



Universidad de Valladolid
Campus de Palencia

**ESCUELA TÉCNICA SUPERIOR
DE INGENIERIAS AGRARIAS**

Titulación
Grado en Enología

**Pasado y presente de la secuenciación
del genoma de la vid**

Alumna: Ana López Montes
Tutora: Elena Hidalgo Rodríguez

SEPTIEMBRE DE 2020

Índice

1. Resumen	2
2. Introducción	3
2.1. La importancia de la secuenciación	3
2.2. Evolución de las técnicas	4
3. Justificación	8
4. Objetivos	9
5. Metodología	10
5.1. Búsqueda y clasificación de la bibliografía	10
5.2. Redacción del texto	10
6. Análisis bibliográfico	10
6.1. Estudios genéticos en vid previos a su secuenciación	10
6.1.1. Mapas genéticos	10
6.1.2. Identificación genética de variedades	11
6.1.3. Estudios de identificación de genes y expresión génica de la vid	13
6.2. Secuenciación del genoma de la vid	16
6.2.1. El genoma de referencia	16
6.2.2. Estudios derivados de la secuenciación y su análisis	18
7. Conclusión	21
8. Bibliografía	22

1. Resumen

La vid tiene una gran importancia socioeconómica a nivel mundial, por lo que no es de extrañar que haya sido el primer genoma secuenciado dentro de los cultivos frutales. Esto ha sido posible gracias a los avances en las tecnologías de secuenciación que se han ido desarrollando desde finales de los años 80, las cuales han permitido iniciar la compleja tarea de la determinación y localización de genes de interés así como de fragmentos de ADN que pueden ser empleados como huella genética, logrando así la caracterización y diferenciación entre variedades y clones.

El estudio del ADN y de la expresión de los genes contenidos en él, resulta indispensable para conocer la base genética de todas las funciones de un organismo, así como para comprender tanto su evolución filogenética, como las relaciones entre las distintas especies y la interacción con el medio ambiente.

Además, permite integrar todo el conocimiento genético previamente acumulado, en particular todo el trabajo llevado a cabo con los mapas genéticos, cuyo desarrollo ha sido posible gracias al desarrollo y aplicación masiva de los distintos marcadores moleculares.

Al comparar las secuencias nucleotídicas de distintas variedades, ya publicadas, se observa una gran heterogeneidad, cuyas causas y cuya amplitud no han sido aún clarificadas.

Todo este conocimiento constituye la base de futuros proyectos de genética en la vid para desarrollar cepas con las características vitícolas y enológicas deseadas.

Abstract

Grapevine is an important crop worldwide and it has been one of the first plants to be sequenced. The evolution in the sequencing technologies since the 80s make this possible, and with them, genes of interest have been determined and localized along the genome, as well as DNA fragments that have been used as fingerprints, that has lead to establish the differences between varieties.

DNA encodes all the necessary functions for an organism, so its knowledge is essential for the study of its evolution as the relationship between species, as well as its response to the environment.

It also helps to collect all the previous studies, in particular those based on genetic maps which have been filled up with different molecular markers.

The comparison of grapevine genome sequences shows heterogeneity among them, and the reasons are not yet clear.

This knowledge will be needed to breed new grapevines with the desired characteristics through genetic improvement.

2. Introducción

2.1. La importancia de la secuenciación

La secuenciación consiste en determinar el orden de los nucleótidos que conforman la molécula de ADN, donde se almacena toda la información genética de los individuos. El objetivo final es disponer de toda la información genética inscrita en el genoma, lo que permitirá, una vez analizada en profundidad, identificar los genes presentes en el genoma, comparar las distintas versiones de los mismos (sus alelos) en diferentes especies, poblaciones o individuos y, comparando estos resultados con su efecto fenotípico en distintas circunstancias, comprender cada vez mejor la función que cumplen en el individuo objeto de estudio.

Las posibilidades que esto ofrece tanto en la medicina como en la agronomía se han traducido en un incremento espectacular de nuestro conocimiento genético de las especies. En cuestión de 35 años hemos pasado de conocer unos pocos genes responsables de algunos caracteres concretos, a tener acceso público a genomas completos de organismos de diversas categorías taxonómicas. Dado que los genes son los responsables de determinar las proteínas encargadas de llevar a cabo todas las funciones de las células, conocer la secuencia nucleotídica de estos permitirá reconocer los genes y las proteínas derivadas de ellos, así como las funciones que realizan. Además, existen genes con funciones similares en distintas especies, lo que ayuda a determinar su posición en el genoma, y a extrapolar la información obtenida en las especies mejor conocidas.

Los estudios de genomas completos permiten un importante avance en el establecimiento de las relaciones filogenéticas y una mejor comprensión de los fenómenos evolutivos. Por último, las nuevas técnicas basadas en la secuenciación, como la edición de genes, junto con las técnicas de transgénesis desarrolladas a finales del siglo XX permitirán la mejora de individuos que podrán modificados de forma muy específica para mejorar alguna de sus características sin afectar al resto de su comportamiento fisiológico.

En cuanto a la vid, el conocimiento de su genoma completo permite tener un genoma de referencia con el que poder comparar las distintas especies del género *Vitis* y las diferentes variedades y clones de *Vitis vinifera*, para conocer el comportamiento de las distintas versiones (alelos) de cada gen o región genómica.

Además, se ha demostrado que los genes suelen ocupar posiciones similares en distintos genomas (lo que se denomina sintenia) lo que facilita la búsqueda de genes concretos ya que estos se encuentran en zonas determinadas de los cromosomas. Tanto la secuenciación de diferentes especies como la resecuenciación de individuos de la misma especie, ayudan a una mejor comprensión de la función de los genes y permiten comparar su estructura y expresión, ya que según se ha comprobado, las diferentes especies comparten muchos genes y sus funciones concretas como se ilustra en la Figura 1. En la imagen se observa que casi 8000 genes de los 29585 encontrados en la vid en 2007, son comunes en cuatro especies vegetales distintas y más de 9000 aparecían en las otras especies secuenciadas (*Oryza sativa*, *Populus trichocarpa*, *Arabidopsis thaliana*).

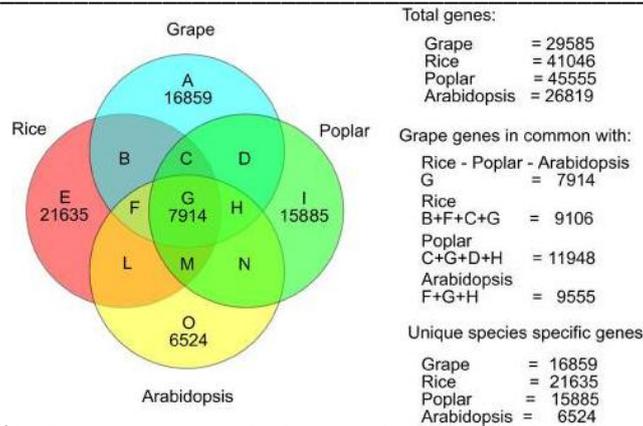


Figura 1: Comparación de genes por homología de cuatro genomas de plantas, considerando como homología secuencias de longitud parecida con más del 60% de similitud a nivel proteico. (Fuente: Velasco *et al.*, 2007)

La comparación de las secuencias nucleotídicas de variedades refleja las variaciones entre ellas. Las principales variaciones estructurales (SVs) en las plantas se deben a transposones o genes saltarines, que son secuencias de ADN que pueden moverse a lo largo del genoma, INDELs (inserciones/delecciones) y a reorganizaciones cromosómicas. El cotejo de las SVs ayuda a su vez a determinar zonas con posibles genes que deben analizarse en mayor profundidad.

Por último, y para una mejor comprensión del genoma y la función de las distintas secuencias, hay que estudiar la expresión de los genes y lo que deriva de ello, a través del estudio global de todo el ARNm sintetizado a partir de los genes activos del genoma en cada situación (transcriptoma) y de las proteínas (proteoma) y metabolitos (metaboloma) derivados de ellos.

También hay que tener en cuenta los factores epigenéticos, ya que pueden alterar la expresión de los genes. Estos son una serie de modificaciones heredables (metilaciones o acetilaciones) en las regiones reguladoras-cis que no afectan a la secuencia de nucleótidos, pero que influyen en la forma de unión del ADN a las histonas (proteínas en las que se ensambla la molécula de ADN) facilitando la expresión o la inactividad de los genes (Gayon, 2016).

2.2. Evolución de las técnicas

En 1944 Oswald Avery, Colin McLeod y Maclyn McCarty demostraron que la información genética reside en el ADN y no en las proteínas como se pensaba. Desde entonces se inició una carrera por descubrir su estructura. La estructura de doble hélice la propusieron en 1953 James Watson, Francis Crick, Rosalind Franklin y Maurice Wilkins. Con los trabajos de Severo Ochoa, Marshall W. Nirenberg, Har Gobind Khorana y Robert W. Holley se descubre en 1961 el código genético, en el que se determina que la secuencia de nucleótidos de los genes activos en cada momento se transcribe a una molécula de ARN que a su vez es traducida en una proteína. El estudio de las secuencias derivó en el "Principio de colinealidad", en el que se relaciona un triplete o codón con la síntesis de un aminoácido determinado. A partir de

entonces, muchas investigaciones se centran en conocer el orden de los nucleótidos (la secuencia) para así conocer la información genética.

En 1977 dos técnicas de secuenciación fueron desarrolladas de forma paralela. Por un lado, el método químico de Maxam y Gilbert, que consiste en fragmentar las cadenas de ADN, previamente marcadas mediante reacciones químicas específicas para cada base (purinas A-G y pirimidinas C-T) y después, separarlas por electroforesis en función de su tamaño en geles de alta resolución, típicamente de poliacrilamida (De Necochea *et al.*, 2004). Por otro lado, Frederick Sanger fue quien desarrolló el método que lleva su nombre, también denominado el método de terminadores de cadena. Este consiste en sintetizar una cadena complementaria a la que se usa de molde mediante la ADN polimerasa. En cada reacción, además de los cuatro nucleótidos, se añaden unos terminadores de cadena, que carecen del grupo hidroxilo necesario para elongar la copia de ADN. Cada vez que se incorpora un dideoxinucleótido (uno para cada base), estos impiden la continuación de la cadena y generan fragmentos de distintos tamaños que son separados por electroforesis. Este método se fue mejorando en los años siguientes para mejorar su eficacia y abaratar su precio, además de reducir el tiempo requerido para su desarrollo.

Algunos de los avances que permitieron la automatización del método de Sanger fueron (i) el desarrollo de la técnica PCR en 1989 por Mullis, que permite amplificar una pequeña muestra de ADN y que supuso incorporar el uso de la enzima Taq-polimerasa, una enzima termoestable que mantiene la actividad a altas temperaturas; (ii) el uso de sistemas de separación de moléculas más eficaz: la electroforesis capilar, que permite lecturas más largas y (iii) el uso de terminadores marcados con fluorescencia, lo que permite que las cuatro reacciones pudieran hacerse en un único tubo y no en cuatro como hasta entonces.

El primer secuenciador automático (ABI 370A) salió al mercado en 1986 a través de la empresa Applied Biosystems que permitía realizar lecturas de hasta 200 pb por hora y muestra, pudiendo cargar hasta 96 muestras (Hutchison, 2007; Pareek *et al.*, 2011).

Ya en las últimas décadas del siglo XX, los investigadores y las empresas biotecnológicas empezaron a estar interesados por las posibles aplicaciones de la secuenciación en medicina y se inició el Proyecto de secuenciación del Genoma Humano. En los años 90 se presentó el PGH (Proyecto del Genoma Humano), con una estimación en la duración de 15 años y un coste de 3 mil millones de dólares. Distintos centros en varios países trabajaban con secuenciadores capilares, método automatizado de Sanger, que permitía una lectura más rápida y de fragmentos de mayor tamaño (Kchouk *et al.*, 2017). De forma paralela, Craig Venter con su empresa Celera, comenzó a secuenciar el genoma humano mediante la técnica *Whole Genome Shotgun sequencing*, en la que el ADN es fragmentado, secuenciado y ensamblado mediante un programa informático a partir del alineamiento de *contigs*, que son secuencias de parte del genoma (Weissenbach, 2016). En 2001 se publicó el primer borrador del genoma humano ensamblando los datos obtenidos por Celera con una profundidad (número medio de veces que cada nucleótido ha sido secuenciado en distintas lecturas o *reads*) de 5X junto con los datos públicos, aumentándola en 3X.

Los avances tecnológicos dieron lugar a las técnicas de NGS (*Next Generation Sequencing*), también llamada secuenciación masiva paralela, que permite secuenciar miles de fragmentos a un precio muy bajo, aunque el tamaño de las lecturas es más pequeño que el que se consigue en las otras técnicas de secuenciación. Con la

tecnología de NGS se consigue una mayor profundidad, lo que aumenta el grado de fiabilidad de las secuencias (Rodríguez-Santiago *et al.*, 2012). Desde 2005 se desarrollaron nuevos secuenciadores basados en diferentes técnicas (Tabla 1). El primer secuenciador de segunda generación que estuvo en el mercado fue desarrollado por 454 Life Sciences y se basaba en la pirosecuenciación, en la que se detecta la liberación de pirofosfato cuando se agrega un nuevo nucleótido (Hutchison, 2007; Kchouk *et al.*, 2017). Por otro lado, Life Technologies sacó al mercado el método IonTorrent, que en vez de usar la fluorescencia como en la anterior, mide los cambios de pH que se producen en cada adición de nuevos nucleótidos. Estos métodos eran más eficaces que los anteriores pero seguían teniendo limitaciones para organismos complejos principalmente por el tamaño de las lecturas que podían producir, del orden de unos cientos de bases. Por ello se fueron mejorando hasta llegar a los que se conocen como los de tercera generación, que reducen el tiempo y no necesitan amplificaciones, además permiten realizar lecturas de miles de bases. La tercera generación también se denomina “Tecnología de secuenciación de molécula única” (Pareek *et al.*, 2011), en ella destaca el secuenciador SMRT (*single-molecule real-time*) de Pacific Biosciences por detectar en tiempo real la fluorescencia desprendida en el proceso. Las técnicas NGS tienen un porcentaje de error más alto que la técnica de Sanger por tener una menor precisión, pero esto se ve solventado por el número de lecturas que son capaces de realizar, la reducción en el tiempo de trabajo (ha pasado de días a horas), la mínima cantidad de muestra de ADN que requieren para el análisis así como la evolución de las herramientas informáticas (Kchouk *et al.*, 2017; López de Heredia, 2016), como el uso del programa BLAST (Basic Local Alignment Search Tool) para la comparación de nuevas secuencias con las guardadas en bases de datos (Hutchison, 2007).

Actualmente conviven todas estas técnicas de secuenciación, y sus variantes, y es frecuente que en cada proyecto de secuenciación se utilicen varias de ellas combinadas.

Tabla 1: Evolución de las técnicas de secuenciación. (Fuente: Elaboración propia a partir de Pareek *et al.*, 2011; Hutchison, 2007; De Necochea *et al.*, 2004 y Kchouk *et al.*, 2017)

Generación	Plataforma	Tamaño de lectura (pb)	Año
1ª	ABI PRISM	200-900	1986-2002
2ª	454	100-700	2005-2014
	Illumina	150-300	2011-2014
	SOLiD	75	2011-2013
	IonTorrent	200-400	2011-2015
3ª	PacBio	13000	2011-2016
	OxfordNanopore	9500-9850	2015-2016

Para la secuenciación de fragmentos grandes de ADN se pueden seguir dos estrategias. La primera es la denominada “*Chromosome walking*” (Figura 2), que consiste en fragmentar el ADN para introducir los fragmentos resultantes en un vector de clonación, crear un banco de moléculas e ir secuenciando cada fragmento que se vaya solapando hasta conseguir la secuencia completa. Esta estrategia fue la empleada en el PGH. Sin embargo, es una estrategia costosa y muy laboriosa, pues son necesarios muchos iniciadores dado que cada fragmento es único.

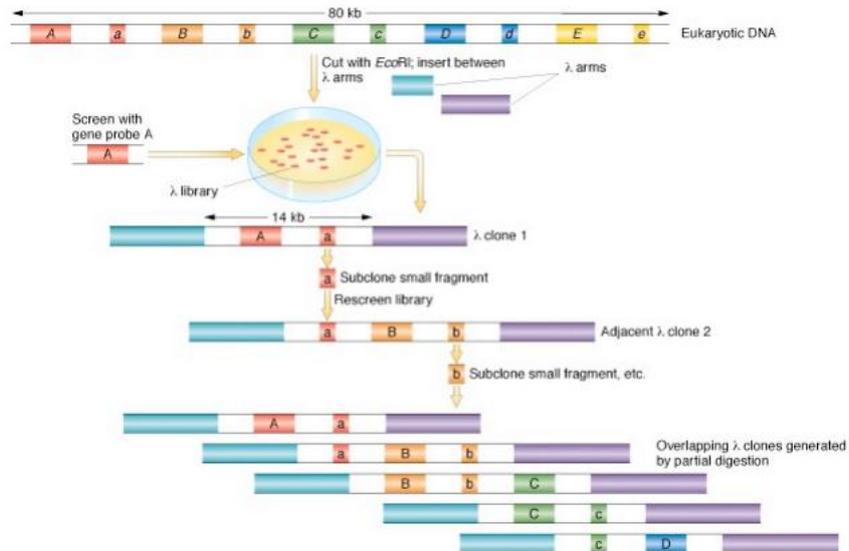


Figura 2: Ilustración sobre "Chromosome walking". (Fuente: De Necochea *et al.*, 2004)

La otra estrategia se llama "Shotgun", en ella el ADN es fragmentado al azar, y se ensambla mediante alguno de los programas disponibles en el mercado buscando coincidencias en las secuencias para reconstruir el genoma (Figura 3). Entre las ventajas frente a la otra se encuentran su rapidez y el menor número de iniciadores requeridos. Sin embargo, en este caso, es necesario secuenciar varias veces el ADN para lograr completar la secuencia y requiere un mayor esfuerzo bioinformático para ensamblar los fragmentos, ya que estos se secuencian sin seguir un orden determinado. Esta fue la técnica empleada por Venter para secuenciar el genoma humano de forma paralela al proyecto anunciado en los años 90 (De Necochea *et al.*, 2004).

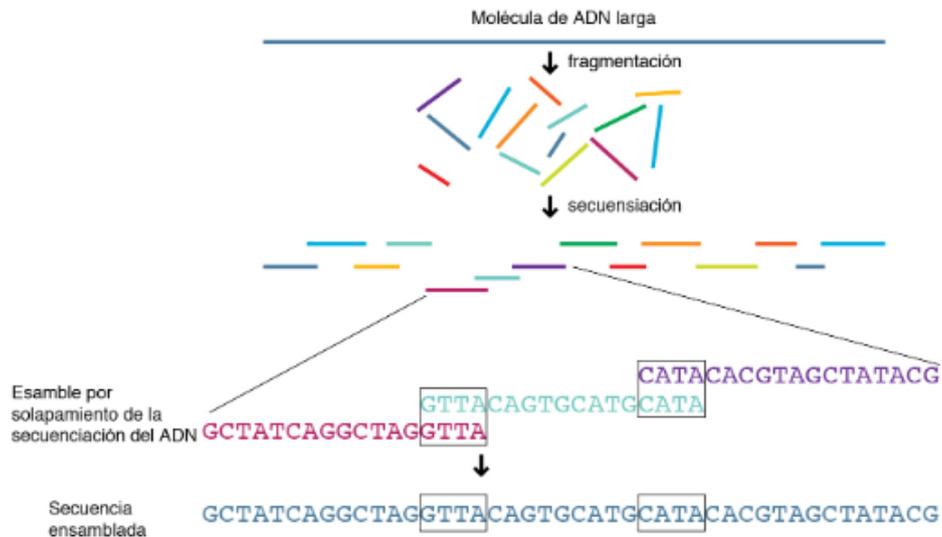


Figura 3: Ilustración de la estrategia Shotgun (Fuente: NHGRI)

El desarrollo de la genómica no hubiera sido posible sin el desarrollo paralelo de la bioinformática, imprescindible para almacenar, ensamblar y comparar la ingente información producida por los millones de lecturas actualmente conocidas y almacenadas en los bancos de información genética (Weissenbach, 2016). Tanto las mejoras en las técnicas de secuenciación como en los programas informáticos hicieron posible el estudio de un organismo completo, lo que ha propiciado la aparición de nuevas áreas de la genética, las ciencias “-ómicas”: genómica, proteómica, metabolómica, transcriptómica y epigenómica (Frigolet *et al.*, 2017), que permiten el análisis del conjunto del genoma, las proteínas, los metabolitos y las modificaciones epigenéticas de un organismo.

Los avances en las tecnologías de secuenciación han permitido pasar de la genética directa, es decir, buscar los genes que determinan una característica determinada, a la genética inversa, en la que el punto de partida es la secuencia genómica, en la cual se intenta identificar los genes presentes e investigar su función.

3. Justificación

La vid tiene una gran importancia socioeconómica a nivel mundial: el territorio destinado a viñedos es de 7,4 millones de hectáreas (Figura 4); la producción de vino, en 2018 ascendió a 292 millones de hectolitros; las exportaciones, en 2018, movieron un volumen aproximado de 108 millones de hectolitros, alcanzando un valor de 31,4 billones de euros (OIV, 2019).

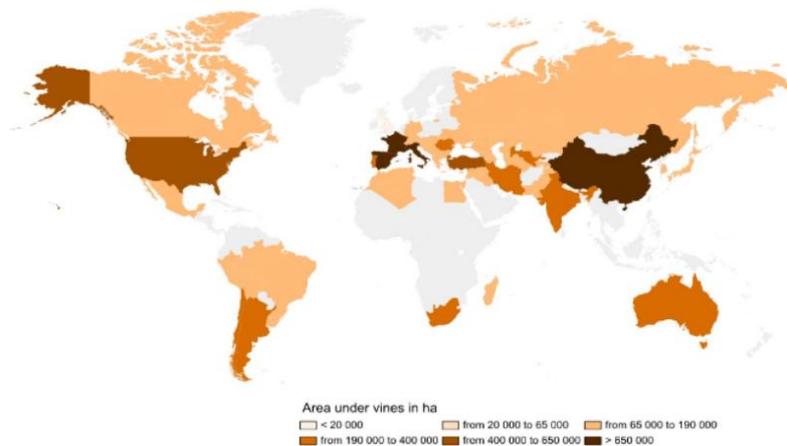


Figura 4: Superficie destinada a viñedo en ha a nivel mundial. (Fuente: Balance OIV, 2019)

Por consiguiente, no es extraño que haya sido el cuarto genoma secuenciado dentro de las plantas con flor, el segundo de las leñosas y el primero de los cultivos frutales (Jaillon *et al.*, 2007).

Desde finales de los 70 hasta día de hoy, se han ido estableciendo una notable cantidad de organizaciones y bases de datos con el objetivo de recopilar y compartir información acerca de la vid: desde descriptores morfológicos a genéticos para la caracterización de especies de vid. Así, en la 2ª edición de la “Lista de Descriptores OIV para Variedades de Vid y especies de *Vitis*” se añadieron como descriptores marcadores moleculares de tipo SSR; también se han realizado varios mapas

genéticos con mayor profundidad cada vez (Lodhi *et al.*, 1995; Riaz *et al.*, 2006; Marguerit *et al.*, 2009; Zhu *et al.*, 2018).

Finalmente, en 2007 se publicó el primer genoma de vid secuenciado (Jaillon *et al.*, 2007) que ha servido de referencia para numerosos trabajos posteriores centrados en la búsqueda de genes responsables de características de interés.

El contar con un genoma de referencia permite la comparación de los resultados obtenidos, lo que incrementa la información sobre la evolución y diferenciación entre especies.

Además, el estudio de la secuencia del genoma de la vid ayudará a mejorar su gestión agronómica, consiguiendo una mayor adaptabilidad al medio, explotando los genes involucrados en la resistencia a patógenos que tienen algunas especies y ayudando a reducir el uso de productos fitosanitarios, lo que reducirá su impacto ambiental (Meredith, 2001), tema a día de hoy que genera controversia, no sólo por la salud ambiental, también por la personal.

Los recientes estudios sobre la secuencia del genoma de la vid son una consecuencia lógica a la gran acumulación de información genética que se había generado en los años anteriores, y merecen una revisión bibliográfica que ayude a entender los avances más importantes en este ámbito.

4. Objetivos

Los objetivos de esta revisión bibliográfica son los que citan a continuación:

- a. Describir la secuenciación nucleotídica y dar a conocer la importancia de este conocimiento en el caso del genoma de la vid.
- b. Revisar cómo han evolucionado las técnicas de secuenciación, desde sus inicios hasta día de hoy.
- c. Recopilar y ordenar los primeros estudios genómicos, los conocimientos y las técnicas en las que se basaban y los principales resultados que se obtuvieron.
- d. Reflejar los resultados obtenidos en distintos trabajos que secuenciaron el genoma de la vid.
- e. Dar a conocer las posibilidades que ofrece la resecuenciación así como las aplicaciones de la ingeniería genética contando con la secuencia del genoma de la vid.

5. Metodología

5.1. Búsqueda y clasificación de la bibliografía

La búsqueda de información la he realizado en distintos portales. Por un lado, en algunos generales como Google y Google Académico para buscar estudios y tesis que recogieran las palabras clave y, por otro lado, en páginas concretas tales como ScienceDirect, Researchgate, Nature, NCBI, BMC y OIV.

Los artículos los he encontrado siguiendo dos estrategias. La primera, realizando una búsqueda con las palabras clave: “*Vitis genome*”, “*grapevine sequence*”, “*sequencing techniques*”, “*DNA sequencing*”. Para la segunda, he buscado artículos a partir de la bibliografía de estudios encontrados previamente, empleando aquellos con mayor concordancia con el tema de mi trabajo, cuyos autores aparecen de forma recurrente en artículos de viticultura y genética, así como los de las fechas más actuales.

Para ampliar información sobre algunos conceptos de la genética he recurrido también a libros de la biblioteca de la Universidad.

5.2. Redacción del texto

Para redactar este trabajo he seguido las directrices recogidas en el documento “Normas de Estilo y Formato de TFG Enología”, así como en el de “Recomendaciones e Índice TFGs Revisión Bibliográfica Grado Enología”, ambos de la página web de la Escuela Técnica Superior de Ingenierías Agrarias. También me he apoyado en el documento “Para empezar a entendernos” de Fernández *et al.* para la construcción de esta revisión bibliográfica.

6. Análisis bibliográfico

6.1. Estudios genéticos en vid previos a su secuenciación

6.1.1. Mapas genéticos

Según el NHGRI (National Human Genome Research Institute), un mapa genético “es un tipo de mapa cromosómico que muestra la ubicación relativa de los genes y otras características importantes. El mapa se basa en el concepto de ligamiento, el cual significa que cuanto más cerca estén dos genes en el cromosoma, mayor será la probabilidad de que se hereden juntos”

El primer mapa genético de la vid lo realizó Lodhi en 1995 a partir de una población de 60 ejemplares de “Cayuga White” x “Aurore”, desarrollada por Bruce I. Reisch en la Estación Experimental de Agricultura de Nueva York en 1981. Se construyó en base a 422 marcadores RAPD, 16 RFLP e isoenzimas. Se estimó un tamaño de 475M pb, resultando 1cM a 300 kb y con una distancia media entre marcadores de 6.1 cM (Lodhi *et al.*, 1995).

Posteriormente, alrededor del año 2000, diferentes grupos de investigación interesados en el estudio de la genética de la vid desarrollaron mapas con el objetivo

de localizar QTLs (*Quantitative Trait Loci*) de interés agronómico. Así, 19 grupos coordinados por Agrogene SA se coordinaron en un consorcio denominado *Vitis Microsatellite Consortium* (VMC). Este consorcio desarrolló 371 microsatélites, que fueron usados para la construcción de diversos mapas los primeros años del siglo XXI, así como 170 microsatélites desarrollados en Francia nombrados VVI. En un estudio de 2004 se lograron mapear 123 nuevos marcadores moleculares, con lo que se establecieron cuatro mapas a partir de las variedades “Syrah”, “Garnacha”, “Riesling” y otro mapa consenso de las variedades “Syrah” y “Garnacha” (Adam-Blondon *et al.*, 2004).

Previamente, desde los años 70 y hasta el primer mapa genético de Lodhi, se habían desarrollado numerosos trabajos sobre la vid buscando las bases genéticas de las características que resultaban más interesantes. Buena parte de estos trabajos incluían el desarrollo de marcadores moleculares, fragmentos de ADN repartidos por el genoma cuya distancia relativa puede ser determinada y que permiten construir mapas genéticos, que se han saturado (Figura 5). La disponibilidad de mapas permite la localización de zonas intergénicas que pueden usarse como huellas genéticas específicas de variedades y especies, lo que conlleva una mejora en la identificación y certificación varietal y clonal. Y, por otro lado, permiten localizar y caracterizar genes relacionados con caracteres específicos de alto interés vitícola o enológico (Vélez T, 2007).

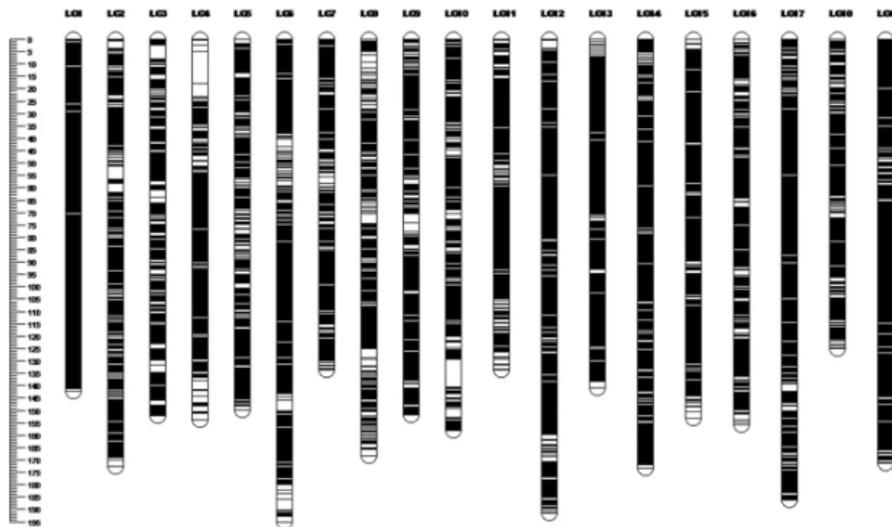


Figura 5: Mapa integrado que muestra el tamaño y distribución de marcadores moleculares (líneas negras) a lo largo de los 19 grupos de ligamiento. La distancia en la escala vertical viene dada en cM (centimorgans). (Fuente: Zhu *et al.*, 2018)

6.1.2. Identificación genética de variedades

En 1984, se estableció el VIVC (*Vitis International Variety Catalogue*), con el fin de recopilar la información repartida por el mundo sobre las diferentes variedades para conocer mejor los cultivares y controlar mejor la erosión genética (Maul E. *et al.*, 2015). En paralelo, y a la par que avanzaban las técnicas, se fueron realizando muchos trabajos con el objetivo de encontrar los responsables de ciertas características de la vid, entre ellas las asociadas con lo apirenía, el color y tamaño de las bayas, brotación, envero o maduración, así como las relaciones entre variedades.

Para identificar vides se han ido desarrollando distintos marcadores moleculares, como RFLP, RAPD, AFLP, microsatélites y SNPs, según se ampliaba la información genética de la vid, complementando a la ampelografía que durante tanto tiempo se empleó como única estrategia (This *et al.* 2004).

Desde finales de los 80 se empezaron a utilizar isoenzimas y RFLP (*Restriction Fragment Length Polymorphisms*) para la diferenciación. Las isoenzimas son indicadores directos del genotipo y están sujetos a cambios ambientales y a estados de desarrollo (Botta R. *et al.*, 1995). Sin embargo, hay descritas pocas isoenzimas que se puedan utilizar, por lo que su cobertura es baja y, aunque se mejoraron algunas etapas (sustituyendo los geles de almidón por geles de poliacrilamida, de mayor resolución), fueron rápidamente sustituidas por otros marcadores (Royo B.J. *et al.*, 1997). Por otro lado, los RFLPs son metodológicamente complejos y tienen un alto coste en reactivos y en tiempo, además requieren técnicos especializados, por lo que presentan grandes limitaciones (Narváez *et al.*, 2000; Vélez, 2007).

Con la invención de la PCR (*Polymerase Chain Reaction*) se desarrollaron gran número de nuevos marcadores moleculares que, además de permitir la caracterización de individuos, permitieron establecer los primeros estudios del conjunto del genoma de las especies, incluida la vid. Los distintos marcadores moleculares que se mencionan a continuación hicieron posible la saturación de los mapas genéticos.

El uso de RAPDS (*Random Amplified Polymorphic DNA*) para la identificación varietal fue muy utilizado en los 90, por el alto polimorfismo y el bajo coste (Moreno *et al.*, 1995). Los primeros resultados con RAPDS demostraron buena capacidad para la identificación pero cuentan con la desventaja de tener una baja reproducibilidad, lo que supone un problema a la hora de compartir datos entre laboratorios (Bowers *et al.*, 1996). Otro grupo de marcadores iniciales fueron los AFLP (*Amplified Fragment Length Polimorphism*) que presentan mayor reproducibilidad, por el contrario, conllevan una mayor complejidad técnica y resultan más lentos y más caros.

Los microsatélites o SSR (*Simple Sequence Repeat*) son reproducibles entre laboratorios y abundantes en el genoma, sobre todo los que tienen estructura de dinucleótidos (GA) y (GT), características por las que se empezaron a usar ampliamente. Por esta razón, resultan útiles para identificar el material vegetal tanto en viñedo como en bancos de germoplasma, así como para el estudio de pedigrí. Los microsatélites se componen de secuencias simples entre 1 y 6 nucleótidos repetidos, en tándem y a lo largo del genoma, un número muy variable de veces, lo que proporciona una fuente de *loci* polimórficos, que pueden usarse para la discriminación entre variedades (Sefc *et al.*, 2001). El cuello de botella para su uso es la necesidad de conocer las secuencias colindantes de las regiones microsatélites y desarrollar, a partir de ellas, *primers* específicos. Después de desarrollar varios cientos de microsatélites con valor de identificación, la OIV acabó por admitir seis sitios microsatélites para la identificación de todas las variedades de vid como metodología común para la comparación entre laboratorios de los resultados (Moreno *et al.*, 1995; Vélez, 2007). Actualmente los SNPs (*Single Nucleotide Polimorphism*) son los marcadores más empleados dada su abundancia y poder discriminatorio, así como su posible automatización en forma de chips de ADN.

Uno de los primeros estudios en los que se utilizaron los microsatélites en la identificación de variedades fue el llevado a cabo por la Universidad de California en Davis. En este trabajo se desarrollaron cuatro SSR y se utilizaron para el estudio de 77

cultivares de *Vitis vinifera*. Cada locus se nombró con las siglas “VVMD” y un número, haciendo referencia a “*Vitis vinifera* Microsatellite, Davis”. Todos los cultivares se clasificaron con esos cuatro marcadores a excepción de dos grupos, uno formado por uvas de vinificación, entre los que se encontraban “Pinot noir”, “Pinot blanc”, “Pinot gris” y “Meunier”, entendiéndose como variaciones de un mismo cultivar por mutaciones somáticas, y otro formado por variedades de uva de mesa, “Keshmesh” y “Thompson Seedles” que se consideraron sinónimos de un mismo cultivar, también llamados “Sultanina” y “Sultana” (Bowers *et al.*, 1996). Por otro lado, en un estudio apoyado por el Consiglio Nazionale delle Ricerche, no permitieron distinguir entre las variedades “Pinot Noir”, “Pinot blanc” y “Pinot gris”, ya que mostraron las mismas bandas (el mismo genotipo) en los cinco microsatélites estudiados, lo que reforzó la creencia de que las dos últimas surgieron por uno o varios genes relacionados con el color. También se encontraron diferencias en el locus VVS5 indicando que “Cabernet franc” y “Carménère” tienen relación, aun siendo cultivares diferentes, lo que apoyaba anteriores estudios basados en patrones de isoenzimas. Dentro del grupo “Refosco” se identificaron diferencias basadas en microsatélites, concluyendo que “Refosco di Faedis” y “Refoscone” son homonimias (Cipriani *et al.*, 1994).

Con la intención de coordinar esfuerzos, entre los años 1997 y 2002 se llevó a cabo el proyecto europeo GENRES CT96 No.81 (European Network for Grapevine Genetic Resources Conservation and Characterization), que reunió a multitud de ampelógrafos con el objetivo de recopilar, estandarizar e intercambiar información para crear una base de datos sobre descriptores apropiados sobre características morfológicas y agronómicas de la vid, así como el uso de microsatélites para la identificación entre variedades. Todo ello siguiendo las recomendaciones de “FAO/IPGRI Multi-crop Passport Descriptor”. Este inventario era necesario para comparar y actualizar los descriptores que se habían establecido años antes con la OIV (International Organisation of Vine and Wine), UPOV (Union for the Protection of New Varieties of Plants) e IBPGR (International Board for Plant Genetic Resources). Se seleccionaron como los marcadores más informativos VVS2, VVMD5, VVMD7, VVMD27, VrZAG62 y VrZAG79. Este proyecto se completó con otro proyecto europeo, el GrapeGen06, que se llevó a cabo entre los años 2007 y 2010 en el que a la lista de los 6 marcadores anteriores, se añadieron otros tres: VVMD25, VVMD28, VVMD32 (Maul *et al.*, 2008; Maul *et al.*, 2015).

Pronto se puso de manifiesto que, además de su valor en la identificación de variedades, los microsatélites (también llamados SSR: *Simple Sequences Repeats*) podían tener validez en estudios de parentesco. En un proyecto realizado en 2001 sobre el origen de la variedad “Cabernet Sauvignon” se confirmó mediante el estudio de microsatélites, que sus progenitores eran “Cabernet franc” y “Sauvignon blanc” (Sefc *et al.*, 2001). Por otro lado, en un estudio posterior sobre 115 individuos, se investigaron 12 loci, y aunque no sirvieron para relacionar variedades entre sí, permitió descartar parentescos (Galbács *et al.*, 2009).

6.1.3. Estudios de identificación de genes y expresión génica en vid

En 1999 se llevó a cabo una de las primeras identificaciones génicas en vid: se determinaron tres secuencias de ADNc (ARN copiado) codificantes de proteínas ligadas al transporte de sacarosa en bayas de la variedad Syrah 10 semanas tras la floración, coincidiendo con las secuencias encontradas en otras plantas. Las secuencias se incluyeron en las bases de datos DDBJ (DNA Data Bank of Japan), EMBL (European Molecular Biology Laboratory) y GenBank (Davies *et al.*, 1999). Esta

forma de identificar genes se basa en su expresión (analizando sus ARN mensajeros), que se denominaron ESTs (*Expressed Sequence Tag*) y han sido empleados, tanto para comparar genomas o descubrir genes como para generar una anotación en los mapas genéticos de las especies. Combinando estos ESTs se establecen librerías de ADNc que se utilizan para identificar genes cuya actividad está estrechamente relacionada con factores ambientales. Las bases de datos con estas secuencias se han ampliado enormemente, pasando de 400 en la base de datos de GenBank en 2001 a 146,075 en el NCBI (National Center for Biotechnology Information) en 2003 (Goes *et al.*, 2005). Actualmente existen más de 11700 (Figura 6) trabajos sobre la vid en la página, desde análisis filogenéticos a estudios más específicos de familias de genes o sobre genes concretos involucrados en distintas características y sus variaciones entre variedades.

NCBI Resources How To

PMC vitis sequencing

US National Library of Medicine National Institutes of Health

Create alert Journal List Advanced

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: <https://www.ncbi.nlm.nih.gov/sars-cov-2/>

Article attributes Associated Data Author manuscripts Digitized back issues MEDLINE journals Open access Preprints Retracted Text availability Include embargoed articles Publication date 1 year 5 years 10 years Custom range... Research Funder NIH AHRQ ACI

Display Settings: Summary, 20 per page, Sorted by Default order Send to:

Search results

Items: 1 to 20 of 11786

<< First < Prev Page 1 of 590 Next > Last >>

[A phylogenetic analysis of the grape genus \(*Vitis* L.\) reveals broad reticulation and concurrent diversification during neogene and quaternary climate change](#)
1. Yizhen Wan, Heidi R Schwaninger, Angela M Baldo, Joanne A Labate, Gan-Yuan Zhong, Charles J Simon
BMC Evol Biol. 2013; 13: 141. Published online 2013 Jul 5. doi: 10.1186/1471-2148-13-141
PMCID: PMC3750556
[Article](#) [PubReader](#) [PDF-2.5M](#) [Citation](#)

[Vitis Phylogenomics: Hybridization Intensities from a SNP Array Outperform Genotype Calls](#)
2. Allison J. Miller, Naim Matasci, Heidi Schwaninger, Mallikarjuna K. Aradhya, Bernard Prins, Gan-Yuan Zhong, Charles Simon, Edward S. Buckler, Sean Myles
PLoS One. 2013; 8(11): e78680. Published online 2013 Nov 13. doi: 10.1371/journal.pone.0078680

Figura 6: Imagen que muestra el número de trabajos sobre la secuenciación de la vid en el NCBI. (Fuente: NCBI)

En 2001 se constituyó el IGGP (International Grapevine Genome Program) con el objetivo de promover el estudio y conocimiento de la vid a nivel internacional. Este programa ha ido produciendo ESTs (*Expressed Sequence Tags*), y ha permitido desarrollar SNPs (*Single Nucleotide Polymorphisms*) que ayudarán a ampliar el conocimiento sobre el genoma de la vid dada su abundancia y la facilidad de su manejo, cuando se combina con la tecnología del “microarray” o microchips. La información depositada en bases de datos internacionales como el NCBI de acceso libre han facilitado el desarrollo de *primers* para amplificar las regiones de interés (This *et al.*, 2006).

El desarrollo de los SNP, abundantes en el genoma de la vid, ha permitido la saturación de los mapas y ha facilitado la localización de genes en los mismos. El Instituto Agrario di San Michele All’ Adige-Fundazione Edmund Mach trabajó para identificar y localizar un gran número de estos, volcando los datos en la página

<http://genomics.research.iasma.it/iasma/>. Por otro lado, utilizando la tecnología NGS o de secuenciación masiva, se identificaron muchos más SNP, cuya localización se puede visualizar en la base de datos Gramene (<https://www.gramene.org/>).

Durante muchos años los EST se han volcado en bases de datos como UniGene y se han usado para la caracterización de genomas, así como para la identificación de genes. Con el desarrollo de las ciencias –“ómicas”, así como a la secuencia completa del genoma, se integraron piezas de información de origen diverso, poniendo las bases para un mejor y más completo conocimiento de las bases genéticas del comportamiento de la vid.

Por ejemplo, en 2008 se construyó un mapa a partir de marcadores moleculares de tipo SSR, en el que se mapeó una región genómica involucrada en el color de las bayas, lo que coincidía con estudios anteriores. Previamente, en un estudio realizado en 2004 por Kobayashi se determinó que las variedades blancas contaban con una inserción del retrotransposón Gret1 en una región de VvmybA1, que anulaba el gen UFGT, relacionado con la síntesis de antocianinas (Salmaso *et al.*, 2008).

Otra línea de investigación fue la de buscar marcadores de tipo SSR, CAPS (*Cleaved Amplified Polymorphic Sequence*) y SCARs (*Sequence Characterized Amplified Regions*) con el fin de buscar genes relacionados con la resistencia a enfermedades. El estudio se realizó sobre 119 individuos derivados del cruzamiento entre “Regent” y VHR 3082-1-42, siendo esta última el resultado del cruzamiento entre *Muscadinia rotundifolia* y *Vitis vinifera*, que ha sido cruzada a su vez cuatro veces con *Vitis vinifera*, y que contiene los genes Run1, responsable de la resistencia al oídio y relacionado con el marcador CAPS GLP1-12 así como el gen Rpv-1, responsable de la resistencia al mildiu (ambos introducidos por *Muscadinia rotundifolia*). “Regent” por otro lado es una variedad desarrollada en Alemania en 1996 con resistencia a estas dos enfermedades. 20 de los individuos mostraron resistencia a ambas enfermedades y de estos, cuatro mostraron haber heredado los genes de ambos progenitores (Figura 7), conteniendo los alelos relacionados con la resistencia. El estudio concluyó que los marcadores VMC_4d9.2a, UDV_015b, VVlv67 y ScOR A7-760 están relacionados con la resistencia al oídio, mientras que los marcadores UDV_130, VMCNG2_f12 y UDV_108 con la resistencia al mildiu (Eibach *et al.*, 2007).

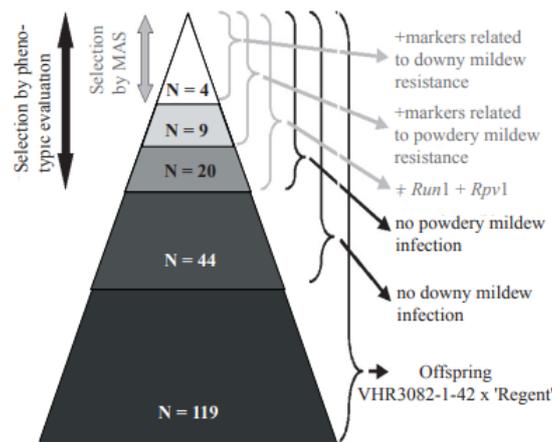


Figura 7: Estudio piramidal para seleccionar genes relacionados con la resistencia al oídio y al mildiu mediante la combinación de la evaluación fenotípica y MAS (*Marker Assisted Selection*) (Fuente: Eibach *et al.*, 2007)

En otro estudio, llevado a cabo en 2006, se mapeó el gen PdR1 involucrado en la resistencia a la enfermedad de Pierce, causado por la bacteria *Xylella fastidiosa*, en el grupo de ligamiento (cromosoma) 14. Se encontraron marcadores en común con distintos mapas realizados anteriormente. Entre estos, 116 coincidieron con el mapa elaborado con las variedades Riesling x Cabernet Sauvignon, 94 del mapa consenso de Riesling x Garnacha y 42 con el mapa de Riesling (Riaz *et al.*, 2006).

En 2009 realizaron un mapa de ligamiento a partir de 138 individuos derivados del cruzamiento entre Cabernet Sauvignon (*V. vinifera*) y Gloire de Montpellier (*V. riparia*) con el objetivo de localizar QTLs relacionados con la morfología de la flor y de la inflorescencia, así como la resistencia al mildiu mediante marcadores moleculares de tipo SSR y SSCP (*Single Strand Conformation Polymorphisms*). Este trabajo sirvió para apoyar la hipótesis que se manejaba entonces sobre el sexo de la flor, siendo esta que el sexo de la flor es un carácter monogénico, controlado por un *locus* con tres alelos. Varios QTLs relacionados con la morfología de la flor y la fecha de floración se encontraron en el cromosoma 2. También se concluyó, al analizar la secuencia de algunos *primers* que amplificaron distintos *loci*, que algunos marcadores empleados eran idénticos, estos fueron: VMC2G2/VMC2H9 en el cromosoma 6, VMC3D12/VVIC51 en el 13 y VVIN74/VVIP17 en el 19. Los marcadores están recogidos en la base de datos UniSTS del NCBI (Marguerit *et al.*, 2009).

Con la llegada de los métodos NGS se han podido desarrollar muchos más SNPs mediante la técnica RAD-seq (*Restriction.site associated DNA sequencing*), lo que permite construir mapas de mucha mejor calidad de manera simple y bajo coste. En un estudio realizado en 2014 se ligaron a cromosomas 65229 SNPs y 4832 INDELS con el que saturaron el mapa (Zhu *et al.*, 2018).

6.2. La secuenciación del genoma de la vid

6.2.1. El genoma de referencia

Conforme bajaban los precios de las tecnologías de secuenciación y se acumulaba información genética, se empezó a abordar la secuenciación de genomas completos de diferentes especies que sirvieran de modelo y así poder seguir extrayendo información a través de la extrapolación de las secuencias previamente secuenciadas en distintos organismos. En este caso concreto, el contar con mapas genéticos saturados como se ha abordado en el epígrafe anterior, facilitó la secuenciación del genoma de la vid.

El genoma completo de la vid fue secuenciado en el año 2007 por Jaillon *et al.* y es considerado el genoma de referencia. Fue realizado por un consorcio franco-italiano, usando un genotipo altamente homocigótico proveniente de sucesivas autofecundaciones de la variedad "Pinot noir", la cepa PN40024.

La versión de Jaillon *et al.* se realizó a partir de la estrategia *shotgun* y la tecnología Sanger con secuenciadores ABI3730xl, con lo que se consiguió una profundidad (media de las veces que se ha secuenciado cada nucleótido) de 8.4X y estimando un tamaño del genoma de 487Mb, muy ajustado al tamaño de 475 Mb estimado por Lodhi. En él se identificaron 30,434 genes codificantes de proteínas, un número menor que los 45,555 genes identificados en *Populus trichocarpa* que tiene un tamaño

similar (485 Mb), y que los 37,544 genes identificados en *Oryza sativa*, con un tamaño de 389 Mb. En este estudio se determinó que alrededor de un 41,4% del genoma está compuesto por elementos repetitivos y transponibles (TEs). En *Vitis vinifera*, al igual que en *Populus trichocarpa* están repartidos de forma heterogénea mientras que en *Arabidopsis thaliana* y *Oryza sativa*, los genes están repartidos de forma más homogénea.

Un primer análisis del proteoma señaló que existe una gran familia de genes relacionados con las características del vino. Se identificaron 43 genes involucrados en la codificación de la enzima estilbeno sintasa (STS), cuya actividad deriva, entre otras, a la síntesis de la fitoalexina conocida como resveratrol. De los 43 genes, alrededor de 20 se han visto expresados anteriormente tras la infección con *Plasmopara viticola*. También se encontraron 89 genes y 27 pseudogenes relacionados con la terpeno sintasa (TPS), casi el triple que en otras especies secuenciadas, como *Arabidopsis thaliana* u *Oryza sativa*. Su actividad depende de la relación planta-ambiente y es responsable de la producción de terpenos, abundantes en los aromas y aceites esenciales, relacionados con el perfil aromático del vino (Jaillon *et al.*, 2007).

De forma paralela, en 2007 Velasco *et al.* publicaron un estudio sobre la secuenciación del genoma del clon ENTAV 115 de "Pinot Noir". Se empleó el método WGS y la técnica de Sanger, que ofreció una profundidad de 6.5X, y que fue complementada para tramos entre *contigs* con SBS (*Sequencing By Synthesis*), permitiendo ampliar la cobertura 4.2X. Se estimó un tamaño del genoma de la vid de 504.6Mb, y se identificaron un total de 29,585 de genes, de los cuales el 96,1% fueron asignados a los distintos cromosomas.

Entre las conclusiones de este estudio, cabe destacar que la vid cuenta con numerosos genes relacionados con la respuesta frente a enfermedades: se encontraron 431 genes NBS (*Nucleotide Binding Site*), en comparación con los 207 de *Arabidopsis thaliana* y 398 de *Populus trichocarpa*. Estos interfieren en las proteínas R que están involucradas en la respuesta frente a patógenos. Otros genes involucrados en la defensa son: EDS1, PAD4, MPK4, JAR1, ETR1, NDR1, NPR1, RAR1 y EIN2, localizados en mayor parte en los cromosomas 12, 14, 15 y 18. También se identificaron 24 genes relacionados con los carotenoides, 24 con la síntesis del ABA (ácido abscísico) y 10 con las giberelinas (Velasco *et al.*, 2007).

El 6.7% del genoma corresponde a genes relacionados con los factores de transcripción, de los que el 80% se encontraban en *metacontigs* previamente señalados. Estos están distribuidos en 62 grupos, un número parecido a los encontrados en *Arabidopsis thaliana*, con 64, los 62 en *Oryza sativa* y los 63 en *Populus trichocarpa*. Su número y distribución parecen indicar que los factores de transcripción tienen un alto factor de conservación en los genomas de las plantas. En la vid, estos se localizaron en mayor proporción en los cromosomas 7 y 18. Uno de los grupos más abundantes es el MYB junto con el de MADS-box. El primero está relacionado con la acumulación de metabolitos en la baya, mientras que el segundo interviene en el desarrollo de las estructuras reproductivas y la floración (Velasco *et al.*, 2007).

6.2.2. Estudios derivados de la secuenciación y su análisis

Tras las primeras secuenciaciones del genoma completo de la vid, se llevaron a cabo diferentes consorcios cuyo objetivo ha sido reunir y ampliar la información con la comparación e integración de los resultados obtenidos de la secuenciación del genoma completo y los mapas. Esto se llevó a cabo mediante el uso de programas informáticos de predicción de genes como JIGSAW o EUGENE así como la construcción de mapas con otras variedades que se saturaron con marcadores SNP (Canaguier *et al.*, 2017).

Los principales objetivos del estudio de la secuenciación del genoma de la vid son determinar la función biológica de aquellos genes de interés relacionados con las variaciones fenotípicas dado que ellas son la base para la mejora genética de individuos. El contar con un genoma de referencia permite conocer casi por completo las secuencias de los genes, así como de sus variantes. Esto ha sido posible gracias a la secuenciación de una línea homocigótica con la que se pueden comparar los fragmentos de vides estudiadas, así como diseñar *primers* específicos para regiones concretas.

A través de las resecuenciaciones de distintas variedades de vid se ha comprobado que las diferencias entre cepas suponen en torno al 11%, lo cual es muy superior al 4% observado en *Arabidopsis thaliana*. Se necesita un mayor estudio sobre el pangenoma (conjunto de genes de una especie) de la vid para averiguar por qué existen tan grandes diferencias, especialmente, en relación con los desequilibrios de ligamiento que permitirán estudiar cómo ha sido la evolución de la vid, y también para entender la heterogeneidad de los genes repartidos por su genoma. Por otro lado, el hecho de que las distintas especies del género *Vitis* sean inter-fértiles hace posible la transmisión de características útiles para la mejora del material vegetal, por lo que es necesario un mayor conocimiento de la vid en sus distintos niveles de estudio. (Martínez-Zapater *et al.*, 2010).

Las nuevas técnicas de secuenciación masiva han permitido llevar a cabo la secuenciación del genoma completo de otras variedades de vid, como la variedad “Chardonnay” (clon FPS 04), que ha sido secuenciada mediante la tecnología SMRT de Pacific Biosciences. Este estudio arrojó un tamaño del genoma de 605 Mb, en el que se encontraron 37,244 genes codificantes de proteínas previamente señaladas y un 47.3% de TEs (Zhou *et al.*, 2019). Pero también han sido útiles para el estudio de la vid a diferentes niveles, dado que refleja la actividad de los genes durante el desarrollo de la vid, el transcriptoma (López, 2016). En un estudio realizado en 2010 se secuenció el ADNc de la variedad “Corvina” durante el desarrollo de las bayas: se produjeron más de 59 millones de lecturas entre 36 y 44 pb y se detectó la expresión de 17324 genes, 6695 de los cuales fueron específicos de una de las etapas de engrosamiento de la baya, reflejando así la complejidad durante los estados de desarrollo que tiene la vid. En este estudio se elaboraron tres librerías de ARNm, correspondiendo a muestras recolectadas en diferentes estados de desarrollo (5, 10 y 15 semanas tras la floración). Estas muestras se compararon con el genoma de referencia de “Pinot Noir” de Jaillon. Entre estas, el 82,4% coincidieron en localizaciones únicas (66,6%) o múltiples (15,8%) en el genoma. Posteriormente, con el algoritmo Python se asignaron a exones (71,5%) e intrones (16,1%) (Zenoni *et al.*, 2010).

En otro estudio realizado sobre la variedad “Tannat”, se encontraron 280 genes comunes con el clon ENTAV 115 de “Pinot Noir” que estudiaron Velasco *et al.* En este trabajo se realizó un análisis del transcriptoma en el que se consideraron 1873 genes específicos de la variedad al no ser compartidos con el genoma de referencia. También se identificaron 141 genes codificantes de enzimas relacionados con la producción de polifenoles (Da Silva *et al.*, 2013).

En 2014 se llevó a cabo la secuenciación *de novo* del genoma de la variedad “Sultanina”, en el que se descubrieron variaciones estructurales respecto al genoma de referencia. Estas variaciones corresponden en su mayoría a elementos transponibles, que a su vez conforman la mayor parte de las repeticiones en el genoma, a INDELS (inserciones/delecciones) y reorganizaciones cromosómicas durante la mitosis. El tamaño del genoma se estimó en 466 Mb, localizando el 82% de los genes presentes en el genoma secuenciado en 2007. Además, se identificaron 240 nuevos genes apoyándose en la base de datos NCBI EST. De estos, 130 correspondían con transposones, 88 con posibles genes, 13 asociados con la resistencia a patógenos y otros con la síntesis de antocianos. Se observó que 2000 genes estaban afectados por variantes estructurales (SVs), con lo que se concluyó que posiblemente estos genes estén relacionados con la apirenia. El 70% de INDELS y SNPs se encontraron en regiones intergénicas e intrones (secuencias que conforman los genes y no se traducen, al contrario que los exones). También se encontró que los SVs más abundantes fueron los compuestos por menos de 50 pb. El análisis de SVs se encontró una inserción de 15 pb en el locus VvAGL11 (Figura 8), del cromosoma 18 relacionado con la apirenia, que no estaba en el de referencia. Otros 13 genes se encontraron y que habían sido previamente mapeados en descendientes del cruzamiento “Ruby seedless” x “Sultanina”, lo que apunta a su relación con este carácter tan apreciado en las uvas de mesa (Di Genova *et al.*, 2014).

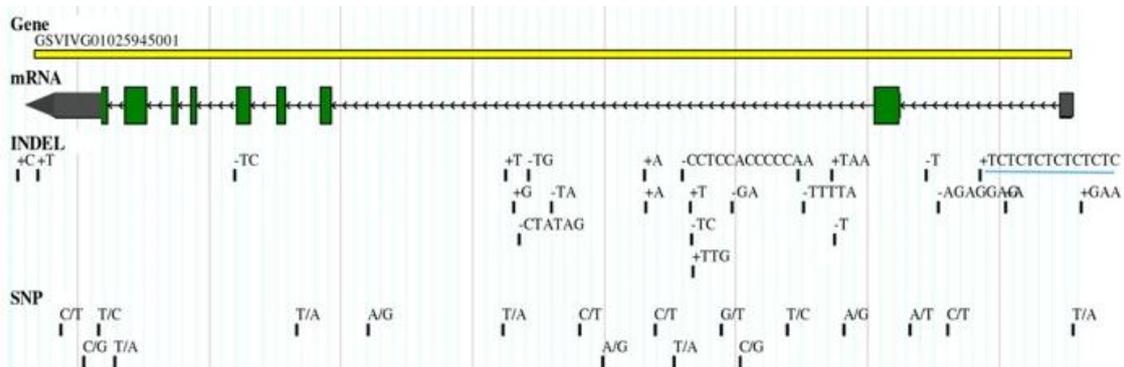


Figura 8: Estructura del gen GSVIVT01025945001 involucrado en la apirenia. Exones y UTRs se muestran en verde y gris respectivamente. Las barras en negro corresponden a los distintos SVs presentes en el genoma de “Sultanina”. El INDEL marcado con un “+” interrumpe una zona rica en GC conocida como una región reguladora-cis, responsable de la expresión del gen. (Fuente: Di Genova *et al.*, 2014).

En un trabajo se compararon cuatro variedades de mesa (“Autumn royal”, “Italia”, “Red globe” y “Thompson seedless”) con el genoma de referencia para el estudio de variaciones entre los genomas. El resultado fue que un 8% del genoma está afectado por variaciones a razón de 1 SNV (*Single Nucleotide Variants*) por cada 100 pb. En las cuatro variedades se observó la duplicación de algunas familias multigénicas como

MYB o TPS, que juegan un papel importante en la calidad de la uva. En concreto, la cantidad de CNVs en algunas familias como TPS pueden ser responsables de las diferencias en algunas rutas metabólicas, dando lugar a perfiles sensoriales diferentes entre variedades. También se encontró un alto grado de variabilidad en algunos genes relacionados con la respuesta frente al estrés como los NBS-LRR (Cardone *et al.*, 2016) así como que el gen MYB1 interactúa con otro gen (STS2) que activa la respuesta frente a patógenos (Yu *et al.*, 2019). Lo expuesto en este apartado, pone de manifiesto la alta plasticidad que tiene la vid.

En un estudio realizado con tres clones de la variedad “Nebbiolo” se detectaron 665,561 SNVs, un número en concordancia con los encontrados en la variedad “Sultanina” en el trabajo publicado en 2014. De los SNVs encontrados, el 95.6% resultaron comunes en los clones. De estos, 157,574 eran comunes a los localizados en regiones codificantes de proteínas en el genoma de referencia. Por otra parte, 5,458 genes no coincidieron con los establecidos en PN40024 ni con las variedades estudiadas “Tannat” o “Corvina”. En particular, dos marcadores fueron útiles para diferenciar un clon frente a los demás, lo que implica que el uso de SNVs podrían ser una herramienta eficaz a la hora de diferenciar entre clones (Gambino *et al.*, 2017).

Contar con un genoma de referencia permite el estudio de las secuencias que conforman las distintas familias de genes y sus variaciones, centrando futuros estudios en las funciones de los que son responsables. La resecuenciación de zonas determinadas en varios estudios apuntan a que dos cepas elegidas al azar podrían diferir entre 10 y 16 nucleótidos por cada mil, cifras hasta tres veces más que en *Arabidopsis thaliana* (Martínez-Zapater *et al.*, 2010).

La secuenciación no atañe únicamente a las unidades básicas que conforman el genoma, la secuenciación del ARN mediante RNA-seq permite conocer variaciones en las regiones codificantes, claves en las diferencias entre clones. Conocer las secuencias concretas que determinan ciertas variaciones genéticas en las vides es útil para proyectos de mejora genética ya sea mediante la transgénesis, la introducción de genes de interés en aquellos individuos que se quiera o la tecnología CRISPR/Cas (*Clustered Regularly Interspaced Short Palindromic Repeats/CRISPR-associated protein*), que aunque no ha sido la primera en desarrollarse, se ha impuesto en la edición de genes (Yang *et al.*, 2019). Esta es mucho más precisa ya que permite la modificación de bases en la secuencia de interés, además se puede aplicar también en marcas epigenéticas.

En 2016 se realizó un trabajo con el objetivo de crear una base de datos con secuencias altamente específicas para promover futuros estudios sobre las posibilidades de la modificación de ciertas regiones de interés mediante la tecnología CRISPR/Cas (Wang *et al.*, 2016).

7. Conclusión

El conocimiento sobre la genética de la vid ha ido estrechamente relacionado con la evolución en las distintas técnicas, que han revolucionado la forma de estudiar, identificar y diferenciar individuos de la misma especie así como relacionar funciones entre ellas. En este contexto, el desarrollo y uso de los marcadores moleculares, no sólo ha permitido saturar mapas ubicando regiones de interés y producir *primers* para secuenciar fragmentos concretos, sino que también se han empleado a modo de código de barras para identificar variedades.

Con la secuenciación del genoma la vid por primera vez en 2007, se obtuvo un genoma de referencia con el que poder comparar con otras variedades, facilitó el estudio de zonas de interés y resultó un apoyo con el que desarrollar iniciadores específicos. La secuenciación no sólo permite el estudio filogenético, sino que hace posible el estudio de la genética funcional, con cuyo estudio, a distintos niveles, se pueden identificar y estudiar los genes de interés y su expresión.

Conocer la secuencia nucleotídica del genoma de la vid ha propiciado y complementado el estudio de la expresión de los distintos genes en diferentes fases de crecimiento de la planta. Además, la comparación con otros genomas objeto de estudio, así como de regiones específicas ha permitido conocer algunos de los elementos responsables de la variación entre variedades.

El mayor conocimiento sobre el genoma de vid, junto con la experiencia adquirida durante años sobre ciertas técnicas de mejora genética, han permitido que tanto en la actualidad como en el futuro se puedan desarrollar proyectos de mejora genética con los que hacer frente a cambios climáticos y en el mercado.

8. Bibliografía

Adam-Blondon AF y Merdinoglu D. Mapping 245 SSR markers on the *Vitis vinifera* genome: A tool for grape genetics. *Theo Appl Gen.* 2004; 109(5): 1017-27.

NHGRI (*National Human Genome Research Institute*).

Botta R, Scott SN, Eynard I y Thomas RM. Evaluation of microsatellite sequence-tagged site markers for characterizing *Vitis vinifera* cultivars. *Vitis.* 1995; 34(2): 99-102.

Bowers JE, Dangl G S, Vignani R y Meredith PC. Isolation and characterization of new polymorphic simple sequence repeat loci in grape (*Vitis vinifera* L.). *Genome.* 1996; 39(4): 628-33.

Canaguier A, Grimplet J, Di Gaspero G, Scalabrin S, Duchêne E, Choisine N *et al.* A new versión of the grapevine reference genome assembly (12X. v2) and of its annotation (VCost. V3). *Genomics Data.* 2017; 14: 56-62.

Cardone M, D'Addabbo P, Alkan C, Bergamini C, Catacchio C, Anaclerio F *et al.* Inter-varietal structural variation in grapevine genomes. *The Plant Journal.* 2016; 88(4): 648-661.

Cipriani G, Frazza G, Peterlunger E y Testolin R. Grapevine fingerprinting using microsatellite repeats. *Vitis.* 1994; 33(4): 211-215.

Davies C, Wolf T y Robinsons P. S. Three putative sucrose transporters are differentially expressed in grapevine tissues. *Plant Science.* 1999; 147(2): 93-100.

Da Silva C, Zamperin G, Ferrarini A, Minio A, Dal Molin A, Venturini L *et al.*, The high polyphenol content of grapevine cultivar Tannat berries is conferred primarily by genes that are not shared with the reference genome. *Plant Cell.* 2013; 25(12): 4777-4788.

De Necochea R y Canul JC. Métodos fisicoquímicos en biotecnología: Secuenciación de ácidos nucleicos. Instituto de biotecnología-UNAM. 2004.

Di Genova A, Almeida MA, Muñoz-Espinosa C, Vizoso P, Travisany D., Moraga C *et al.* Whole genome comparison between table and wine grapes reveals a comprehensive catalog of structural variants. *BMC Plant Biology.* 2014; 14(1):7.

Eibach R, Zyprian E, Welter L, Töpfer R. The use of molecular markers for pyramiding resistance genes in grapevine breeding. *Vitis.* 2007; 46(3):120-124.

Frigolet EM, Gutiérrez-Aguilar R. Ciencias "ómicas", ¿cómo ayudan a las ciencias de la salud? UNAM. 2017; 18(7).

Galbács F, Molnár S, Halász G, Kozma P, Hoffman S, Kovács L *et al.* Identification of grapevine cultivar using microsatellite-based DNA barcodes. *Vitis.* 2009; 48(1): 17-24

Gambino G, Dal Molin A, Boccacci P, Minio A, Chitarra W, Avanzato G *et al.* Whole-genome sequencing and SNV genotyping of “Nebbiolo” (*Vitis vinifera* L.) clones. *Sci Rep.* 2017; 7: 17294

Gayon J. From Mendel to epigenetics: History of genetics. *Comp Ren Biol.* 2016; 339(7-8): 225-230.

Goes da Silva F, Iandolo A, Al-Kayal F, Bohlmann CM, Cushman M, Lim H *et al.* Characterizing the grape transcriptome. Analysis of Expressed Sequence Tags from multiple *Vitis* species and development of a compendium of gene expression during berry development. *Plant Physiol.* 2005; 139(2): 574-597.

Hutchison C. DNA sequencing: bench to bedside and beyond. *Nucleic Acids Res.* 2007; 35(18): 6227-6237.

Jaillon O, Aury J, Noel B *et al.* The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature.* 2007; 449: 463-467.

Kchouk M, Gibrat JF, Elloumi M. Generations of sequencing technologies: from first to next generation. *Biol Med.* 2017; 9(3).

Lodhi M, Daly MJ, Ye GN *et al.* A molecular marker based linkage map of *Vitis*. *Genome.* 1995; 38(4): 786-794.

López de Heredia U. Las técnicas de secuenciación masiva en el estudio de la diversidad biológica. *Munibe Cienc nat.* 2016; 64: 7-31.

Marquerit E, Boury C, Manicki A, Donnart M, Butterlin G, Némorin A *et al.* Genetic dissection of sex determinism, inflorescence morphology and downy mildew resistance in grapevine. *Theor Appl Genet.* 2009; 118: 1261-1278

Martínez-Zapater JM, Carmona MJ, Díaz-Riquelme J, Fernández L y Lijavetzky D. Grapevine genetics after the genome sequence: Challenges and limitations. *Australian Journal of Grape and Wine Research.* 2010; 16: 33-46.

Maul E, Dias EJ, Kaserer H, Lacombe T, Ortiz JM, Schneider A *et al.* Report of a working group on *Vitis*. First meeting, 12-14 June 2003, Palić, Serbia and Montenegro. *Biodiversity International.* 2008.

Maul E y Töpfer R. *Vitis International Variety Catalogue (VIVC): A cultivar database referenced by genetic profiles and morphology.* EDP Sciences. 2015. *BIO Web of Conferences* 5: 01009.

Meredith PC. Grapevine Genetics: Probing the past and facing the future. *Agriculturae Conspectus Scientificus.* 2001; 66(1): 21-25.

Moreno S, Gorgocena Y y Ortiz MJ. The use of RAPD markers for identification of cultivated grapevine (*Vitis vinifera* L.). *Scientia Horticulturae.* 1995; 62: 237-243.

Narváez RC, Valenzuela BJ, Muñoz SC y Hinrichsen RP. Comparison of RAPD and AFLP as methods for genetics identification on *Vitis* based on the analysis of anonymous genomic sequences. *Agric. Téc.* 2000; 60(4).

OIV. Balance 2019 de la OIV sobre la situación vitivinícola mundial. 2019; 2ª edición de la Lista de Descriptores OIV para Variedades de Vid y especies de *Vitis*. 2009.

Pareek SC, Smoczynski R y Tretyn A. Sequencing technologies and genome sequencing. *J Appl Genet.* 2011; 52(4): 413-435.

Riaz S, Krivanek FA, Xu K y Walker AM. Refined mapping of the Pierce's disease resistance locus, PdR1, and Sex on an extended genetic map of *Vitis rupestris* x *V. arizonica*. *Theor Appl Genet.* 2006; 113(7):1317-29.

Rodríguez-Santiago B y Armengol L. Tecnologías de secuenciación de nueva generación en diagnóstico genético pre- y postnatal. 2012; 23(2):56-66.

Royo JB, Cabello F, Miranda S, Gogorcena Y, González J, Moreno S *et al.* The use of isoenzymes in characterization of grapevines (*Vitis vinifera* L.). Influence of the environment and time of sampling. *Scientia Horticulturae.* 1997; 69(3-4): 145-155.

Salmaso M, Malacarne G, Troggio M, Faes G, Stefanini M, Grando MS *et al.* A grapevine (*Vitis vinifera* L.) genetic map integrating the position of 139 expressed genes. *Theor Appl Gen.* 2008; 116(8): 1129-43.

Sefc KM, Pejic I, Maletic E, Thomas MR y Lefort F. Microsatellite markers for grapevine: a state of the art. *Researchgate.* 2001

This P, Jung A, Boccacci P, Borrego J, Botta R, Constantini L *et al.* Development of a standard set of microsatellite reference alleles for identification of grape cultivars. *Theor Appl Genet.* 2004; 109(7): 1448-1458.

This P, Lacombe T y Thomas RM. Historical origins and genetic diversity of wine grapes. *Trends in Genetics.* 2006; 22(9): 511-519.

Velasco R, Zharkikh A, Troggio M, Cartwright A. D, Cestaro A, Pruss D *et al.* A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PloS One.* 2007; 2(12): e1326.

Vélez TM. Estudio de un sistema de marcadores microsatélites para la protección y defensa legal de variedades de vid (*Vitis vinifera* L.) [tesis doctoral]. Madrid: Departamento de biología celular y genética, Universidad de Alcalá; 2007.

Wang Y, Liu X, Ren C, Zhong G, Li S y Liang Z. Identification of genomic sites for CRISPR/Cas9-based genome editing in the *Vitis vinifera* genome. *BMC Plant Biol.* 2016; 16: 96

Weissenbach J. The rise of genomics. *Comptes Rendus Biologies.* 2016; 339(7-8): 231-239.

Yang G y Huang X. Methods and applications of CRISPR/Cas system for genome editing in stem cells. *Cell Regen (Lond)*. 2019; 8(2): 33-41.

Yu Y, Guo D, Li G, Yang Y, Zhang G, Li S *et al*. The grapevine R2R3-type MYB transcription factor VdMYB1 positively regulates defense responses by activating stilbene synthase gene 2 (VdSTS2). *BMC Plant Biol*. 2019; 19: 478.

Zenoni A, Ferrarini A, Giacomelli E, Xumerle L, Fasoli M, Malerba G *et al*. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. *Plant Physiol*. 2010; 152(4): 1787-1795.

Zhou Y, Minio A, Massonnet M, Solares E, Lv Y, Beridze T *et al*. The population genetics of structural variants in grapevine domestication. *Nature*. 2019; 5:965-979

Zhu J, Guo Y, Su K, Liu Z, Ren Z, Li K *et al*. Construction of a highly saturated genetic map for *Vitis* by Next-generation restriction site-associated DNA sequencing. *BMC Plant Biol*. 2018; 18:347.