Escola Tècnica Superior
d'Enginyeria Agronòmica i del Medi Natural

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

ESCUELA TÉCNICA SUPERIOR DE INGENIERIA AGRONÓMICA Y DEL MEDIO NATURAL

Organización genómica y análisis bioinformático de genes de toxinas de veneno de serpiente mediante diversas plataformas de secuenciación.

TRABAJO DE FIN DE GRADO EN BIOTECNOLOGÍA

CURSO 2016/2017

AUTOR: **Enric Vercher Herráez**

TUTORA: **María Pilar López Gresa**

DIRECTOR: **Juan José Calvete Chornet**

VALENCIA, JULIO 2017

Resumen

Valencia, Julio 2017

El envenenamiento por mordedura de serpiente es un problema de salud pública que afecta a muchas de las comunidades rurales más pobres del mundo, principalmente las que practican la agricultura de subsistencia y las actividades de pastoreo en regiones tropicales y subtropicales de África, Asia, América Latina y Oceanía (Harrison *et al.*, 2009). El envenenamiento ofídico es considerado por la Organización Mundial de la Salud (OMS) una enfermedad desatenidida. Por ello, su estudio es importante para desarrollar antivenenos que neutralicen los efectos tóxicos de los venenos, así como para aprovechar muchas de sus propiedades para el desarrollo de nuevos fármacos. Los análisis proteómicos han mostrado que los venenos de serpientes están generalmente constituídos por un número reducido de familias de toxinas multigénicas, donde las metaloproteasas representan uno de los componentes mayoritarios de los venenos de las familias *Crotalidae* y *Viperidae* (Markland y Swenson, 2013). La falta de datos genómicos impide llevar a cabo un estudio de genómica comparada entre las distintas especies de serpientes, por lo que no se conoce muy bien la estructura y distribución de los genes de las distintas subfamilias de metaloproteasas y disintegrinas. En este trabajo hemos amplificado, a partir del DNA genómico, genes de disintegrinas y metaloproteasas del ofidio *Bitis arietans* y hemos obtenido sus secuencias mediante secuenciación Sanger iterativa, lo que ha resultado en la caracterización de una subunidad de una nueva disintegrina dimérica. Así mismo, mediante la plataforma de secuenciación de tercera generación MiniON, secuenciamos ocho BACs (*bacterial artificial chromosome*) que incluían fragmentos genómicos del clúster de genes de metaloproteasas de la serpiente *Bothrops jararaca*. Mediante análisis bioinformático se anotó un *contig* codificante de 4 genes de metaloproteasas tipo PIII y 2 tipo PII dispuestos en tándem.

Palabras clave: *Bitis arietans*, *Bothrops jararaca*, disintegrina, metaloproteasa, MiniON, secuenciación masiva.

Autor: Enric Vercher Herráez

Tutora: Dra. Maria Pilar López Gresa

Directores: Profesor Dr. Juan José Calvete Chornet

Abstract

Valencia , July 2017

Snake bite envenoming represents a public health issue affecting many of the world's poorest rural communities, especially those involved in subsistence farming and grazing activities in tropical and subtropical regions of Africa, Asia, Latin America and Oceania (Harrison *et al.*, 2009) The World Health Organization (WHO) has adopted snakebite envenoming as a 'category A' neglected tropical disease (NTD)—the WHO's highest possible ranking for an NTD. Research on venoms and their toxins is relevant to develop antivenoms that neutralize their toxic effects, but also to develop new therapeutics. Proteomic analyzes of snake venoms that snake venoms are generally made up of a reduced number of multigenic toxin families. Metalloproteinases represent one of the major components of the venoms of the *Crotalidae* and *Viperidae* families(Markland and Swenson, 2013).The lack of comprehensive genomic data has precluded comparative snake genomic studies and thus the genomic structure and distribution of genes of the different subfamilies of metalloproteases remain poorly investigated. In this work, using genomic DNA from *Bitis arietans*, we have amplified disintegrin and metalloprotease genes and have unveiled their sequences through iterative Sanger sequencing, resulting in the characterization of a novel subunit of a dimeric disintegrin. Also, using the third generation sequencing MiniON platform, we have sequenced 8 BACs (bacterial artificial chromosome) that included a genomic fragment encoding part of the metalloprotease gene cluster of *Bothrops jararaca*. By means of bioinformatic analysis, a contig including 4 PIII-class and 2 PII-class tandemly arranged metalloprotease genes was annotated.

Key words: *Bitis arietans*, *Bothrops jararaca*, metalloproteinase, disintegrin, MiniON mass sequencing.

Author: Enric Vercher Herráez

Tutoress: Dra. Maria Pilar López Gresa

Directors: Professor Dr. Juan José Calvete Chornet

Agradecmientos

Ha llegado el día en el que este ‘’pequeño’’ proyecto ha quedado acabado y me gustaría agradecer a todos y cada uno de los miembros del laboratorio mi más sincera gratitud por estos siete meses tan intensos.

En primer lugar, a Alicia por haberme enseñado, no sólo a manejarme técnicamente en el laboratorio, sino también a entender el por qué de cada proceso y que al fin y al cabo cada detalle es importante para obtener un buen resultado final. Por nuestros momentos de reflexiones, por compartir nuestras visiones de la vida y nuestro apoyo moral mútuo siempre presente fuera cual fuese el momento.

A Libia por responder y aguantar mis miles de preguntas sobre las serpientes y su seductor mundo, del que prácticamente empecé ignorante y del que ahora parece que tengo algo de idea. Por preocuparte siempre por mí cuando me veías que estaba estresado o un poco enfermo, por animarme y hacerme ver que la vida es una montaña rusa y que hay que saber donde se está en cada momento.

A Jordi por reconciliarme con la bioinformática y aclarar mi relación amor-odio con ella. Y a Vincent, por sus consejos sobre el mundo de la investigación y su ímpetu por la ciencia.

A Juanjo, por enseñarme que la ciencia y la investigación son campos duros y muy competitivos y que con ganas y esfuerzo puedes conseguir lo que al principio parece imposible, aunque la frustración sea el plato principal.

A Yania, que con nuestros cuchicheos hemos hecho de los momentos más duros algo agradable y soportable, con tus consejos prácticos sobre la vida y la gente y que el *Carpe diem* debe ser una pieza fundamental en nuestras vidas y a Sarai, que me enseñó que la vida no es tan rosa como parece y que el esfuerzo y la constancia son muy importantes para crecer como personas.

A Davinia por estar dispuesta a ayudarme en cualquier momento, sea el tema que sea y ser tan agradable conmigo.

A Elena, que juntos hemos podido acabar la carrera y ha sido un placer conocerte un poco mejor.

A Diana, que por conseguir tu sueño, eres capaz de atravesar un océano entero y dejar tu tierra por un tiempo. Me has enseñado que la disciplina y la constancia son muy importantes. Tus consejos y puntos de vista me han ayudado mucho.

Muchas gracias a todos por ser mi segunda familia durante este tiempo y por escuchar mis disparatadas y algo extrañas visiones de la vida así como acompañarme en mis primeros pasos como científico. Espero que siempre os quede una parte de mi en vosotros!!

Indice General

Índice de Figuras

1. Introducción

El envenenamiento por mordedura de serpiente representa actualmente un importante problema de salud pública en regiones tropicales y subtropicales de todos los continentes, a excepción de la Antártida. Estos problemas se agravan al afectar a zonas pobres y subdesarrolladas, donde el trabajo de subsistencia agrícola y ganadera es común. (Kasturiratne *et al.*, 2008). Por ello, ha sido señalada como 'enfermedad desatendida' porque sus mayores consecuencias tienen lugar en el tercer mundo, donde los recursos económicos son escasos y donde los intereses económicos y farmacéuticos son escasos (Gutiérrez *et al.*, 2015; Harrison and Gutiérrez, 2014; Gutiérrez *et al.*, 2014). El número de casos anuales a nivel mundial es aproximadamente de 5 millones, de los cuales unos 85000 son mortales y unos 250000 acaban en amputaciones o deformidades permanentes (Harrison *et al.*, 2009; Warrell, 2013). Es por ello que el estudio de la composición de los venenos de serpiente es crucial para el desarrollo de antivenenos que, de manera significativa, puedan reducir dichas graves consecuencias a nivel mundial. A su vez, los efectos biológicos de algunos de los componentes del veneno pueden servir para el desarrollo de fármacos y para el diagnóstico clínico, con lo que, en este sentido, empresas farmacéuticas o biotecnológicas tiene grandes perspectivas futuras en este campo (Harvey, 2014; King, 2015).

1.1 LAS SERPIENTES

Las serpientes pertenecen al suborden de las *Serpentes,* dentro del orden de los *Squamata* y a su vez a la clase *Reptilia*. Representan alrededor de 3619 especies distribuidas por todos los continentes (a excepción de la Antártida) poblando tanto la tierra como los mares. No obstante, del total de especies, solamente unas 775 representan un potencial peligro para los humanos (*reptile-database.org*).

El grupo de las serpientes evolucionó a partir de una familia de lagartos durante el periodo Jurásico hace unos 200 millones de años (Vidal and Hedges, 2005). Estas serpientes eran boídos (ancestros de boas, pitones y anacondas), no venenosas y que mataban por constricción. Más tarde, en el Oligoceno (hace 35-25 millones de años), a causa de la separación de los continentes y del cambio climático, la superfamilia *Colubroidea* (colúbridos), que hasta entonces había representado una minoría respecto a los boídos, accedieron a nuevos nichos ecológicos, expandiendose rápidamente por todo el planeta, representando hoy en día más de dos tercios del total de serpientes existentes (Greene, 2000). La superfamilia *Colubroidea* incluye las serpientes actuales venenosas, agrupadas en las familias *Viperidae* (vipéridos: víboras y serpientes de cascabel), *Elapidae* (elápidos: serpiente de coral, mambas y cobras), y las familias *Colubridae*, *Lamprophiidae* y *Natricidae*. Dentro de la familia *Viperidae* se encuentran las subfamilias *Crotalinae* y *Viperinae*, las más estudiadas y conocidas (Figura 1). (http://www.reptile-database.org)

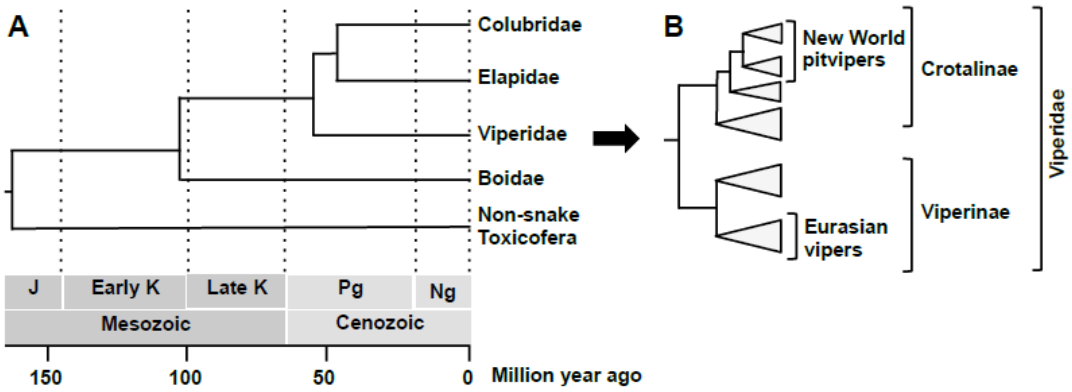

Figura 1: A. Representación filogenética de la evolución de las serpientes. B Distribución de la familia Viperidae, sin representación del tiempo evolutivo. Figura extraída de la tesis doctoral de Raquel Sanz Soler 2016. Abreviaturas: J (jurásico), K (cretácico), Pg (Paleógeno), y Ng (neógeno)

1.2 Evolución de los venenos

Los venenos de serpiente son mezclas de proteínas biológicamente activas, sales y moléculas orgánicas como poliaminas, aminoácidos, y neurotransmisores (Casewell, 2012). No obstante, la definición más actual de veneno es 'una secreción producida en un tejido especializado (generalmente encapsulado en una glándula) de un animal e inyectado en la presa a través de la generación de una herida. El veneno debe contener además moléculas que alteran el proceso fisiológico o bioquímico normal para facilitar la alimentación y / o la defensa del animal (Fry *et al.*, 2012). El veneno tiene muchos papeles diferentes entre los que se encuentran: inmovilizar y paralizar, matar y finalmente devorar a la presa (Calvete, 2017).

Hay controversia acerca del origen de los venenos. Por una parte, existe la hipótesis de que hubo un ancestro común a todos los organismos del clado Toxicofera, ''los que llevan toxinas'' (serpientes, lagartos, iguanas) que portaba los primeros genes de veneno. En este caso, el ancestro reclutó los genes de toxinas de veneno en un primer momento y a partir de ahí, a medida que iban surgiendo diferentes especies, iban apareciendo las diferentes familias de toxinas de veneno. Por tanto, la hipótesis defiende un origen común de los venenos de serpientes y lagartos (Fry *et al.*, 2006).

Por otra parte existe la hipótesis según la cual el veneno evolucionó múltiples veces en diferentes clados del reino animal (Hargreaves *et al.*, 2014). Se trataría, pues, de procesos de convergencia evolutiva, mientras que la hipótesis del origen único arriba mencionada implica mecanismos de evolución divergente.

Las toxinas del veneno evolucionaron a partir de proteínas con una función fisiológica normal, cuya función fue reclutada en la glándula de Duvernoy (donde se produce y almacena el veneno) antes de la diversificación de las serpientes actuales, concretamente en la base de la expansión de la superfamilia Colubroidea (Fry *et al.*, 2006). La diversidad molecular se debe principalmente a la duplicación génica. Esto permitió que una copia génica se expresara selectivamente en la glándula, adquiriendo nuevas características funcionales por selección Darwiniana positiva (Gibbs and Rossiter, 2008). Esto es lo que se conoce como evolución 'birth-and-death' en la cual nuevos genes son creados a partir de duplicaciones génicas. Algunos de estos, llegan a desarrollar nuevas funciones y pueden ser la fuente de otra duplicación y neofuncionalización y

así sucesivamente. En cambio, muchos otros genes son eliminados del genomas o se convierten en pseudogenes (Fry *et al.*, 2003).

La aparición del veneno permitió a las serpientes venenosas subyugar a presas, a menudo más grandes que ellas mismas, de manera más efectiva que la constricción, siendo esta innovación clave para la expansión de las serpientes venenosas respecto a las no venenosas.

1.2 Composición de los venenos de serpiente

La composición del veneno de serpientes varía mucho entre familias, géneros, especies e incluso entre individuos. Dependiendo de su principal efecto tóxico en el animal envenenado, los venenos se pueden clasificar en neuro-/miotóxicos y hemorrágicos/citotóxicos (Ménez, 2002).

Los primeros aparecen principalmente en la familia *Elapidae*, cuyo veneno contiene péptidos neurotóxicos los cuales actúan sobre los canales iones de las neuronas y las conexiones sinápticas. En este grupo encontramos, entre otras, las serpientes australianas reconocidas como las más tóxicas del planeta. (Fry, 1999).

Los segundos son típicos de la familia *Viperidae*. En este caso, el veneno degrada la matrix extracelular de capilares sanguíneos produciendo hemorrágia e interfiere sistémicamente con la cascada de coagulación, los sistemas hemostático y cardiovascular y la reparación de tejidos. (Kini, 2006).

Los venenos de *Viperidae* contienen gran cantidad de isoformas de proteínas, pertenecientes a unas pocas familias estructurales, pudiendo clasificarse como proteínas enzimáticas (serinproteasas, metaloproteasas dependientes de Zn^2, L-aminoácidos oxidasas, PLA_2 (fosfolipasas), nucleotidasas) o proteínas no enzimáticas (lectinas tipo C, péptidos natriuréticos, ohaninas myotoxinas, CRISP (cysteine-rich secretory proteins), disintegrinas o inhibidores de proteasas tipo Kunitz). Los venenos de diferentes serpientes expresan diferente perfil proteico, produciendo un número variable de dichos componentes, tanto en cantidad como en clase proteica (Gutiérrez *et al.*, 2009).

1.2.1 Las metaloproteasas

Las metaloproteasas de venenos de serpiente (SVMPS, Snake Venom MetalloProteases) son proteínas abundantes de los venenos de víboras y serpientes de cascabel. Son endoproteasas en cuyo centro activo hay un sitio de unión a un átomo de Zn^{2+}, que afectan y degradan la matriz extracelular (fibronectina, colágeno ,etc) produciendo hemorragia local, inflamación, y necrosis, así como hipotensión e hipovolemia (Fox and Serrano, 2005).

Estructuralmente, se incluyen con las ADAMs celulares (A disintegrin and metalloprotease) en la familia M12 (reprolisinas) de las metaloproteasas. Evolutivamente las SVMPs provienen de dicha familia génica y como las ADAMs (proteínas transmembrana), son proteínas multidominio. En este caso, las SVMPs evolucionaron a partir del gen codificante de la región extracelular asociada a la proteína transmembrana ADAM que fue reclutado en el proteoma de la glándula de veneno de serpiente después de la divergencia de reptiles squamata en lagartos y serpientes. Con todo esto, la proteína ADAM precursora pasó de ser una proteína transmembrana a ser un proteína soluble con capacidad de ser secretada a la matriz extracelular, al perder el dominio transmembrana, convirtiéndose en una SVMP tipo PIII (Carbajo *et al.*, 2015)

Existen unos 40 genes de ADAMs cuya principal función es la de participar en la remodelación controlada de la región extracelular (ectodominio) de proteínas transmembrana en procesos relacionados con el desarrollo, o la adhesión, migración y señalización celular (Seals and Courtneidge, 2003)

La principal diferencia entre las diferentes clases de SVMPs recae esencialmente en los dominios que poseen. Se clasifican en PI (sólo el dominio metaloproteasa), PII (metaloproteasa más dominio disintegrina C-terminal) y PIII(metaloproteasa más dominio similar a disintegrina y dominio rico en cisteína)(Fox and Serrano, 2005).

El ancestro de las SVMP, como ya se ha comentado, fue un gen precursor de la ADAM 28 del cual surgieron las metaloproteasas tipo III. Las PII y PI son resultado de la pérdida sucesiva del dominio rico en cisteínas y del de disintegrina, respectivamente, (Sanz and Calvete, 2016) como se observa en la Figura 2 (Sanz-Soler, 2016).

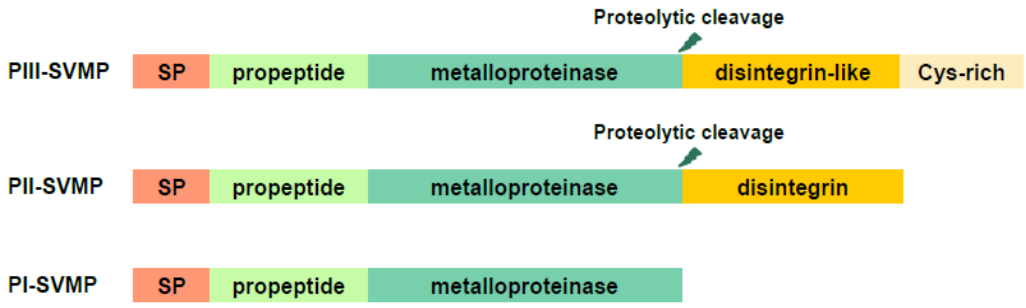

Figura 2: Representación de los diferentes tipos de SVMPs. SP, péptido señal.

1.2.2 Las disintegrinas

Las disintegrinas de los venenos de las subfamilias *Viperinae* y *Crotalinae*, son polipéptidos de entre 41-84 aminoácidos, cuyo principal papel subyace en el hecho de ser antagonistas de receptores de la familia de las integrinas, mimetizando el mecanismo de unión de la integrina a su ligando natural (Wierzbicka-Patynowski *et al.*, 1999; Xiong *et al.*, 2002). Las integrinas, a su vez, son una superfamilia de glicoproteínas, receptores transmembrana que participan en las interacciones célula-matriz extracelular y célula-célula y en procesos como angiogénesis, formación de trombos, integridad de la piel o en el sistema inmunitario (Hynes, 2002). Volviendo a las disintegrinas, éstas son proteínas ricas en cisteínas. Por ello, los puentes disulfuro, posteriormente formados en la maduración de la proteína, juegan un papel muy importante en la estructura final activa de la molécula como se verá más adelante.

Las disintegrinas son sintetizadas o bien a partir de una pequeña molécula de mRNA directamente (Okuda *et al.* 2002), o bien a través del procesamiento proteolítico de una SVMP-PII (Kini and Evans, 1992) (Figura 3).

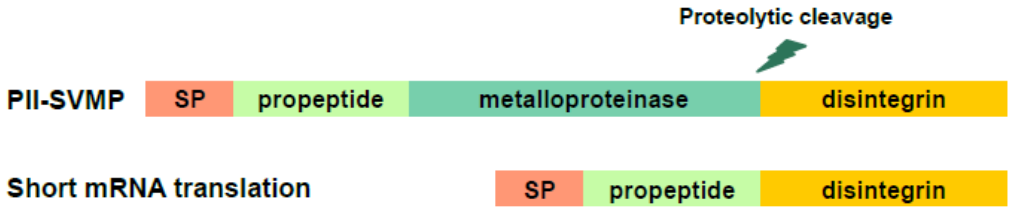

Figura 3: Dibujo que representa los diferentes precursores de las disintegrinas en los venenos de Viperidae (Sanz-Soler, 2016).El sitio de corte se indica en la SVMP-PII. SP, péptido señal.

Las disintegrinas han evolucionado a través de la adaptación de un epítopo conformacional, consistente en un tripéptido característico que reconoce los diferentes tipos de integrinas. Un ejemplo es la tríada RGD (arginina-glicina-glutámico), presente en todas las disintegrinas largas y en la mayoría de medias y cortas. Este tripéptido actúa como origen de los demás tipos de tríadas, que han evolucionado a través de mutaciones puntuales (Calvete, 2010).

En cuanto a la clasificación de las disintegrinas, se clasifican de acuerdo a su longitud y al número de puentes disulfuro en disintegrinas largas (84 aminoácidos (aa) y 7 puentes disulfuro intracatenarios (SS)), medias (70 aa y 6 SS intracatenarios), diméricas –homo- o heterodímeros de subunidades de 63 aa incluyendo 4 SS en cada subunidad y 2 SS intercatenarios y disintegrinas cortas (40-49 aa y 4 SS intracatenarios). La familia de las disintegrinas se originó mediante un mecanismo de minimizacón de la estructura primaria y pérdida sucesiva de enlaces disulfuro, los que se ha denominado un proceso de ‘’ingeniería de enlaces disulfuro’’ (Calvete *et al.*, 2003) (Figura 4)

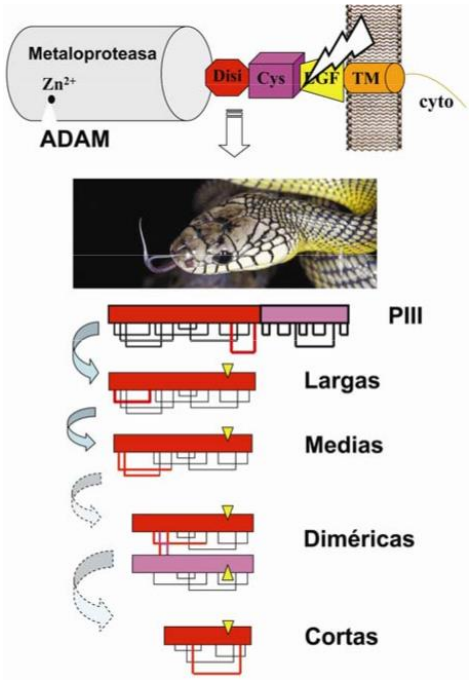

Figura 4: Esquema de la diversificación de estructural de las disintegrinas por el mecanismo de pérdida sucesiva de los enlaces disulfuro. Los triángulos amarillos indican la posición de las secuencias de unión a integrinas. (Calvete et al., 2003)

Las disintegrinas tienen un gran interés biomédico en el estudio de diversos procesos biológicos en los que las integrinas juegan un papel fundamental. Así, a partir de la estructura del tripéptido RGD se han diseñado peptidomiméticos que antagonizan especificamente a la integrina plaquetaria αIIbβ3, con objeto de prevenir procesos tromboembolíticos. (Marcinkiewicz, 2005).

1.3 Métodos de secuenciación

Las técnicas de secuenciación de DNA han evolucionado de manera muy rápida en los últimos años (Bainbridge *et al.*, 2006). El método descrito por Sanger en el año 1977 basado en el uso de dideoxinucleótidos terminales, fue el método de secuenciación de DNA dominante y la técnica que se usó en el proyecto de secuenciación del genoma humano (Venter *et al.*, 2001).

Sin embargo, en los último 10 años, el escenario ha cambiado radicalmente gracias a la aparición de las técnicas NGS (*next generation sequencing)*, llevando a cabo una mejora respecto a la longitud de lectura obtenida y al rendimiento total del proceso (número total de bases secuenciadas), dando lugar a lecturas cada vez más largas y con un mayor rendimiento a través de la obtención de una gran cantidad de datos. Presentan además la ventaja de poder analizar de manera masiva varias muestras en paralelo y un precio mucho menor respecto a las anteriores plataformas de secuenciación (Metzker, 2010). Las nuevas plataformas han superado de una manera asombrosa anteriores problemas, ofreciendo ventajas respecto a la longitud de la lectura, el precio y el tiempo, como se aprecia en la Tabla 1 (Liu *et al.*, 2012).

(a)

Sequencer	454 GS FLX	HiSeq 2000	SOLiDv4	Sanger 3730xl
Sequencing mechanism	Pyrosequencing	Sequencing by synthesis	Ligation and two-base coding	Dideoxy chain termination
Read length	700 bp	50SE, 50PE, 101PE	50 + 35 bp or 50 + 50 bp	400~900 bp
Accuracy	99.9%*	98%, (100PE)	99.94% *raw data	99.999%
Reads	1 M	3 G	1200~1400 M	—
Output data/run	0.7 Gb	600 Gb	120 Gb	1.9~84 Kb
Time/run	24 Hours	3~10 Days	7 Days for SE 14 Days for PE	20 Mins~3 Hours
Advantage	Read length, fast	High throughput	Accuracy	High quality, long read length
Disadvantage	Error rate with polybase more than 6, high cost, low throughput	Short read assembly	Short read assembly	High cost low throughput

(b)

Sequencers	454 GS FLX	HiSeq 2000	SOLiDv4	3730xl
Instrument price	Instrument $500,000, $7000 per run	Instrument $690,000, $6000/(30x) human genome	Instrument $495,000, $15,000/100 Gb	Instrument $95,000, about $4 per 800 bp reaction
CPU	2* Intel Xeon X5675	2* Intel Xeon X5560	8* processor 2.0 GHz	Pentium IV 3.0 GHz
Memory	48 GB	48 GB	16 GB	1 GB
Hard disk	1.1 TB	3 TB	10 TB	280 GB
Automation in library preparation	Yes	Yes	Yes	No
Other required device	REM e system	cBot system	EZ beads system	No
Cost/million bases	$10	$0.07	$0.13	$2400

Tabla 1: Comparación de las principales características entre la tecnología de primera generación (Sanger 3730xl) y las de segunda generación (el resto). Se puede apreciar el gran salto en cuanto a la longitud de las lecturas, los datos de salida (output data/run) o el tiempo. También remarcar el abismal descenso del coste de secuenciación por millón de bases. Tabla proveniente de Liu et al. 2012.

Las secuenciación por Sanger es conocida como técnica de primera generación. Las técnicas de NGS, en cambio, se han clasificado como técnicas de segunda (Illumina, Roche 454, Ion torrent) y tercera generación (PacBio y Nanopore). La principal diferencia entre la segunda y la tercera

generación es que en la segunda se tiene que amplificar primero la muestra por PCR para poder tener una señal significativa, mientras que en la de tercera generación se usa la técnica SMRT (single-molecule real-time), donde una molécula de DNA es directamente secuenciada a tiempo real, mediante detección de fluorescencia (Pacbio®, Pacific Biosciences, California) o corriente eléctrica (Nanopore®, Oxford Nanopore)(Deamer *et al.*, 2016; Magi *et al.*2017) .La longitud de la lectura es mucho más larga en tercera generación, llegando a varias kb de una sola lectura.

A continuación se explicará con más detalle la plataforma de tercera generación MinION de la empresa Nanopore Oxford, ya que es la que se ha usado en este Trabajo Final de Grado.

1.3.1 MinION

La plataforma de secuenciación MinION entra dentro de la denominada tercera generación de secuenciadores, y está revolucionando el campo de la genómica. Se trata de un sistema de tamaño un poco menor que la palma de la mano (Feng *et al.*, 2015)(Figura 5), lo que facilita su transporte y su manejabilidad (www.nanoporetech.com).

Figura 5: Tamaño del MinION (Feng et al., 2015)

El MinION es capaz de generar lecturas largas, de decenas de kilobases, a bajo coste y a una alta velocidad con una mínima preparación de la muestra usando una instrumentación sencilla (Laszlo *et al.*, 2014)

A grandes rasgos, la secuenciación del DNA tiene lugar cuando se añade la muestra de DNA a la célula de flujo (Flowcell). Cuando la molécula de DNA pasa a través del nanoporo, hay un cambio en la magnitud de la corriente que atraviesa dicho poro, medido por un sensor (figura 6). El flujo de datos es posteriormente analizado por los softwares ASIC y MinKNOW (Lu *et al.*, 2016)

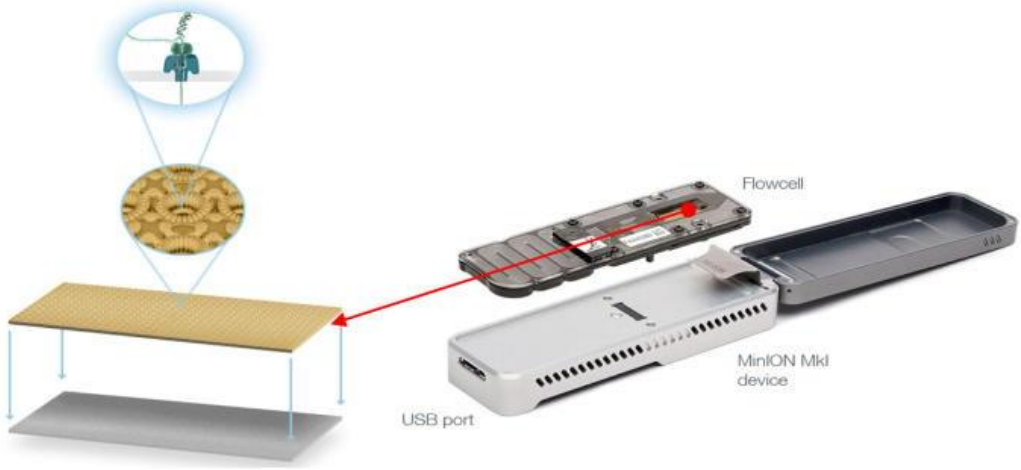

Figura 6: Representación del sistema de secuenciación MinION, donde se aprecia la célula de flujo así como una representación de los poros por los que pasará una única hebra de DNA (Lu et al., 2016)

En la célula de flujo, una fina membrana de grafeno divide en dos parte una disolución salina: una parte cis (la superior) y una parte trans (la inferior). En la membrana hay poros embebidos de tamaño nanométrico que conectan la parte *cis* con la *trans* (Laszlo *et al.*, 2016). Cuando se aplica un voltaje de manera transversal a la membrana, se crea un flujo de corriente de iones a través del poro. Esta corriente crea la señal primaria. Una vez se aplica la muestra, las moléculas de DNA, cargadas negativamente, son atraídas hacia el interior de los poros electroforéticamente y empiezan a travesarlos monocatenariamente en dirección cis-trans. En el proceso de paso del DNA a través del poro, se bloquea una fracción de la corriente iónica. El valor de la fracción de corriente bloqueada depende en última estancia de la identidad del nucleótido que en ese momento atraviesa el interior del poro y se detecta como una señal característica (Manrao *et al.*, 2011) (Figura 7).

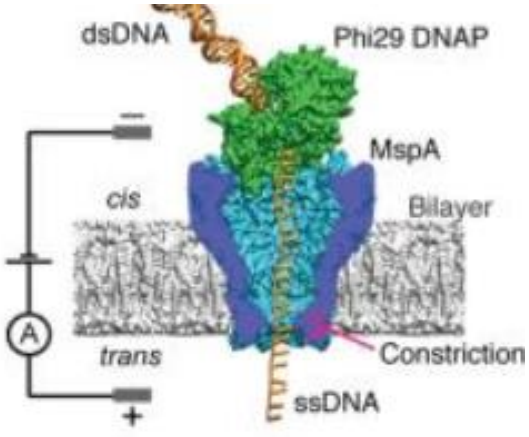

Figura 7: Sección transversal de la lámina con el poro embebido. Se aprecian las partes cis y trans, el voltaje aplicado y el DNA monocatenario atravesando el poro. (Laszlo et al., 2014)

Actualmente, el poro que se usa es una lipoproteína de canal transmembrana llamada Holey Gray 9 (CsgG Porina de *E.coli* altamente modificada) porque tiene un tamaño tal que permite el paso justo de una hebra monocatenaria de DNA. A este poro se unirá en su parte cis la helicasa junto con el DNA de doble cadena, lo que permitirá que por el poro sólo pase DNA monocatenario ya que la helicasa cataliza la separación de las dos hebras.

En concreto, no se mide nucleótido a nucleótido de manera individual, sino combinaciones de éstos, normalmente de 5 en 5, que bloquean la corriente de una manera característica y única. Centenares de poros trabajan en paralelo y a un ritmo diferente, y cada señal que se genera en cada poro se va registrando en el sistema.

Por debajo del poro hay circuitos de detección muy sensibles que transmiten la magnitud del cambio en la corriente eléctrica a una velocidad muy rápida (figura 8). Estas señales se registran como cambios en la intensidad de corriente a través del inductor analizando estadísticamente la amplitud y la duración de cada interrupción del flujo iónico(Figura 9) (Feng *et al.*, 2015).

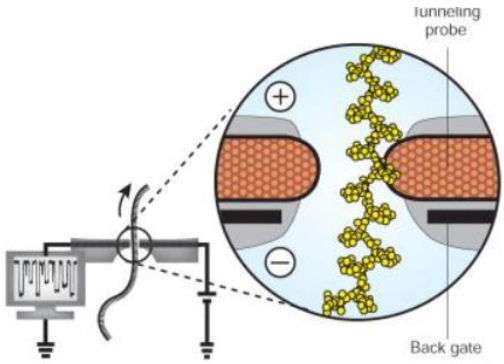

Figura 8: En la imagen de la izquierda se observa cómo se registra la señal generada por la disrupción de la corriente iónica a lo largo del poro. (Branton et al., 2008)

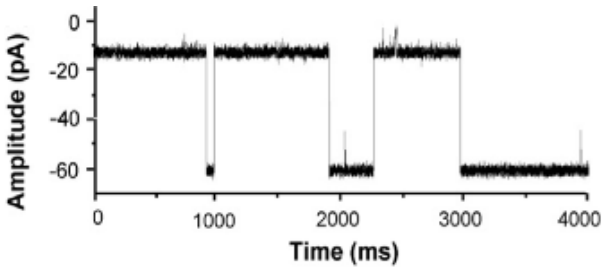

Figura 9: Representación esquemática del cambio de intensidad debido al paso de la hebra monocatenaria. Cada amplitud (intensidad en picoamper) y duración (tiempo) es característica de un conjunto de cinco nucleótidos. (Feng et al., 2015)

Como resumen, el MinION ofrece ventajas respecto a otras plataformas de secuenciación como son la ausencia de marcaje del DNA, lecturas muy largas (10^4-10^6), alto rendimiento y bajo material de partida así como su bajo coste (entre 500-1000$ una secuencia genómica completa) (Feng *et al.*, 2015).

2. OBJETIVO

El objetivo de este Trabajo Final de Grado consistió en caracterizar genes de metaloproteasas y disintegrinas nuevas de dos especies de serpientes mediante dos métodos de secuenciación diferentes:

- Secuenciación mediante Sanger iterativo de genes de disintegrinas de la especie *Bitis arietans* (vipérido). A partir de los datos obtenidos se realizó un estudio comparativo de los intrones y exones de las disintegrinas frente a datos existentes en las bases de datos y se caracterizó la proteína codificada.
- Secuenciación mediante la plataforma MinION de genes de metaloproteasas de la especie *Bothrops jararaca* para elucidar la secuencia de los genes a través del ensamblaje de las diferentes lecturas mediante análisis bioinformático y caracterización de la estructura genómica y distribución en tándem de los genes.

3. MATERIALES Y MÉTODOS

3.1 Secuenciación por Sanger Iterativo

3.1.1 DNA genómico

Se extrajo DNA genómico de hígado de *Bitis arietans* (Nigeria). El hígado fue molido a polvo fino con un mortero cerámico tras su congelación en nitrógeno líquido y se obtuvo el DNA genómico usando el protocolo procedente de Molecular cloning: a laboratory manual (Sambrook J, 1989) conteniendo tampón de lisis con SDS (concentración final de SDS de 0,5%) y proteinasa K (0,5 mg por mL de tampón de lisis). El homogeneizado se incubó a 37ºC toda la noche. Después se añadió el mismo volumen de fenol: cloroformo: isoamílico (25:24:1), se mezcló y se centrifugó 10 minutos a 2500 r.p.m, recogiendo la fase acuosa. Luego se precipitó la muestra con 1/10 de volumen de acetato sódico 3M pH= 5,4 y dos volúmenes de etanol absoluto y se centrifugó 2 minutos a 4500 r.p.m. El precipitado resultante se lavó con etanol al 70%, se centrifugó 2 min a 4500 r.p.m, y se dejó secar, se disolvió en 500 µL de tampón TE y se midió concentración en NanoDrop™ (ThermoFisher) y se caracterizó la integridad del DNA en un gel de agarosa del 0,8%.

3.1.2 Amplificación por PCR de los fragmentos de interés.

Se siguieron dos estrategias diferentes para amplificar genes de disintegrinas y metaloproteinasas a partir de DNA genómico de *Bitis arietans*: usando directamente el genómico o mediante la aproximación denominada "*genome walking*".

A partir de los datos de las secuencias de metaloproteasas de *Bitis arietans* registrados en la base de datos del NCBI, se diseñaron los cebadores. Las secuencias encontradas fueron un RNA mensajero de metaloproteasa MP-2 (código de acceso del GenBank AY885661) y del RNA mensajero del pseudogen BA-5A (AM117393).

Se diseñó el cebador directo a partir de la secuencia del péptido señal de BA-5A, denominado P120 y el reverso a partir de la secuencia del final de la secuencia MP-2, BarietStop_Rv (Tabla 1).

Para la reacción de la PCR se usó el kit comercial Phusion High-Fidelity PCR Master Mix®(Thermo Scientific Referencia F-532S). De los 50 µL de reacción, 25 µL eran de la solución 2XMáster Mix® (dNTPs, polimerasa Phusion y buffer conteniendo el $MgCl_2$), 3% de DMSO, 0,5 µM de cada cebador, 100 ng de DNA y agua. El programa del termociclador incluyó 98ºC de desnaturalización 30 segundos, seguido de 35 ciclos de 98ºC durante 10 segundos (desnaturalización), 60ºC por 20 segundos (anillamiento de los cebadores) y 72ºC durante 7 minutos (extensión), como último paso, una extensión final a 72ºC durante 5 minutos.

La aproximación por *genome walking* se llevó a cabo mediante GenomeWalker™ Universal Kit®(Clontech). Este abordaje comprende dos amplificaciones sucesivas (PCR anidada) de los fragmentos amplificados en una primera PCR, con la creación de una librería Genome Walker previa. Esta librería comprende fragmentos de DNA previamente digeridos por una enzima de restricción específica, obteniendo extremos romos, a los que se le añaden y unen adaptadores a ambos extremos.

Para ello se digirieron 2,5 µg de DNA genómico en cuatro tubos, cada uno de los cuales llevaba un enzima de restricción PvuII, EcoRV, StuI y DraI distinto. En todos ellos se llevó a cabo el mismo

procedimiento. Se procedió a la digestión del DNA con 80 unidades del enzima correspondiente, tampón 10X, agua libre de nucleasas y se dejó digerir 18 horas a 37ºC. Una vez digerido, se lleva a cabo la purificación del DNA, primero con 95 µL de fenol y, tras la recuperación de la fase acuosa, se añadieron 95 µL de cloroformo. Con este tratamiento se consigue eliminar el enzima. Después, tras otra recuperación de fase acuosa, se añadieron 190 µL (2V) de etanol al 95%, 9,5µL de acetato sódico 3M y 20 µg de glucógeno y se centrifugó 10 min a 15000 r.p.m. Con esto se consigue que el DNA precipite. Finalmente, tras descartar el sobrenadante, se lavó el precipitado con etanol frío al 80% y se dejó secar para posteriormente disolverlo en 20 µL de TE y cuantificar su concentración por NanoDrop™ (ThermoFisher).

A continuación, se ligaron los adaptadores en cada uno de los cuatro tubos. Para ello, en un volumen total de reacción de 8 µL, se añadieron 4 µL del digerido anterior, 1,9 µL de adaptadores, tampón 10X y 0,5 µL de la ligasa T4 y, tras una incubación de 18h, se desactivó la reacción incubando los tubos 5 minutos a 70ºC. Finalmente se añadieron 72µL de TE y se obtuvo la librería para cada una de las cuatro reacciones.

Se llevó a cabo la primera PCR, en la que se usó un cebador AP1 del kit GenomeWalking, que hibrida con el adaptador y los cebadores directo p120 y el reverso BarietaMet_Rev2. Se añadieron, en ocho tubos diferentes, combinaciones de AP1 y uno de los cebadores (el directo o reverso, 2 reacciones para cada tipo de enzima). La mezcla resultante fue de 12,5 µL de MasterMix 2X, 3% de DMSO, 1µL de librería, 0,5 µM del cebador directo o reverso, 0,5 µM de AP1 y agua. El termociclador se programó para 7 ciclos de 98ºC durante 30 segundos (desnaturalización) y 72ºC durante 5 minutos para la extensión y luego 32 ciclos de 98ºC durante 30 segundos y 67ºC durante 5 minutos, y como último 67ªC durante 7 minutos para la extensión final. Para la segunda PCR, 1 µL de la anterior reacción en dilución 1:25 se volvió a amplificar excepto que se programaron 5 y 25 ciclos y la extensión final tuvo una duración de 7 minutos. Para esta reacción, se utilizó el primer AP2, incluído también en el GenomeWalker kit, que hibrida también con el adaptador pero más hacia el interior del fragmento al que está unido. En la mezcla final se mezclaron 25 µL de MasterMix 2X, 2% de DMSO, 0,5 µM del primer p120 y AP2, 1 µL resultado de la amplificación de la primera PCR y agua.

CEBADOR	**SECUENCIA DE DNA 5'→3'**
P120	ATGATGCAAGTTCTCTTAGTAACTATATGCTTAGC
Barietstop_rv	GTAGGCTGTATTCACATCAACACAC
BarietaMet_Rev2	ATGGCCTACCATTGCAAGTACAG
Adaptador	GTAATACGACTCACTATAGGGCACGCGTGGTCGACGGCCCGGGCTGGT
AP1	GTAATACGACTCACTATAGGGGC
AP2	ACTATAGGGCACGCGTGGT
pJET T7 DIRECTO	CGACTCACTATAGGGAGAGCGGC
pJET REVERSO	AAGAACATCGATTTTCCATGGCA

Tabla 2: Cebadores usados en ambas estrategias para la amplificación de genes de interés

3.1.3 Purificación y clonación de los productos de PCR

Las muestras se separaron por electroforesis en un gel de agarosa del 0,8%. Las bandas se cortaron y se purificaron usando el kit QIAEX II (QIAGEN) para bandas <10 kb. Los fragmentos purificados se insertaron en el plásmido pJET usando la ligasa T4 del kit CloneJET™ PCR (ThermoScientific) y transformando en bacterias electrocompetentes de *Escherichia coli* cepa DH5α por electroporación a 1700 V. Las células transformadas se resuspendieron en 200 µL de medio LB e incubadas 1 hora a 37ºC. Tras esto, fueron cultivadas en placas de LB agar/ampicilina para seleccionar los clones positivos. La presencia del fragmento insertado en el plásmido se verificó a través de una PCR de 20 colonias que habían crecido en la placa, con los cebadores pJET T7 directo y pJET reverso en una reacción de volumen total de 15 µL con MasterMix 2X, 0,5 µM de cada uno de los cebadores anteriores y agua. No se añade DMSO. El programa fue de 98ºC durante 30 segundos y 35 ciclos de 98ºC por 10 segundos, 60ºC por 20 segundos y 72ºC durante 1 minuto y para finalizar 5 minutos a 72ºC. Las colonias positivas se volvieron a cultivar en LB toda la noche a 37ºC y se extrajo el plásmido con el kit Wizard® Plus SV Minipreps DNA purification system (Promega). El inserto se secuenció en el sistema de secuenciación de Applied Biosystems 377 usando T7 directo y T7 reverso como cebadores (tabla 1).

3.1.4 Proceso iterativo y análisis de secuencias.

En el proceso de secuenciación, debido al gran tamaño del inserto, se tuvieron que diseñar varios cebadores directos y reversos, a medida que se iba conociendo para cubrir la totalidad del inserto e ir uniendo posteriormente unas secuencias con otras hasta tener la secuencia completa del inserto (Tabla 2). Para ello, se unían las secuencias a las siguientes basándose en la calidad de cada nucleótido del electroferograma a través del software Chromas.

CEBADOR	SECUENCIA DNA
BarietIntr1-Fw	GACACTCAATTGATTCAGAAGTTACC
BarietIntr-Rv	TCTGAGTGGTTCCTAGAAGTCC
BarietEx_2Fw	ACGAAGGGAGCTCTATAATCCTGG
BarietaEx_3_Rev	CAGTTACGACAACATGGTCCAG

Tabla 3: Cebadores usados para la secuenciación del fragmento de 6Kb

3.2 Secuenciación con MinION

3.2.1 Origen de las muestras

El material de partida fueron 8 fragmentos de DNA de un tamaño de entre 150-200 kb insertados en BACs de tipo pBeloBAC11 (cromosoma bacteriano artificial). Éstos se transformaron en la cepa DH5α de *E.coli* procedentes del Instituto Butantan (São Paulo, Brasil). A pesar de que la secuencia era desconocida, se sabía que contenía genes y regiones intergénicas de metaloproteasas, ya que habían sido seleccionadas previamente de una librería de BACs del genoma de la serpiente *Bothrops jararaca* mediante PCR.

3.2.2 Preparación de la librería

Se cultivaron las bacterias en 500 mL de medio de cultivo y se extrajeron los BACs utilizando el kit NucleoBond® Xtra BAC (Macherey-Nagel) y se determinó la concentración de cada BAC con el fluorímetro Qubit (ThermoFisher). Después se preparó la librería para cargar los BACs en el MinION con el kit Ligation Sequencing Kit 1D R9 version de referencia SQK-LSK108 (Oxford

Nanopore). Para ello, se transfirieron 500 ng de cada BAC (4 µg de DNA total) a un tubo de 1,5 mL y se ajustó a un volumen total de 80 µL de agua libre de nucleasas. A continuación, se transfirió la dilución a un tubo Covaris® (Eppendorf) y se centrifugó a 5790 g, donde los BACs se fragmentaron.

A continuación se procedió a reparar las posibles muescas o nicks que puediera haber en los fragmentos de DNA con el kit NEBNext® FFPE DNA Repair Mix (NewEngland BioLabs) y los extremos de los fragmentos para que quedaran romos con el kit NEBNext® Ultra™ II (NewEngland BioLabs). Luego se purificó el DNA con el uso de bolas magnéticas con el kit AMPure XP® (Beckman Coulter). Seguidamente, se añadieron los adaptadores y se ligaron con el kit Quick Ligation Module® (NewEngland BioLabs) y se volvió a purificar los fragmentos con las anteriores bolas magnéticas. La librería quedó preparada para cargarla en el MinION

3.2.3 Carga de la librería en la célula de flujo y secuenciación.

Se ensambló la célula de flujo al MinION, y se prepararó el software MinKNOW para adquirir e interpretar las señales generadas por el MinION durante la secuenciación. Se añadió medio tamponante a la célula de flujo y se preparó la librería para cargarla con el kit Library Loading Bead Kit R9 version EXP-LLB001. Finalmente se añadieron 75 µL, gota a gota, de la librería y empezó a secuenciar durante 48 horas.

3.2.4 Análisis bioinformático de las secuencias

El análisis de secuencias se llevó a cabo mediante línea de comandos de Linux. Las secuencias se convirtieron a formato Fastq a partir de los datos crudos del MinKNOW de formato Fast5 con el programa PORETOOLS mediante el comando *poretools fastq rawdata/ > out.fastq*. A continuación se eliminaron las secuencias de los adaptadores con el programa PORECHOP mediante el comando *Porechop –i out.fastq –b trimmed > & porechop.log*. Se eliminaron las secuencias procedentes del DNA genómico de *Escherichia coli* consecuencia de contaminación durante el proceso de extracción de los BACs. Para ello se utilizó el software BWA y SAMTOOLS a través de mapeo contra el genómico procedente de la cepa K12, previamente descargada y guardada con el comando *curl-s http://hypervolu.me/%7Eerik/genomes/E.coli_K12_MG1655.fa >E.coli_K12_MG1655.fa*. Luego se indexó y se ensamblaron contra el genoma de referencia y se eliminaron los reads que contenían secuencias del genómico de *E.coli*, obteniendo los reads finales. Para mapear se procedió con el comando *bwa mem -t 3 -x ont2d E.coli_K12_MG1655.fa trimmed/none.fastq > map/Bjararaca_3_restart_Ecoli.sam.* Posteriormente, se creó un archivo con las secuencias que no mapean con nada de *E. coli* con el comando *samtools view -f 4 Bjararaca_3_restart_Ecoli.sam > Bjararaca_3_restart_unmapped.sam*. Más tarde se transforma el archivo generado con formato SAM en formato BAM mediante el comando *samtools view -bS Bjararaca_3_restart_unmapped.sam | samtools sort -o Bjararaca_3_restart _unmapped.sorted* para finalmente transformarlo en formato Fastq mediante el comando *samtools bam2fq Bjararaca_3_restart_unmapped.sorted > Bjararaca_3_restart_unmapped.fastq*

Se eliminaron las secuencias del vector del BAC utilizando el comando *bwa index gi_157065011_gb_EU140750.1_8128.fa bwa mem -t 5 -x ont2d /home/enric/Desktop/Bjararaca_3_restart/map/gi_157065011_gb_EU140750.1_8128.fa Bjararaca_3_restart_unmapped.fastq > Bjararaca_3_restart_unmapped_novector.fastq samtools view –bS Bjararaca_3_restart_unmapped_novector.sam| samtools sort - Bjararaca_3_restart_unmapped_novector.sorted*

Mediante BLAST se eliminaron los reads que todavía contenían algún resto de vector mediante el comando *formatdb -t vector_BAC -i gi_157065011_gb_EU140750.1_8128.fa -p F -n vector_BAC*, por una parte, y después, *blastall -p blastn -d /home/enric/Desktop/Bjararaca_3_restart/map/vector_BAC -i Bjararaca_3_restart_unmapped_novector.fasta -e 0.0001 -m 7 -o Bjararaca_3_restart_unmapped_results_novector.xml* .

El archivo FASTA anterior viene del Bjararaca_3_restart_unmapped_novector.fastq a través del comando *Seqret –sequence reads.fastq –outseq reads.fasta*.

Se eliminaron las secuencias que contenían todavía algún rastro de vector con el script Archivo_1.py a través de Python. El archivo sin genómico de *E.coli* y sin vector se llama fasta_for_enric.fa

Se ensamblaron las secuencias mediante el programa CANU a través del comando *canu –p Canu_assembly_Bjararaca_3_restart -d canu_assembly corMinCoverage=0 contigFilter="2 1000 1.0 1.0 2" errorRate=0.035 genomeSize=2,4m -nanopore-raw fasta_for_enric.fa*

Se mapearon los contings obtenidos contra los reads anteriores para determinar la cobertura utilizando el bwa. Concretamente el archivo canu_assembly_Bjararaca_3_restart.contings.fasta contra los reads fasta_for_enric.fa.

Finalmente se usó el programa nanopolish para pulir los contings anteriores. Este programa mapea los contings contra los reads anteriores para asegurar la secuencia de aquellos cuya cobertura era muy baja.

Los contigs resultantes se alinearon mediante BLAST y se anotaron las secuencias. Para la anotación se realizó un alineamiento múltiple entre los exones de metaloproteasas y la secuencia del contig.

4. RESULTADOS Y DISCUSIÓN

4.1 Estructura de un gen de disintegrina dimérica

El resultado final de la PCR con los primers P120 y BarietStop_RV fue un fragmento de DNA de entre 4-6 Kb, como se observa en el gel de agarosa de la Figura 10.

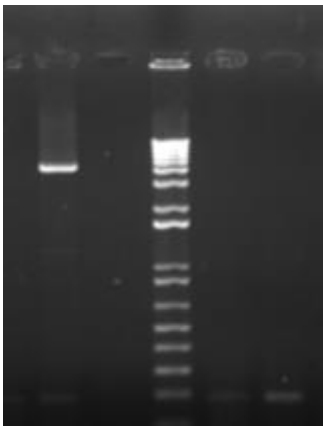

Figura 10: Gel de agarosa con el fragmento amplificado por PCR dos carreras a la izquierda del marcador de pesos moleculares.

Tras la secuenciación Sanger iterativa con los cebadores anteriormente indicados, se observó que la secuencia tiene un tamaño de 6459 nucleótidos, constituida por 4 intrones y 5 exones (Figura 11).

```
ATG ATG CAA GTT CTC TTA GTA ACT ATA TGC TTA GCA TTA TTT CCA TAT CAA G
                    P120
 M   M   Q   V   L   L   V   T   I   C   L   A   L   F   P   Y   Q
gta aga tgt tct ctt tgg ttc cct tgt tca gaa tct tac tgc taa aag act ctt gca ccc
aac aga ttg cta aat tgt tgt ttt tgt ttt tat ttt tga caa tta atc aaa att tgc tcc
act tca gtt tct aca gat tta gca aaa gaa tga tct caa gga tca cat ttt ttt cta aag
tga cat aaa ttc ttt ctt ggt att tgc tta aat ttt gtc aag cca cga gca gaa tct aag
aaa gaa aag ttc taa gat ctt cag aag aaa agc aag ttc aag ttg gag aga aca ttt tct
aat ggt gtg agt ttt tta aaa agg ctt aga aga ctg aag agg ata gag aga gtt tgt cac
aaa aaa aat tca aaa agc aaa aga aaa aaa aat agg ttt ggt ttt aat ttt ctt tca gta
            Simple repeats
ggt caa tta gac atc ctg taa tta aac atg cat cat att tat cca aat aaa agc tag tgc
tat tgg tgt tac aga gaa tat ttg aac tgt act ctg tga gct aag ata cat ttt aca tgc
cag ttg acc agt ctg tgg ctc ttc tcc ccg tct ccc aag tca ttc ccc cat atc att aca
aac aat gac act caa ttg att cag aag tta cca aac aca gtt aag ggt cag agg gct ctg
          BarietIntr1-Fw
ttg gtt tat ttc ttt ttt aaa tat taa gga aag atg ggg aat acc tta cca gta gta att
tat ggg gaa caa tgt ggg aaa gaa gaa aga gag aca gaa ggg aaa ttg ctg ctt att ctg
tgt tct aag caa gag agg aat agt aag ctt gct tgc tag tag gtg agt tga atc ctt ccc
acc aat ctg gga gca cca ggt aca aat ggg tag gta cgt aaa tag gta tga gga ata tgt
att ctt cta tca ggc tgt gca tcc aaa acg acc tgc agg tgt cag aat gcc aca ttt aca
aaa taa gat aag gaa caa ata tca ggt gac cat gca ata tcc ata tgg gga cca act cta
aat ttt tca tct act aat caa att ctg aaa tca cca gat tat ttg cat gtt gtc tgg tag
tcc agt aat agt tat ggt gct gtg ggc tgt cca tat att ctt ttg aaa gtg aaa gaa atg
gaa gtc agc aaa att att agg atg gca gtg aaa tag cat ttc aac tga aaa gaa ttc cta
ata agc tgc att tgt att ccc aga gaa aat ctg gat ttc tgt gtt gtg tcc cac tag aaa
aag att cca ttc ttc act gat aac tga ata aca aat tgt gtt gtg gat gtg aaa atg gta
gaa taa ctt cac aaa gta aag caa acc atc ttt tct ttt tac tta cga ag  GG  AGC TCT
                                                                  G   S   S
ATA ATC CTG GAA TCT GGG AAT GTT AAT GAT TAT GAA ATA GTG TAT CCA AAA AAA GTC GCT
 BarietEx_2Fw
 I   I   L   E   S   G   N   V   N   D   Y   E   I   V   Y   P   K   K   V   A
GTG TTG CCC ACA GGA GCA ATG gta aga aaa gat ctt tgt ttc aac aac aaa tta ttt ctt
```

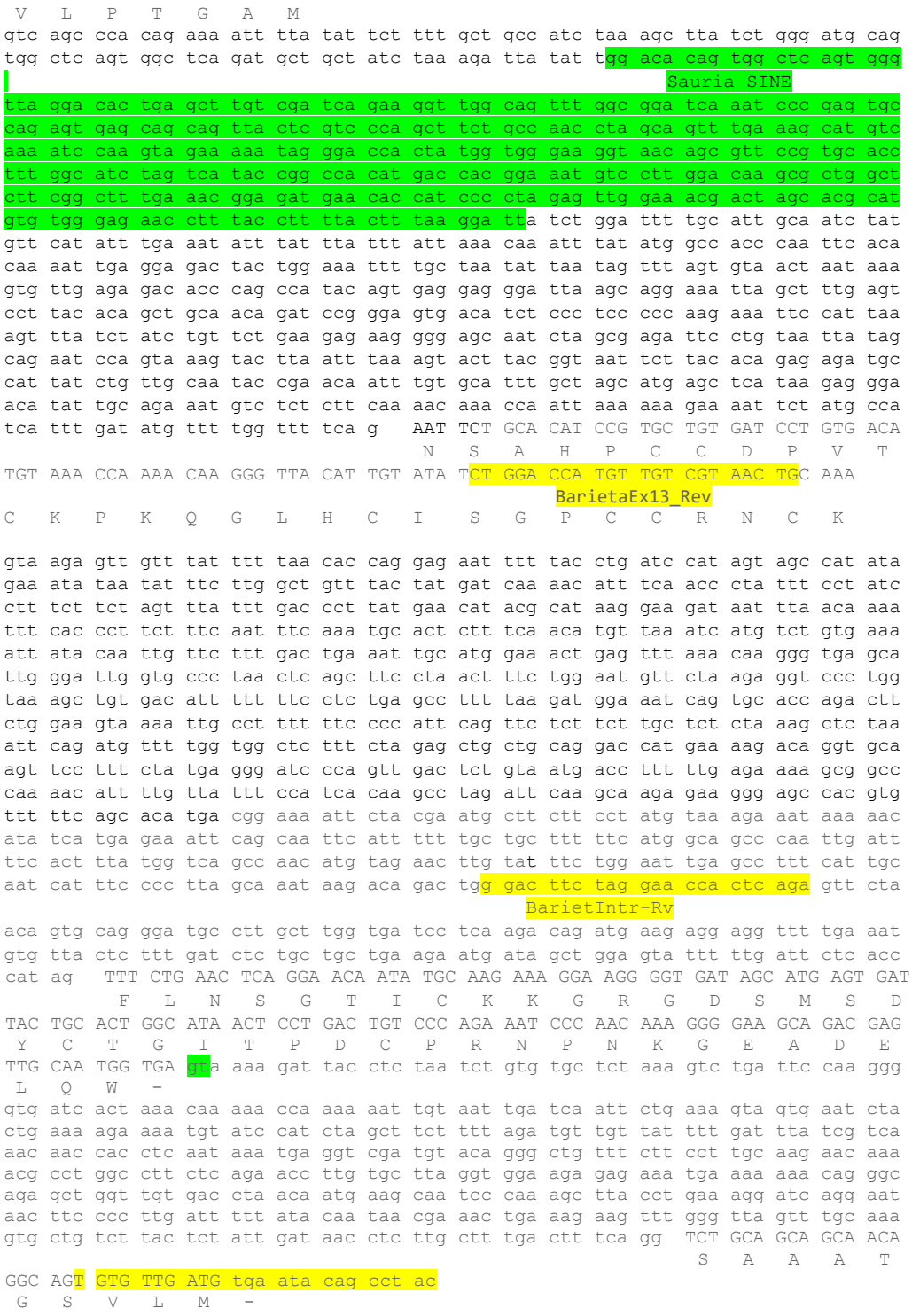

Figura 11: Resultado de la secuenciación Sanger iterativa. Se observan en mayúscula los exones y en minúscula los intrones. En amarillo aparecen los cebadores usados durante la secuenciación.

Al alinear la secuencia con la base de datos del NCBI mediante BLAST, la mayor identidad correspondió a una subunidad de disintegrina dimérica de *Macrovipera lebetina* con número de acceso AM261812. De aquí se deduce, a partir de su parálogo, que esta secuencia corresponde a una subunidad de disintegrina dimérica (homo- o heterodimérica) de *Bitis arietans.*

En cuanto a su estructura génica, el intrón 1 es de fase 1 y el resto son de fase 0. El programa RepeatMasker identificó en el intrón 1 una secuencia repetitiva simple y en el intrón 2 una secuencia Sauria SINE. Además se observa que las secuencias de procesamiento de intrones (5'-GT(donante)/3'-AG(aceptor)) están conservadas.

La secuencia Sauria SINE ha sido caracterizada en la mayoría de serpientes y lagartos y son clasificados como retrotransposones. Se cree que su origen proviene de la fusión de una secuencia similar a un gen de tRNA y el final de la secuencia 3' de una secuencia LINE. Los Sauria SINE utilizan una endonucleasa y una transcriptasa reversa de una secuencia LINE para retrotranscibirse y expandirse por el genoma. Las secuencias repetitivas Sauria SINE están distribuídas por todo el genoma y se cree que actúan como sitios de recombinación homóloga entre SINEs, lo que llevaría a una serie de reordenamientos genómicos como duplicaciones, deleciones y traslocaciones que proporcionarían un mecanismo de diversidad genética a los reptiles. Además, los análisis filogenéticos de los Sauria SINEs indican que su origen y expansión coinciden con la evolución de los sistemas de veneno de serpiente y han cobrado importancia en la evolución de las SVMPs y las disintegrinas (Piskurek and Okada, 2007).

Se sabe también que la evolución de las disintegrinas, incluyendo el mecanismo molecular de diversificación estructural y funcional de esta familia de proteínas, engloba no sólo una sucesión de deleciones y mutaciones puntuales sino también una pérdida de secuencias intrónicas (Bazaa *et al.*, 2007) mediante un mecanismo de minimización, como se ilustra en el esquema de la Figura 12.

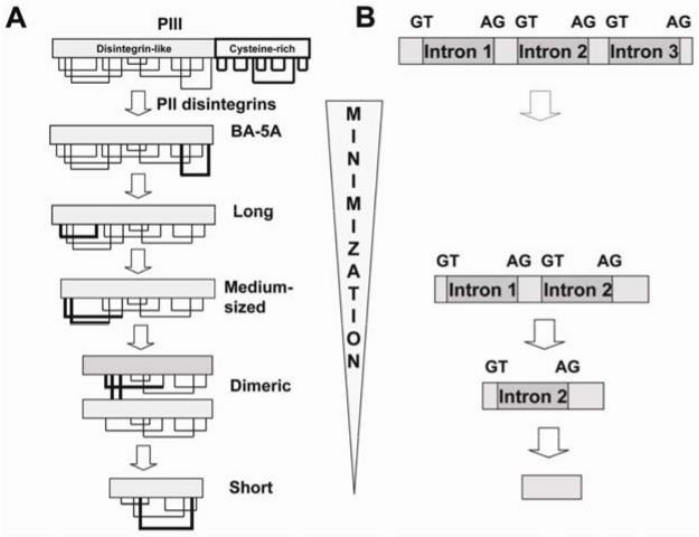

Figura 12: A. Evolución y aparición de los distintos tipos de disintegrinas a partir de una metaloproteasa PIII. B Esquema de la organización exón-intrón conservada y evolución por pérdida de intrones (Calvete et al., 2003)

En cuanto a la estructura de los exones que conforman la secuencia, no todas dan lugar a la proteína madura. El primer exón forma la secuencia del péptido señal (SP) que dirige la proteína al espacio extracelular para su secreción en el lúmen de la glándula de Duvernoy. El segundo exón lo conforma el propéptido, que será escindido proteolíticamente. El tercer y la mitad del cuarto exón forman la subunidad de la disintegrina dimérica en sí y la otra mitad del cuarto exón y el último forman la parte final de la pre-pro-proteína y también es escindida.

4.2 Estructura proteica de la disintegrina

Se unieron las secuencias de aminoácidos de los diferentes exones y se reconstruyó la disintegrina como se observa en la figura 13.

M M Q V L L V T I C L A L F P Q G S S I L E S G N V
N D Y E I V Y P K K V A V L P T G A M N S A H P C C D
P V T C K P K Q G L H C I S G P C C R N C K F L N S G
T I C K K G R G D S M S D Y C I A S G I T P D C P R N
P N K G E A D E L Q W S A A A T G S V L

Figura 13: Estructura primaria de la disintegrina dimérica. Subrayado aparece el péptido señal; en turquesa el propéptido; en verde la parte activa de la disintegrina; en gris el tripéptido RGD; y en blanco el resto de la proteína

Si se analiza con más detalle la estructura primaria, se observa que la disintegrina está formada por 64 aminoácidos y contiene 10 cisteínas. Esto corrobora que se trata de una disintegrina dimérica (Calvete *et al.*, 2003). De las 10 cisteínas, 4 forman enlaces disulfuro intracatenarios y las otras dos cisteínas enlaces disulfuro intercatenarios, aunque a partir de la estructura primaria no podemos saber si esta proteína formará homo- o heterodímero.

Se observa el motivo RGD característico de la mayoría de las disintegrinas. Dicho tripéptido presenta especificidad de unión por integrinas cuyos ligandos naturales también incluyen RGD (Calvete *et al.*, 2003) , como $\alpha_8\beta_1$, cuyo ligando Tensacina C es una proteína que se encuentra en la matriz extracelular y que participa en procesos de inflamación y reparación (Midwood and Orend, 2009). Otra integrina RGD-dependiente es $\alpha_{IIb}\beta_3$, responsable de la agregación plaquetaria. En este caso el dominio RGD de la disintegrina bloquea estericamente la interacción entre la integrina y el fibrinógeno, como se esquematiza en la Figura 14.

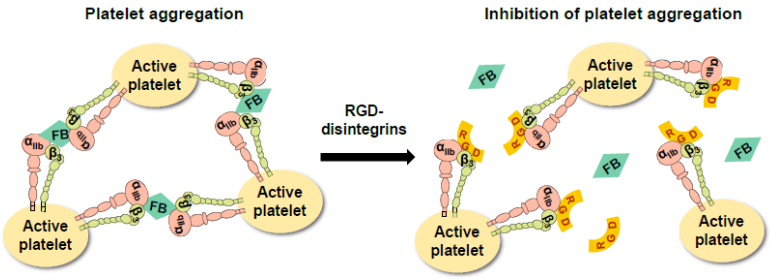

Figura 14: Inhibición de la agregación plaquetaria provocada por el motivo RGD de la disintegrina. FB corresponde al fibrinógeno(Calvete et al., 1994).

De hecho, se ha aprovechado esta característica para diseñar dos fármacos peptidomiméticos antagonistas de la integrina plaquetaria: el Tirofiban (Agrastat®) y Eptifibatide (Integrillin®). Ambos inhiben la agregación plaquetaria y se administran a pacientes que presentan un potencial riesgo de sufrir trombos (Marian *et al.*, 2017).

Las disintegrinas se encuentran ampliamente distribuidas entre las especies de *Viperinae* y *Crotalinae*. Hasta la fecha, se ha observado que las subunidades α de disintegrinas diméricas son codificadas por mARN cortos que no incluyen el dominio de metaloproteasa, presente en los genes de las subunidades β. Ello explica (Calvete *et al.*, 2005) que sea posible sintetizar homodímeros de subunidades α y heterodímeros $\alpha\beta$, pero no homodimeros $\beta\beta$ debido a problemas estéricos entre los dominios de metaloproteinasa.

Por último, es importante destacar también que dicha proteína no se ha encontrado en el proteoma del veneno de *Bitis arietans* ni tampoco en ningún acceso en la base de datos.

A pesar de que se conoce el genoma de cuatro especies de serpientes, en concreto, de la boa (*Boa constrictor*), de una pitón (*Python bivittatus*), de un vipérido (*Deinagkistrodon acutus*), y de un elápido (*Ophiophagus hannah)*, las regiones donde se encuentran las metaloproteinasas y las disintegrinas no están muy bien anotadas y descritas (Yin *et al.*, 2016) por lo que quedan interrogantes aún sobre el número de genes, su compartimentación, la regulación de la expresión o la estructura de los promotores. Tampoco se conocen bien los mecanismos moleculares de reclutamiento, transformación de proteínas ordinarias en toxinas y de la evolución acelerada de toxinas de familias multigénicas. Es importante pues, disponer de la secuencia completa del genoma de diferentes especies de serpientes y compararlas con otros genomas conocidos de especies filogenéticamente cercanas como *Anolis carolinensis* e incluso más alejadas como el ser humano o la gallina, para tratar de elucidar los interrogantes anteriormente citados o la historia evolutiva de dichos genes.

4.3 Metaloproteasa de la estrategia "Genome Walking"

La reacción de PCR que amplificó algún fragmento de tamaño significativo para secuenciar fue aquella en la que se usó el enzima de restricción DraI. El alineamiento por BLAST dio como resultado un fragmento de la primera parte de una secuencia de una metaloproteasa tipo I (figura 14) de *Echis ocellatus* con número de acceso EOC00006. Sólo se observó el péptido señal (muy conservado) y el primer intrón. No obstante, parece que también podría tratarse de una PII o una PIII puesto que no se obtuvieron secuencias discriminadoras de cada clase de metaloproteasa. Parece además plausible que el DNA genómico del que se partió estaba un poco degradado.

```
ATG ATG CAA GTT CTC TTA GTA ACT ATA TGC TTA GCA GTT TTT CCA TAT CAA G

 M   M   Q   V   L   L   V   T   I   C   L   A   V   F   P   Y   Q

gta aga tgt ttt ctt tgg ttc cct tgt tca gtt tct tac tgc taa aaa act att gca tcc aag
aga ttg tta tat tgt tgg ttt tgt ttt ggg ttt tat ttt tga gca aaa gaa tgt ctc aag gat
cac att tgt tct aaa gtg aca taa atg gtt tct tgg tat ttg ctt aaa ctt tgt caa gcc aca
aac aga tcc taa gag aga aaa gtc cta aga ttg tac att att cag cca aac aaa agc tag tgc
tat tgt tgt tac aga gaa tct tca acc tgt act cta gat aac agg aca cta act ttt ctc aat
gca aga ttc tga att atc tta aat tat tta ttc cct cct ttc att tta ctc tgc aga aaa cca
ggg agg tcc ctt ctg aaa ttc aag ttc tcc ttc ctt cag tgg gct gag ata ctt ttt aga tgc
cag ttg atc agt cta tgg ctc atc ttc ctg tct ccc aag tca ttc ccc cat ttc att aca aat
gat act caa ttg att cag aag ttc cca aac aca att aag ggt cag agg ctc act tca gta gcc
ctc ttt gtt gtt gtt ttt ttt tac cag ccc ggc ccg tcg acc acg cgtg ccc tat agt
```

Figura 15: Resultado de la secuenciación Sanger a partir de la obtención del fragmento por Genome Walking con el enzima de restricción DraI.

4.4 Resultados de secuenciación y análisis de los datos generados en el MinION

Los resultados obtenidos y las características de los mismos se exponen en la Tabla 2 y se obtuvieron a partir del procesamiento de los datos el MinKNOW mediante el programa estadístico R, por una parte, y los resultados del procesamiento de los datos.

Reads totales	**Pares de bases totales**	**Media**	**Mediana**	**Mínimo**	**Máximo**	**N50**
28712	175.6 Mb	6117.21	4151	29	806017	10231
Reads tras Trimm	**Reads>2000nt**	**Reads>5000nt**	**Reads>10000nt**	**Reads>20000nt**	**Reads>30000nt**	**Reads>40000nt**
28709	20099	12972	5815	857	165	49
No *E.coli*	**%recuperados**	**contigs**	**unitigs**			
22092	76,95	12	28			

Tabla 4: Resultados obtenidos tras secuenciación y tras el procesamiento de los datos.

Se obtuvieron lecturas muy largas en comparación con las secuenciación anteriores (tabla 1). En concreto, se obtuvo una lectura inusualmente larga, de unos 806 kb. En total se obtuvieron 175.6 megabases a partir de 28712 lecturas de 8 BACs con insertos de 300 kb. Al analizar los contigs y los unitigs, se prefirió anotar y buscar genes de metaloproteasas a partir de los contigs. Aunque el número de contigs sea menor, estos tienen mejor calidad y son más largos, y además porque muchos unitigs estaban ya incluidos en los contigs.

Se caracterizó y se anotó el contig 1 del total de 12 contigs. Como se trata de un proyecto que actualmente está en marcha, la secuencia de 273 kb de longitud no está disponible en ningún banco de datos. Se caracterizó a lo largo del contig 1 la cobertura de cada nucleótido como se muestra en la figura 15.

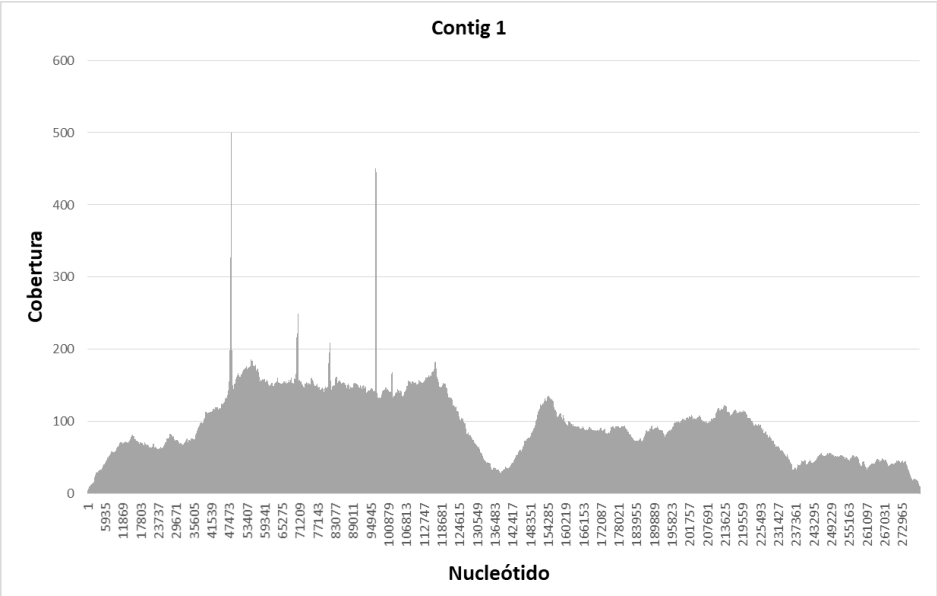

Figura 16: Gráfica donde se representa la cobertura de cada nucleótido en el contig1.

En este caso se observan zonas donde la cobertura cae drásticamente, ej. alrededor del nucleótido 136000, y luego vuelve a valores anteriores. Nuestra hipótesis es que esta caída de cobertura separa dos BACs diferentes, es decir, que el algoritmo de ensamblaje unió en un mismo contig lecturas pertenecientes a diferentes BACs.

Mediante alineamiento de los mRNA de metaloproteasas con el contig 1, se obtuvo el esquema mostrado en la Figura 16 que refleja la organización en tándem de los genes de metaloproteasas de tipo II y III en el contig 1. Las metaloproteasas tipo III tienen en su estructura 17 exones, mientras que las tipo II tienen 15, debido a la pérdida de los dos exones que codifican para la región rica en cisteína. Ambas tienen secuencias muy parecidas en sus últimos exones por lo que el exón 15 de la PII es muy similar al exón 17 de la PIII y se llamarán región Z (15Z y 17Z).

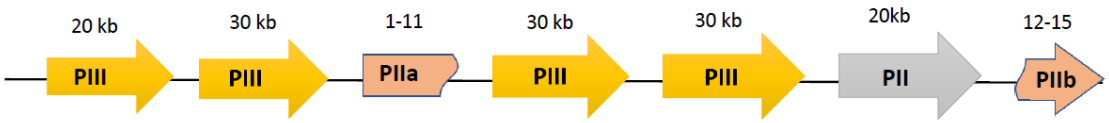

Figura 17: Metaloproteasas contenidas en el contig 1. Cada flecha corresponde a un gen de metaloproteasa diferente indicándose el tipo y el tamaño del mismo en la parte superior. En el gen partido PIIa y PIIb se indica en la parte de arriba qué exones hay en cada uno sin indicar tamaño

En la región 5' del contig 1 encontramos 4 genes de metaloproteasas tipo III y dos de tipo II, cuyo tamaño varía entre 20 y 30 kb. Uno de los genes tipo II está partido en dos fragmentos, uno que contiene los exones 1-11 y otro que contiene los exones 12-15, posiblemente como consecuencia de la inserción de un segmento "PIII-PIII-PII" (Figura 17).

Se confirma que los genes se distribuyen en agrupaciones (clusters) en tándem como se ha predicho en varios artículos (Sanz *et al.*, 2012). Este hecho apoya la idea de que las metaloproteasas evolucionaron por duplicación génica yuxtapuesta y neofuncionalización, como la PII que apareció a partir de la neofuncionalización de la PIII por pérdida de los dos exones anteriormente citados. Otro ejemplo que se observa y que puede apoyar dicha hipótesis de duplicación del gen e inserción en una región adyacente del genoma es el hecho de que haya una PII dividia por tres metaloproteasas. Esto pudo ocurrir cuando se duplicó una PIII y se insertó dentro de otra metaloproteasa. Luego este gen insertado se volvió a duplicar íntegro dos veces y uno de ellos sufrió una serie de mutaciones que la llevaron a convertirse en una PII.

Del análisis de cada gen se puede advertir que la longitud de los intrones entre los exones 14 y 15Z (en PII) y entre 14 y 17Z (en PIII), está conservado, con longitudes de 440 nucleótidos y 4-5 kb, respectivamente. Esto apoya la hipótesis de que las inserciones y deleciones de regiones intrónicas desempeñaron funciones clave a lo largo de la vía evolutiva que dio paso a la actual diversidad de loci de SVMP (Sanz and Calvete, 2016).

No obstante, sería muy conveniente, para consolidar estas hipótesis, analizar los 11 contigs restantes e interpretarlos como se ha hecho con el contig 1.

El método descrito (mapeo de exones en el contig) sirve para poder localizar y clasificar las metaloproteasas. No obstante, si se quisiera anotar con exactitud cada uno de los genes, hay que tener en cuenta que la secuenciación MinION lleva asociado un error del 5-10% por lo que al final no sería del todo fiable la secuencia obtenida y habría que combinarla con otra técnica

para aumentar su cobertura y así su fiabilidad. Por ejemplo, se podría aumentar la calidad de la secuencia si se combinasen las lecturas largas de los contigs del MiION con lecturas cortas de Illumina®. Sin embargo, en este trabajo final de grado no se llevó a cabo dicho abordaje, con lo que para futuras anotaciones sería muy recomendable hacerlo para así obtener resultados más exactos.

Siguiendo con el problema de la cobertura, no sería útil dirigir la plataforma MinION para, por ejemplo, la detección de SNPs o microsatélites en muestras humanas, donde se hace necesario conocer con exactitud la variante alélica que provoca una patología en concreto. Este es un reto al que se enfrenta la plataforma MinION y que tendrá que mejorar para poder ampliar sus aplicaciones.

En cuanto al precio, aunque es barato empezar con el MinION ($1000), puede resultar caro si se utiliza muchas veces para intentar elucidar un genoma, debido a la calidad de la secuencia, puesto que cada célula de flujo se puede utilizar sólo dos o tres veces y al final no resultaría rentable la relación calidad de la secuencia/precio, sobre todo si su objetivo es aplicarlo al diagnóstico en humanos.

4.5 Perspectivas

En este trabajo final de grado se han buscado genes de disintegrinas y metaloproteasas a través de dos métodos de secuenciación completamente diferentes: por la vía más clásica y antigua como es la secuenciación de primera generación Sanger y mediante la secuenciación de tercera generación MinION.

La secuenciación Sanger se puede aplicar para la búsqueda de genes concretos donde se necesita conocer con seguridad alguna variación alélica o SNP y cuyas regiones flanqueantes se conocen de antemano. Esto permite diseñar cebadores específicos para amplificar el fragmento por PCR y posteriormente secuenciarlo por Sanger.

Pero sin duda los secuenciadores de tercera generación están ganando en estos momentos, al poder secuenciar genomas completos en un tiempo récord y a un precio hasta hace unos años inimaginable.

Conocer el genoma completo de organismos cercanos y no tan cercanos filogenéticamente y poder compararlos permitirá entender con mayor profundidad la evolución de sus genes y genomas. Si a esto le añadimos estudios de transcriptómica y proteómica, se podrá conocer en detalle la relación entre dichos genes y sus respectivas proteínas.

En cuanto a los nuevos secuenciadores, la cantidad de datos producidos será el pilar central para avanzar en todos los interrogantes que todavía rodean al funcionamiento y regulación de los genes. También se convertirá en una herramienta insustituíble en biología evolutiva que permitirá conocer en última instancia el origen de la vida y los primeros genes y cómo la evolución de éstos dio lugar a la explosión de diversidad actual (Knapp and Hofreiter, 2010). La plataforma de secuenciación MinION deberá superar el inconveniente de la baja calidad de algunas secuencias así como mejorar la vida útil de las células de flujo y así aumentar significativamente su campo de aplicación, ya que actualmente sólo se usa para investigación (Jain *et al.*, 2016).

Finalmente, un problema importante al que se enfrenta la genómica y la secuenciación masiva en general es la inexistencia de computadoras con una capacidad de procesamiento equiparable a la astronómica cantidad de datos generados en cada secuenciación.

5. CONCLUSIONES

- La disintegrina dimérica secuenciada está formada por 64 aminoácidos y contiene 10 cisteínas que (por homología) forman 4 enlaces disulfuro intracatenarios y 2 intercatenarios. Presenta el tripéptido RGD que se ha sido mimetizado en el desarrollo de varios medicamentos.

- Los genes de metaloproteasas se agrupan en clusters genómicos y de manera yuxtapuesta, sugeriendo que evolucionaron por duplicación y posterior neofuncionalización por pérdida o reducción de regiones intrónicas. En este caso se anotó un contig de 273 Kb conteniendo 4 genes de metaloproteasas tipo III y dos de tipo II, cuyo tamaño varía entre 20 y 30 kb.

- La secuenciación MinION genera gran cantidad de lecturas, algunas de ellas de centenares de kilobases, circunstancia inédita en otros métodos de secuenciación. No obstante, hay que mejorar el problema de la cobertura para llevar sus aplicaciones más allá de la investigación.

6 BIBLIOGRAFÍA

Bainbridge, M.N., Warren, R.L., Hirst, M., Romanuik, T., Zeng, T., Go, A., Delaney, A., Griffith, M., Hickenbotham, M., Magrini, V., *et al.* (2006). Analysis of the prostate cancer cell line LNCaP transcriptome using a sequencing-by-synthesis approach. BMC Genomics *7*, 246.

Bazaa, A., Juárez, P., Marrakchi, N., Lasfer, Z.B., Ayeb, M.E., Harrison, R.A., Calvete, J.J., and Sanz, L. (2007). Loss of Introns Along the Evolutionary Diversification Pathway of Snake Venom Disintegrins Evidenced by Sequence Analysis of Genomic DNA from *Macrovipera lebetina* transmediterranea and *Echis ocellatus*. J. Mol. Evol. *64*, 261–271.

Branton, D., Deamer, D.W., Marziali, A., Bayley, H., Benner, S.A., Butler, T., Di Ventra, M., Garaj, S., Hibbs, A., Huang, X., *et al.* (2008). The potential and challenges of nanopore sequencing. Nat. Biotechnol. *26*, 1146–1153.

Calvete, J.J. (2010). Brief History and Molecular Determinants of Snake Venom Disintegrin Evolution. In Toxins and Hemostasis, R.M. Kini, K.J. Clemetson, F.S. Markland, M.A. McLane, and T. Morita, eds. (Springer Netherlands), pp. 285–300.

Calvete, J.J. (2017). Venomics: integrative venom proteomics and beyond. Biochem. J. *474*, 611–634.

Calvete, J.J., Mclane, M.A., Stewart, G.J., and Niewiarowski, S. (1994). Characterization of the Cross-Linking Site of Disintegrins Albolabrin, Bitistatin, Echistatin, and Eristostatin on Isolated Human Platelet Integrin GpIIb/IIIa. Biochem. Biophys. Res. Commun. *202*, 135–140.

Calvete, J.J., Moreno-Murciano, M.P., Theakston, R.D.G., Kisiel, D.G., and Marcinkiewicz, C. (2003). Snake venom disintegrins: novel dimeric disintegrins and structural diversification by disulphide bond engineering. Biochem. J. *372*, 725–734.

Calvete, J.J., Marcinkiewicz, C., Monleón, D., Esteve, V., Celda, B., Juárez, P., and Sanz, L. (2005). Snake venom disintegrins: evolution of structure and function. Toxicon *45*, 1063–1074.

Carbajo, R.J., Sanz, L., Perez, A., and Calvete, J.J. (2015). NMR structure of bitistatin – a missing piece in the evolutionary pathway of snake venom disintegrins. FEBS J. *282*, 341–360.

Casewell, N.R. (2012). On the ancestral recruitment of metalloproteinases into the venom of snakes. Toxicon Off. J. Int. Soc. Toxinology *60*, 449–454.

Deamer, D., Akeson, M., and Branton, D. (2016). Three decades of nanopore sequencing. Nat. Biotechnol. *34*, 518–524.

Feng, Y., Zhang, Y., Ying, C., Wang, D., and Du, C. (2015). Nanopore-based Fourth-generation DNA Sequencing Technology. Genomics Proteomics Bioinformatics *13*, 4–16.

Fox, J.W., and Serrano, S.M.T. (2005). Structural considerations of the snake venom metalloproteinases, key members of the M12 reprolysin family of metalloproteinases. Toxicon *45*, 969–985.

Fry, B.G. (1999). Structure-function properties of venom components from Australian elapids. Toxicon Off. J. Int. Soc. Toxinology *37*, 11–32.

Fry, B.G., Wüster, W., Kini, R.M., Brusic, V., Khan, A., Venkataraman, D., and Rooney, A.P. (2003). Molecular evolution and phylogeny of elapid snake venom three-finger toxins. J. Mol. Evol. *57*, 110–129.

Fry, B.G., Vidal, N., Norman, J.A., Vonk, F.J., Scheib, H., Ramjan, S.F.R., Kuruppu, S., Fung, K., Hedges, S.B., Richardson, M.K., *et al.* (2006). Early evolution of the venom system in lizards and snakes. Nature *439*, 584–588.

Fry, B.G., Casewell, N.R., Wüster, W., Vidal, N., Young, B., and Jackson, T.N.W. (2012). The structural and functional diversification of the Toxicofera reptile venom system. Toxicon *60*, 434–448.

Gibbs, H.L., and Rossiter, W. (2008). Rapid evolution by positive selection and gene gain and loss: PLA(2) venom genes in closely related Sistrurus rattlesnakes with divergent diets. J. Mol. Evol. *66*, 151–166.

Greene, H.W. (2000). Snakes: The Evolution of Mystery in Nature (University of California Press).

Gutiérrez, J.M., Lomonte, B., León, G., Alape-Girón, A., Flores-Díaz, M., Sanz, L., Angulo, Y., and Calvete, J.J. (2009). Snake venomics and antivenomics: Proteomic tools in the design and control of antivenoms for the treatment of snakebite envenoming. J. Proteomics *72*, 165–182.

Gutiérrez, J.M., Burnouf, T., Harrison, R.A., Calvete, J.J., Brown, N., Jensen, S.D., Warrell, D.A., Williams, D.J., and Initiative, G.S. (2015). A Call for Incorporating Social Research in the Global Struggle against Snakebite. PLoS Negl. Trop. Dis. *9*, e0003960.

Hargreaves, A.D., Swain, M.T., Logan, D.W., and Mulley, J.F. (2014). Testing the Toxicofera: Comparative transcriptomics casts doubt on the single, early evolution of the reptile venom system. Toxicon *92*, 140–156.

Harrison, R.A., and Gutiérrez, J.M. (2016). Priority Actions and Progress to Substantially and Sustainably Reduce the Mortality, Morbidity and Socioeconomic Burden of Tropical Snakebite. Toxins *8*, 351.

Harrison, R.A., Hargreaves, A., Wagstaff, S.C., Faragher, B., and Lalloo, D.G. (2009). Snake envenoming: a disease of poverty. PLoS Negl. Trop. Dis. *3*, e569.

Harvey, A.L. (2014). Toxins and drug discovery. Toxicon Off. J. Int. Soc. Toxinology *92*, 193–200.

Hynes, R.O. (2002). Integrins: bidirectional, allosteric signaling machines. Cell *110*, 673–687.

Initiative, for the G.S. (2014). A multicomponent strategy to improve the availability of antivenom for treating snakebite envenoming. Bull. World Health Organ. *92*, 526–532.

Jain, M., Olsen, H.E., Paten, B., and Akeson, M. (2016). The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. *17*, 239.

Kasturiratne, A., Wickremasinghe, A.R., de Silva, N., Gunawardena, N.K., Pathmeswaran, A., Premaratna, R., Savioli, L., Lalloo, D.G., and de Silva, H.J. (2008). The global burden of snakebite: a literature analysis and modelling based on regional estimates of envenoming and deaths. PLoS Med. *5*, e218.

King, G. (2015). Venoms to Drugs: Venom as a Source for the Development of Human Therapeutics (Royal Society of Chemistry).

Kini, R.M. (2006). Anticoagulant proteins from snake venoms: structure, function and mechanism. Biochem. J. *397*, 377–387.

Kini, R.M., and Evans, H.J. (1992). Structural domains in venom proteins: evidence that metalloproteinases and nonenzymatic platelet aggregation inhibitors (disintegrins) from snake venoms are derived by proteolysis from a common precursor. Toxicon Off. J. Int. Soc. Toxinology *30*, 265–293.

Knapp, M., and Hofreiter, M. (2010). Next Generation Sequencing of Ancient DNA: Requirements, Strategies and Perspectives. Genes *1*, 227–243.

Laszlo, A.H., Derrington, I.M., Ross, B.C., Brinkerhoff, H., Adey, A., Nova, I.C., Craig, J.M., Langford, K.W., Samson, J.M., Daza, R., *et al.* (2014). Decoding long nanopore sequencing reads of natural DNA. Nat. Biotechnol. *32*, 829–833.

Laszlo, A.H., Derrington, I.M., and Gundlach, J.H. (2016). MspA nanopore as a single-molecule tool: From sequencing to SPRNT. Methods San Diego Calif *105*, 75–89.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012). Comparison of next-generation sequencing systems. BioMed Res. Int. *2012*.

Lu, H., Giordano, F., and Ning, Z. (2016). Oxford Nanopore MinION Sequencing and Genome Assembly. Genomics Proteomics Bioinformatics *14*, 265–279.

Magi, A., Semeraro, R., Mingrino, A., Giusti, B., and D'Aurizio, R. Nanopore sequencing data analysis: state of the art, applications and challenges. Brief. Bioinform.

Manrao, E.A., Derrington, I.M., Pavlenok, M., Niederweis, M., and Gundlach, J.H. (2011). Nucleotide discrimination with DNA immobilized in the MspA nanopore. PloS One *6*, e25723.

Marian, M.J., Alli, O., Al Solaiman, F., Brott, B.C., Sasse, M., Leesar, T., Prabhu, S.D., and Leesar, M.A. (2017). Ticagrelor and Eptifibatide Bolus Versus Ticagrelor and Eptifibatide Bolus With 2-Hour Infusion in High-Risk Acute Coronary Syndromes Patients Undergoing Early Percutaneous Coronary Intervention. J. Am. Heart Assoc. *6*.

Markland, F.S., and Swenson, S. (2013). Snake venom metalloproteinases. Toxicon Off. J. Int. Soc. Toxinology *62*, 3–18.

Ménez, A. (2002). Perspectives in Molecular Toxinology (John Wiley & Sons).

Metzker, M.L. (2010). Sequencing technologies — the next generation. Nat. Rev. Genet. *11*, 31–46.

Midwood, K.S., and Orend, G. (2009). The role of tenascin-C in tissue injury and tumorigenesis. J. Cell Commun. Signal. *3*, 287–310.

Piskurek, O., and Okada, N. (2007). Poxviruses as possible vectors for horizontal transfer of retroposons from reptiles to mammals. Proc. Natl. Acad. Sci. *104*, 12046–12051.

Sanz, L., and Calvete, J.J. (2016). Insights into the Evolution of a Snake Venom Multi-Gene Family from the Genomic Organization of *Echis ocellatus* SVMP Genes. Toxins *8*.

Sanz-Soler, R. (2016). Molecular and functional approaches to understand the natural history of snake short disintegrins.

Seals, D.F., and Courtneidge, S.A. (2003). The ADAMs family of metalloproteases: multidomain proteins with multiple functions. Genes Dev. *17*, 7–30.

Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., *et al.* (2001). The Sequence of the Human Genome. Science *291*, 1304–1351.

Vidal, N., and Hedges, S.B. (2005). The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. C. R. Biol. *328*, 1000–1008.

Wierzbicka-Patynowski, I., Niewiarowski, S., Marcinkiewicz, C., Calvete, J.J., Marcinkiewicz, M.M., and McLane, M.A. (1999). Structural Requirements of Echistatin for the Recognition of αvβ3 and α5β1Integrins. J. Biol. Chem. *274*, 37809–37814.

Xiong, J.-P., Stehle, T., Zhang, R., Joachimiak, A., Frech, M., Goodman, S.L., and Arnaout, M.A. (2002). Crystal Structure of the Extracellular Segment of Integrin αVβ3 in Complex with an Arg-Gly-Asp Ligand. Science *296*, 151–155.

Yin, W., Wang, Z., Li, Q., Lian, J., Zhou, Y., Lu, B., Jin, L., Qiu, P., Zhang, P., Zhu, W., *et al.* (2016). Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. Nat. Commun. *7*.