgo de enfermedad celíaca e intolerancia a la lactosa.

## RESUMEN

Las intolerancias alimentarias son relativamente frecuentes en la población adulta e infantil, siendo las intolerancias al gluten (celiaquía o enfermedad celíaca (EC)) y a la lactosa (no persistencia de la lactasa) las más frecuentes. Existe muy poca información sobre su prevalencia individual y conjunta, tanto en población europea como en España en particular, debido a que su diagnóstico sigue siendo en muchos casos todo un desafío. Los individuos afectados presentan, en ambos casos, signos y síntomas comunes con otras patologías, dificultando su identificación precisa. Resulta, por tanto, muy interesante estudiar los marcadores genéticos de riesgo descritos para ambas intolerancias en población española, para lo cual la tecnología de Secuenciación de Nueva Generación (NGS) se presenta como una opción altamente ventajosa, al ofrecer una gran resolución, sensibilidad y eficiencia en la identificación de variantes genéticas, incluso en regiones complejas o hipervariables. Se conoce que la presencia de determinados haplotipos de los genes *DQA1* y *DQB1* del sistema HLA confieren riesgo de desarrollar EC, y que 5 SNPs de la zona 5' del gen *LCT* (codificante de la lactasa) confieren riesgo de desarrollar intolerancia a la lactosa. Por consiguiente, el análisis de estos alelos resulta fundamental para poder estudiar su frecuencia y la implicación de cada uno de ellos en sus respectivas intolerancias y posible interacción. El análisis de amplicones cortos mediante la tecnología de Secuenciación de Nueva Generación (NGS) se presenta como una buena opción para llevar a cabo este tipo de estudios. Por esta razón, el presente trabajo experimental tiene como objetivo principal la optimización de un procedimiento de NGS por amplicones cortos para la identificación de los alelos descritos en literatura científica como alelos de riesgo para EC e intolerancia a la lactosa. Para ello, se procede al diseño de cebadores específicos que permitan amplificar las regiones de interés de los genes *LCT*, *HLA-DQA1* y *HLA-DQB1* para su posterior secuenciación. Con el propósito de obtener una cobertura homogénea para todos los fragmentos amplificados, resulta necesario optimizar las condiciones de generación de librerías. Para ello, se realizan diversas pruebas utilizando diferentes combinaciones de cebadores, concentraciones y aditivos, determinando, a partir de los resultados obtenidos, las condiciones óptimas que permiten la correcta identificación de las variantes de interés. De este modo, la secuenciación masiva de amplicones cortos ha demostrado ser una herramienta altamente eficiente para la detección de variantes alélicas de riesgo en las patologías estudiadas, pudiendo en un futuro implantarse en el ámbito clínico, brindando resultados que faciliten un preciso diagnóstico. Este trabajo se relaciona con los siguientes ODS de la Agenda 2030: Salud y Bienestar (ODS 3) y, en menor medida, con Hambre Cero (ODS 2) y Reducción de las Desigualdades (ODS 10).

**Palabras clave:** NGS, amplicones, celiacía, intolerancia a la lactosa, HLA, coberturas.

**Autora:** Miriam Millán Castillo.

**Localidad y fecha:** Valencia, 3 de julio de 2023.

**Tutor:** José Gadea Vacas.

**Cotutor externo:** Felipe Javier Chaves Martínez.

**Directora experimental:** Ana Bárbara García García.

## TITLE

Optimization of a Next-Generation Sequencing procedure for short amplicon for the identification of risk alleles for celiac disease and lactose intolerance.

## ABSTRACT

Food intolerances are relatively common in both adults and children population, being gluten intolerance (celiac disease (CD)) and lactose intolerance (lactase non-persistence) the most frequent. There is very little information on the individual and combined prevalence of these intolerances, both in the European population and particularly in Spain. Their diagnosis remains challenging in many cases. Affected individuals present, in both cases, common signs and symptoms that overlap with other pathologies, making their accurate identification difficult. It is therefore very interesting to study the genetic risk markers described for both intolerances in the Spanish population, for which the New Generation Sequencing (NGS) technology is presented as a highly advantageous option, as it offers high resolution, sensitivity, and efficiency in the identification of genetic variants, even in complex or hypervariable regions. It is known that the presence of certain haplotypes of the *DQA1* and *DQB1* genes of the HLA system confer a risk of developing CD, and that 5 SNPs in the 5' zone of the *LCT* gene (coding for lactase) confer a risk of developing lactose intolerance. Therefore, the analysis of these alleles is essential to study their frequency and their association with their respective intolerances, as well as possible interactions. The utilization of New Generation Sequencing (NGS) technology for analyzing short amplicons is presented as a good option for conducting such studies. For this reason, the aim of this experimental study is to optimize an NGS protocol using short amplicons for the identification of alleles described in scientific literature as risk alleles for CD and lactose intolerance. To achieve this, specific primers are designed to amplify the target regions of the *LCT*, *HLA-DQA1* and *HLA-DQB1* genes for subsequent sequencing analysis. To obtain a homogeneous coverage for amplified fragments, it is essential to optimize the conditions for library generation. To accomplish this, multiple tests are conducted, employing different combinations of primers, concentrations, and additives. Based on the obtained results, the optimal conditions are determined, ensuring accurate identification of the targeted variants of interest. In this way, the utilization of high-throughput sequencing of short amplicons has proven to be a highly efficient tool for detecting risk allelic variants in the studied pathologies. This approach holds potential for future implementation in the clinical setting, providing results that facilitate an accurate and precise diagnosis. This work is related to the following SDGs of the 2030 Agenda: Health and Well-being (SDG 3) and, to a lesser extent, Zero Hunger (SDG 2) and Reduction of Inequalities (SDG 10).

**Key words:** NGS, amplicons, celiac disease, lactose intolerance, HLA, coverages.

**Author:** Miriam Millán Castillo.

**Location and date:** Valencia, 3 de julio de 2023.

**Tutor:** José Gadea Vacas.

**External cotutor:** Felipe Javier Chaves Martínez.

**Experimental director:** Ana Bárbara García García.

## **AGRADECIMIENTOS**

En primer lugar, quiero expresar mi profundo agradecimiento a Javier por brindarme la oportunidad de embarcarme en este apasionante proyecto.

Gracias Ana Bárbara, por confiar en mí, guiarme, ayudarme en todo momento y responder a mis innumerables dudas.

A Fran y Elena, gracias por todo lo que me habéis enseñado, por vuestra infinita paciencia, vuestra dedicación y por todo vuestro apoyo; sin vuestra ayuda este trabajo no hubiera sido posible. Os deseo lo mejor en vuestros futuros proyectos.

En general, a todo el equipo de la UGD, gracias por haberme acogido tan bien y hacerme la estancia tan agradable.

Gracias a mi familia, por todo su apoyo durante todos estos años y por haberme dado la oportunidad de estudiar esta carrera tan bonita y que tantas cosas buenas me ha traído.

A mis amigas, porque a pesar de estar lejos he sentido todo vuestro apoyo igual de cerca que siempre. Gracias por toda vuestra confianza en mí.

Y, por último, a lo más bonito que me llevo de esta carrera, los que empezaron siendo amigos y han terminado siendo mi familia en Valencia, gracias a todos por esos buenos momentos, risas, comidas en familia, viajes y también por todos esos malos momentos vividos, pero que nos unirán por siempre.

# ÍNDICE

1.	INTRODUCCIÓN.....	1
1.1.	ENFERMEDAD CELIACA (EC).....	1
1.1.1.	Patogenia de la enfermedad celíaca (EC).....	1
1.1.2.	Signos y síntomas.....	3
1.1.3.	Diagnóstico y tratamiento.....	4
1.2.	INTOLERANCIA A LA LACTOSA.....	5
1.2.1.	Patogenia de la intolerancia a la lactosa.....	6
1.2.2.	Signos y síntomas.....	7
1.2.3.	Diagnóstico y tratamiento.....	7
1.3.	TECNOLOGÍA NGS.....	9
2.	OBJETIVOS.....	10
3.	MATERIALES Y MÉTODOS.....	11
3.1.	DISEÑO EXPERIMENTAL.....	11
3.2.	OBTENCIÓN DE MUESTRAS.....	11
3.3.	DISEÑO DE CEBADORES.....	11
3.3.1.	Regiones amplificadas.....	13
3.4.	GENERACIÓN Y OPTIMIZACIÓN DE LIBRERÍAS.....	13
3.4.1.	Optimización PCR multiplex (PCR 1).....	13
3.4.2.	Optimización PCR anidada (PCR 2).....	14
3.4.3.	Electroforesis capilar.....	16
3.5.	PURIFICACIÓN DE LIBRERÍAS.....	16
3.6.	CUANTIFICACIÓN.....	17
3.7.	SECUENCIACIÓN.....	18
3.8.	ANÁLISIS BIOINFORMÁTICO.....	19
3.9.	OPTIMIZACIÓN.....	20
4.	RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS.....	20
4.1.	DISEÑO DE CEBADORES.....	20
4.2.	OPTIMIZACIÓN DEL PANEL.....	21
4.2.1.	Prueba 1.....	22
4.2.2.	Prueba 2.....	24
4.2.3.	Prueba 3.....	27
4.2.4.	Prueba 4.....	28
4.2.5.	Prueba 5.....	31
4.2.6.	Prueba 6.....	34

4.2.7. Prueba 7 .....	35
5. CONCLUSIONES.....	38
6. REFERENCIAS BIBLIOGRÁFICAS.....	39
7. ANEXOS.....	41

## ÍNDICE DE FIGURAS

Figura 1. Factores genéticos y ambientales propuestos como contribuyentes al desarrollo de la enfermedad celíaca. ....	2
Figura 2. Distribución geográfica mundial de las frecuencias de la no persistencia de la lactasa (LNP).....	6
Figura 3. Principales métodos actuales para el diagnóstico de la intolerancia a la lactosa, así como las ventajas e inconvenientes que cada uno de ellos presenta.....	8
Figura 4. Diagrama del diseño experimental llevado a cabo para el proceso de optimización. ....	11
Figura 5. Representación de la división de cebadores contiguos en diferentes pools.....	12
Figura 6. Representación del proceso de generación de librerías. ....	15
Figura 7. Representación del proceso de purificación de la librería mediante el kit de purificación Magsi-NGS PREP® .....	17
Figura 8. Ejemplo de recta patrón para la cuantificación de librerías.....	18
Figura 9. Diagrama del proceso de secuenciación en el sistema MiSeq de Illumina.. ....	19
Figura 10. Sección de la secuencia de referencia del exón 2 de los genes HLA-DQA1 y HLA-DQB1. ....	21
Figura 11. SNPs identificados en el exón 2 del gen HLA-DQA1. ....	21
Figura 12. Resultados de la electroforesis capilar para la prueba 1.....	23
Figura 13. Distribución de lecturas mínimas obtenidas por amplicón secuenciado en el panel DQ-LAC de la prueba 1. ....	24
Figura 14. Distribución de lecturas mínimas obtenidas por amplicón secuenciado en el panel DQ de la prueba 1.. ....	24
Figura 15. Resultados de la electroforesis capilar para la condición 1 de la prueba 2.....	26
Figura 16. Resultados de la electroforesis capilar para la condición 2 de la prueba 2. ....	26
Figura 17. Distribución de lecturas mínimas obtenidas por amplicón secuenciado según las combinaciones de cebadores utilizadas en el panel DQ-LAC de la prueba 2.....	27
Figura 18. Distribución de lecturas mínimas obtenidas por amplicón secuenciado según las combinaciones de cebadores utilizadas en el panel DQ de la prueba 2. ....	27
Figura 19. Resultados de la electroforesis capilar para la prueba 3.....	28
Figura 20. Resultados de la electroforesis capilar para la prueba 4.....	30
Figura 21. Distribución de lecturas mínimas obtenidas por amplicón secuenciado según las combinaciones de cebadores utilizadas en el panel DQ-LAC para la prueba 4.. ....	30
Figura 22. Resultados de la electroforesis capilar para las combinaciones 1.1, 1.2 y 1.3 de la prueba 5.....	32
Figura 23. Resultados de la electroforesis capilar para las combinaciones 2.1, 2.2 y 2.3 de la prueba 5.....	32
Figura 24. Resultados de la electroforesis capilar para las combinaciones 5.1, 5.2 y 5.3 de la prueba 5.....	32
Figura 25. Distribución de lecturas mínimas obtenidas por amplicón secuenciado para las combinaciones de cebadores 1.1, 2.1 y 5.1 utilizadas en la condición 1 del panel DQ-LAC.....	33



Figura 26. Distribución de lecturas mínimas obtenidas por amplicón secuenciado para las combinaciones de cebadores 1.2, 2,2 y 5.2 utilizadas en la condición 2 del panel DQ-LAC.....	33
Figura 27. Distribución de lecturas mínimas obtenidas por amplicón secuenciado para las combinaciones de cebadores 1.3, 2.3 y 5.3 utilizadas en la condición 3 del panel DQ-LAC.....	34
Figura 28. Resultados de la electroforesis capilar para la prueba 6.....	35
Figura 29. Distribución de lecturas mínimas obtenidas por amplicón secuenciado para las condiciones del panel DQ-LAC utilizadas en la prueba 6.....	35
Figura 30. Resultados de la electroforesis capilar para la prueba 6.....	36
Figura 31. Distribución de lecturas mínimas obtenidas por amplicón secuenciado para las condiciones del panel DQ-LAC utilizadas en la prueba 7.....	37
Figura 32. Calidad de secuenciación por base y muestra secuenciada obtenida mediante el análisis en el programa FastQC..	37

## ÍNDICE DE TABLAS

Tabla 1. Factores genéticos asociados a la aparición de la enfermedad celíaca. ....	3
Tabla 2. Diversas variantes genéticas estudiadas asociadas a la persistencia de lactasa (LP).....	7
Tabla 3. Variantes genéticas del gen LCT estudiadas con su nomenclatura tradicional y su equivalencia en la nomenclatura actual. ....	13
Tabla 4. Reactivos de partida utilizados para la realización de la PCR multiplex (PCR 1).....	14
Tabla 5. Condiciones iniciales utilizadas en la PCR multiplex (PCR 1). ....	14
Tabla 6. Reactivos utilizados para la realización de la PCR anidada (PCR 2). ....	16
Tabla 7. Condiciones utilizadas en la PCR anidada (PCR 2). ....	16
Tabla 8. Distribución de cebadores utilizada en la prueba 1.. ....	22
Tabla 9. Condiciones de amplificación de la PCR 1 utilizadas en la prueba 1. ....	22
Tabla 10. Condiciones de amplificación de la PCR 2 utilizadas en la prueba 1. ....	22
Tabla 11. Distribución de cebadores utilizada en la condición 1 de la prueba 2. ....	25
Tabla 12. Distribución de cebadores utilizada en la condición 2 de la prueba 2. ....	25
Tabla 13. Condiciones probadas en la prueba 4 según las posibles combinaciones de cebadores con la pareja de cebadores correspondiente al gen HLA-DQB1-E2B. ....	28
Tabla 14. Nuevas parejas de cebadores diseñadas para la amplificación del gen HLA-DQB1-E2B. ....	31
Tabla 15. Condiciones analizadas en la prueba 5 en función de las combinaciones de cebadores y la concentración de estos.....	31
Tabla 16. Condiciones analizadas en la prueba 6 partiendo de la combinación 1.3 de la prueba anterior....	34
Tabla 17. Condiciones analizadas en la prueba 7 utilizando la combinación 1.3 de la prueba 5. ....	36

## **ABREVIATURAS**

APC: Célula presentadora de antígenos

CU: Cola universal

DGP: Péptido de gliadina desamidada

dNTP: Desxirribonucleótido trifosfato

EC: Enfermedad celíaca

EmA: Anticuerpos antiendomiso

ESPGHAN: Sociedad Europea de Gastroenterología, Hepatología y Nutrición Pediátrica

GWAS: Estudio de Asociación de Genoma Completo

HBT: Prueba de hidrógeno en el aliento

HLA: Antígenos Leucocitarios Humanos

IgA: Inmunoglobulina A

IL: Interleucina

MHC: Complejo Mayor de Histocompatibilidad

NGS: Secuenciación de Nueva Generación

LD: Desequilibrio de ligamiento

LNP: Lactasa no persistente

LP: Lactasa persistente

LPH: Lactasa-floricina hidrolasa

PCR: Reacción en Cadena de la Polimerasa

RS-id: Identificador SNP de referencia

SNP: Polimorfismo de un solo nucleótido

TG2: Transglutaminasa tisular

## 1. INTRODUCCIÓN

Las intolerancias alimentarias constituyen un conjunto de patologías cada vez más presente en nuestra sociedad, estimándose que estas afectan hasta a un 20% de la población mundial. No obstante, los mecanismos de acción de muchas de ellas todavía no se conocen al completo, resultando en un difícil diagnóstico, que en muchos de los casos supone todo un desafío (Tuck *et al.*, 2019).

Entre los alimentos asociados al desarrollo de intolerancias, caben destacar los productos lácteos, por su contenido en lactosa, y el gluten (Muthukumar *et al.*, 2020). Un diagnóstico incorrecto puede derivar en efectos perjudiciales en la salud del paciente, afectando a su nutrición y a su calidad de vida. En muchas ocasiones, la eliminación injustificada de ciertos alimentos de la dieta puede hacer que el paciente pierda su tolerancia, aumentando el riesgo de desarrollo de la enfermedad (Hon & Gupta, 2021).

### 1.1. ENFERMEDAD CELÍACA (EC)

La intolerancia al gluten o enfermedad celíaca (EC) es una afección autoinmune sistémica, crónica e inflamatoria del intestino delgado desencadenada por la ingesta de gluten en individuos genéticamente susceptibles (Serena *et al.*, 2019).

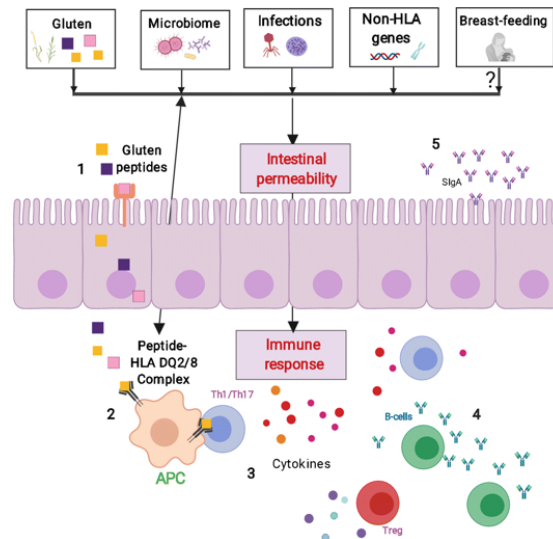
El gluten es una proteína de almacenamiento presente en diferentes cereales, compuesta por monómeros de gliadina y polímeros de glutenina, los cuales no se degradan fácilmente en el tracto gastrointestinal. Constituye la fracción proteica predominante del trigo, presente también en centeno y cebada, donde las prolaminas de este pueden desencadenar inflamación en individuos con EC (Verdu & Danska, 2018).

La EC destaca como una de las pocas enfermedades autoinmunes en las que se ha establecido la relación existente entre el factor ambiental desencadenante de la enfermedad, el gluten, y la predisposición genética, la cual viene conferida por los haplotipos HLA-DQ2 y HLA-DQ8, siendo estas combinaciones de alelos de los genes *HLA-DQA1* y *HLA-DQB1* que coexisten en un mismo cromosoma (Serena *et al.*, 2019; Snyder *et al.*, 2015). Esta puede manifestarse a cualquier edad, tanto en la infancia como en la edad adulta, presentando dos momentos destacados de inicio: durante los primeros años de vida, tras la introducción del gluten en la dieta, así como en la segunda o tercera década de vida (Caio *et al.*, 2019). Su prevalencia está experimentando un aumento en todo el mundo, con tasas que oscilan entre el 0,5% y el 2%, con un promedio mundial de alrededor del 1%. Se observa con mayor frecuencia en mujeres, con una proporción de dos mujeres por cada hombre afectado. En cambio, su frecuencia se ve disminuida en regiones donde los cereales con gluten no constituyen una parte fundamental de la dieta y la presencia de los haplotipos HLA-DQ2 y HLA-DQ8 es baja, como en el este de Asia y parte de África subsahariana. Diversos estudios han planteado la posibilidad de que este incremento en la prevalencia sea consecuencia de un mayor consumo de gluten, así como del cambio de proteínas de este. Sin embargo, la falta de datos sobre la incidencia anual dificulta llegar a una conclusión definitiva (Catassi *et al.*, 2022; Serena *et al.*, 2019).

#### 1.1.1. Patogenia de la enfermedad celíaca (EC)

En la patogenia de la EC están involucrados los sistemas inmunitarios innato y adaptativo. Tras la ingesta de gluten, este es digerido en el lumen por medio de proteasas humanas y bacterianas, resultando en la formación de péptidos proteicos. Estos son dirigidos hacia la lámina propia intestinal, donde son desamidados por acción del enzima transglutaminasa tisular (TG2) y reconocidos por las células presentadoras de antígenos (APCs) a través de las moléculas HLA de tipo II, específicamente por medio

de su unión a los heterodímeros HLA-DQ2 o HLA-DQ8. Esto desencadena una respuesta inmunitaria aberrante mediada por las células T CD4+, lo que conduce a una producción aumentada de al menos 15 citocinas plasmáticas, tales como las interleucinas (IL) *IL-15*, *IL-8*, *IL-2* o *IL-10*, entre otras, reclutando a neutrófilos y resultando en la expansión de los linfocitos intraepiteliales (Catassi *et al.*, 2022; Serena *et al.*, 2019; Verdu & Danska, 2018). Además del gluten, diversos factores adicionales también están siendo estudiados como posibles causas de la EC, entre lo que se incluyen la lactancia materna, ciertas infecciones virales o el microbioma (Figura 1) (Serena *et al.*, 2019).



**Figura 1. Factores genéticos y ambientales propuestos como contribuyentes al desarrollo de la enfermedad celíaca.** Extraído de (Serena *et al.*, 2019)

Los genes *HLA-DQ* (*HLA-DQA1* y *HLA-DQB1*), situados en el cromosoma 6p21, representan el 40% de la variación genética de la EC. Estos codifican heterodímeros de moléculas del complejo mayor de histocompatibilidad (MHC), los cuales se componen de dos subunidades: una subunidad  $\alpha$  y otra  $\beta$ , responsables de la unión a los péptido proteicos, al expresarse en la superficie de las células presentadores de antígenos (APCs) (Choung *et al.*, 2020; Lebwohl & Rubio-Tapia, 2021; Serena *et al.*, 2019).

Los haplotipos HLA-DQ2 y HLA-DQ8 son los encargados de la presentación de antígenos del gluten por las APCs, activando la respuesta inmunitaria y el desarrollo de la EC (Lebwohl & Rubio-Tapia, 2021). El haplotipo HLA-DQ2 consta de 2 heterodímeros distintos, como resultado de la diferente combinación de alelos: HLA-DQ2.2 y HLA-DQ2.5. Aproximadamente el 90% de individuos con enfermedad celíaca expresan el heterodímero HLA-DQ2.5, el cual está formado por la combinación de los alelos *HLA-DQA1\*05:01* y *HLA-DQB1\*02:01*, representando el heterodímero de mayor riesgo. Los pacientes restantes, a excepción de algunos casos mínimos, expresan el heterodímero HLA-DQ2.2, aproximadamente en un 5% de los casos, resultante de la combinación de alelos *HLA-DQA1\*02:01* y *HLA-DQB1\*02:02* o HLA-DQ8, en otro 5% de los casos, como resultado de los alelos *HLA-DQA1\*03:01* y *HLA-DQB1\*03:02* (Catassi *et al.*, 2022; Choung *et al.*, 2020).

La presencia de uno de estos haplotipos es necesaria para manifestar la EC, indicando una mayor predisposición, pero no necesariamente desarrollo de la enfermedad (Serena *et al.*, 2019). Concretamente, solo un 3-4% de los individuos con dichos haplotipos que ingieran gluten desarrollarán la enfermedad, sugiriendo que los factores de riesgo ambientales poseen un papel fundamental (Verdu & Danska, 2018). Se ha observado que el riesgo de padecer la enfermedad se ve incrementado en pacientes en los que el heterodímero HLA-DQ2 se encuentra en configuración homocigota, así como

en aquellas personas con antecedentes familiares de EC, quienes presentan un riesgo de entre 10-15% de desarrollar la enfermedad a lo largo de su vida (Catassi *et al.*, 2022). Diversos estudios también han demostrado que la presencia de uno u otro haplotipo podría tener relación con una aparición temprana (HLA-DQ2) o tardía (HLA-DQ8) de la enfermedad (Serena *et al.*, 2019).

Recientemente, otros genes también se han asociado con la patogénesis de la EC. Estudios de asociación del genoma completo (GWAS) han revelado que alrededor del 18% de la susceptibilidad a la enfermedad celíaca tiene un componente genético, identificando además de los genes de la región HLA de clase II, otros 18 factores de riesgo en la región MHC y 39 genes no HLA que podrían estar involucrados en la patogénesis de la enfermedad, entre los que se encuentran genes involucrados en la maduración y diferenciación de las células T, así como en la respuesta proinflamatoria, como son los genes que codifican para citocinas plasmáticas, *IL-2* e *IL-21*, entre otros (Tabla 1) (Serena *et al.*, 2019).

**Tabla 1. Factores genéticos asociados a la aparición de la enfermedad celíaca.** Extraída de (Serena *et al.*, 2019).

FACTOR GENÉTICO	GENES CANDIDATOS	PAPEL EN LA PATOGENIA DE LA ENFERMEDAD
Genes MHC (HLA)	<ul style="list-style-type: none"> <li>HLA clase I: <i>DQB*08:01</i></li> <li>HLA clase II: <i>DQA1*05; DQB1*02; DQA1*02; DQB1*02; DQA1*03; DQB1*03.</i></li> </ul>	<ul style="list-style-type: none"> <li>Detección de péptidos específicos.</li> <li>Presentación de células T.</li> </ul>
Genes no MHC (no HLA)	<ul style="list-style-type: none"> <li><i>IL2; IL21</i></li> <li><i>IL18R1; IL18RAP</i></li> <li><i>MYO9B; HP2; PARD3; MAGI2.</i></li> </ul>	<ul style="list-style-type: none"> <li>Activación y proliferación de células B y T.</li> <li>Participación en la señalización proinflamatoria.</li> <li>Control de la permeabilidad intestinal.</li> </ul>

### 1.1.2. Signos y síntomas

La EC es una patología que afecta a múltiples sistemas, la cual se caracteriza principalmente por la aparición de síntomas gastrointestinales, pudiendo originar en ocasiones signos y síntomas extraintestinales que incluyen dermatitis, osteoporosis, neuropatía periférica o dolores de cabeza. Diversos estudios también han relacionado dicha afección con anemia y un crecimiento deficiente en niños. No obstante, cabe destacar que algunos individuos con dicha enfermedad pueden no presentar ninguno de estos síntomas, siendo completamente asintomáticos (Serena *et al.*, 2019).

Según la clasificación de Oslo, y en función de los signos y síntomas que manifieste el paciente en el momento del diagnóstico, la EC se clasifica en forma clásica, no clásica, subclínica, potencial y refractaria (Caio *et al.*, 2019; Catassi *et al.*, 2022).

La forma clásica, también denominada intestinal, se presenta con mayor frecuencia como una enfermedad pediátrica, la cual causa diarrea crónica, pérdida de apetito y de peso, distensión abdominal y crecimiento deficiente. En estos casos, un diagnóstico tardío puede derivar en graves afecciones, pudiendo ser incluso mortal (Catassi *et al.*, 2022). La forma no clásica constituye la forma más común de EC, caracterizada por síntomas intestinales inespecíficos, tales como dolor y distensión

abdominal, y episodios de diarrea y estreñimiento. En este caso, también es común que se den síntomas extraintestinales, tanto en adultos como en niños, entre los que se encuentran déficits nutricionales, afecciones a la piel, como dermatitis u osteoporosis (Caio *et al.*, 2019; Catassi *et al.*, 2022). Por su parte, la EC subclínica constituye la forma asintomática de la enfermedad, la cual se diagnostica gracias a los programas de detección que se llevan a cabo en grupos de riesgo, como son aquellos que presentan un factor genético hereditario. La EC potencial se caracteriza por positividad en las pruebas serológicas, sin presentar atrofia en las vellosidades, mientras que la EC refractaria se define como una complicación de la patología en la que los síntomas persiguen a pesar de mantener una dieta libre de gluten (Catassi *et al.*, 2022).

### **1.1.3. Diagnóstico y tratamiento**

Los avances en el diagnóstico de la EC se deben principalmente a la actual disponibilidad de pruebas de elevada sensibilidad y especificidad, lo que ha llevado a que el número de individuos diagnosticados sea cada vez mayor (Caio *et al.*, 2019).

El método de referencia para efectuar el diagnóstico más preciso requiere que se cumplan los siguientes factores: una serología positiva y atrofia de la mucosa duodenal, detectada por biopsia del intestino delgado (Catassi *et al.*, 2022). Esto es debido a que ninguna de las pruebas serológicas actuales presenta la suficiente especificidad y sensibilidad como para no necesitar complementación de otra prueba diagnóstica (Caio *et al.*, 2019). Sin embargo, según la ESPGHAN, la biopsia duodenal puede ser omitida en pacientes pediátricos con elevados niveles de anticuerpos séricos (IgA anti-tTG2, IgA anti-EmA) y predisposición genética (Caio *et al.*, 2019).

La presencia de marcadores serológicos es una de las principales características de la EC, como son los anticuerpos IgA anti-tTG2, IgA antiendomiso (EmA) y anticuerpos IgA anti-DGP o péptido de gliadina desamidada (Caio *et al.*, 2019). Concretamente, la existencia de anticuerpos IgA anti-transglutaminasa 2 (TG2) en individuos que padecen la enfermedad y todavía mantienen el gluten en su dieta constituye el principal método de detección de la enfermedad, siendo estos anticuerpos los encargados de reconocer al principal autoantígeno del gluten, relacionados también con las alteraciones de la mucosa duodenal. Su determinación se realiza actualmente por medio de diferentes métodos, tales como ELISA, radioinmunoensayo, quimioluminiscencia o inmunoensayo fluorescente (Catassi *et al.*, 2022).

La enteropatía celíaca puede variar en función de la gravedad de la enfermedad, pudiendo originar desde un aumento de linfocitos intraepiteliales hasta una atrofia total de las vellosidades. No obstante, aunque dichas alteraciones no son exclusivas de esta patología, sí que constituyen una segunda prueba para corroborar el correcto diagnóstico (Catassi *et al.*, 2022).

A pesar de que las pruebas genéticas son un indicativo de predisposición a desarrollar la enfermedad, estas brindan otra alternativa de diagnóstico al permitir descartar la enfermedad en situaciones en las que los signos y síntomas pueden estar también relacionados con otras patologías, con un valor predictivo negativo mayor al 99%. Además, presentan mucha utilidad en el diagnóstico de aquellos pacientes en cuya dieta el gluten ya no está presente, así como en el caso de obtener discrepancia en los resultados serológicos o histológicos o en individuos con posible EC refractaria (Lebwohl & Rubio-Tapia, 2021).

Se ha comprobado que un diagnóstico tardío de la enfermedad está relacionado con un aumento de mortalidad de esta, pudiendo derivar, aunque es poco frecuente (1% de pacientes con EC) en patologías de mayor gravedad, como enteropatía celíaca refractaria, linfomas o adenocarcinomas intestinales (Caio *et al.*, 2019). Además, la detección de los marcadores genéticos HLA-DQ2 y HLA-DQ8 puede facilitar el diagnóstico en determinadas situaciones, como en el caso del riesgo familiar, pudiendo

proporcionar información relevante acerca de los mecanismos que controlan la pérdida de tolerancia al gluten (Catassi *et al.*, 2022; Serena *et al.*, 2019). Diversos estudios han demostrado que el riesgo de padecer EC es prácticamente nulo en hermanos de pacientes afectados que carecen de dichos genes, aumentando entre un 10-15% en el caso de poseer una sola copia de alguno de los genes, y pudiendo ascender a más del 20% en aquellos homocigotos para HLA-DQ2, siendo los polimorfismos en *HLA-DQB1* (variantes *A1\*05* y *B1\*02*) y *HLA-DQA1* (variantes *A1\*03* y *B1\*0302*) los alelos de mayor riesgo (Catassi *et al.*, 2022; Lebwohl & Rubio-Tapia, 2021).

Hasta el momento, no existe ninguna terapia aprobada para tratar la EC, siendo la dieta sin gluten de manera estricta la única alternativa para resolver los síntomas asociados a la enfermedad y restaurar la arquitectura de las vellosidades intestinales (Serena *et al.*, 2019). No obstante, el llevar una dieta de este tipo no resulta nada sencillo, ya que esta supone un gran impacto en la calidad de vida del paciente, relacionándose en diversos estudios con una ingesta insuficiente de ciertos nutrientes. Aunque cada vez existen más opciones de productos sin gluten, el riesgo de contaminación sigue siendo elevado, y en la mayoría de los casos, dichos productos no presentan opciones muy saludables (Caio *et al.*, 2019; Catassi *et al.*, 2022).

Pese a los diferentes métodos de diagnóstico disponibles, lograr un correcto diagnóstico sigue siendo un desafío y, a pesar del gran aumento en el número de diagnosticados, muchos individuos permanecen todavía sin diagnosticar. Ello subraya la necesidad de una estandarización global del mismo, esperando disponer en un futuro de nuevas herramientas terapéuticas y preventivas que permitan un diagnóstico mucho más temprano, basado en la evidencia y sin necesidad de realizar pruebas invasivas, que permita prevenir o mejorar la progresión y la calidad de vida de aquellas personas con EC o con predisposición a padecerla (Caio *et al.*, 2019; Lebwohl & Rubio-Tapia, 2021; Serena *et al.*, 2019). Por ello, resulta de gran importancia el llevar a cabo un estudio genético de predisposición en pacientes de riesgo, junto con estudios genómicos y metabolómicos del microbioma o diversos factores ambientales que puedan lograr una mayor comprensión del desarrollo de los mecanismos de resistencia al gluten, pudiendo obtener una información clínica más precisa que la únicamente basada en predisposición genética (Caio *et al.*, 2019).

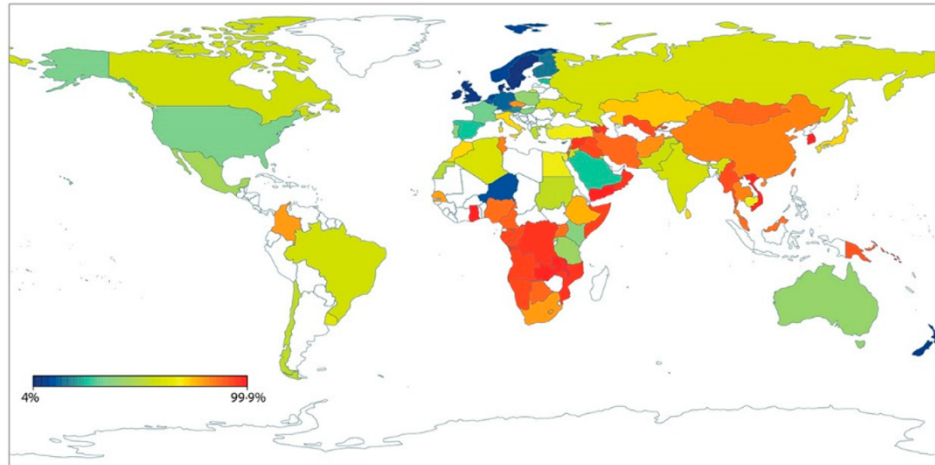
## **1.2. INTOLERANCIA A LA LACTOSA**

La intolerancia a la lactosa es una condición patológica en la cual los individuos que la padecen presentan dificultad para digerir y metabolizar la lactosa tras el consumo de diversos productos lácteos (Szilagyi & Ishayek, 2018).

La lactosa es el principal disacárido de la leche, el cual está formado por los monosacáridos D-glucosa y D-galactosa, unidos mediante enlace glucosídico  $\beta$ -1,4, constituyendo la principal fuente de energía en mamíferos durante el periodo de lactancia (Anguita-Ruiz *et al.*, 2020; Szilagyi & Ishayek, 2018). Para obtener dicha energía, la lactosa es hidrolizada en glucosa y galactosa, los cuales pueden ser absorbidos en el tracto intestinal. En humanos, la capacidad de digestión recae en el enzima lactasa-floricina hidrolasa (LPH),  $\beta$ -galactosidasa presente en la mucosa intestinal y la cual está codificada en el propio gen de la lactasa (LCT), situado en el cromosoma 2q21. No obstante, tras la lactancia, la actividad del enzima LPH en la mayoría de los mamíferos se ve disminuida, pudiendo llegar a ser ausente, derivando en una incapacidad de digerir la lactosa, que recibe el nombre de lactasa no persistente (LNP). Se estima que alrededor de dos tercios de la población mundial son LNP, ya que, aunque la prevalencia mundial de casos confirmados es de un 57%, se calcula que la prevalencia real supera el 66%, con una distribución muy desigual (*Figura 2*) (Anguita-Ruiz *et al.*, 2020; Catanzaro *et al.*, 2021).



Sin embargo, en el tercio restante parece ser que la capacidad de digerir la lactosa o persistencia de la lactasa (LP) se mantiene en la edad adulta, siendo frecuente en aquellos descendientes de poblaciones que históricamente se han dedicado a la domesticación del ganado. Esta presenta una elevada prevalencia en el norte de Europa, mientras que disminuye en el sur de Europa y Oriente Medio, con una muy baja frecuencia en poblaciones asiáticas y africanas no pastoriles (Anguita-Ruiz *et al.*, 2020).



**Figura 2. Distribución geográfica mundial de las frecuencias de la no persistencia de la lactasa (LNP).**  
Extraída de (Szilagyi & Ishayek, 2018).

Cuando se refiere a LNP, resulta de gran importancia diferenciarla de otras patologías, como son la deficiencia congénita de lactasa y la hipolactasia secundaria. La primera se trata de una enfermedad rara causada por mutaciones en ambos alelos del gen *LCT*, y se manifiesta en los primeros días de vida con la lactancia materna. Por otro lado, la hipolactasia secundaria es una disminución de la actividad de la enzima lactasa no relacionada con factores genéticos, causada por alteraciones de la mucosa intestinal debido a otras patologías, como la EC (Anguita-Ruiz *et al.*, 2020).

### 1.2.1. Patogenia de la intolerancia a la lactosa

El gen *LCT* humano es un gen de 49 kb y 17 exones situado en el cromosoma 2q21, que codifica para transcritos de ARN mensajero de unas 6 kb y 1927 aminoácidos (Anguita-Ruiz *et al.*, 2020). Por su parte, el gen de la LPH, también localizado en el cromosoma 2q21, se encuentra regulado por una región promotora situada 14000 pb aguas arriba del gen *LCT* y codifica para una glicoproteína transmembrana de 160KDa cuyo extremo N-terminal se encuentra en la superficie de las microvellosidades de los enterocitos (Anguita-Ruiz *et al.*, 2020; Szilagyi & Ishayek, 2018).

En la actualidad, existe una gran evidencia de que la persistencia a la lactasa presenta una base genética, siguiendo un patrón de herencia dominante (Anguita-Ruiz *et al.*, 2020). Diversos estudios han postulado que la LP podría estar originada por la existencia de cinco (o más) polimorfismos de un solo nucleótido (SNPs) en la región reguladora del gen *LCT*, situada a 14 kb de este (Anguita-Ruiz *et al.*, 2020). Parece ser que la existencia de dichos SNP conduce a una vía alternativa, proporcionando nuevos sitios de unión a factores de transcripción, en la cual la expresión de *LCT* no está modulada a la baja, a diferencia de la vía original.

La primera variante descrita fue -13910:C>T (rs4988235), la cual se encuentra presente en gran parte de la población europea. Otras variantes también han sido identificadas, entre las que se encuentran: -13907:C>G (rs41525747), -13915:T>G (rs41380347), -14009:T>G (rs869051967) y -14010:G>C (rs145946881) cuyas frecuencias varían en poblaciones de Oriente Medio y África. En total, se conocen

hasta la fecha veintitrés SNPs que confieren LP, las cuales habrían surgido de forma independiente en diferentes poblaciones en el mismo periodo de tiempo, proporcionando un claro ejemplo de evolución convergente y coevolución gen-cultura (Tabla 2). Sin embargo, las causas que conllevan a la disminución de la actividad de la LPH en la vida adulta siguen sin estar claras (Anguita-Ruiz *et al.*, 2020).

**Tabla 2. Diversas variantes genéticas estudiadas asociadas a la persistencia de lactasa (LP).** Extraída de (Anguita-Ruiz *et al.*, 2020).

SNP	RS-id	Información adicional
-14010:G>C **	rs145946881	Ampliamente estudiado y asociado
-14009:T>G **	rs869051967	
-13915:T>G **	rs41380347	
-13910:C>T **	rs4988235	
-13907:C>G **	rs41525747	
-22.018:G>A	rs182549	En LD completo con la causal -14010: G>C
-13914:G>A		Variante rara (q < 5%)
-14062:G>A		
-14028:T>C	rs759157971	
-13753:C>T		
-13693:G>A		
-13806:A>G	ss820496565	
-13964:C>A		
-13771:A>G		
-14010 G>A	rs145946881	
-13926 A>C		

La evidencia de la expresión funcional de LCT se indica mediante: \*\*: SNPs funcionales validados; \*: alguna evidencia para SNPs funcionales. LD: desequilibrio de ligamiento; RS-id: identificador SNP de referencia; SNP: polimorfismo de un solo nucleótido.

### 1.2.2. Signos y síntomas

La incapacidad de digerir la lactosa cuando esta es ingerida conlleva al desarrollo de una variada sintomatología, en la que destacan principalmente los síntomas gastrointestinales. Esto es debido a que el disacárido regresa al intestino, impulsando consigo los fluidos hacia la luz intestinal, donde es fermentado por las bacterias intestinales, ocasionando la producción de gases, dolores y calambres abdominales, hinchazón, diarrea, náuseas y vómitos (Anguita-Ruiz *et al.*, 2020; Szilagyi & Ishayek, 2018). Son varias las variables que pueden influir en la gravedad de la sintomatología, como la dosis de lactosa ingerida, la expresión de LPH, el tiempo de tránsito intestinal o la fermentación por bacterias intestinales (DI Costanzo & Berni Canani, 2019).

### 1.2.3. Diagnóstico y tratamiento

La intolerancia a la lactosa o la LNP puede ser directamente diagnosticada mediante una biopsia de la mucosa duodenal, en la cual se miden las enzimas intestinales presentes, comprobando su actividad enzimática mediante ensayo enzimático (Catanzaro *et al.*, 2021). No obstante, esta resulta una técnica invasiva, por lo que se han desarrollado pruebas menos drásticas, que permiten realizar un diagnóstico indirecto. Una de ellas consiste en la medición de hidrógeno en el aliento tras la ingesta de lactosa, denominada prueba de hidrógeno en el aliento (HBT), como resultado de la fermentación del disacárido en el caso de poseer el fenotipo LNP. La segunda prueba, o test de tolerancia a la lactosa, consiste en la medición de glucosa en sangre tras la ingesta de 50g de lactosa, donde valores inferiores a 1,1 y 1,4 mmol/l de glucosa se asocian a una mala digestión de la lactosa. No obstante, esta presenta

muchos casos de falsos positivos y falsos negativos, por lo que está siendo reemplazada por el HBT (DI Costanzo & Berni Canani, 2019). En niños, una acidificación de las heces también puede ser indicativo de una deficiencia en la absorción de los carbohidratos, pudiendo ser indicativo de un fenotipo LNP (*Figura 3*) (Anguita-Ruiz *et al.*, 2020; Szilagyí & Ishayek, 2018).

Las pruebas genéticas también pueden facilitar el diagnóstico mediante la identificación del SNP asociado a la LNP. No obstante, su identificación no implica una aparición inmediata de la intolerancia, sino que esta puede desarrollarse a lo largo de la vida del paciente (*Figura 3*) (DI Costanzo & Berni Canani, 2019).

MÉTODO DE DIAGNÓSTICO	DESCRIPCIÓN	VENTAJAS	DESVENTAJAS
<b>PRUEBA GENÉTICA</b>	Comprobación de la presencia de los SNP 13910 C>T y 22018 G>A en el gen LCT mediante análisis de ADN en muestra de sangre.	<ul style="list-style-type: none"> <li>• Permite confirmar o excluir una forma primaria de intolerancia.</li> <li>• Mínimamente invasivo.</li> </ul>	<ul style="list-style-type: none"> <li>• No permite identificar formas secundarias.</li> <li>• La presencia de polimorfismos no implica necesariamente el desarrollo de la intolerancia.</li> </ul>
<b>PRUEBA DE HIDÓGENO EXHALADO O THB</b>	Medición del hidrógeno exhalado antes y después de la ingesta de 25-50g de lactosa. Valores de aumento de 20 ppm desde la línea base confirmar una mala digestión.	<ul style="list-style-type: none"> <li>• No invasivo.</li> <li>• Alta sensibilidad y especificidad.</li> <li>• Fácil de realizar e interpretar.</li> <li>• Bajo coste.</li> </ul>	<ul style="list-style-type: none"> <li>• Larga duración (3-6h).</li> <li>• Obtención de falsos positivos y falsos negativos.</li> </ul>
<b>PRUEBA RÁPIDA DE LACTASA</b>	Prueba de actividad enzimática de la lactasa en muestras de biopsia de la mucosa duodenal. Actividad enzimática ausente o reducida pueden confirmar la intolerancia.	<ul style="list-style-type: none"> <li>• Diagnóstico de certeza de hipolactasia.</li> </ul>	<ul style="list-style-type: none"> <li>• Técnica invasiva.</li> <li>• Personal cualificado debe estar presente para realizar la prueba tras la endoscopia.</li> <li>• Coste elevado.</li> </ul>
<b>PRUEBA DE TOLERANCIA A LA LACTOSA</b>	Medición de glucosa en sangre tras 30, 60 y 120 min de la ingesta de lactosa. La ausencia de cambios en la glucemia puede confirmar la intolerancia.	<ul style="list-style-type: none"> <li>• Mínimamente invasivo.</li> <li>• Bajo coste.</li> </ul>	<ul style="list-style-type: none"> <li>• Baja sensibilidad y especificidad.</li> <li>• Existencia de una estrecha relación con las características del paciente (tiempo de vaciado gástrico, etc.)</li> </ul>

**Figura 3. Principales métodos actuales para el diagnóstico de la intolerancia a la lactosa, así como las ventajas e inconvenientes que cada uno de ellos presenta. Adaptado de (Catanzaro *et al.*, 2021).**

La mayoría de los tratamientos para las diferentes intolerancias alimentarias se basan en la eliminación de los alimentos causales de la dieta. Sin embargo, diversos estudios han demostrado que, en el caso de la intolerancia a la lactosa, una reducción de la ingesta de lactosa sería suficiente para no presentar sintomatología, en lugar de una exclusión total de la dieta, pudiendo ingerir, individuos adolescentes o adultos, hasta 12g de lactosa sin presentar sintomatología (DI Costanzo & Berni Canani, 2019).

La relación entre la microbiota y el microbioma humano con la fermentación de la lactosa en individuos LNP ha llevado a proponer la alimentación y el cuidado de la microbiota como un posible tratamiento para paliar los síntomas originados por dicha intolerancia. De este modo, el consumo de probióticos de cepas concretas ha resultado ser un factor significativo, pudiendo observar una mejoría de la sintomatología en función de la presencia y abundancia de ciertas cepas de bacterias en el colon (Anguita-Ruiz *et al.*, 2020; DI Costanzo & Berni Canani, 2019). El reemplazo enzimático se presenta

también como otro posible enfoque terapéutico, el cual ha mostrado mejorías en la sintomatología y niveles menores de hidrógeno en la prueba HBT, aunque son necesarios más datos sobre su eficacia (DI Costanzo & Berni Canani, 2019).

### **1.3. TECNOLOGÍA NGS**

En los últimos años, la genética clínica ha adquirido gran importancia para el diagnóstico, pronóstico y prevención de muchas enfermedades. Las tecnologías de Secuenciación de Nueva Generación (NGS, dadas sus siglas en inglés) han supuesto un gran revolución en muchas áreas de investigación científica, transformando la genética clínica y la investigación biomédica gracias al análisis de una gran cantidad de genes en tiempo reducido, al ser una técnica eficiente y con un coste mucho menor al de otras técnicas de secuenciación convencionales (Pereira *et al.*, 2020). Los avances en el desarrollo y comprensión de los procesos de secuenciación han permitido que actualmente se pueda realizar la secuenciación de paneles de genes, lo cual supone un gran avance en el diagnóstico de ciertas enfermedades específicas o con fenotipos similares. Todo ello es debido a que las técnicas NGS han reducido considerablemente los costos de secuenciación, generando un gran impacto en los campos de la genética y la genómica (Schuler *et al.*, 2022).

Diversos estudios han propuesto la tecnología NGS como una técnica de alta resolución, que presenta gran precisión y exactitud, para la secuenciación de genes completos, tales como los genes de la región HLA, incluidos los exones 2 y 3. Dicha tecnología ofrece plataformas de multiplexado que proporcionan un alto rendimiento a bajo coste, con una redundancia de fragmentos superpuestos que permiten mapear la secuencia de interés obteniendo una profundidad de lectura (Bravo-Egana *et al.*, 2021). A pesar de que la secuenciación de genoma completo también se puede utilizar en la tipificación del gen HLA, su secuenciación presenta mejores resultados mediante el enriquecimiento específico de la secuencia por PCR o captura híbrida, siendo por ello la metodología de tipificación para dicho gen más utilizada (Bravo-Egana *et al.*, 2021).

Las tecnologías NGS utilizan, generalmente, la secuenciación por síntesis, la cual se basa en la adición de dNTPs fluorescentes que se unen secuencialmente a la cadena a secuenciar por complementariedad, los cuales son posteriormente leídos por imagen, resultando en secuencias de entre 100 y 200 pb. No obstante, los protocolos a seguir son laboriosos y requieren de una validación por personal cualificado, ya que muchas veces la resolución de regiones complejas resulta problemática, puesto que las lecturas cortas pueden ser difíciles de alinear con el genoma de referencia, requiriendo la identificación de ciertas variantes varios procesos de análisis para llegar a un correcto diagnóstico (Schuler *et al.*, 2022). Además, es común que en la realización de dicha metodología se produzcan errores, como la pérdida de ciertos alelos, que llevan a una mala tipificación. Dichos errores pueden proceder de errores en la amplificación por la ADN polimerasa, sesgos de amplificación por PCR, ciclos de secuenciación u otros errores durante el propio proceso de secuenciación. Por tanto, es necesario tener un gran conocimiento sobre la metodología llevada a cabo, con la finalidad de que el proceso de secuenciación se lleve a cabo satisfactoriamente, siendo todavía necesaria mucha más investigación para poder solventar dichos errores (Bravo-Egana *et al.*, 2021).

Es por ello por lo que resulta imprescindible el desarrollo de una metodología sencilla y optimizada que permita la identificación de los alelos de riesgo de muchas enfermedades, siendo en este caso la enfermedad celíaca (EC) y la intolerancia a la lactosa las patologías estudiadas. Para ello, el presente trabajo experimental se basa en la optimización de un proceso de NGS que permita identificar de manera eficiente y precisa dichas variantes, el cual podría en un futuro implantarse en la clínica obteniendo resultados que permitan llevar a cabo un correcto diagnóstico.

## 2. OBJETIVOS

La intolerancia al gluten (celiaquía o enfermedad celíaca (EC)) y la intolerancia a la lactosa (o no persistencia de la lactasa) son algunas de las intolerancias alimentarias más frecuentes en la población, pudiendo llegar a afectar hasta a un 20% de la población mundial. No obstante, su diagnóstico sigue resultando en muchos casos todo un desafío, al presentar los individuos afectados, en ambos casos, signos y síntomas también presentes en muchas otras patologías.

Por ello, resulta de interés realizar un estudio genético que permita la identificación de los alelos de riesgo que predisponen al desarrollo de dichas patologías, para lo cual la tecnología de Secuenciación de Nueva Generación (NGS) resulta altamente beneficiosa.

De este modo, el presente Trabajo de Fin de Grado tiene como objetivo principal la optimización de un procedimiento de NGS por amplicones cortos que permita la identificación de los alelos descritos en literatura científica como alelos de riesgo para EC e intolerancia a la lactosa.

Con esta finalidad principal, se plantean además los siguientes objetivos parciales:

- Diseñar cebadores que permitan la amplificación de un panel de genes constituido por los genes *HLA-DQA1*, *HLA-DQB1* y *LCT* y sus alelos de riesgo con los correspondientes polimorfismos asociados a la EC e intolerancia a la lactosa.
- Utilizar dicho panel con el objetivo de optimizar el proceso, variando la distribución de los correspondientes cebadores y las condiciones del proceso de amplificación.
- Seleccionar la mejor condición obtenida con el fin de llevar a cabo el estudio planificado posteriormente.

### 3. MATERIALES Y MÉTODOS

#### 3.1. DISEÑO EXPERIMENTAL

La totalidad del procedimiento experimental fue realizada en los laboratorios de la Unidad de Genómica y Diabetes (UGD) del Instituto de Investigación Sanitaria INCLIVA.

El presente diseño experimental se llevó a cabo con el objetivo de lograr la optimización de un procedimiento de NGS dirigido a la identificación de los alelos de riesgo en EC e intolerancia a la lactosa, mediante el análisis de los genes *LCT*, *HLA-DQA1* y *HLA-DQB1*, utilizando la técnica de secuenciación masiva de amplicones cortos.

La metodología utilizada en dicho proceso de optimización se basó en la preparación del material genómico de interés y en el diseño de cebadores que permitieran la amplificación del mismo mediante dos etapas de amplificación consecutivas, utilizando la técnica de reacción en cadena de la polimerasa (PCR): una primera PCR selectiva, que permitió seleccionar y amplificar los fragmentos de interés, añadiendo además una cola universal de nucleótidos, seguida de una PCR anidada o nominativa que permitió identificar cada una de muestras mediante el uso de secuencias identificadoras o *barcodes*. Los productos amplificados fueron posteriormente purificados y cuantificados. Finalmente, se realizó la secuenciación de los amplicones obtenidos y el posterior análisis de las coberturas mediante el uso de una *pipeline* disponible en el laboratorio. De este modo, se evaluó la cobertura y calidad de los fragmentos secuenciados, comprobando si la cobertura era homogénea o si el proceso requería una mayor optimización (*Figura 4*).



**Figura 4. Diagrama del diseño experimental llevado a cabo para el proceso de optimización. Elaborado mediante Biorender.**

#### 3.2. OBTENCIÓN DE MUESTRAS

La optimización del procedimiento se llevó a cabo utilizando ADN genómico, extraído previamente en el laboratorio a partir de muestras de sangre mediante el sistema Maxwell (Promega®, USA). Para ello, se utilizó el kit Maxwell® 16 LEV Blood DNA. Posteriormente, las muestras obtenidas fueron cuantificadas y se realizaron las diluciones oportunas para lograr en todas ellas una concentración de 10 ng/μl, adecuada para la generación de librerías.

#### 3.3. DISEÑO DE CEBADORES

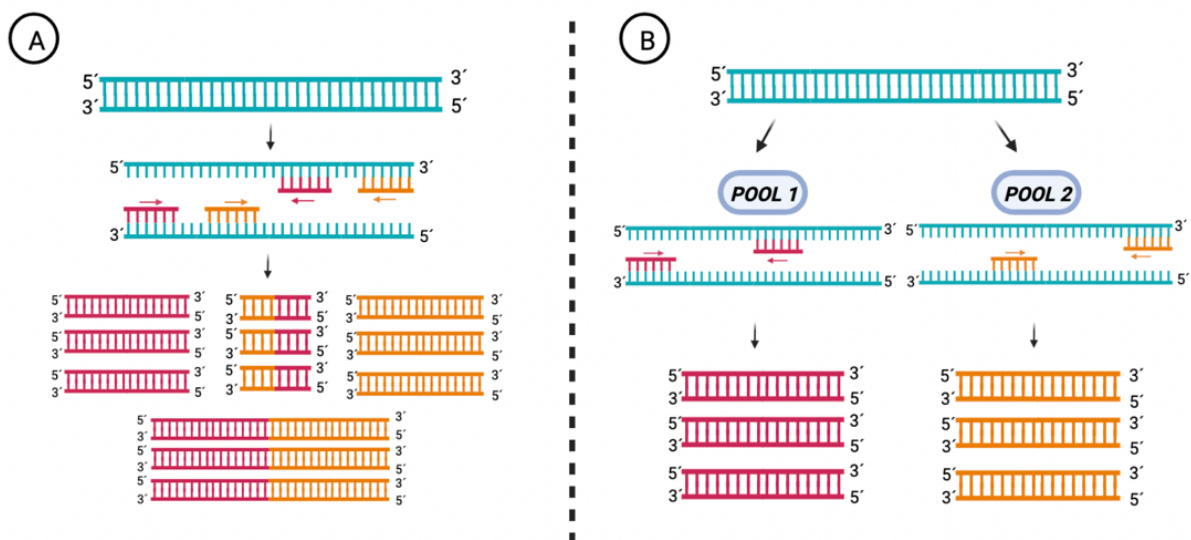
La generación de librerías requiere de cebadores o *primers* que posibiliten la amplificación de los fragmentos de ADN de interés al permitir la unión de la ADN polimerasa. El diseño de estos se llevó a cabo mediante el software *Primer 3* (<https://primer3.ut.ee/>), en el cual se introdujo la secuencia de la

región de interés, la cual fue obtenida previamente de la base de datos *Ensembl* (<https://www.ensembl.org/index.html>), utilizando el genoma GRCh38/hg38 como genoma de referencia. Mediante el uso de dicha *software*, se obtuvieron las secuencias de los pares de cebadores que permitirían amplificar los fragmentos de interés (cebador directo en el extremo 5' y cebador reverso en el extremo 3'), siendo estas complementarias a las regiones de interés de la secuencia de referencia. Además, en el diseño de estos se añadió una secuencia de ácidos nucleicos denominada cola universal, la cual resulta de gran importancia en los pasos posteriores al permitir la unión de las secuencias identificadoras o *barcodes*.

Dado que la región HLA es una región hipervariable, para la amplificación de algunos de los alelos de este y el correcto genotipado de todos los individuos resultó necesario diseñar más de una pareja de cebadores por amplicón, al darse la hibridación de algunos de los cebadores en zonas con presencia de SNPs. Ello fue fundamental para lograr amplificar todos los fragmentos, independientemente de los polimorfismos presentes en la zona de hibridación del cebador.

El diseño fue realizado teniendo en cuenta los siguientes parámetros: longitud del producto a amplificar (secuencia junto con los cebadores), el cual debe encontrarse alrededor de 150 pares de bases (pb); longitud de los cebadores, de entre 18-22 pares de bases; contenido en GC (%) de entre un 20-80%, y una temperatura de *melting* (TM) de entre 58-62°C. Una vez realizado el diseño, se comprobó que este era el correcto y que la hibridación se iba a dar de forma satisfactoria, para lo cual se utilizaron diferentes herramientas bioinformáticas. Una de ellas fue UCSC In-Silico PCR, la cual realiza una PCR *in silico*, comprobando si los cebadores permiten la amplificación del producto deseado. Por otra parte, también se utilizó FastPCR, la cual permite comprobar si durante la amplificación se pueden formar productos inespecíficos como *self-dimers* y *cross-dimers*.

Finalmente, en función del diseño, los cebadores fueron divididos en diferentes *pools* o multiplexes, con la finalidad de evitar el solapamiento de amplicones y la formación de productos inespecíficos como *cross-dimers* (Figura 5).



**Figura 5. Representación de la división de cebadores contiguos en diferentes pools.** A) Pares de cebadores contiguos son adicionados en una misma reacción de PCR, originando amplicones de tamaño no deseado. B) Los cebadores contiguos son divididos en dos pools: pool 1 y pool 2, amplificando todo el fragmento de interés en dos reacciones independientes, evitando la formación de amplicones no deseados. Diseñado mediante Biorender.

Dicha separación de cebadores se debe al tipo de secuenciación realizada (150X), en la cual la enzima de secuenciación solo es capaz de leer 150 pb por cada extremo del fragmento. Esto implica que la amplificación de la secuencia de interés debe realizarse mediante amplicones solapantes, para lo cual resulta necesario diseñar cebadores contiguos que amplifiquen la totalidad del fragmento, al tener este un mayor tamaño. No obstante, estos cebadores contiguos pueden amplificar fragmentos más cortos o largos de lo deseado si se encuentran en una misma reacción, provocando un sesgo en el número de lecturas durante el posterior proceso de secuenciación. Por ello, son separados en diferentes pools o reacciones de PCR independientes (*Figura 5*). Por otra parte, el *software* FastPCR también permitió comprobar la posible existencia de *cross-dimers* entre diferentes cebadores, siendo estos también separados en diferentes *pools* o reacciones de PCR independientes.

### 3.3.1. Regiones amplificadas

Para la generación de la librería, los exones 2 y 3 de los genes *HLA-DQA1* y *HLA-DQB1* fueron amplificados, ya que es en estos exones donde se localizan las posiciones que originan los diferentes alelos de riesgo de estos genes.

En el caso de la no persistencia a la lactasa, las variantes estudiadas fueron las indicadas en la *Tabla 3*.

**Tabla 3. Variantes genéticas del gen LCT estudiadas con su nomenclatura tradicional y su equivalencia en la nomenclatura actual.** Adaptada de (Anguita-Ruiz et al., 2020)

Nomenclatura tradicional	Rs- id (nomenclatura de referencia NCBI)	Alelo asociado a la persistencia de la lactasa
-14010:G>C **	rs145946881	C
-13915:T>G **	rs41380347	G
-13910:C>T **	rs4988235	T
-13907:C>G **	rs41525747	G
-22.018:G>A	rs182549	A

La primera columna muestra la nomenclatura tradicional para los Polimorfismos de un Solo Nucleótido (SNP), mientras que la segunda hace referencia a la nomenclatura actual estipulada por el NCBI. En la tercera se muestra qué alelo de cada SNP se asocia a la persistencia de la lactasa. Los \*\* hacen referencia a: SNPs funcionales validados.

## 3.4. GENERACIÓN Y OPTIMIZACIÓN DE LIBRERÍAS

Tras el diseño de los cebadores, la generación de librerías se llevó a cabo mediante dos amplificaciones sucesivas: una PCR multiplex y una PCR anidada, las cuales se explican a continuación.

### 3.4.1. Optimización PCR multiplex (PCR 1)

La PCR multiplex es una reacción de amplificación en la cual el conjunto de cebadores diseñados para la amplificación de diferentes fragmentos de ADN se incluye en la misma reacción de PCR, permitiendo así la amplificación simultánea de varias secuencias de interés en una única reacción (Shen, 2019).

En este caso, la PCR 1 es una PCR multiplex selectiva, realizada con el objetivo de amplificar la región de interés. Para ello, la misma muestra fue amplificada en dos reacciones de PCR 1 independientes: una con el *pool* de cebadores 1 y otra con el *pool* 2, llevadas a cabo en las mismas condiciones. De este



modo, los cebadores fueron divididos en función de su posible solapamiento, evitando además la formación de *cross-dimers*, tal y como se ha explicado previamente en el diseño de cebadores (Figura 5).

La amplificación se llevó a cabo en placa de 384 pocillos, utilizando unas condiciones iniciales en las cuales se partió de 2  $\mu\text{L}$  de ADN genómico (10  $\text{ng}/\mu\text{L}$ ), junto con 7,5  $\mu\text{L}$  de mix de reactivos de PCR (ADN polimerasa + *buffer*), 2  $\mu\text{L}$  de la mix de cebadores específicos y 3,50  $\mu\text{L}$  de agua ultrapura (Tabla 4). No obstante, dado que en el presente trabajo experimental se llevó a cabo un protocolo de optimización del proceso, las concentraciones y volúmenes de algunos de los reactivos previamente mencionados sufrieron variaciones en las diferentes pruebas realizadas, pudiendo diferir de las cantidades indicadas, con el objetivo de conseguir las condiciones óptimas para el mismo.

Del mismo modo, el programa de PCR escogido fue modificado en las diferentes pruebas, variando tiempos, temperaturas o número de ciclos, hasta dar con el que permitió obtener mejores resultados. En este caso, el escogido en las condiciones iniciales partió de una activación inicial de la ADN polimerasa, la cual tiene lugar a 95°C durante 15 min, seguida de entre 30 y 40 ciclos consecutivos de: desnaturalización de las hebras de ADN (95°C durante 30 segundos), alineamiento o *annealing* (60°C durante 30 segundos) y extensión (72°C durante 30 segundos). La extensión final se realizó a 72°C durante 5 minutos, seguida de un descenso de la temperatura y mantenimiento a 4°C (Tabla 5).

**Tabla 4. Reactivos de partida utilizados para la realización de la PCR multiplex (PCR 1).**

Reactivos	Volumen ( $\mu\text{L}$ )	[ $\mu\text{M}$ ] EN PCR
Máster mix PCR	7,50	1,00
Cebadores (2,63 $\mu\text{M}$ )	2,00	0,35
<i>Agua</i>	3,50	
DNA (10 $\text{ng}/\mu\text{L}$ )	2,00	0,67
<i>Volumen total</i>	15,00	

**Tabla 5. Condiciones iniciales utilizadas en la PCR multiplex (PCR 1).**

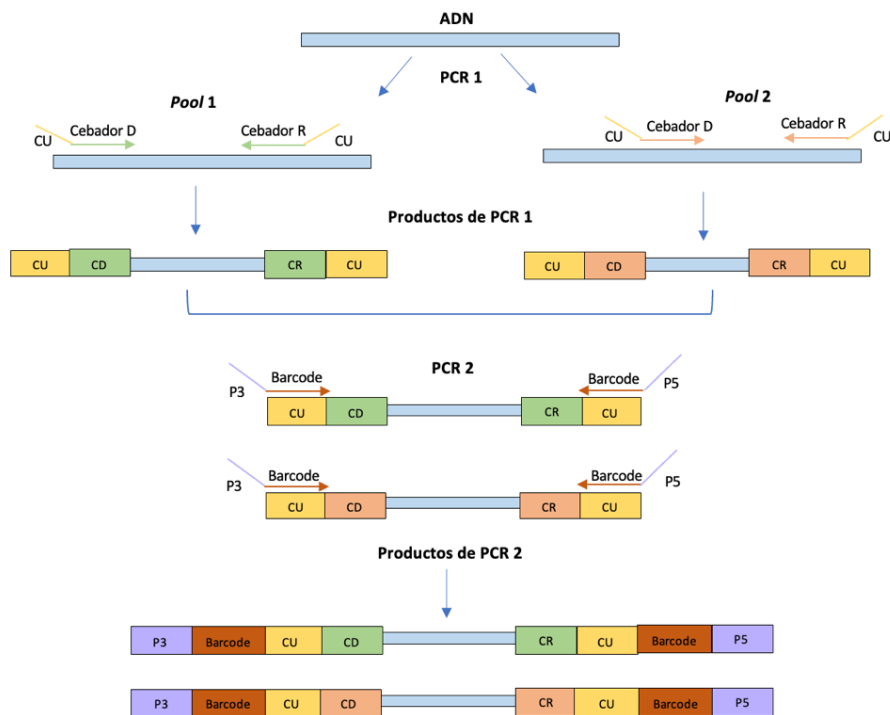
Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95°C	15 min	
Desnaturalización	95 °C	30 seg	40X
Annealing	60°C	30 seg	
Extensión	72 °C	30 seg	
Extensión final	72 °C	5 min	
Hold	4°C	$\infty$	

### 3.4.2. Optimización PCR anidada (PCR 2)

La PCR 2, o PCR anidada, consiste en una segunda reacción de amplificación del producto ya amplificado, cuyo propósito se basa en la adición de adaptadores necesarios para el proceso de secuenciación, junto con secuencias que permitirán identificar las muestras tras la secuenciación, denominadas secuencias identificadoras o *barcodes*. La adición de dichas secuencias es posible gracias

a que este segundo conjunto de cebadores posee una secuencia complementaria a la cola universal (CU), otorgada a la secuencia amplificada por el primer grupo de cebadores, permitiendo así su hibridación (Shen, 2019).

De este modo, en la PCR 2, los productos previamente amplificados (*pool 1* y *pool 2*) fueron utilizados como ADN de interés, siendo amplificados de forma conjunta en una única reacción de PCR, al no ser ya necesario hacer una separación de las parejas de cebadores solapantes (*Figura 6*). Como resultado de la adición de los *barcodes* y los adaptadores de secuenciación, se obtuvo un tamaño final de librería de unas 300 pares de bases (pb), resultado de la suma del fragmento previamente amplificado y dichas secuencias.



**Figura 6. Representación del proceso de generación de librerías.** En la PCR1 se produce la amplificación del fragmento de interés a partir de los cebadores previamente diseñados, los cuales poseen colas universales (CU) para su hibridación con la ADN polimerasa. En la PCR 2 se adicionan los barcodes, que permiten la posterior identificación de las muestras. Estos poseen una secuencia complementaria a las CU adicionadas en la PCR1, permitiendo su hibridación. Por su otro extremo, estos poseen los adaptadores de secuenciación P3 y P5, necesarios para el proceso de secuenciación. CD: cebador directo; CR: cebador reverso; CU: cola universal. Diseñada mediante Microsoft Word.

Al igual que en la PCR 1, con el fin de realizar dicha amplificación, se seleccionaron unas condiciones iniciales, las cuales fueron variando en función de las pruebas realizadas. De este modo, se partió de 2  $\mu\text{L}$  de producto de PCR1 (1 $\mu\text{L}$  producto de PCR *pool 1* + 1  $\mu\text{L}$  producto de PCR *pool 2*) junto con 7,5  $\mu\text{L}$  de mix de reactivos de PCR (ADN polimerasa + *buffer*), 2  $\mu\text{L}$  de *barcodes* específicos y 3,50  $\mu\text{L}$  de agua ultrapura (*Tabla 6*). Sin embargo, en este caso, el programa de PCR utilizado fue el mismo para todas las pruebas, el cual comenzó con una activación inicial de la ADN polimerasa a 95 $^{\circ}\text{C}$  durante 15 minutos, seguida de 38 ciclos consecutivos de: desnaturalización de las hebras de ADN (98 $^{\circ}\text{C}$  durante 20 segundos), alineamiento o *annealing* (60 $^{\circ}\text{C}$  durante 30 segundos) y extensión (72 $^{\circ}\text{C}$  durante 1 minuto). La extensión final se realizó a 72 $^{\circ}\text{C}$  durante 10 minutos, seguida de un descenso de la temperatura y mantenimiento a 4 $^{\circ}\text{C}$  (*Tabla 7*).

**Tabla 6. Reactivos utilizados para la realización de la PCR anidada (PCR 2).**

Reactivos	Volumen (μL)	[μM] EN PCR
Máster mix PCR	7,50	1,00
Cebadores <i>barcode</i> (5μM)	2,00	0,67
<i>Agua</i>	3,50	
Producto PCR1	2,00	
<i>Volumen total</i>	15,00	

**Tabla 7. Condiciones utilizadas en la PCR anidada (PCR 2).**

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95°C	15 min	
Desnaturalización	98 °C	20 seg	38x
Annealing	60°C	30 seg	
Extensión	72 °C	1 min	
Extensión final	72 °C	10 min	
Hold	4 °C	∞	

### 3.4.3. Electroforesis capilar

Una vez realizadas las correspondientes PCR, se comprobó si la amplificación de los fragmentos se había realizado de forma correcta, obteniendo fragmentos del tamaño deseado. Dicha comprobación se llevó a cabo por medio de la técnica de electroforesis capilar, la cual se basa en la separación de los fragmentos de ADN en función de su carga y tamaño cuando se les aplica un campo eléctrico.

El sistema utilizado para realizar dicha técnica fue el sistema **QIAxcel (QIAxcel DNA Screening Kit, Qiagen®, Alemania)**, el cual utiliza cartuchos de gel, sustituyendo a la tradicional electroforesis en gel de agarosa. Este es un sistema mucho más eficiente, dada su velocidad, y que proporciona mejores resultados, con una elevada resolución. Con el fin de conocer el tamaño de los fragmentos amplificados se utilizó un marcador de tamaños, siendo en este caso el marcador escogido de 25-500 pb, dado que se esperaba obtener fragmentos de 300 pb. A su vez, resultó necesario utilizar un marcador de alineamiento, que indicara el mínimo y el máximo de bases que se podían detectar, siendo en este caso utilizado el alineamiento de entre 15-1000 pb.

En cada uno de los pocillos se adicionó un volumen total de 10 μL, correspondiente a 2 μL de muestra y 8 μL de tampón o *buffer*, siendo en el caso de los pocillos sobrantes la totalidad del volumen correspondiente al tampón o *buffer*.

### 3.5. PURIFICACIÓN DE LIBRERÍAS

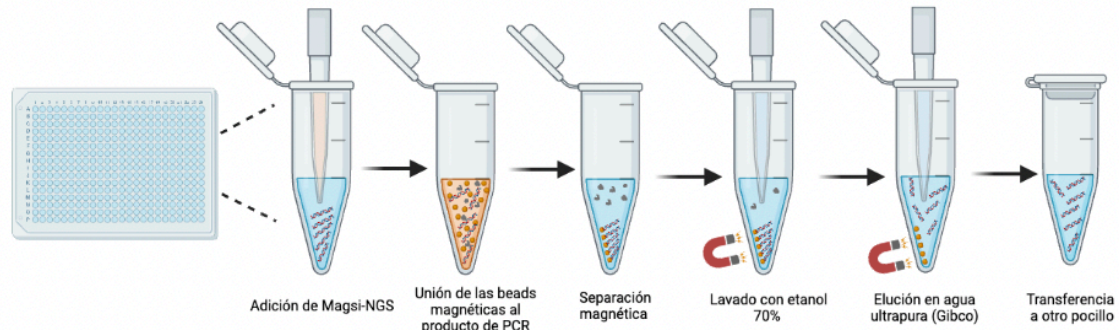
La purificación de librerías tiene como objetivo la eliminación de productos de PCR no deseados, menores de 200 pb, tales como restos de reactivos, productos de PCR inespecíficos, así como cebadores que han dimerizado entre sí (*primer dimer*) o que no han hibridado.

Para ello, se utilizó el kit de purificación **Magsi-NGS PREP® (MagnaMedics Diagnostics B.V., The Netherlands)**, el cual fue optimizado previamente, permitiendo la eliminación de los productos

indeseados. Dicho kit se basa en el uso de *beads* o bolas magnéticas que permiten la correcta purificación de las librerías al estar cargadas positivamente, favoreciendo la unión del ADN (cargado negativamente).

El protocolo indicado por el fabricante (*Figura 7*) es el siguiente:

1. Adición de Magsi-NGS PREP a una concentración de 0,6X, siendo dicho volumen el limitante que determinará que únicamente los fragmentos de ADN del tamaño deseado se unan a las *beads*, eluyendo los fragmentos menores de 200 pb.
2. Tapar la placa, vortexear durante 5'' y dejar 5' a temperatura ambiente.
3. Centrifugar 15'' a 1000 rpm. Poner sobre la placa magnética durante 5', de modo que se produzca la unión del ADN a las *beads*.
4. Sin retirar de la placa magnética, retirar el sobrenadante por medio de golpes poca abajo.
5. Adicionar 20  $\mu$ l de etanol al 70% y mantener en la placa magnética durante 30''.
6. Retirar de nuevo el sobrenadante y volver a adicionar etanol al 70 % (pasos 4 y 5).
7. Eliminar todo el sobrenadante, retirar de la placa magnética y dejar secar durante 10' para evaporar todo el etanol restante.
8. Adicionar agua ultrapura (Gibco), de modo que el volumen sea igual al volumen de muestra a purificar. Vortexear 5'' y centrifugar durante 15'' a 1000 rpm. Incubar en el agua 5' a temperatura ambiente.
9. Poner de nuevo sobre la placa magnética 5'.
10. Finalmente, tomar todo el sobrenadante (en el cual se encuentra el ADN) sin retirar de la placa magnética y pasarlo a nuevos pocillos, de modo que las *beads* queden en los pocillos anteriores.



**Figura 7. Representación del proceso de purificación de la librería mediante el kit de purificación Magsi-NGS PREP®. Diseñada mediante Biorender.**

### 3.6. CUANTIFICACIÓN

Tras la purificación, resulta necesario cuantificar las librerías con el objetivo de equilibrar las concentraciones de cada una de las muestras en el *pool* final que se desea secuenciar, de modo que todos los fragmentos obtengan un número de lecturas homogéneo.

La cuantificación se llevó a cabo por fluorescencia mediante el uso de un espectrofluorímetro utilizando el kit **Quantifluor® dsDNA System**. Este permitió interpolar la concentración de las librerías en una recta patrón previamente realizada (*Figura 8*), gracias a los datos de fluorescencia emitidos como resultado del uso de un colorante de fluorescencia de unión al dsDNA, con rangos de absorbancia y emisión de entre 510nm y 570 nm. La recta patrón se realizó a partir de diluciones seriadas de una

concentración de ADN del fago  $\lambda$  (dsDNA) en TE 1X como tampón de dilución y como blanco, a la cual se le adicionó en última instancia el colorante **QuantiFluor**®.

El proceso detallado es el siguiente:

1. Preparación de la mix QuantiFluor.
  - a. Dilución del fluoróforo en tampón TE 1X en una proporción 1:200.
2. Curva patrón:
  - a. En un tubo eppendorf, adicionar a 240  $\mu$ l de tampón TE 1X 10  $\mu$ l de fago  $\lambda$  (dilución 1:25).
  - b. Traspasar 200  $\mu$ l de la mix preparada en el paso 2a a un nuevo eppendorf y adicionar 200  $\mu$ l de tampón TE 1X.
  - c. A partir de la dilución del paso anterior, realizar diluciones seriadas 1:2 en tampón TE 1X, a excepción del octavo tubo, el cual únicamente contiene tampón TE 1X y servirá como blanco.
3. En una placa de cuantificación, adicionar 100  $\mu$ l de cada una de las diluciones previamente elaboradas, cuya concentración servirá para elaborar la recta patrón.
4. En el resto de los pocillos, se procede a adicionar 2  $\mu$ l de cada una de las muestras junto con 98  $\mu$ l de TE 1X.
5. Finalmente, se adicionan en todos los pocillos, tanto a los correspondientes a la curva como a los que contienen muestra, 100  $\mu$ l de la mix QuantiFluor, elaborada en el primer paso.

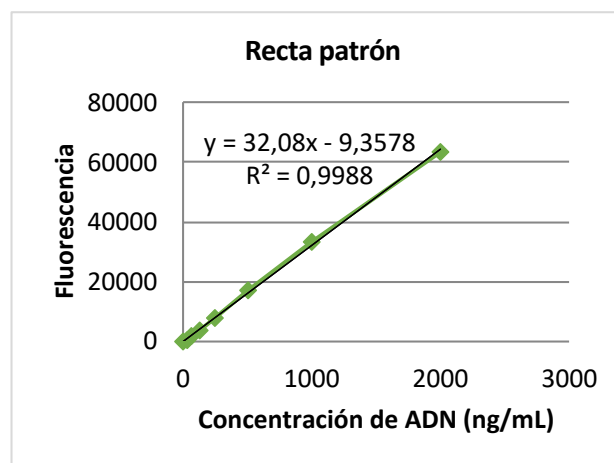


Figura 8. Ejemplo de recta patrón para la cuantificación de librerías.

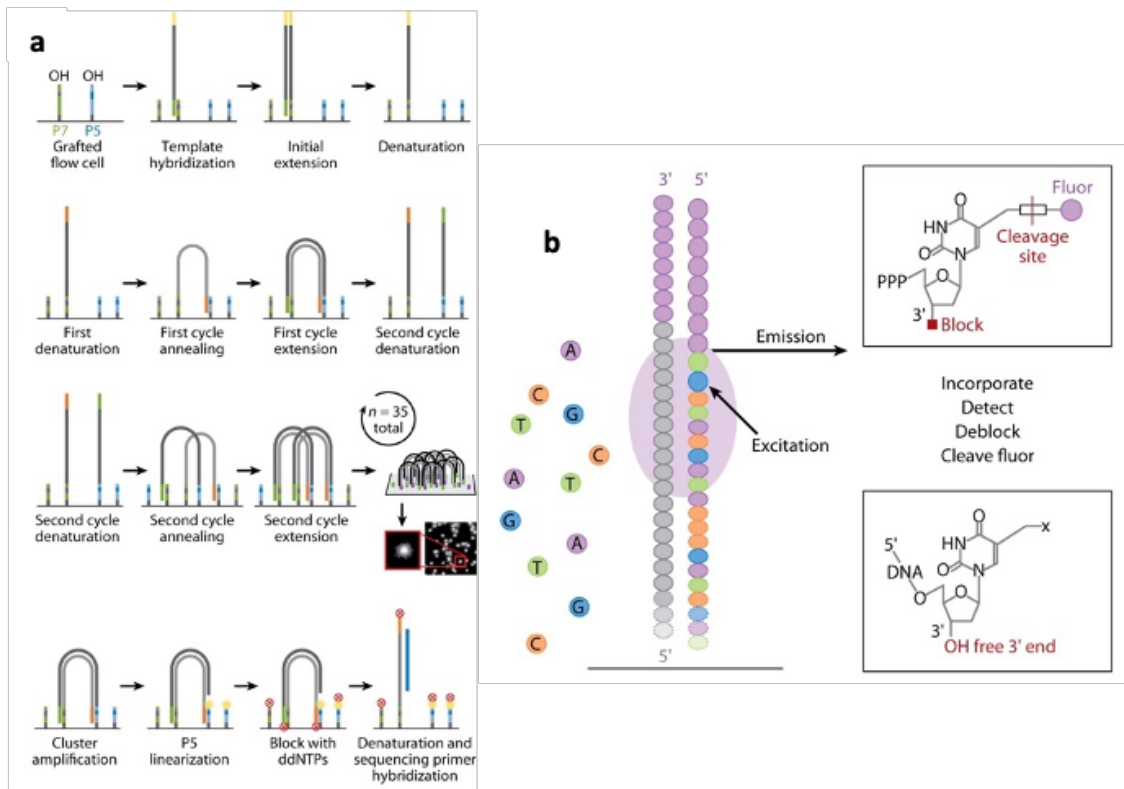
### 3.7. SECUENCIACIÓN

La secuenciación se llevó a cabo utilizando el sistema de secuenciación MiSeq® (Illumina®, California) de INCLIVA, el cual se basa en una secuenciación por síntesis mediante la amplificación en puente.

El procedimiento estándar de Illumina comienza con la preparación de la celda de flujo, la cual permite realizar miles de reacciones de secuenciación de forma simultánea, al separar espacialmente los diferentes sitios de unión de los diferentes fragmentos del material genómico a secuenciar. Para ello, en primer lugar, se produce una agrupación de los fragmentos genómicos en grupos o *clusters*, al poseer dichos fragmentos secuencias complementarias a las secuencias adaptadoras presentes en la celda de flujo del secuenciador, las cuales fueron añadidas junto con los *barcodes* durante la amplificación por PCR anidada, produciéndose una unión covalente entre estas. A continuación, la ADN polimerasa sintetiza la cadena complementaria, produciéndose una posterior desnaturalización. El extremo de la hebra sintetizada que queda libre también posee dichas secuencias adaptadoras,

pudiendo interactuar con cebadores cercanos, formando así una estructura en forma de puente y su posterior amplificación por PCR puente (*Figura 9a*) (Mardis, 2013; Metzker, 2010).

La secuenciación propiamente dicha comienza tras la eliminación de la hebra reversa, mediante reacciones de terminación cíclica reversible (CRT). Este enfoque se basa en el uso de desoxinucleótidos (dNTPs) fluorescentes cuyo grupo 3'-OH ribosa ha sido bloqueado por un O-azidometilio, impidiendo de este modo la elongación de la cadena. De este modo, durante cada ciclo, los fragmentos de cada *cluster* incorporan un solo nucleótido, al haberse impedido la extensión, siendo los dNTPs no incorporados eliminados por lavado. A continuación, dicho dNTP es identificado, al obtener una imagen de la celda de flujo con las señales de excitación de este. Finalmente, el grupo bloqueante es desbloqueado químicamente, repitiendo de nuevo el proceso hasta obtener la secuencia del fragmento de interés. De este modo, se generan una gran cantidad de datos que requieren su posterior análisis bioinformático (*Figura 9b*) (Mardis, 2013; Metzker, 2010).



**Figura 9. Diagrama del proceso de secuenciación en el sistema MiSeq de Illumina.** A) Generación de clusters por amplificación en puente. B) Secuenciación por síntesis con terminadores fluorescentes reversibles. Adaptada de (Mardis, 2013).

### 3.8. ANÁLISIS BIOINFORMÁTICO

El análisis bioinformático fue llevado a cabo por el personal bioinformático de la Unidad de Genómica y Diabetes, con el objetivo de evaluar la calidad de las coberturas obtenidas para los diferentes amplicones tras el proceso de secuenciación. Para ello se utilizó una *pipeline* bioinformática desarrollada en la UGD, basada en el uso de diferentes *softwares* y *scripts*.

Dicho análisis comenzó con el proceso de demultiplexado, el cual consiste en identificar qué secuencias corresponden a cada una de las muestras mediante el uso de los *barcodes* añadidos en el proceso de amplificación. De este modo, los datos obtenidos de la secuenciación fueron procesados y almacenados como archivos FASTQ, generándose dos archivos FASTQ por fragmento secuenciado,

dada la naturaleza del método de secuenciación de Illumina. Una vez completado este paso, los archivos FASTQ obtenidos fueron sometidos a un control de calidad utilizando el programa Fastqc (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>), el cual aplica diversos filtros que determinan si las secuencias poseen la calidad y la puntuación adecuadas, expresando los posibles errores de secuenciación en función de la puntuación Phread (probabilidad de error logarítmica), siendo Q20 la probabilidad de que una base sea incorrecta en cada 100 bases secuenciadas (Pereira *et al.*, 2020).

Posteriormente se realizó el proceso de *trimming* o recorte de extremos, el cual implica la eliminación de regiones no deseadas, como adaptadores, y secuencias de baja calidad. Una vez completado este paso, los datos obtenidos se sometieron a un segundo control de calidad, para garantizar que el proceso de *trimming* se hubiese realizado adecuadamente. Finalmente, se llevó a cabo un alineamiento o mapeo de las secuencias contra el genoma de referencia (GRCh38/hg38), permitiendo determinar la ubicación precisa de las lecturas obtenidas en el genoma de referencia.

Después de realizar el análisis, se obtuvieron indicadores de profundidad y amplitud de cobertura para los amplicones secuenciados, indicando las coberturas mínimas y máximas obtenidas para cada amplicón. Utilizando estos datos de cobertura, se procedió al análisis de estos, para lo cual se generaron gráficas representando la cobertura mínima obtenida para cada uno de ellos. Estas gráficas fueron comparadas entre sí para determinar las condiciones que mostraron mejores resultados en términos de cobertura, evaluando la calidad y la eficiencia del proceso de secuenciación.

### **3.9. OPTIMIZACIÓN**

Una vez obtenidos los resultados tras el análisis bioinformático, se evaluó si las coberturas de secuenciación resultaban ser las adecuadas para los diferentes amplicones, corroborando si el número de lecturas era similar para todos los amplicones, o si la cobertura obtenida para alguno de ellos no era la adecuada, requiriéndose en este caso una optimización de las condiciones del panel. Dicha optimización requirió modificar las concentraciones de los diferentes cebadores en la PCR multiplex (PCR 1), de modo que se produjese un mayor producto de los fragmentos amplificados por dichos cebadores, a fin de que estos obtuvieran un mayor número de lecturas durante el proceso de secuenciación. De este modo, el objetivo de la optimización fue obtener una cobertura homogénea para los diferentes amplicones. También fue necesario modificar el volumen de otros reactivos, así como de las condiciones de la PCR1. En el presente diseño experimental se fueron modificando dichas condiciones, así como la distribución de los cebadores en los diferentes *pools*, con el objetivo de reducir la presencia de productos inespecíficos, hasta lograr obtener la condición idónea que permitiera una cobertura homogénea de todos los amplicones, la cual pudo utilizarse posteriormente para el análisis de los alelos de riesgo de una determinada población.

## **4. RESULTADOS Y DISCUSIÓN DE LOS RESULTADOS**

### **4.1. DISEÑO DE CEBADORES**

Los cebadores se diseñaron para la amplificación de los genes *HLA-DQA1*, *HLA-DQB1* y *LCT*, partiendo de la secuencia genómica de los mismos (*Figura 10*). En el caso de los genes *HLA-DQA1* y *HLA-DQB1*, un total de 22 cebadores (directos y reversos) fueron diseñados para llevar a cabo su amplificación, siendo necesario en algunos de los casos más de una pareja de cebadores por fragmento, dada la gran cantidad de SNPs que estos presentan, dificultando su correcta hibridación (*Figura 11*). En cuanto a los genes de la persistencia a la lactasa (*LCT*), fueron 4 los cebadores diseñados.





### 4.2.1. Prueba 1

La prueba 1 se realizó con el objetivo de comparar los resultados de coberturas de los distintos fragmentos amplificados con o sin los cebadores diseñados para la detección de las variantes de la no persistencia a la lactasa (*LAC-13900* Y *LAC-22000*).

Para ello, se prepararon cuatro *pools* de cebadores, necesarios para llevar a cabo la primera PCR, en los cuales la separación de estos se realizó según las condiciones previas utilizadas en la UGD, teniendo en cuenta el diseño de estos y el posible solapamiento (*Tabla 8*).

En este caso, todas las parejas de cebadores se adicionaron a la misma concentración en cada una de las condiciones, realizándose para cada una de ellas tres pruebas a distintas concentraciones: 1X (o concentración inicial) y diluciones 1/10X y 1/20X de la *mix* original. La amplificación de los fragmentos se llevó a cabo en dos reacciones sucesivas (PCR 1 y PCR 2) bajo distintas condiciones (*Tabla 9, 10*).

**Tabla 8. Distribución de cebadores utilizada en la prueba 1.** Para cada una de las condiciones (con o sin los cebadores para el gen de la lactasa), los cebadores fueron divididos en distintos pools (pool 1 y pool2).

POOLS CON LACTASA		POOLS SIN LACTASA	
POOL 1 DQ-LAC	POOL 2 DQ-LAC	POOL 1 DQ	POOL 2 DQ
DQA1-E2A-5	DQA1-E2B-5	DQA1-E2A-5	DQA1-E2B-5
DQA1-E2A-3	DQA1-E2B-3	DQA1-E2A-3	DQA1-E2B-3
DQA1-E2C-5S	DQA1-E3b-5	DQA1-E2C-5S	DQA1-E3b-5
DQA1-E2C-5C	DQA1-E3b-3	DQA1-E2C-5C	DQA1-E3b-3
DQA1-E2C-3	DQA1-E3B-5	DQA1-E2C-3	DQA1-E3B-5
DQA1-E3A-5	DQA1-E3B-3	DQA1-E3A-5	DQA1-E3B-3
DQA1-E3A-3	DQB1-E2B-5	DQA1-E3A-3	DQB1-E2B-5
DQB1-E2A-5n	DQB1-E2B-3S	DQB1-E2A-5n	DQB1-E2B-3S
DQB1-E2Ab-3	DQB1-E2B-3J	DQB1-E2Ab-3	DQB1-E2B-3J
DQB1-E3-5	DQB1-E2C-3	DQB1-E3-5	DQB1-E2C-3
DQB1-E3-3	DQB1-E2C-5	DQB1-E3-3	DQB1-E2C-5
LAC-13900-5			
LAC-13900-3			
LAC-22000-5			
LAC-22000-3			

**Tabla 9. Condiciones de amplificación de la PCR 1 utilizadas en la prueba 1.**

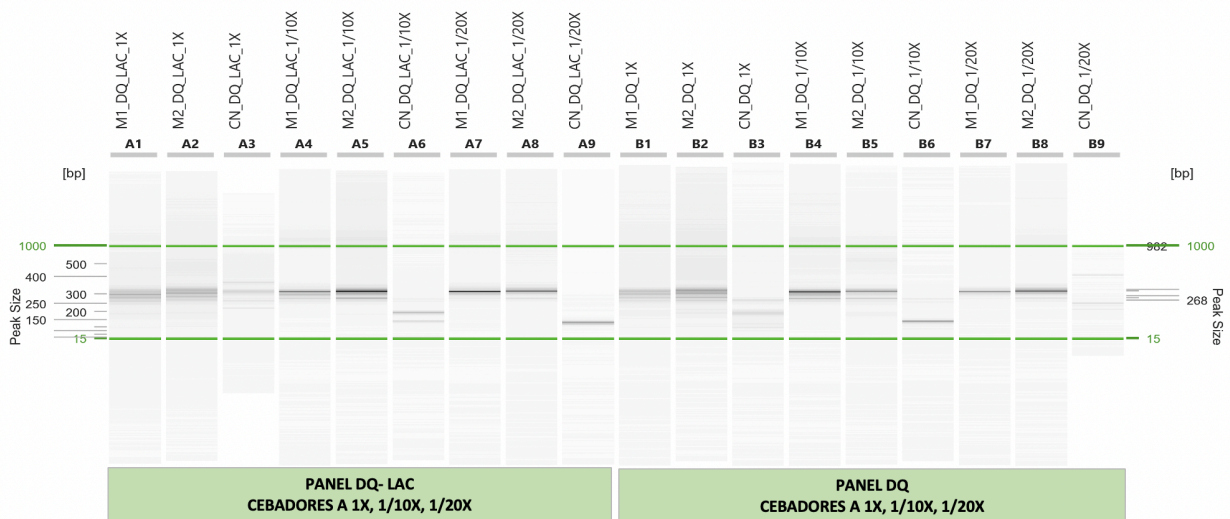
Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95°C	15 min	
Desnaturalización	95 °C	30 seg	40x
Annealing	60°C	30 seg	
Extensión	72 °C	30 seg	
Extensión final	72 °C	5 min	
Hold	4°C	∞	

**Tabla 10. Condiciones de amplificación de la PCR 2 utilizadas en la prueba 1.**

Programa	Temperatura	Tiempo	Ciclos
Activación inicial	95°C	15 min	
Desnaturalización	98 °C	20 seg	38x
Annealing	60°C	30 seg	
Extensión	72 °C	1 min	
Extensión final	72 °C	10 min	
Hold	4 °C	∞	

Posteriormente, se evaluó, mediante electroforesis capilar, el tamaño de los fragmentos obtenidos tras la amplificación, comprobando que la amplificación se había realizado correctamente y que las muestras no habían sido contaminadas. En los resultados obtenidos, se observaron bandas de mayor intensidad entorno a las 300pb, como resultado de la suma del fragmento amplificado y los distintos cebadores utilizados para llevar a cabo las dos amplificaciones. Las bandas entorno a 150 pb correspondieron a la formación de dímeros entre los distintos adaptadores durante el proceso de amplificación, siendo las inferiores a estas resultado de productos inespecíficos, los cuales fueron posteriormente eliminados mediante el proceso de purificación (Figura 12).

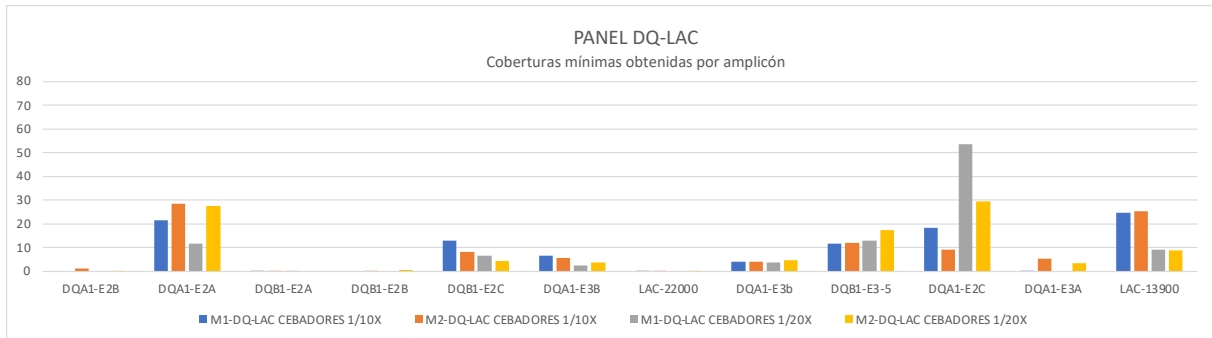
Se observaron también bandas en los pocillos correspondientes a los controles negativos, resultado de la interacción existente entre algunos de los cebadores. No obstante, también aparecieron bandas cercanas a 300 pb en los controles negativos de la condición 1X, por lo que dicha condición no fue secuenciada, al poderse tratar de una posible contaminación (Figura 12).



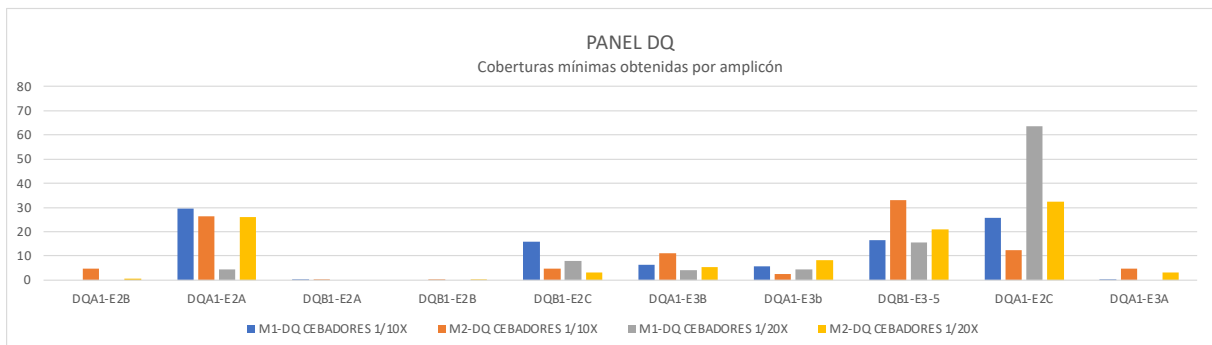
**Figura 12. Resultados de la electroforesis capilar para la prueba 1.** Se muestran las bandas resultantes de la amplificación de la muestra 1 (M1), muestra 2 (M2) y control negativo (CN) con una concentración de cebadores a 1X, 1/10X y 1/20X para los paneles DQ-LAC y DQ respectivamente.

En este caso, dado que las bandas correspondientes a las condiciones de los cebadores a 1X en ambos paneles mostraban mucha menor nitidez y aumentaba la presencia del dímero o de una posible contaminación en los respectivos controles negativos, estas fueron descartadas y, por tanto, no fueron secuenciadas. El resto de las muestras correspondientes a las condiciones a 1/10X y 1/20X fueron purificadas y cuantificadas, juntándolas posteriormente en un nuevo *pool* a una concentración equimolar, el cual se secuenció siguiendo el procedimiento estándar de *Illumina*.

Tras la secuenciación, se llevó a cabo el análisis bioinformático por personal del laboratorio, obteniendo de este modo los siguientes porcentajes de coberturas mínimas obtenidas por fragmento amplificado, en función de las muestras y las concentraciones de los cebadores utilizados para los distintos paneles (Figura 13, 14).



**Figura 13. Distribución de lecturas mínimas obtenidas por amplicón secuenciado en el panel DQ-LAC de la prueba 1.** Los distintos colores hacen referencia a los fragmentos amplificados con los cebadores a una concentración de 1/10X y 1/20X para las muestras 1 y 2 respectivamente.



**Figura 14. Distribución de lecturas mínimas obtenidas por amplicón secuenciado en el panel DQ de la prueba 1.** Los distintos colores hacen referencia a los fragmentos amplificados con los cebadores a una concentración de 1/10X y 1/20X para las muestras 1 y 2 respectivamente.

Como puede observarse, las coberturas obtenidas en esta prueba no resultaron ser homogéneas, sino que algunos de los fragmentos recibieron muy pocas lecturas, lo cual pudo deberse a diferentes factores durante la reacción de amplificación (secuencia e hibridación de los cebadores, interacciones o uniones inespecíficas de los cebadores o incluso de los amplicones entre sí, condiciones de la PCR, etc.), siendo necesario continuar con el proceso de optimización.

Por su parte, tampoco se observó una diferencia notable entre los diferentes paneles (con o sin el gen de la persistencia a la lactasa), por lo que los cebadores para el gen de la persistencia a la lactasa no parecieron interactuar con el resto de los cebadores de manera general. No obstante, dado que el amplicón *LAC-22000* no obtuvo apenas lecturas, sí que podría darse el caso de que los cebadores para el mismo hubiesen interactuado con otros cebadores, disminuyendo así el producto amplificado, resultando necesarias más pruebas para obtener una conclusión.

#### 4.2.2. Prueba 2

En la prueba 2 se probaron nuevas combinaciones de cebadores, aumentando además la concentración de los correspondientes a aquellos fragmentos con una menor cobertura (*Tabla 11, 12*). Para ello, se probaron otras dos condiciones, con y sin los cebadores para el gen de la persistencia a la lactasa, en las cuales se intercambiaron de *pool* los cebadores de los fragmentos correspondientes al exón 2 del gen *HLA-DQB1*, separando *DQB1-E2A* de *DQB1-E2B* y *DQB1-E2C*, con el objetivo de comprobar si existía algún tipo de interacción entre ellos que hubiese condicionado el resultado de las coberturas obtenidas en la prueba anterior. Además, se añadió un 50% más del volumen de estos, así como de los cebadores para el fragmento correspondiente a *HLA-DQA1-E3A*, al haberse obtenido

también una baja cobertura. Por tanto, de todos los cebadores se añadieron 15  $\mu\text{L}$ , excepto de los cebadores mencionados, de los cuales se añadieron 22,5  $\mu\text{L}$ .

Con respecto a la prueba 1, se separaron también en *pools* diferentes los cebadores correspondientes a los fragmentos *LAC-22000* y *LAC-13900*, pero la concentración de estos se mantuvo como la del resto.

La amplificación de los fragmentos tuvo lugar bajo las mismas condiciones de PCR que las utilizadas para la prueba 1. Posteriormente, la electroforesis capilar se realizó para evaluar el tamaño de los fragmentos obtenidos, observando bandas en el tamaño esperado, cercanas a 300 pb (*Figura 15,16*).

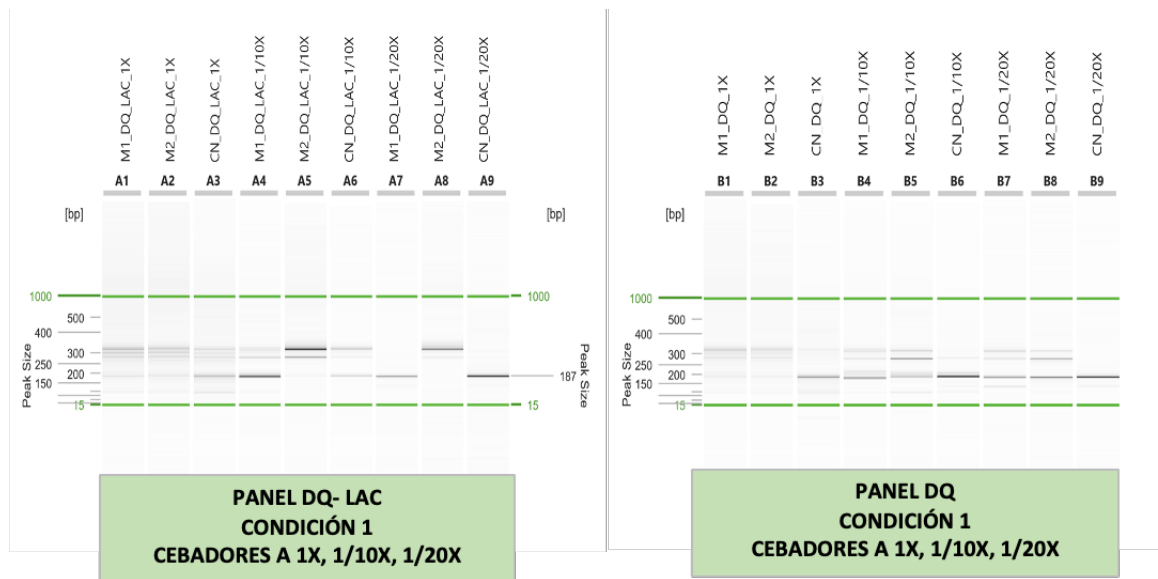
**Tabla 11. Distribución de cebadores utilizada en la condición 1 de la prueba 2.**

CONDICIÓN 1			
POOLS CON LACTASA		POOLS SIN LACTASA	
POOL 1 DQ-LAC	POOL 2 DQ-LAC	POOL 1 DQ	POOL 2 DQ
DQA1-E2A-5	DQA1-E2B-5	DQA1-E2A-5	DQA1-E2B-5
DQA1-E2A-3	DQA1-E2B-3	DQA1-E2A-3	DQA1-E2B-3
DQA1-E2C-5S	DQA1-E3b-5	DQA1-E2C-5S	DQA1-E3b-5
DQA1-E2C-5C	DQA1-E3b-3	DQA1-E2C-5C	DQA1-E3b-3
DQA1-E2C-3	DQA1-E3B-5	DQA1-E2C-3	DQA1-E3B-5
DQA1-E3A-5	DQA1-E3B-3	DQA1-E3A-5	DQA1-E3B-3
DQA1-E3A-3	DQB1-E2A-5n	DQA1-E3A-3	DQB1-E2A-5n
DQB1-E2B-5	DQB1-E2Ab-3	DQB1-E2B-3J	DQB1-E2Ab-3
DQB1-E2B-3J	LAC-22000-5	DQB1-E2B-5	
DQB1-E2B-3S	LAC-22000-3	DQB1-E2B-3S	
DQB1-E2C-3		DQB1-E2C-3	
DQB1-E2C-5		DQB1-E2C-5	
DQB1-E3-5		DQB1-E3-5	
DQB1-E3-3		DQB1-E3-3	
LAC-13900-5			
LAC-13900-3			

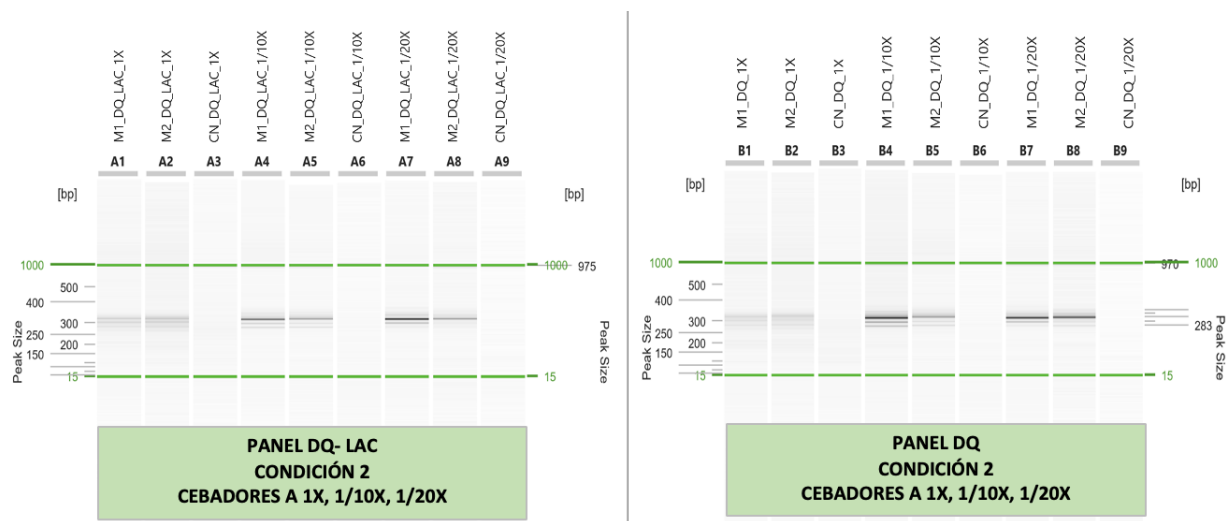
**Tabla 12. Distribución de cebadores utilizada en la condición 2 de la prueba 2.**

CONDICIÓN 2			
POOLS CON LACTASA		POOLS SIN LACTASA	
POOL 1 DQ-LAC	POOL 2 DQ-LAC	POOL 1 DQ	POOL 2 DQ
DQA1-E2A-5	DQA1-E2B-5	DQA1-E2A-5	DQA1-E2B-5
DQA1-E2A-3	DQA1-E2B-3	DQA1-E2A-3	DQA1-E2B-3
DQA1-E2C-5S	DQA1-E3b-5	DQA1-E2C-5S	DQA1-E3b-5
DQA1-E2C-5C	DQA1-E3b-3	DQA1-E2C-5C	DQA1-E3b-3
DQA1-E2C-3	DQA1-E3B-5	DQA1-E2C-3	DQA1-E3B-5
DQA1-E3A-5	DQA1-E3B-3	DQA1-E3A-5	DQA1-E3B-3
DQA1-E3A-3	DQB1-E2B-5	DQA1-E3A-3	DQB1-E2B-3J
DQB1-E2A-5n	DQB1-E2B-3J	DQB1-E2A-5n	DQB1-E2B-5
DQB1-E2Ab-3	DQB1-E2B-3S	DQB1-E2Ab-3	DQB1-E2B-3S
DQB1-E3-5	DQB1-E2C-3	DQB1-E3-5	DQB1-E2C-3
DQB1-E3-3	DQB1-E2C-5	DQB1-E3-3	DQB1-E2C-5
LAC-13900-5	LAC-22000-5		
LAC-13900-3	LAC-22000-3		

En ambas condiciones, de los cebadores marcados en color se adicionó un 50% más de volumen que en el resto de los cebadores, indicando además aquellos marcados en azul y naranja los cebadores que se intercambiaron de pool en la condición 2.



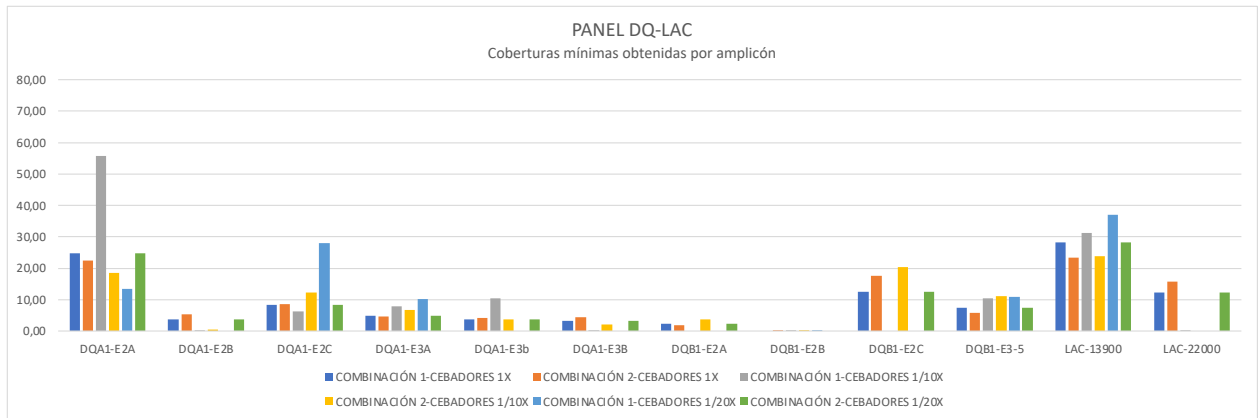
**Figura 15. Resultados de la electroforesis capilar para la condición 1 de la prueba 2.** Se muestran las bandas resultantes de los fragmentos obtenidos tras las dos reacciones de amplificación para la muestra 1 (M1), muestra 2 (M2) y control negativo (CN) con una concentración de cebadores a 1X, 1/10X y 1/20X para el panel DQ-LAC y DQ respectivamente.



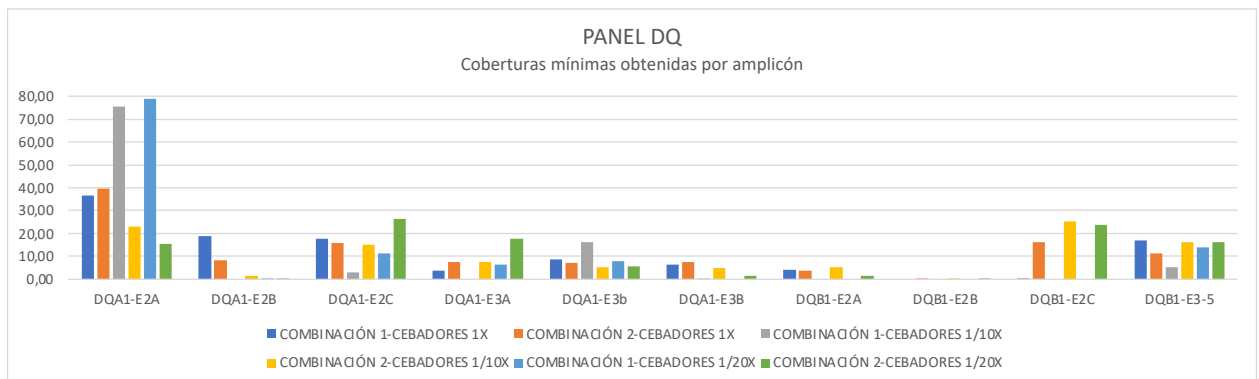
**Figura 16. Resultados de la electroforesis capilar para la condición 2 de la prueba 2.** Se muestran las bandas resultantes de los fragmentos obtenidos tras las dos reacciones de amplificación para la muestra 1 (M1), muestra 2 (M2) y control negativo (CN) con una concentración de cebadores a 1X, 1/10X y 1/20X para el panel DQ-LAC y DQ respectivamente.

Teniendo en cuenta los resultados de coberturas obtenidos para cada uno de los amplicones (Figura 17, 18), se comprobó que la adición de los cebadores seleccionados a una mayor concentración aumentó la cobertura de los correspondientes amplicones, a excepción del exón 2 del gen *HLA-DQB1-E2B*, el cual de nuevo no apareció representado.

No obstante, a pesar de que, en general, se obtuvieron mejores resultados que en la prueba 1, tampoco se observó un condición y/o combinación que resultase mejor opción y en la cual los todos los amplicones se viesen representados.



**Figura 17. Distribución de lecturas mínimas obtenidas por amplicón secuenciado según las combinaciones de cebadores utilizadas en el panel DQ-LAC de la prueba 2.** La gráfica muestra una comparativa del promedio de las lecturas obtenidas para cada uno de los fragmentos tras la secuenciación en función de las combinaciones de cebadores (1 y 2) a concentraciones 1X, 1/10X y 1/20X.



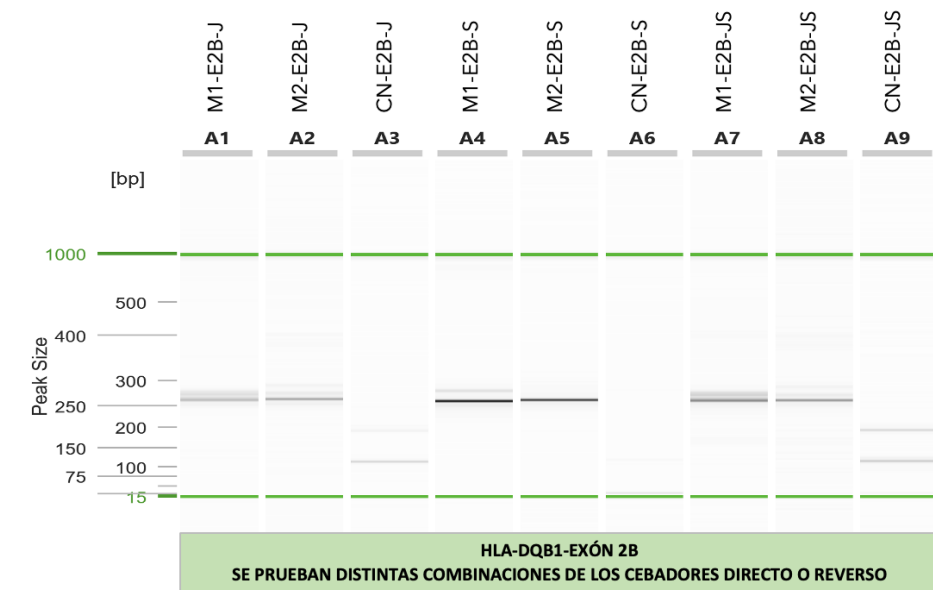
**Figura 18. Distribución de lecturas mínimas obtenidas por amplicón secuenciado según las combinaciones de cebadores utilizadas en el panel DQ de la prueba 2.** La gráfica muestra una comparativa del promedio de las lecturas obtenidas para cada uno de los fragmentos tras la secuenciación en función de las combinaciones de cebadores (1 y 2) a concentraciones 1X, 1/10X y 1/20X.

#### 4.2.3. Prueba 3

La prueba 3 se realizó con el objetivo de comprobar si el fragmento correspondiente al exón 2B del gen *HLA-DQB1* (*HLA-DQB1-E2B*) amplificaba por sí solo con los cebadores diseñados para este, ya que este no se vio representado en los resultados de coberturas de los diferentes paneles (*DQ-LAC* y *DQ*) en las pruebas previas.

Para lograr una mejor amplificación, dados los posibles SNPs de la secuencia de interés, fueron diseñados tres cebadores: un cebador directo (*DQB1-E2B-5*) y dos cebadores reversos (*DQB1-E2B-3J* y *DQB1-E2B-3S*). Para comprobar si todos amplificaban bien el fragmento, se probaron las tres posibles combinaciones: cebador directo con cada uno de los reversos por separado y directo con ambos cebadores reversos en la misma reacción de PCR.

En este caso, solo se realizó la PCR1, en las mismas condiciones que en las pruebas anteriores, ya que con los ciclos de esta era suficiente para corroborar si se daba la amplificación. Mediante electroforesis capilar, se pudo comprobar que la amplificación del fragmento sí se dio con todas las combinaciones de cebadores probadas, concluyendo que este sí amplificaba por sí solo, pudiendo darse algún factor que afectara en su amplificación en PCR multiplex (*Figura 19*).



**Figura 19. Resultados de la electroforesis capilar para la prueba 3.** Se comprueba si la amplificación del fragmento *DQB1-E2B* se da correctamente, observando banda entorna a 250 pb en todas las combinaciones de cebadores probadas: cebador directo *DQB1-E2B-5* con reverso *DQB1-E2B-3J*, *DQB1-E2B-S* y ambos conjuntamente para muestra 1 (M1), muestra 2 (M2) y control negativo (CN), respectivamente.

#### 4.2.4. Prueba 4

La prueba 4 se realizó con la finalidad de obtener lecturas para el fragmento *DQB1-E2B*, ya que, en base a los resultados anteriores, podría existir algún tipo de interacción entre los cebadores diseñados para amplificar dicho fragmento y alguna de las otras parejas de cebadores, impidiendo así la correcta amplificación de este. En dicha prueba se probaron 7 combinaciones de cebadores en tres *pools*, de modo que se probaran todas las posibles combinaciones de los cebadores correspondientes a *DQB1-E2B* con el resto de los cebadores, comprobando si alguna de las combinaciones permitía una adecuada cobertura de este (Tabla 13). En este caso, todos los cebadores se adicionaron a una misma concentración.

La amplificación de los fragmentos se corroboró por electroforesis capilar, observando bandas del tamaño esperado, cercanas a 300 pb (Figura 20).

**Tabla 13. Condiciones probadas en la prueba 4 según las posibles combinaciones de cebadores con la pareja de cebadores correspondiente al gen *HLA-DQB1-E2B*.**

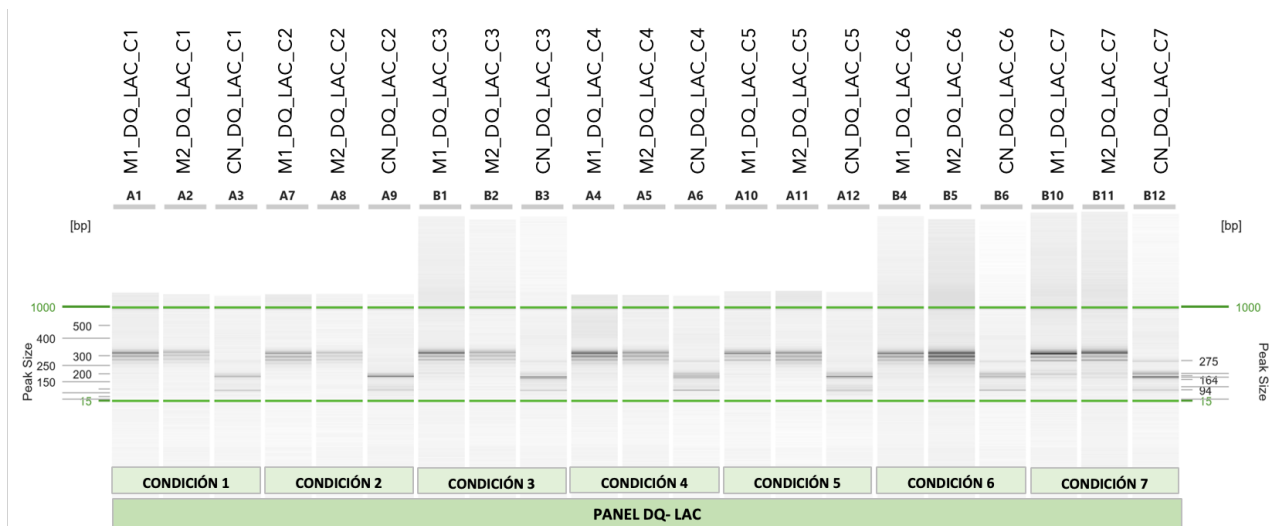
CONDICIÓN 1			CONDICIÓN 2		
POOL 1	POOL 2	POOL 3	POOL 1	POOL 2	POOL 3
<i>DQA1-E2A-5</i>	<i>DQA1-E2B-5</i>	<i>DQA1-E2C-5S</i>	<i>DQA1-E2A-5</i>	<i>DQA1-E2B-5</i>	<i>DQA1-E2C-5S</i>
<i>DQA1-E2A-3</i>	<i>DQA1-E2B-3</i>	<i>DQA1-E2C-5C</i>	<i>DQA1-E2A-3</i>	<i>DQA1-E2B-3</i>	<i>DQA1-E2C-5C</i>
<i>DQA1-E3b-5</i>	<i>DQA1-E3A-5</i>	<i>DQA1-E2C-3</i>	<i>DQA1-E3b-5</i>	<i>DQA1-E3A-5</i>	<i>DQA1-E2C-3</i>
<i>DQA1-E3b-3</i>	<i>DQA1-E3A-3</i>	<i>DQA1-E3B-5</i>	<i>DQA1-E3b-3</i>	<i>DQA1-E3A-3</i>	<i>DQA1-E3B-5</i>
<i>DQB1-E2B-5</i>	<i>DQB1-E2A-5n</i>	<i>DQA1-E3B-3</i>	<i>DQB1-E2C-5</i>	<i>DQB1-E2A-5n</i>	<i>DQA1-E3B-3</i>
<i>DQB1-E2B-3J</i>	<i>DQB1-E2Ab-3</i>	<i>DQB1-E2C-5</i>	<i>DQB1-E2C-3</i>	<i>DQB1-E2Ab-3</i>	<i>DQB1-E2B-5</i>
<i>DQB1-E2B-3S</i>	<i>LAC-13900-3</i>	<i>DQB1-E2C-3</i>	<i>DQB1-E3-3</i>	<i>LAC-13900-3</i>	<i>DQB1-E2B-3J</i>
<i>DQB1-E3-5</i>	<i>LAC-13900-5</i>	<i>LAC-22000-5</i>	<i>DQB1-E3-5</i>	<i>LAC-13900-5</i>	<i>DQB1-E2B-3S</i>
<i>DQB1-E3-3</i>		<i>LAC-22000-3</i>			<i>LAC-22000-5</i>
					<i>LAC-22000-3</i>

CONDICIÓN 3			CONDICIÓN 4		
POOL 1	POOL 2	POOL 3	POOL 1	POOL 2	POOL 3
DQA1-E2A-5	DQA1-E2B-5	DQA1-E2C-5S	DQA1-E2A-5	DQA1-E2B-5	DQB1-E2B-5
DQA1-E2A-3	DQA1-E2B-3	DQA1-E2C-5C	DQA1-E2A-3	DQA1-E2B-3	DQB1-E2B-3J
DQA1-E3b-5	DQA1-E3A-5	DQA1-E2C-3	DQA1-E2C-5S	DQA1-E3B-5	DQB1-E2B-3S
DQA1-E3b-3	DQA1-E3A-3	DQA1-E3B-5	DQA1-E2C-5C	DQA1-E3B-3	DQB1-E2C-5
DQB1-E2A-5n	DQB1-E2B-5	DQA1-E3B-3	DQA1-E2C-3	DQA1-E3b-5	DQB1-E2C-3
DQB1-E2Ab-3	DQB1-E2B-3J	DQB1-E3-5	DQA1-E3A-5	DQA1-E3b-3	DQB1-E3-3
LAC-13900-3	DQB1-E2B-3S	DQB1-E3-3	DQA1-E3A-3	DQB1-E2A-5n	DQB1-E3-5
LAC-13900-5	DQB1-E2C-5	LAC-22000-3	LAC-13900-3	DQB1-E2Ab-3	
	DQB1-E2C-3	LAC-22000-5	LAC-13900-5	LAC-22000-5	
				LAC-22000-3	

CONDICIÓN 5			CONDICIÓN 6		
POOL 1	POOL 2	POOL 3	POOL 1	POOL 2	POOL 3
DQA1-E2A-5	DQA1-E2B-5	DQA1-E3A-5	DQA1-E3b-5	DQB1-E2C-5	DQA1-E2B-5
DQA1-E2A-3	DQA1-E2B-3	DQA1-E3A-3	DQA1-E3b-3	DQB1-E2C-3	DQA1-E2B-3
DQA1-E2C-5S	DQA1-E3b-5	DQB1-E2C-5	DQA1-E3B-5	DQB1-E3-3	DQB1-E2A-5n
DQA1-E2C-5C	DQA1-E3b-3	DQB1-E2C-3	DQA1-E3B-3	DQB1-E3-5	DQB1-E2Ab-3
DQA1-E2C-3	DQA1-E3B-5	DQB1-E3-3	DQA1-E2A-5	LAC-22000-5	DQA1-E3A-5
DQB1-E2B-5	DQA1-E3B-3	DQB1-E3-5	DQA1-E2A-3	LAC-22000-3	DQA1-E3A-3
DQB1-E2B-3J	DQB1-E2A-5n	LAC-22000-5	DQA1-E2C-5S	DQB1-E2B-5	
DQB1-E2B-3S	DQB1-E2Ab-3	LAC-22000-3	DQA1-E2C-5C	DQB1-E2B-3J	
LAC-13900-3			DQA1-E2C-3	DQB1-E2B-3S	
LAC-13900-5			LAC-13900-3		
			LAC-13900-5		

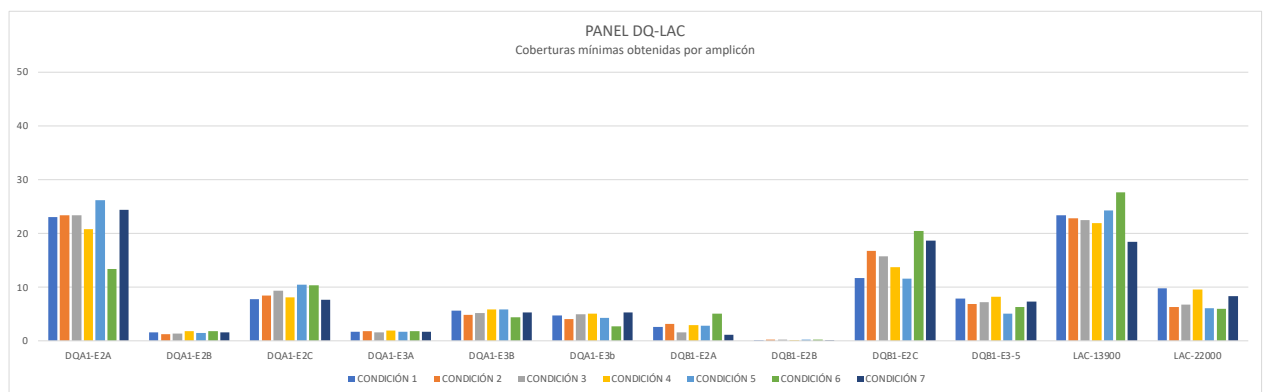
CONDICIÓN 7		
POOL 1	POOL 2	POOL 3
DQA1-E2A-5	DQA1-E3b-5	DQA1-E2B-5
DQA1-E2A-3	DQA1-E3b-3	DQA1-E2B-3
DQA1-E2C-5S	DQA1-E3B-5	DQA1-E3A-5
DQA1-E2C-5C	DQA1-E3B-3	DQA1-E3A-3
DQA1-E2C-3	DQB1-E2C-5	DQB1-E2B-5
DQB1-E2A-5n	DQB1-E2C-3	DQB1-E2B-3J
DQB1-E2Ab-3		DQB1-E2B-3S
DQB1-E3-3		LAC-22000-5
DQB1-E3-5		LAC-22000-3
LAC-13900-3		
LAC-13900-5		





**Figura 20. Resultados de la electroforesis capilar para la prueba 4.** Se muestran las bandas resultantes de los fragmentos obtenidos en las siete condiciones probadas para la muestra 1 (M1), muestra 2 (M2) y control negativo (CN) tras las dos reacciones de amplificación para el panel DQ-LAC y DQ.

Los resultados de coberturas obtenidos para dicho panel mostraron que prácticamente todos los amplicones fueron cubiertos en todas las condiciones, unos mucho más que otros, destacando las condiciones 4 y 5, aunque sin una gran diferencia notable entre ellas. No obstante, el amplicón correspondiente al exón 2 del gen *HLA-DQB1-E2B* no obtuvo prácticamente lecturas en ninguna de las condiciones, requiriendo de nuevo continuar con la optimización del panel (Figura 21).



**Figura 21. Distribución de lecturas mínimas obtenidas por amplicón secuenciado según las combinaciones de cebadores utilizadas en el panel DQ-LAC para la prueba 4.** Para cada una de las condiciones se representa el promedio de lecturas obtenidas por fragmento amplificado tras la secuenciación, utilizando en todos los casos cebadores a la misma concentración.

#### 4.2.5. Prueba 5

En base a los resultados de cobertura obtenidos para el gen *HLA DQB1-E2B*, se realizó un nuevo diseño de los cebadores necesarios para la amplificación de este, con la finalidad de comprobar si alguno de ellos permitía obtener mejores resultados (*Tabla 14*).

**Tabla 14. Nuevas parejas de cebadores diseñadas para la amplificación del gen *HLA-DQB1-E2B*.**

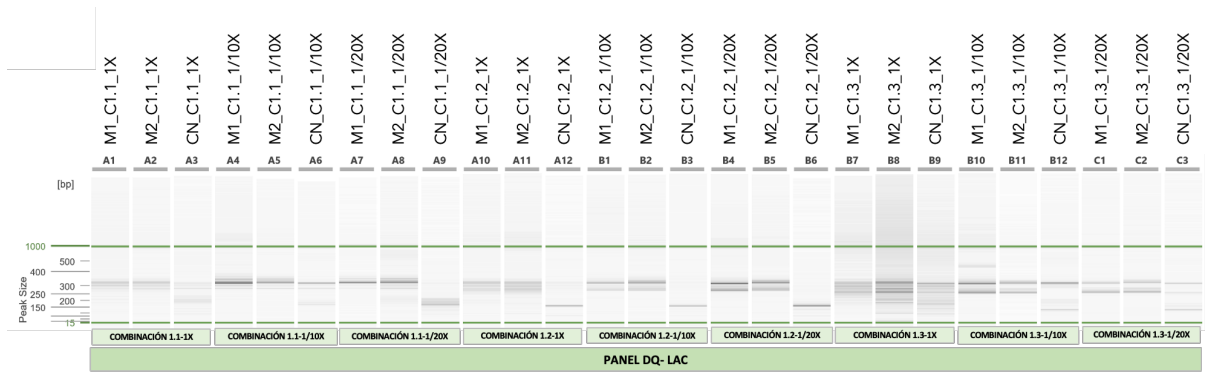
CEBADORES DQB1-E2B	
DIRECTO	REVERSO
<i>DQB1-E2Bnd1-5</i>	<i>DQB1-E2Bnd1-3</i>
<i>DQB1-E2Bnd2-5</i>	<i>DQB1-E2Bnd2a-3</i>
	<i>DQB1-E2Bnd2b-3</i>
<i>DQB1-E2Bnd3a-5</i>	<i>DQB1-E2Bnd2a-3</i>
<i>DQB1-E2Bnd3b-5</i>	<i>DQB1-E2Bnd2b-3</i>

A partir de los nuevos cebadores diseñados, se probaron nuevas condiciones, utilizando las condiciones 1 y 2 de la prueba 2 y la condición 5 de la prueba 4, pero variando para cada una de ellas las nuevas posibilidades de cebadores para el gen *HLA-DQB1-E2B*, reduciendo a 30 los ciclos del programa de PCR 1, al haberse comprobado que eran suficientes para la correcta amplificación. De este modo, para cada una de las condiciones se probaron tres nuevas combinaciones a diferentes concentraciones de cebadores: 1X, 1/10X y 1/20X, manteniendo en cada una de ellas los cebadores a la misma concentración (*Tabla 15*).

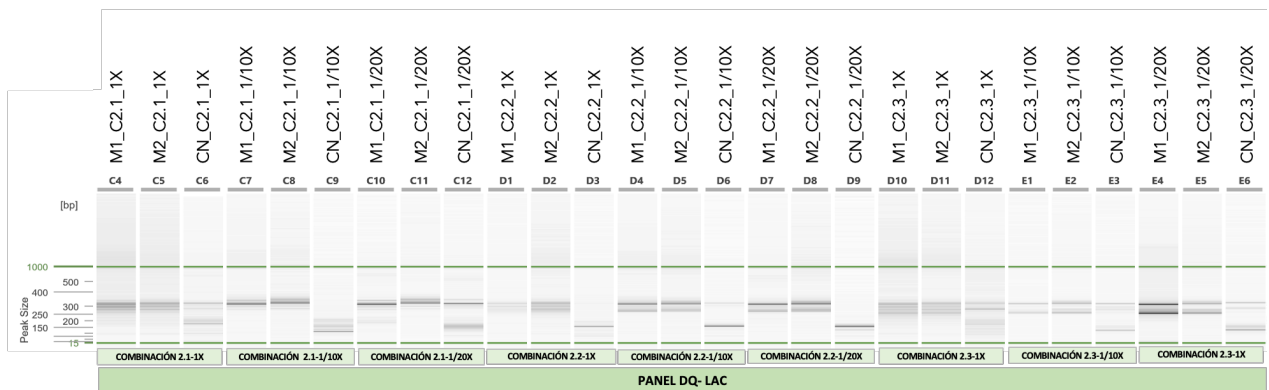
**Tabla 15. Condiciones analizadas en la prueba 5 en función de las combinaciones de cebadores y la concentración de estos.**

		POOLS	CEBADORES DQB1-E2B	CONCENTRACIÓN CEBADORES
CONDICIÓN 1	COMBINACIÓN 1.1	2	<i>DQB1-E2Bnd1-5</i> <i>DQB1-E2Bnd1-3</i>	1X
	COMBINACIÓN 2.1			1/10X
	COMBINACIÓN 5.1	3		1/20X
CONDICIÓN 2	COMBINACIÓN 1.2	2	<i>DQB1-E2Bnd2-5</i> <i>DQB1-E2Bnd2a-3</i> <i>DQB1-E2Bnd2b-3</i>	1X
	COMBINACIÓN 2.2			1/10X
	COMBINACIÓN 5.2	3		1/20X
CONDICIÓN 3	COMBINACIÓN 1.3	2	<i>DQB1-E2Bnd2a-3</i> <i>DQB1-E2Bnd2b-3</i> <i>DQB1-E2Bnd3a-5</i> <i>DQB1-E2Bnd3b-5</i>	1X
	COMBINACIÓN 2.3			1/10X
	COMBINACIÓN 5.3	3		1/20X

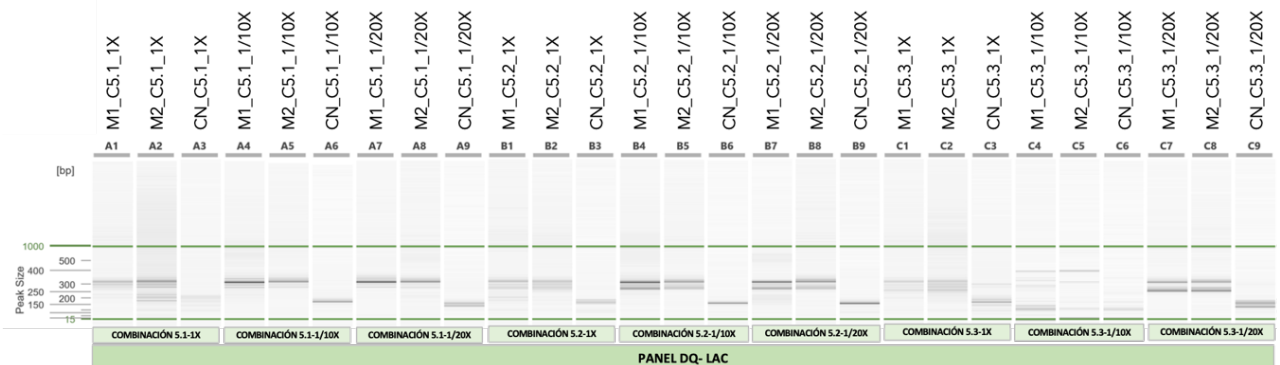
Tras las dos reacciones de amplificación, el tamaño de los fragmentos para cada una de las combinaciones fue evaluado mediante electroforesis capilar (*Figura 22, 23, 24*). No obstante, los resultados de electroforesis capilar obtenidos para la condición 1/10X de la combinación 5.3 no fueron los esperados, por lo que esta condición no pudo ser secuenciada (*Figura 24*).



**Figura 22. Resultados de la electroforesis capilar para las combinaciones 1.1, 1.2 y 1.3 de la prueba 5. Se muestran las bandas resultantes de la amplificación de la muestra 1 (M1), muestra 2 (M2) y control negativo (CN) con una concentración de cebadores a 1X, 1/10X y 1/20 respectivamente.**

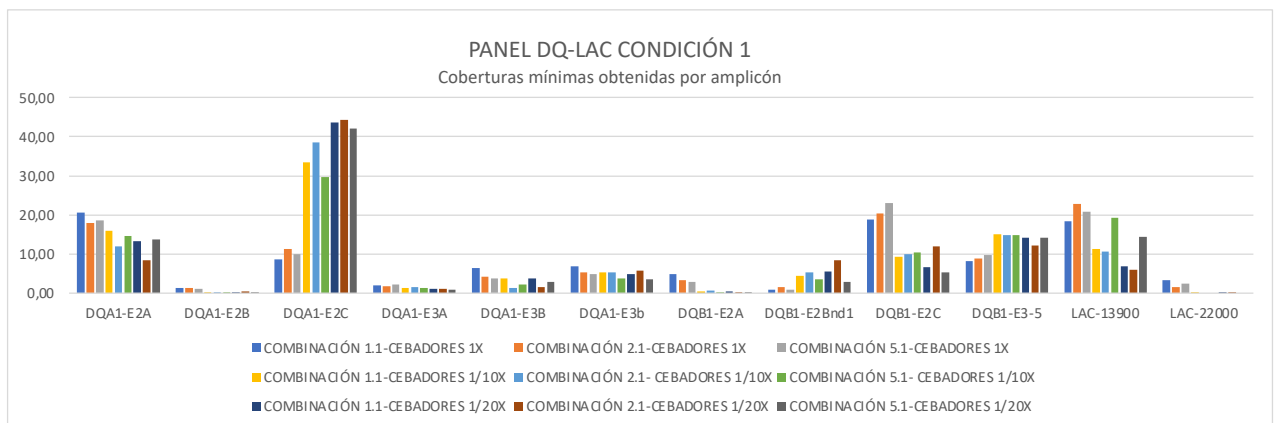


**Figura 23. Resultados de la electroforesis capilar para las combinaciones 2.1, 2.2 y 2.3 de la prueba 5. Se muestran las bandas resultantes de la amplificación de la muestra 1 (M1), muestra 2 (M2) y control negativo (CN) con una concentración de cebadores a 1X, 1/10X y 1/20 respectivamente.**

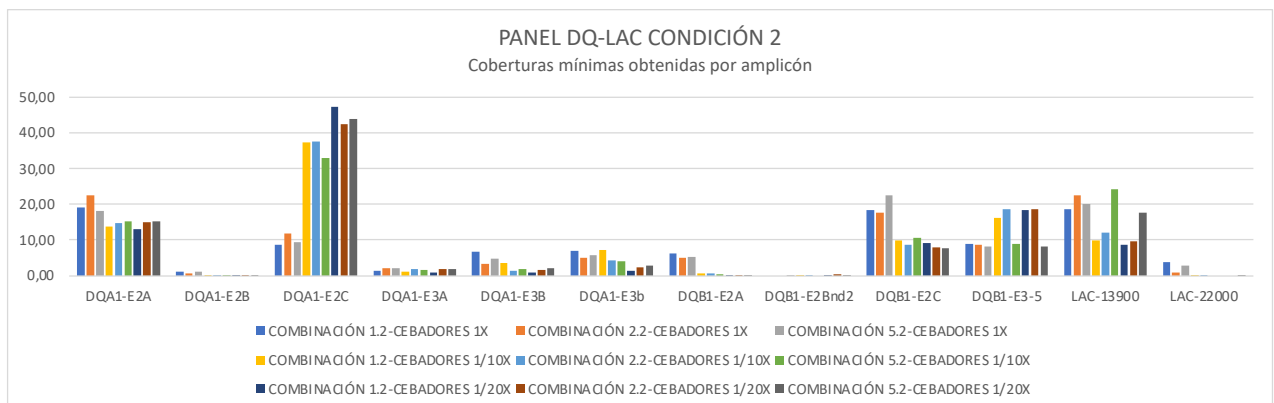


**Figura 24. Resultados de la electroforesis capilar para las combinaciones 5.1, 5.2 y 5.3 de la prueba 5. Se muestran las bandas resultantes de la amplificación de la muestra 1 (M1), muestra 2 (M2) y control negativo (CN) con una concentración de cebadores a 1X, 1/10X y 1/20 respectivamente.**

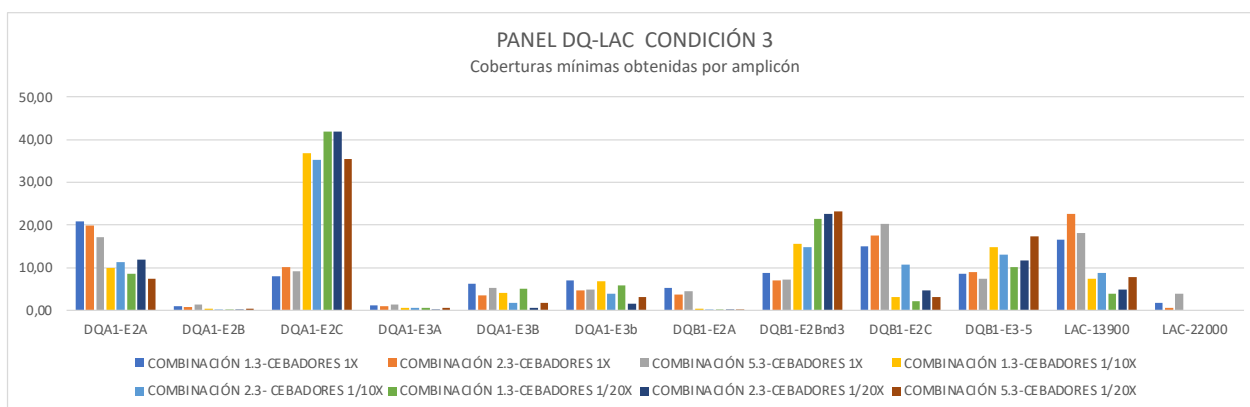
En cuanto a los resultados de coberturas, no se observaron diferencias significativas entre los resultados de cobertura de las combinaciones de cebadores en 2 *pools* y aquellos divididos en 3 *pools*, por lo que se escogió continuar realizando todas las pruebas con 2 *pools*, al implicar un menor tiempo y coste, al requerir una menor cantidad de reactivos. Por otra parte, tampoco se obtuvieron lecturas para el amplicón *DQB1-E2B* con los nuevos cebadores diseñados en ninguna de las combinaciones de la condición 2. En cambio, sí se obtuvo cobertura para este en las condiciones 1 y 3, con un mayor número de lecturas en la condición 3, siendo la combinación 1.3 con los cebadores a 1X la mejor de todas ellas al obtenerse cobertura, aunque no homogénea, para todos los fragmentos (*Figura 25, 26,27*).



**Figura 25. Distribución de lecturas mínimas obtenidas por amplicón secuenciado para las combinaciones de cebadores 1.1, 2.1 y 5.1 utilizadas en la condición 1 del panel DQ-LAC. Cada una de las combinaciones fue probada con diferentes concentraciones de cebadores: 1X, 1/10X y 1/20X.**



**Figura 26. Distribución de lecturas mínimas obtenidas por amplicón secuenciado para las combinaciones de cebadores 1.2, 2.2 y 5.2 utilizadas en la condición 2 del panel DQ-LAC. Cada una de las combinaciones fue probada con diferentes concentraciones de cebadores: 1X, 1/10X y 1/20X.**



**Figura 27. Distribución de lecturas mínimas obtenidas por amplicón secuenciado para las combinaciones de cebadores 1.3, 2.3 y 5.3 utilizadas en la condición 3 del panel DQ-LAC. Cada una de las combinaciones fue probada con diferentes concentraciones de cebadores: 1X, 1/10X y 1/20X, a excepción de la condición 5.3, probada a 1X y 1/20X.**

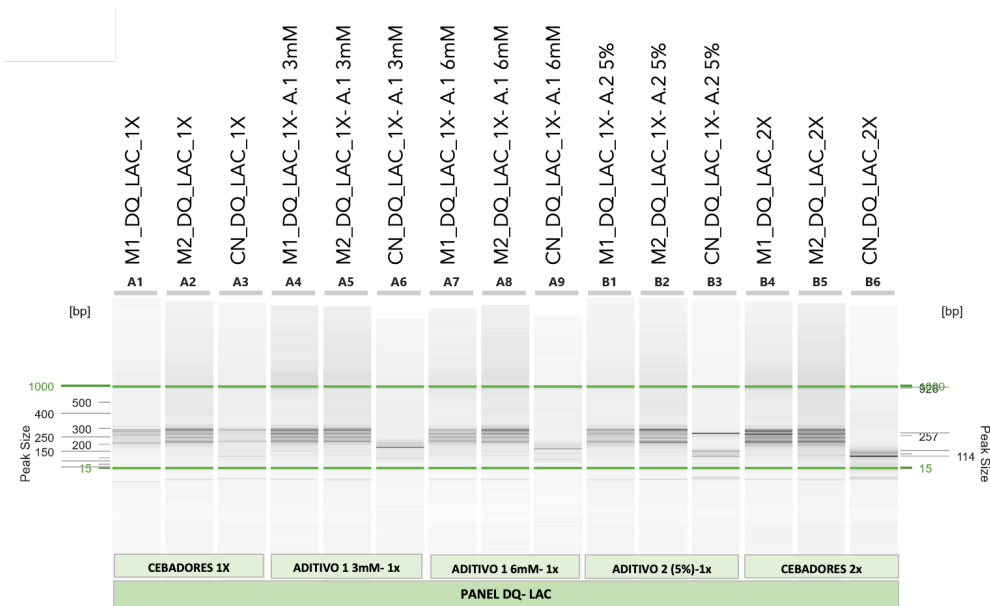
#### 4.2.6. Prueba 6

En base a los resultados obtenidos en la prueba anterior, en la prueba 6 se seleccionó la combinación de cebadores 1.3 de la condición 3 a una concentración de 1X. Esta se realizó con el objetivo de lograr una mejor optimización del procedimiento, que permitiera obtener una cobertura más homogénea para todos los amplicones.

Para ello, se volvieron a probar varias condiciones, añadiendo ciertos aditivos y duplicando la concentración de los cebadores correspondientes a los fragmentos que menos lecturas habían obtenido en la prueba anterior (Tabla 16). Al igual que en las pruebas anteriores, tras las dos reacciones de amplificación, el tamaño de los fragmentos fue evaluado mediante electroforesis capilar, concluyendo que estas se habían dado satisfactoriamente al observar fragmentos del tamaño esperado, cercanos a 300 pb (Figura 28).

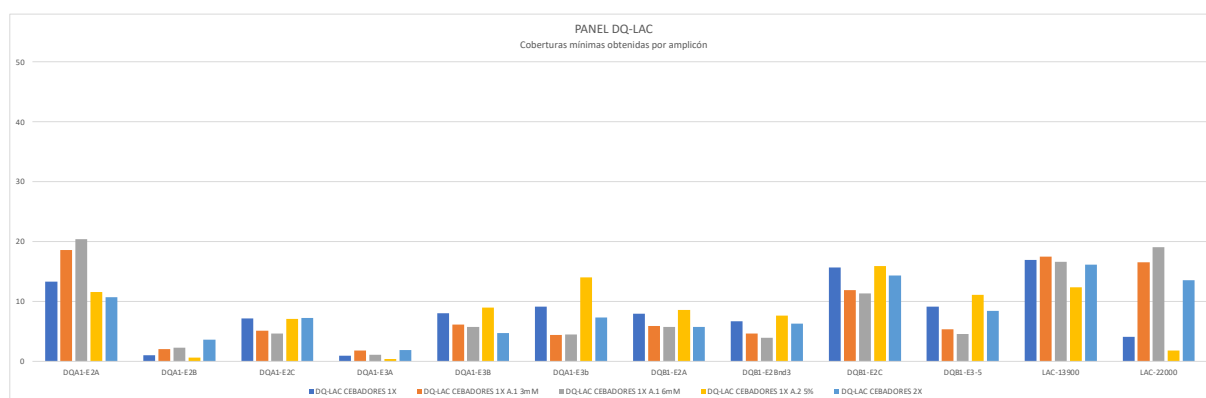
**Tabla 16. Condiciones analizadas en la prueba 6 partiendo de la combinación 1.3 de la prueba anterior.**

COMBINACIÓN 1.3			
	ADITIVO	CONCENTRACIÓN DE CEBADORES	CEBADORES DQB1-E2B
CONDICIÓN 1	-	1X	DQB1-E2Bnd2a-3 DQB1-E2Bnd2b-3 DQB1-E2Bnd3a-5 DQB1-E2Bnd3b-5
CONDICIÓN 2	Aditivo 1 (3mM)		
CONDICIÓN 3	Aditivo 1 (6mM)		
CONDICIÓN 4	Aditivo 2 (5%)		
CONDICIÓN 5	-	2X DQA1-E3A DQA1-E2B LAC-22000	



**Figura 28. Resultados de la electroforesis capilar para la prueba 6.** Se muestran las bandas correspondientes a los fragmentos amplificados para las diferentes condiciones probadas: cebadores a 1X, cebadores a 1X con aditivo 1 a 3mM, cebadores a 1X con aditivo 1 a 6mM, cebadores a 1X con aditivo 2 al 5% y cebadores seleccionados a 2X.

Tras el análisis de las coberturas obtenidas para cada uno de los amplicones, se observó que la condición con los cebadores seleccionados a 2X permitía obtener resultados de cobertura mucho mejores para dichos fragmentos, incluso que con los aditivos seleccionados. En cuanto a la cobertura obtenida para el resto de amplicones, no se percibió una gran diferencia generalizada entre las condiciones con o sin aditivos, por lo que se consideró como la mejor condición aquella con los cebadores seleccionados a 2X, permitiendo obtener una mejor cobertura de sus correspondientes amplicones (Figura 29).



**Figura 29. Distribución de lecturas mínimas obtenidas por amplicón secuenciado para las condiciones del panel DQ-LAC utilizadas en la prueba 6.** Para cada uno de los amplicones se muestran las lecturas mínimas obtenidas según las condiciones: cebadores a 1X, cebadores a 1X con aditivo 1 a 3 mM, cebadores a 1X con aditivo 1 a 6mM, cebadores a 1X con aditivo 2 al 5% y cebadores seleccionados a 2X, manteniendo el resto a 1X.

#### 4.2.7. Prueba 7

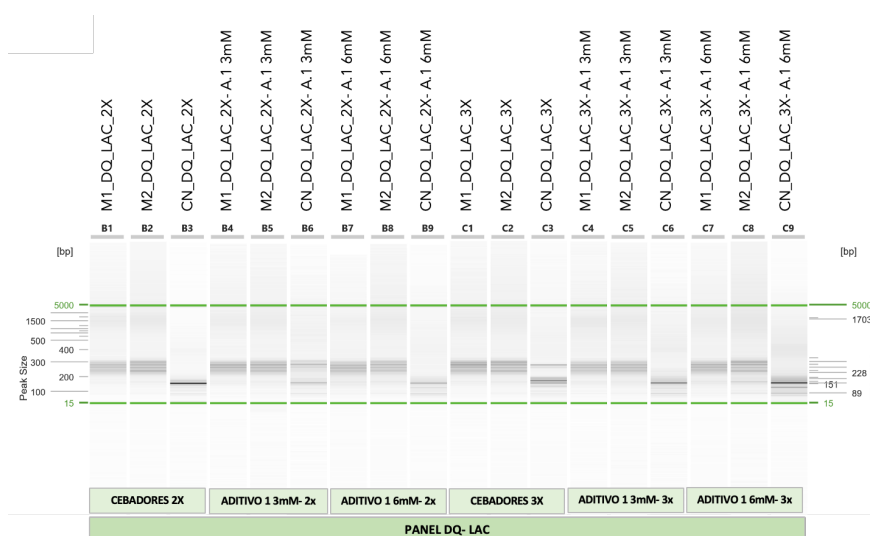
La prueba 7 se realizó partiendo de la condición de cebadores a 2X de la combinación 1.3 de la prueba anterior. En ella, los cebadores correspondientes a aquellos fragmentos que todavía habían obtenido pocas lecturas fueron probados a una concentración de 3X con respecto al resto, que se mantuvieron

a 1X, a excepción de los correspondientes al fragmento *LAC-22000*, los cuales se mantuvieron a 2X al haber obtenido suficientes lecturas (*Tabla 17*).

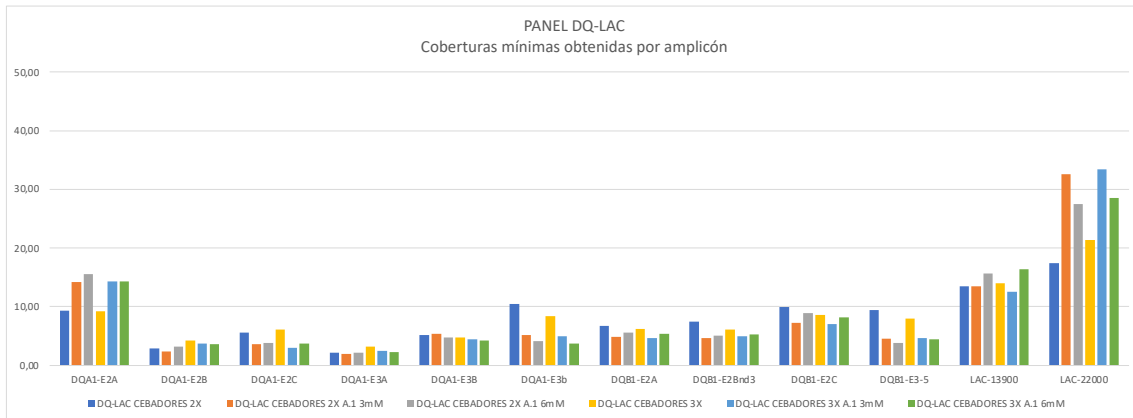
**Tabla 17. Condiciones analizadas en la prueba 7 utilizando la combinación 1.3 de la prueba 5.**

	COMBINACIÓN 1.3		
	ADITIVO	CONCENTRACIÓN DE CEBADORES	CEBADORES DQB1-E2B
CONDICIÓN 1	-	2X <i>DQA1-E3A</i> <i>DQA1-E2B</i> <i>LAC-22000</i>	<i>DQB1-E2Bnd2a-3</i> <i>DQB1-E2Bnd2b-3</i> <i>DQB1-E2Bnd3a-5</i> <i>DQB1-E2Bnd3b-5</i>
CONDICIÓN 2	Aditivo 1 (3mM)		
CONDICIÓN 3	Aditivo 1 (6mM)		
CONDICIÓN 4	-	2X <i>DQA1-E3A</i> <i>DQA1-E2B</i>	
CONDICIÓN 5	Aditivo 1 (3mM)		
CONDICIÓN 5	Aditivo 1 (6mM)	3X <i>LAC-22000</i>	

Tras los resultados de electroforesis capilar obtenidos (*Figura 30*) se llevó a cabo la secuenciación de las muestras. Los resultados de coberturas obtenidos tras la secuenciación en la prueba 7 (*Figura 31*) mostraron una cobertura prácticamente homogénea para todos los amplicones en todas las condiciones probadas, recibiendo algunos de ellos un mayor número de lecturas. A pesar de que no se observó una gran diferencia en el número de lecturas obtenido con las distintas condiciones, se consideró la condición 4, con los cebadores correspondientes para la amplificación de los exones 3A y 2B del gen *DQA1* (*DQA1-E3A* y *DQA1-E2B*) a 3X y los cebadores para el gen *LAC-22000* a 2X como la mejor de las condiciones, al presentar un mayor número de lecturas para los fragmentos que salían más bajos, así como una cobertura más homogénea para todos los fragmentos que con el resto de las condiciones. De este modo, se consiguió optimizar el proceso para que todos los amplicones se viesen cubiertos de forma homogénea en el panel de genes seleccionado, pudiéndose utilizar dichas condiciones para la posterior identificación de alelos de riesgo en determinadas poblaciones con enfermedad celiaca e intolerancia a la lactosa.

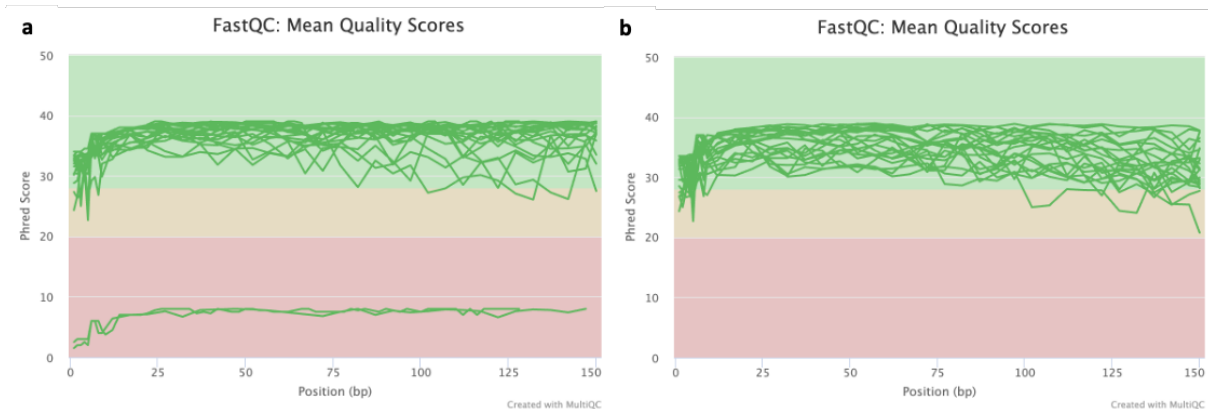


**Figura 30. Resultados de la electroforesis capilar para la prueba 6.** Se muestran las bandas correspondientes a los fragmentos amplificados para las diferentes condiciones probadas: cebadores seleccionados a 2X, con aditivo 1 a 3mM y aditivo 1 a 6mM y cebadores seleccionados a 3X, con aditivo 1 a 3mM y aditivo 1 a 6mM.



**Figura 31. Distribución de lecturas mínimas obtenidas por amplicón secuenciado para las condiciones del panel DQ-LAC utilizadas en la prueba 7.** Para cada uno de los amplicones se muestran las lecturas mínimas obtenidas según las condiciones: cebadores seleccionados 2X, cebadores seleccionados a 2X con aditivo 1 a 3 mM y aditivo 1 a 6mM, cebadores seleccionados a 3X, cebadores seleccionados a 3X con aditivo 1 a 3 mM y aditivo 1 a 6mM. El resto de los cebadores se mantuvieron a su concentración inicial o 1X, a excepción de los cebadores para LAC-22000, que se mantuvieron a 2X en todas las pruebas.

Con el fin de corroborar la calidad de las coberturas obtenidas para las distintas muestras en las condiciones probadas, se llevó a cabo el análisis de los FASTQ obtenidos a partir de la *pipeline* utilizada en el laboratorio, para lo cual se utilizó el programa FastQC. Dicho control de calidad fue realizado antes y después del *trimming*, aplicando varios filtros de calidad que permitiesen eliminar aquellas lecturas cuyo valor en la escala Phread fuese  $Q < 20$  (Figura 32). De este modo, se pudo comprobar que el proceso de *trimming* se había realizado con éxito, eliminando aquellas lecturas de mala calidad, que no cumplieran con los filtros aplicados.



**Figura 32. Calidad de secuenciación por base y muestra secuenciada obtenida mediante el análisis en el programa FastQC.** A) Calidad de las lecturas antes de su procesamiento. B) Calidad tras el proceso de *trimming* o recorte de extremos, eliminando los adaptadores y las secuencias de baja calidad.



## 5. CONCLUSIONES

Basándose en el trabajo experimental realizado, se puede llegar a las siguientes conclusiones:

1. La secuenciación masiva de amplicones cortos ha demostrado ser una herramienta altamente eficiente para la identificación de los alelos de riesgo en enfermedad celíaca e intolerancia a la lactosa.
2. La amplificación de un panel constituido por los genes *HLA-DQA1*, *HLA-DQB1* y *LCT* puede realizarse de forma simultánea satisfactoriamente, ya que los cebadores diseñados para la amplificación de las variantes del gen de la persistencia a la lactasa (*LAC-13900* y *LAC-22000*) no interfieren con el resto de los cebadores, pudiendo ser introducidos en la reacción de PCR multiplex.
3. Una adecuada cobertura del exón *2B* del gen *HLA-DQB1* (*HLA-DQB1-E2B*) es posible a partir de dos de las parejas de cebadores obtenidas tras el segundo diseño realizado: *DQB1-E2Bnd1-5* y *DQB1-E2Bnd1-3*, siendo mejores los resultados obtenidos con las parejas de cebadores *DQB1-E2Bnd2a-3*, *DQB1-E2Bnd2b-3*, *DQB1-E2Bnd3a-5* y *DQB1-E2Bnd3b-5*.
4. Los resultados de coberturas obtenidas tras las amplificaciones con las combinaciones de cebadores en 2 *pools* y aquellos divididos en 3 *pools* no muestran diferencias significativas, continuando el proceso de optimización con 2 *pools* al suponer un ahorro de tiempo y dinero.
5. La mejor condición obtenida y con la cual concluye el proceso de optimización es la condición 4 de la prueba 7, en la cual se utiliza la combinación 1.3 de cebadores a 1X procedente de la prueba 5, pero variando la concentración de los cebadores: duplicando la concentración de los cebadores correspondientes a los genes *HLA-DQA1-E3A* y *HLA-DQA1-E2B* y triplicando la de los cebadores correspondientes a *LAC-22000*.

## 6. REFERENCIAS BIBLIOGRÁFICAS

- ANGUITA-RUIZ, A., AGUILERA, C. M., & GIL, Á. (2020). Genetics of lactose intolerance: An updated review and online interactive world maps of phenotype and genotype frequencies. *Nutrients*, *12*(9), 1–20. <https://doi.org/10.3390/nu12092689>
- BRAVO-EGANA, V., SANDERS, H., & CHITNIS, N. (2021). New challenges, new opportunities: Next generation sequencing and its place in the advancement of HLA typing. *Human Immunology*, *82*(7), 478–487. <https://doi.org/10.1016/j.humimm.2021.01.010>
- CAIO, G., VOLTA, U., SAPONE, A., LEFFLER, D. A., DE GIORGIO, R., CATASSI, C., & FASANO, A. (2019). Celiac disease: A comprehensive current review. *BMC Medicine*, *17*(1). <https://doi.org/10.1186/s12916-019-1380-z>
- CATANZARO, R., SCIUTO, M., & MAROTTA, F. (2021). Lactose intolerance: An update on its pathogenesis, diagnosis, and treatment. *Nutrition Research*, *89*, 23–34. <https://doi.org/10.1016/j.nutres.2021.02.003>
- CATASSI, C., VERDU, E. F., BAI, J. C., & LIONETTI, E. (2022). Coeliac disease. *The Lancet*, *399*(10344), 2413–2426. [https://doi.org/10.1016/S0140-6736\(22\)00794-2](https://doi.org/10.1016/S0140-6736(22)00794-2)
- CHOUNG, R. S., MILLS, J. R., SNYDER, M. R., MURRAY, J. A., & GANDHI, M. J. (2020). Celiac disease risk stratification based on HLA-DQ heterodimer (HLA-DQA1 ~ DQB1) typing in a large cohort of adults with suspected celiac disease. *Human Immunology*, *81*(2–3), 59–64. <https://doi.org/10.1016/j.humimm.2020.01.006>
- DI COSTANZO, M., & BERNI CANANI, R. (2019). Lactose Intolerance: Common Misunderstandings. *Annals of Nutrition and Metabolism*, *73*(Suppl 4), 30–37. <https://doi.org/10.1159/000493669>
- HON, E., & GUPTA, S. K. (2021). Gastrointestinal Food Allergies and Intolerances. *Gastroenterology Clinics of North America*, *50*(1), 41–57. <https://doi.org/10.1016/j.gtc.2020.10.006>
- LEBWOHL, B., & RUBIO-TAPIA, A. (2021). Epidemiology, Presentation, and Diagnosis of Celiac Disease. *Gastroenterology*, *160*(1), 63–75. <https://doi.org/10.1053/j.gastro.2020.06.098>
- MARDIS, E. R. (2013). Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, *6*, 287–303.
- METZKER, M. L. (2010). Sequencing technologies the next generation. *Nature Reviews Genetics*, *11*(1), 31–46. <https://doi.org/10.1038/nrg2626>
- MUTHUKUMAR, J., SELVASEKARAN, P., LOKANADHAM, M., & CHIDAMBARAM, R. (2020). Food and food products associated with food allergy and food intolerance – An overview. *Food Research International*, *138*. <https://doi.org/10.1016/j.foodres.2020.109780>
- PEREIRA, R., OLIVEIRA, J., & SOUSA, M. (2020). Bioinformatics and computational tools for next-generation sequencing analysis in clinical genetics. In *Journal of Clinical Medicine* (Vol. 9, Issue 1). MDPI. <https://doi.org/10.3390/jcm9010132>
- SCHULER, B. A., NELSON, E. T., KOZIURA, M., COGAN, J. D., HAMID, R., & PHILLIPS, J. A. (2022). Lessons learned: next-generation sequencing applied to undiagnosed genetic diseases. *Journal of Clinical Investigation*, *132*(7). <https://doi.org/10.1172/JCI154942>

- SERENA, G., LIMA, R., & FASANO, A. (2019). Genetic and Environmental Contributors for Celiac Disease. *Current Allergy and Asthma Reports*, 19(9). <https://doi.org/10.1007/s11882-019-0871-5>
- SHEN, C.-H. (2019). Amplification of Nucleic Acids. In *Diagnostic Molecular Biology* (pp. 215–247). Elsevier. <https://doi.org/10.1016/b978-0-12-802823-0.00009-2>
- SNYDER, M. W., ADEY, A., KITZMAN, J. O., & SHENDURE, J. (2015). Haplotype-resolved genome sequencing: Experimental methods and applications. *Nature Reviews Genetics*, 16(6), 344–358. <https://doi.org/10.1038/nrg3903>
- SZILAGYI, A., & ISHAYEK, N. (2018). Lactose intolerance, dairy avoidance, and treatment options. *Nutrients*, 10(12). <https://doi.org/10.3390/nu10121994>
- TUCK, C. J., BIESIEKIERSKI, J. R., SCHMID-GRENDELMEIER, P., & POHL, D. (2019). Food intolerances. *Nutrients*, 11(7). <https://doi.org/10.3390/nu11071684>
- VERDU, E. F., & DANSKA, J. S. (2018). Common ground: Shared risk factors for type 1 diabetes and celiac disease. *Nature Immunology*, 19(7), 685–695. <https://doi.org/10.1038/s41590-018-0130-2>

## 7. ANEXOS

### ANEXO I

#### RELACIÓN DEL TRABAJO CON LOS OBJETIVOS DEL DESARROLLO SOSTENIBLE DE LA AGENDA 2030.

Objetivos de Desarrollo Sostenibles	Alto	Medio	Bajo	No procede
ODS 1. Fin de la pobreza.				X
ODS 2. Hambre cero.			X	
ODS 3. Salud y bienestar.	X			
ODS 4. Educación de calidad.				X
ODS 5. Igualdad de género.				X
ODS 6. Agua limpia y saneamiento.				X
ODS 7. Energía asequible y no contaminante.				X
ODS 8. Trabajo decente y crecimiento económico.				X
ODS 9. Industria, innovación e infraestructuras.				X
ODS 10. Reducción de las desigualdades.			X	
ODS 11. Ciudades y comunidades sostenibles.				X
ODS 12. Producción y consumo responsable.				X
ODS 13. Acción por el clima.				X
ODS 14. Vida submarina.				X
ODS 15. Vida de ecosistemas terrestres.				X
ODS 16. Paz, justicia e instituciones sólidas.				X
ODS 17. Alianzas para lograr objetivos.				X

El presente Trabajo de Fin de Grado, se relaciona principalmente con los siguientes Objetivos de Desarrollo Sostenible dados los motivos expuestos a continuación:

- 1. Salud y bienestar (ODS 3):** La optimización de un proceso de identificación de las variantes genéticas asociadas a la EC y la intolerancia a la lactosa contribuye a mejorar la salud y el bienestar de las personas que presentan o son susceptibles de presentar dichas patologías, al permitir un diagnóstico temprano de dichas patologías, de modo que se puedan tomar medidas preventivas o tratamientos para las mismas.
- 2. Hambre cero (ODS 2):** Llegar a una mayor comprensión de la predisposición genética a la celiaquía o a la intolerancia a la lactosa podría ayudar a desarrollar estrategias que permitan hacer frente de forma más efectiva a problemas de nutrición derivados de dichas patologías, garantizando una alimentación segura y saludable para todas las personas que las padecen.
- 3. Reducción de las desigualdades (ODS 10):** La optimización de dicho proceso de identificación de alelos de riesgo puede favorecer su incorporación en la clínica a un menor coste, pudiendo contribuir a reducir las desigualdades en el acceso al diagnóstico y tratamiento de dichas patologías.