



Ainoa Gallego Zaragoza

Secuenciación, ensamblaje *de novo*
y anotación del genoma
del cloroplasto del ajo (*Allium sativum*)

Trabajo tutorizado por: Dr. Héctor Candela Antón
Departamento de Biología Aplicada, Área de Genética

Grado en Biotecnología
Facultad de Ciencias Experimentales
Universidad Miguel Hernández de Elche
Curso 2016-2017

ÍNDICE DE MATERIAS

1. Resumen	4
2. Introducción.....	5
2.1. El ajo cultivado	5
2.1.1. Importancia del cultivo del ajo	5
2.1.2. Herramientas moleculares para la mejora genética del ajo	6
2.2. El cloroplasto.....	6
2.2.1. Función del cloroplasto en la fotosíntesis	8
2.2.2. Teoría endosimbiótica del origen de los cloroplastos	8
2.2.3. El genoma del cloroplasto	9
3. Antecedentes y Objetivos	12
4. Materiales y Métodos	13
4.1. Cultivo en maceta de <i>Allium sativum</i>	13
4.2. Purificación de ADN genómico.....	13
4.3. Comprobación de la cantidad y la calidad del ADNg	14
4.4. Análisis bioinformático.....	14
4.4.1. Ensamblaje del genoma.....	14
4.4.2. Anotación del genoma.....	15
4.4.3. Análisis filogenético.....	15
5. Resultados	17
5.1. Ensamblaje <i>de novo</i> del genoma del cloroplasto del ajo	17
5.2. Anotación del genoma.....	19
5.2.1. Genes necesarios para la expresión génica del genoma del cloroplasto	20
5.2.2. Genes implicados en la fotosíntesis.....	24
5.2.3. Otros genes.....	25
5.3. Análisis filogenético	25
6. Discusión	30
7. Conclusiones y proyección futura	33
8. Bibliografía.....	34

ÍNDICE DE FIGURAS

Figura 1.- Mapa del genoma completo del cloroplasto de <i>Allium sativum</i>	19
Figura 2.- Estructura secundaria de cuatro moléculas de ARNt del cloroplasto de <i>Allium sativum</i> para los aminoácidos metionina, formilmetionina e isoleucina.....	24
Figura 3.- Historia evolutiva de 14 especies de plantas inferida mediante el método de máxima parsimonia.....	26
Figura 4.- Historia evolutiva de 14 especies de plantas inferida mediante el método de unión al vecino.....	27
Figura 5.- Comparación de las uniones entre las repeticiones invertidas (IRa, IRb) y las secuencias de copia única (SSC, LSC) en 11 especies de plantas	28

ÍNDICE DE TABLAS

Tabla 1.- Comparación del tamaño del genoma del cloroplasto en diversas plantas.....	16
Tabla 2.- Frecuencias nucleotídicas del genoma del cloroplasto del ajo.....	18
Tabla 3.- Genes identificados en la secuencia del genoma del cloroplasto del ajo.....	21
Tabla 4.- Anticodones presentes en los ARNt del genoma del cloroplasto del ajo.	23



1. Resumen

En este trabajo, hemos llevado a cabo la secuenciación, el ensamblaje *de novo* y la anotación del genoma completo del cloroplasto del ajo (*Allium sativum* L.). La secuenciación fue realizada con un equipo HiSeq 2500 de Illumina, empleando lecturas emparejadas de 101 nucleótidos de longitud. Hemos realizado el ensamblaje *de novo* de las secuencias obtenidas utilizando el programa Velvet. El genoma del cloroplasto del ajo es una molécula de ADN bicatenario circular, con un tamaño total de 153.131 pares de bases (pb). Está constituido por dos repeticiones invertidas (IRa e IRb; *inverted repeats*), de 26.492 pb cada una, que separan la región de copia única pequeña (SSC; *small single copy*), de 18.042 pb, de la región de copia única grande (LSC; *large single copy*), de 82.105 pb. Hemos determinado que el genoma contiene 136 genes funcionales (incluyendo 90 genes que codifican proteínas, 38 genes de ARN de transferencia y 8 ARN ribosómicos) y 6 pseudogenes. Por último, hemos llevado a cabo un análisis filogenético con las secuencias de aminoácidos de 17 proteínas conservadas (todas ellas codificadas en el genoma del cloroplasto) de 14 especies de plantas monocotiledóneas y dicotiledóneas, para determinar la historia evolutiva de *Allium sativum*.

Palabras clave: Ajo, *Allium sativum*, genoma cloroplástico, análisis filogenético.

In this work, we report the sequencing, *de novo* assembly and annotation of the complete chloroplast genome of garlic (*Allium sativum* L.). The genome was sequenced using the Illumina HiSeq 2500 next-generation sequencing platform, using 101-bp paired-end reads. The sequence was assembled using Velvet software. The chloroplast genome of garlic is a circular, double-stranded DNA molecule with a total length of 153,131 base pairs (bp), which consists of two inverted repeats (IRa and IRb; 26,492 bp each) that are located between a small single copy region (SSC; 18,042 bp) and a large single copy region (LSC; 82,105 bp). The annotation uncovered 136 functional genes (including 90 protein-coding genes, 38 transfer RNA genes, and 8 ribosomal RNA genes) and 6 pseudogenes. Finally, we performed a phylogenetic analysis with the amino acid sequences of 17 conserved, plastid-encoded proteins from 14 different plant species, including monocots and dicots, in order to investigate the evolutionary history of *Allium sativum*.

Keywords: Garlic, *Allium sativum*, chloroplast genome, phylogenetic analysis.

2. Introducción

2.1. El ajo cultivado

El ajo cultivado (*Allium sativum*) es una planta monocotiledónea de la familia Alliaceae, dentro del orden Asparagales. Es una especie originaria de Asia Central que se cultiva mundialmente en zonas de clima templado. El género *Allium* comprende más de 700 especies (Mes, 1999). En las plantas de este género, como la cebolla, la parte basal de las hojas vegetativas se diferencia para formar un bulbo. En el ajo, los meristemos axilares de estas hojas dan lugar a nuevos bulbos (los denominados “dientes”), que quedan recubiertos por las hojas senescentes blanquecinas (la denominada “túnica”). Las plantas producen una inflorescencia (escapo floral) en cuyas flores raramente se desarrollan semillas. En la parte superior de esta inflorescencia se desarrollan frecuentemente los denominados bulbilos (*bulbils*), propágulos que, como los dientes, permiten la propagación vegetativa de la planta. La propagación vegetativa del ajo tiene como consecuencia negativa la acumulación de patógenos, por lo que los productores recurren habitualmente al cultivo *in vitro* de meristemos para su saneamiento y la obtención de plantas libres de los mismos.

2.1.1. Importancia del cultivo del ajo

El principal productor mundial de ajo es China (20.082.000 toneladas en 2012; según se describe en http://www.odepa.cl/wp-content/files_mf/1408548688Ajo201408.pdf), seguido de lejos por la India y Egipto. España es el noveno país productor a nivel mundial, con una producción de 151.900 toneladas en 2012. La mayor parte de esta producción se centra en Las Pedroñeras (Cuenca), donde se tiene la indicación geográfica protegida "Ajo Morado de Las Pedroñeras".

El ajo tiene gran importancia en la gastronomía mediterránea, siendo ampliamente usado como condimento en la cocina de diferentes países. Los compuestos azufrados del ajo son los responsables de su olor y sabor característicos. Uno de estos compuestos es la alicina, que se produce por la actividad de la enzima alinasa en respuesta a daños en el tejido (Borlinghaus *et al.*, 2014). La alicina es una molécula biológicamente activa con actividad antimicrobiana (Benkeblia, 2003; Li *et al.*, 2015). Por estas propiedades antimicrobianas, la industria alimentaria ha utilizado extractos de *A. sativum* para prolongar la conservación de productos cárnicos, lo que ilustra el potencial de esta especie en el desarrollo de productos innovadores (Sung *et al.*, 2014).

Otras especies del género *Allium* poseen gran importancia económica, ya sea porque se cultivan como alimento (como la cebolla, *Allium cepa*) o con propósitos ornamentales.

2.1.2. Herramientas moleculares para la mejora genética del ajo

Los programas de mejora genética del ajo están limitados por su modo de reproducción vegetativa, ya que la mayor parte de las variedades ha perdido la capacidad de reproducirse por medio de semillas (reproducción sexual). Debido a la imposibilidad de realizar cruzamientos, la obtención de nuevas variedades de ajo se ha realizado habitualmente mediante la selección clonal de variantes aparecidas espontáneamente (Peña-Iglesias, 1988).

El genoma del ajo tiene un tamaño cinco veces superior al humano, alcanzando las 15,9 gigabases (Kamenetsky *et al.*, 2015). Si bien este genoma aún no ha sido secuenciado por completo, algunos autores han logrado generar marcadores moleculares de diferentes tipos (microsatélites, AFLPs y otros) (Chand *et al.*, 2015; Ipek *et al.*, 2006; Buso *et al.*, 2008; Cunha *et al.*, 2012; García-Lampasona *et al.*, 2012; Jo *et al.*, 2012). Estos marcadores han sido empleados con fines diversos, como el estudio de las relaciones filogenéticas entre especies del género *Allium*, la caracterización de la diversidad intraespecífica del ajo y la cebolla, o la elaboración de mapas de ligamiento genético de la especie (Ipek *et al.*, 2004; Zewdie *et al.*, 2005). En estos mapas, el número de grupos de ligamiento se corresponde con el número haploide de cromosomas de la especie ($2n=16$). En ausencia de la secuencia completa del genoma, en los últimos años se han realizado algunos estudios que, alternativamente, han permitido obtener información sobre el transcriptoma (la parte expresada del genoma) de diversos tejidos, tales como los meristemos y otros órganos (Kim *et al.*, 2009; Sun *et al.*, 2012; Sun *et al.*, 2013; Kamenetsky *et al.*, 2015; Shemesh-Mayer *et al.*, 2015). A los esfuerzos para caracterizar el contenido en genes del ajo, se suman también otros trabajos realizados para identificar genes en especies filogenéticamente próximas, como la cebolla (*Allium cepa*; Kim *et al.*, 2015a), el puerro (*Allium fistulosum*; Tsukazaki *et al.*, 2015) y el cebollino chino (*Allium tuberosum*; Zhou *et al.*, 2015).

2.2. El cloroplasto

Los cloroplastos son orgánulos presentes en las células vegetales, que contienen la maquinaria necesaria para la fotosíntesis (Sugiura, 1992). En los cloroplastos se lleva a cabo la biosíntesis de almidón, ciertos aminoácidos, lípidos, vitaminas y pigmentos. Asimismo, en los cloroplastos residen varias rutas principales del metabolismo del azufre y del nitrógeno (Yang *et al.*, 2010; Bausher, 2006; Sugiura, 1992).

Los cloroplastos y otros tipos de plastos se diferencian a partir de proplastidios, indiferenciados y de menor tamaño, presentes en las células en división de los meristemos. Durante la diferenciación celular, los proplastidios dan lugar a diferentes tipos de plastos según el tipo celular en que se encuentren (Pyke, 1999). La diferenciación más estudiada es la que conduce de los proplastidios a los cloroplastos en las células fotosintéticas del

mesófilo foliar. Los proplastidios que no han sido expuestos a la luz se encuentran en forma de etioplastos, que carecen de pigmentos activos y son por tanto incoloros.

El tamaño, forma y número de los cloroplastos varía entre especies. Su tamaño habitual es de 2 a 6 μm de diámetro y de 5 a 10 μm de longitud. En las plantas superiores suelen ser ovoides, pero poseen diferente morfología en algunas algas, como *Spirogyra* (cuyos cloroplastos son helicoidales) o *Chlamydomonas* (en forma de copa). Aunque se han descrito células con un único cloroplasto, la mayoría de las células vegetales contienen entre 40 y 200 cloroplastos. Las células con mayor número de cloroplastos se encuentran en las hojas, habiéndose estimado que hay aproximadamente 500.000 cloroplastos por milímetro cuadrado de superficie foliar. Dentro de las células, los cloroplastos están sometidos a movimientos de ciclosis por las corrientes citoplasmáticas y presentan a su vez movimientos activos relacionados con la iluminación, como contracciones y deformaciones (Ohiwa, 1977).

A diferencia de otros orgánulos celulares, los cloroplastos presentan una doble membrana ya que, como propone la teoría endosimbiótica, derivan de una cianobacteria que fue fagocitada por una célula eucariota ancestral (Weeden, 1981). El espacio interno del cloroplasto se denomina estroma, donde se acumulan sustancias como gránulos de almidón y lípidos. El estroma es el compartimento donde reside el genoma de ADN circular bicatenario y tienen lugar la transcripción y la traducción de los genes del mismo. La síntesis de proteínas es llevada a cabo por ribosomas 70 S (denominados a veces plastorribosomas), similares a los ribosomas 70 S de las bacterias.

La membrana externa de los cloroplastos es permeable a iones y otras moléculas, mientras el paso de metabolitos a través de la membrana interna está mediado por proteínas transportadoras. Ambas membranas son muy permeables al CO_2 , que se utiliza en el estroma como sustrato de la fotosíntesis.

Los tilacoides (del griego *thylakos*, saco o bolsa) forman un sistema de membranas en el interior de los cloroplastos, en cuyo estroma se apilan formando los denominados grana. Normalmente, un cloroplasto contiene de 10 a 100 grana, en los que se acumulan diferentes sustancias como pigmentos fotosintéticos (clorofila, carotenoides, xantófilas), lípidos, proteínas y complejos enzimáticos, como el de la ATP sintasa.

En la estructura de los cloroplastos también se pueden encontrar plastoglóbulos, rodeados de una membrana similar a la de los tilacoides, en los que se acumulan moléculas orgánicas y lípidos (Austin *et al.*, 2006).

2.2.1. Función del cloroplasto en la fotosíntesis

Durante la fotosíntesis, las plantas son capaces de producir glucosa a partir de dióxido de carbono (CO₂) y agua, liberando oxígeno O₂ a la atmósfera. La fotosíntesis se divide en dos fases: la fase luminosa y la fase oscura.

La fase luminosa se lleva a cabo en las membranas tilacoidales, que contienen los pigmentos fotosintéticos (clorofilas). Las clorofilas *a* y *b* absorben energía luminosa y poseen espectros de absorción que se complementan. Cuando la luz incide, se llevan a cabo una serie de reacciones en las que la energía lumínica se transforma en energía química, almacenada en forma de ATP y NADPH.

En los tilacoides, los pigmentos están unidos a proteínas y empaquetados formando dos fotosistemas, el fotosistema I y el fotosistema II. Las membranas tilacoidales contienen proteínas que forman parte de los fotosistemas, la cadena fotosintética de transporte electrónico y el complejo de ATP sintasa. Este último tiene una estructura y función semejante al de la mitocondria y produce ATP mediante fotofosforilación.

La fase oscura tiene lugar en el estroma, donde residen las enzimas encargadas de la fijación del CO₂ atmosférico, y consta de un conjunto de reacciones que no requieren la presencia de luz. En esta fase, el ATP y el NADPH producidos en la fase luminosa se utilizan para reducir el CO₂ a glúcidos simples. La más abundante de las enzimas del estroma es la ribulosa 1,5-bisfosfato carboxilasa-oxigenasa (RuBisCO), una enzima del ciclo de Calvin que representa aproximadamente la mitad de las proteínas del cloroplasto.

2.2.2. Teoría endosimbiótica del origen de los cloroplastos

La similitud de plastos y mitocondrias con ciertas bacterias de vida libre se explica mediante la teoría endosimbiótica, que proporciona una explicación generalmente aceptada para el origen de estos orgánulos (Zimorski *et al.*, 2014). La teoría endosimbiótica sostiene que los plastos derivan de una cianobacteria fotosintética ancestral de vida libre que fue capturada por una célula eucariótica no fotosintética, estableciendo una relación de endosimbiosis (Sagan, 1967; Margulis, 1975; Howe *et al.*, 2003). Esta endosimbiosis primaria permitió que estos eucariotas pasaran de ser heterótrofos a autótrofos gracias a su nueva capacidad de aprovechar la energía solar (Wicke *et al.*, 2011). La teoría endosimbiótica explica la presencia de una doble membrana en los cloroplastos como consecuencia de la fagocitosis: la membrana externa derivaría de la vesícula que envolvió originalmente a la cianobacteria.

Numerosas características de los cloroplastos actuales son coherentes con el origen de los mismos a partir de una cianobacteria ancestral. Entre estas características cabe destacar: (1) la presencia de un genoma circular de ADN bicatenario circular, que contiene genes muy similares a los de las cianobacterias; (2) la división de los cloroplastos mediante

un mecanismo de fisión binaria (a diferencia de los eucariotas que se dividen por mitosis); (3) la localización de sus centros de obtención de energía en las membranas, como los procariotas; o (4) la ya mencionada capacidad de sintetizar sus propias proteínas con ribosomas 70 S, semejantes a los bacterianos (a diferencia de los ribosomas 80 S de los eucariotas).

Sin embargo, el genoma del cloroplasto contiene un número de genes (~150; Leister, 2003) muy inferior al de las cianobacterias actuales (2.000-3.000 genes; Martin *et al.*, 2002). Esta diferencia pone de manifiesto una dependencia creciente del cloroplasto hacia el hospedador eucariótico, a cuyo núcleo se han transferido la inmensa mayoría de los genes del genoma organular (Curtis *et al.*, 1984; Whatley, 1981; Weeden, 1981). La transferencia horizontal de secuencias desde el genoma de los plastos al genoma nuclear es un proceso evolutivo que aún continúa (Stern *et al.*, 1986; Nakazono *et al.*, 1993; Sheppard *et al.*, 2009; Guo *et al.*, 2008; Matsuo *et al.*, 2005; Wicke *et al.*, 2011). Por ello, la mayoría de las proteínas que desempeñan su función en este orgánulo son codificadas por genes del genoma nuclear, sintetizadas en el citoplasma y transportadas posteriormente al cloroplasto (Timmis *et al.*, 2004; Wicke *et al.*, 2011). La mayor parte de los genes que se han mantenido en el genoma del cloroplasto desempeñan funciones relacionadas con la expresión de sus genes y con procesos fotometabólicos.

2.2.3. El genoma del cloroplasto

Como se describe en el apartado anterior, los cloroplastos contienen su propio material genético independiente del material genético nuclear celular, que se replica de forma semiautónoma cuando la célula se divide (Palmer, 1991). Se ha estimado que cada célula vegetal contiene entre 400 y 1.600 copias del genoma cloroplástico (Yang *et al.*, 2010).

El genoma del cloroplasto se encuentra altamente conservado (Palmer, 1991). La estructura del genoma cloroplástico (tamaño, contenido y orden de los genes) evoluciona mucho más lentamente que el genoma nuclear (Tanksley y Pichersky, 1988) e incluso el mitocondrial, que varía mucho (Palmer, 1990). Pese a que hay especies con estructuras únicas debido a adaptaciones a condiciones específicas (Wicke *et al.*, 2011), el genoma del cloroplasto de las plantas terrestres es el genoma eucariota más conservado evolutivamente. Esta alta conservación evolutiva podría deberse a la organización de sus genes en operones, como ocurre en los genomas de las cianobacterias y los cloroplastos de algas verdes y plantas terrestres (Wicke *et al.*, 2011). Paradójicamente, las tasas de sustitución nucleotídicas son en promedio varias veces mayores en el genoma del cloroplasto que en el mitocondrial (Palmer, 1990).

En las angiospermas, el genoma del cloroplasto es una molécula de ADN circular bicatenario de entre 120 y 160 kb, formada por cuatro regiones claramente delimitadas (Asano *et al.*, 2004): una región grande de copia única (LSC, del inglés *large single copy*; 80-90 kb), una región pequeña de copia única (SSC, *small single copy*; 16-27 kb) y dos repeticiones invertidas idénticas (IRa e IRb, *inverted repeats*; 20-28 kb). La variación en el tamaño del genoma del cloroplasto de distintas especies se debe, a menudo, a contracciones y expansiones de las repeticiones invertidas (Yang *et al.*, 2010; Do *et al.*, 2013). El genoma del cloroplasto del tabaco (*Nicotiana tabacum*) fue el primero en ser completamente secuenciado (Shinozaki *et al.*, 1986), por lo que ha servido de referencia con la que se han comparado otros genomas cloroplásticos secuenciados posteriormente. Los genomas de los cloroplastos suelen contener unos 150 genes, que corresponden a aproximadamente 80 proteínas, 30 ARN de transferencia (ARNt) y 4 ARN ribosómicos (ARNr) (Palmer, 1991; Chumley *et al.*, 2006; Yang *et al.*, 2010). Algunos genes que codifican funciones relacionadas con la fotosíntesis y la expresión del genoma se han perdido en el genoma de los cloroplastos de plantas parásitas, que ya no los necesitan por obtener sus nutrientes necesarios de su hospedador (De Pamphilis y Palmer, 1990; Wolfe *et al.*, 1992; Funk *et al.*, 2007).

2.2.3.1. Secuenciación y anotación de genomas de plastos

En la actualidad, la base de datos *Organelle Genome Resources* del *National Center for Biotechnology Information* (NCBI; Do *et al.*, 2013) contiene información sobre más de 250 genomas de cloroplastos completamente secuenciados. Un gran número de estos genomas han sido secuenciados rápidamente gracias a las nuevas tecnologías de secuenciación masivamente paralelas (*massively parallel sequencing technologies*), que presentan un alto rendimiento y suponen un importante ahorro de tiempo y dinero (Yang *et al.*, 2010). El avance paralelo de los métodos bioinformáticos y moleculares ha permitido interpretar y extender la investigación genómica a plantas no modelo y con genomas de gran tamaño, como el ajo, que sin embargo poseen un gran interés para la agricultura. La aplicación de nuevas herramientas moleculares a plantas cultivadas permitirá acelerar los programas de mejora genética.

Se denomina anotación de un genoma a la tarea de identificar y asignar funciones a los distintos elementos presentes en la secuencia genética de un organismo (Stein, 2001). La anotación puede realizarse de forma automática o manual. Si bien la anotación manual es más laboriosa, es a menudo más precisa y produce menos errores que la automática (Rouze *et al.*, 1999). Los métodos de anotación basados en homología se basan en la comparación de las secuencias de distintas especies, con el fin de inferir sus correspondientes funciones. Estos métodos asumen que secuencias muy similares deben

desempeñar las mismas funciones en especies distintas, ya que derivan de una secuencia ancestral común y han sido sometidas a los mismos procesos evolutivos (Koonin, 2001). Los métodos *ab initio*, sin embargo, intentan identificar los genes a partir de características de la propia secuencia, como la existencia de las pautas abiertas de lectura (ORF, *open reading frames*).

2.2.3.2. Uso del genoma del cloroplasto en estudios filogenéticos

Los estudios filogenéticos llevados a cabo con los genomas organulares permiten comprender mejor las relaciones evolutivas entre los distintos grupos de plantas (Cattolico, 1985), así como avanzar en el conocimiento de los procesos evolutivos que han dado lugar a cambios en el contenido y orden de genes (Gray, 1999), tales como pérdidas de genes (Mardanov, 2008; Chang, 2006), inserciones y deleciones (Ingvarsson *et al.*, 2003) y transferencias de genes al genoma nuclear (Bungard, 2004; Pires, 2006). Las secuencias de los genomas de los cloroplastos también han permitido establecer relojes moleculares y estimar los tiempos de eventos clave en la evolución de las plantas (Do *et al.*, 2013; Savard *et al.*, 1994; Clegg, 1993).



3. Antecedentes y Objetivos

Con el propósito de desarrollar herramientas genéticas y moleculares para el estudio del ajo cultivado, en el laboratorio del tutor (Dr. Héctor Candela) se ha iniciado la caracterización del genoma de los orgánulos (cloroplasto y mitocondria) de esta especie. El estudio a nivel molecular del ajo cultivado permitirá abordar el estudio de la base genética y molecular de algunos de los problemas a los que se enfrenta la mejora del cultivo, como la infertilidad de las flores o las infecciones por patógenos como nematodos o virus. La fertilidad de las flores ya ha sido lograda en algunas variedades y algunos estudios transcriptómicos han intentado determinar la base molecular de esta infertilidad (Shemesh Mayer *et al.*, 2013; 2015). Algunos autores han utilizado información similar para examinar la relación entre temperaturas de almacenamiento y la formación del bulbo en *Allium sativum* (Shalom *et al.*, 2015). Por su interés básico y aplicado, desde el año 2015, en el laboratorio del tutor se vienen realizando experimentos para contribuir al conocimiento del genoma del ajo, incluyendo la secuenciación, ensamblaje *de novo* y anotación del transcriptoma de varios tejidos del ajo. El objetivo principal de este Trabajo de Fin de Grado ha sido caracterizar molecularmente el genoma del cloroplasto de *Allium sativum*. Este objetivo ha sido concretado en los siguientes objetivos particulares, que se corresponden con las etapas seguidas en la realización del trabajo:

1. Poner a punto un método para la purificación de ADN total a partir de tejido fresco de hojas de plantas jóvenes, con calidad suficiente para su secuenciación.
2. Realizar la secuenciación y el ensamblaje *de novo* del genoma completo del cloroplasto mediante el uso de herramientas bioinformáticas adecuadas.
3. Anotar los genes presentes en la secuencia del genoma del cloroplasto combinando métodos *ab initio* y métodos basados en homología para inferir la función de estos genes.
4. Utilizar la secuencia del genoma cloroplástico obtenida para determinar la historia evolutiva y la posición filogenética de *Allium sativum* respecto de otras especies de plantas monocotiledóneas y dicotiledóneas.

4. Materiales y Métodos

4.1. Cultivo en maceta de *Allium sativum*

El cultivo de las plantas se realizó en bandejas de plástico de 28×50 cm, con hasta 15 macetas de 11×11,5 cm cada una. Las macetas se rellenaron con una mezcla de perlita, vermiculita y turba en proporción 2:2:1 (v/v/v) que previamente había sido esterilizada en autoclave. Se sembraron dos dientes de ajo por maceta y se irrigaron con una disolución nutritiva comercial (abono Universal Carrefour) que contiene nitrógeno (N) total: 6% (incluye N amoniacal: 2,8%; N ureico: 2,2%; y N nítrico: 1,0%); pentóxido de fósforo (P₂O₅): 6%; óxido de potasio (K₂O): 6%; micronutrientes: cobre (Cu): 0,002%; hierro (Fe): 0,025%; manganeso (Mn): 0,014%; molibdeno (Mo): 0,001%; y zinc (Zn) quelado con EDTA: 0,004%.

Las plantas se irrigaron semanalmente con agua y cada dos semanas con una dilución al 0,5% (v/v) de la disolución nutritiva descrita, manteniendo un nivel de 2 o 3 cm de líquido en las bandejas. Los cultivos en maceta se mantuvieron a 23±1°C en una cámara climática bajo iluminación continua de 5000 lx suministrada por tubos fluorescentes F28 T5/D y Leuci T5-28W.

4.2. Purificación de ADN genómico

La purificación de las muestras de ADNg se llevó a cabo con el kit para plantas *GeneJet Plant Genomic DNA Purification Mini Kit*. El procedimiento seguido con el mismo fue el indicado por el fabricante (Thermo Scientific), y se sintetiza a continuación.

Dos muestras de tejido foliar de plantas jóvenes de la variedad “Spring blanco” (140 mg y 200 mg, respectivamente), que fueron suministradas por la empresa COOPAMAN, se trituraron en morteros de porcelana estériles en presencia de nitrógeno líquido. El material vegetal se transfirió a tubos *ependorf*, a los que se añadieron 350 µL de tampón de lisis A. Tras agitar durante 10-15 s con un vórtex, se añadieron 50 µL de tampón de lisis B y 20 µL de RNasa A. Los tubos se incubaron, con agitación periódica, en un baño a 65°C durante 10 min. A continuación, añadimos 130 µL de disolución de precipitación a cada tubo, que agitamos por inversión e incubamos durante 5 min en hielo. Tras esto, se centrifugaron los tubos durante 5 min a 13.300 rpm (16.300g). Transferimos los sobrenadantes a tubos limpios, a los que añadimos 400 µL de disolución de unión (*binding solution*) y 400 µL de etanol al 96%. Transferimos parte del volumen (600-700 µL) a las columnas suministradas con el kit y centrifugamos los tubos a 8.000 rpm (5.900g) durante 1 min. La operación se repitió con el volumen restante, descartando cada vez el sobrenadante tras su paso por la columna. Las columnas se lavaron dos veces con 500 µL de tampón de lavado 1. El tampón de lavado se eliminó centrifugando durante 1 min a 10.000 rpm (9.200g). Por último, eluimos el ADNg retenido en la columna a tubos *ependorf* estériles añadiendo 100 µL de tampón de elución al centro de la columna, incubando a temperatura

ambiente durante 5 min y centrifugando a 10.000 rpm durante 5 min. Las muestras se conservaron a 4°C hasta el momento de su secuenciación.

4.3. Comprobación de la cantidad y la calidad del ADNg

A fin de comprobar la calidad del ADN purificado, se llevó a cabo una electroforesis en gel de agarosa al 1% (m/v) en tampón de electroforesis TAE 1x (0,5 g de agarosa en 50 mL de TAE 1x), que se preparó a partir de una disolución madre a concentración 50x (Tris-HCl 2 M; ácido acético 5,71%; EDTA 50 mM; pH 8,0). La tinción del gel se realizó con 1,25 µL de SafeView, que se añadieron a la disolución de agarosa antes de que se solidificase. Además, en la primera y última calle de cada gel se cargaron 4 µL de marcador de peso molecular (*1 Kb DNA Ladder* de GeneCraft a una concentración de 0,1 µg/µL). Las muestras de ADN purificado se prepararon con 1 µL de ADN, 8 µL de agua desionizada y 1 µL de tampón de carga 10x (*DNA Loading Buffer 10x* de 5prime). Por último, la electroforesis se llevó a cabo con una fuente de alimentación a 90 V (LabNet) durante 30 min.

4.4. Análisis bioinformático

4.4.1. Ensamblaje del genoma

Para la secuenciación del ADN genómico, recurrimos a los servicios de la empresa STAB VIDA (Caparica, Portugal), que dispone de un secuenciador masivamente paralelo Illumina HiSeq 2500. Los fragmentos se secuenciaron por ambos extremos con lecturas de 101 nucleótidos (*paired-end reads*). Para evaluar la calidad de las lecturas obtenidas y determinar la presencia en las mismas de secuencias derivadas de los adaptadores añadidos durante la preparación de las muestras, utilizamos el programa FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>).

El ensamblaje *de novo* de las lecturas se llevó a cabo con el programa Velvet (versión 1.2.10) (Zerbino y Birney, 2008), con las opciones que se detallan en la sección de Resultados. El alineamiento de las lecturas obtenidas al genoma del cloroplasto se llevó a cabo con el programa Bowtie2 (versión 2.1.0), que emplea un algoritmo de alineamiento muy eficiente basado en la transformación de Burrows-Wheeler (Langmead *et al.*, 2009; Langmead y Salzberg, 2012). El índice del genoma de referencia se preparó con el programa bowtie2-build, especificando su secuencia con la opción -f. El alineamiento de las lecturas emparejadas se llevó a cabo mediante el programa bowtie2, con las siguientes opciones: -x (especifica el índice del genoma al que deben alinearse las lecturas, preparado anteriormente con bowtie2-build), --qc-filter (descarta las lecturas de baja calidad), -p (especifica el número de procesadores disponibles; en nuestro caso es igual a 8), --no-discordant (descarta las parejas de lecturas cuyos alineamientos no son concordantes),

--no-mixed (descarta las lecturas cuya pareja no ha sido eliminada) y --no-una1 (descarta aquellas lecturas para las que no se detectan alineamientos).

Los archivos SAM resultantes, que contienen la descripción de los alineamientos de las lecturas, se convirtieron a formato BAM mediante el comando `view` de samtools (Li *et al.*, 2009). Los archivos BAM correspondientes se ordenaron según la posición de las lecturas alineadas en el genoma de referencia mediante el comando `sort` de samtools. El archivo resultante se indizó con el comando `index` de samtools y se visualizó con ayuda del programa Tablet (Milne *et al.*, 2013).

4.4.2. Anotación del genoma

Para la identificación de genes correspondientes a moléculas de ARN no codificante (ARNt y ARNr) utilizamos los programas ARAGORN (Laslett y Canback, 2004), tRNA-scan (Lowe y Eddy, 1997; Schattner *et al.*, 2005) y RNAmmer 1.2 Server (Lagessen *et al.*, 2007), con sus opciones por defecto. Identificamos pautas de lectura abiertas (ORF; *open reading frames*) con ayuda de la aplicación *ORFfinder* (<http://www.ncbi.nlm.nih.gov/orffinder>). La anotación funcional basada en homología se basó en alineamientos realizados con la familia de programas BLAST (Altschul *et al.*, 1997; 2010). Los límites de las regiones repetidas se identificaron mediante BLAST 2 SEQUENCES (Tatusova y Madden, 1999). Para su depósito en Genbank, la secuencia obtenida ha sido procesada con el programa Seqin (<https://www.ncbi.nlm.nih.gov/Seqin>). El archivo resultante se procesó con el programa *OrganelleGenomeDRAW* (OGDRAW; Lohse *et al.*, 2007; 2013) para generar un mapa del genoma.

4.4.3. Análisis filogenético

Las secuencias de aminoácidos de los distintas proteínas consideradas en el análisis filogenético se alinearon por medio del programa T-Coffee (Notredame *et al.*, 2000; Di Tommaso *et al.*, 2011; <http://tcoffee.crg.cat>). El análisis filogenético se realizó mediante el programa MEGA7 (Kumar *et al.*, 2016; <http://www.megasoftware.net>), que implementa diversos métodos cladísticos y fenéticos de reconstrucción filogenética. En este trabajo hemos obtenido árboles filogenéticos mediante los métodos de unión al vecino (*neighbor-joining*; Saitou y Nei, 1987) y máxima parsimonia (Nei y Kumar, 2000), ambos implementados en MEGA7.

La Tabla 1, de la página siguiente, contiene los números de acceso e información adicional sobre los genomas cuyas secuencias se han considerado en este estudio para la realización de análisis filogenéticos.

Tabla 1.- Comparación del tamaño del genoma del cloroplasto en diversas plantas

Especie	Número de acceso	Tamaño (pb)			
		LSC	SSC	IRa/IRb	Total
<i>Phoenix dactylifera</i> ¹	NC_013991	86.198	17.712	27.276	158.462
<i>Cocos nucifera</i> ²	NC_022417	84.230	17.391	26.555	154.731
<i>Allium cepa</i>	NC_024813	82.695	17.923	26.460	153.538
<i>Zea mays</i> ³	NC_001666	82.353	12.537	22.747	22.747
<i>Oryza sativa</i> ⁴	KM103368	80.548	12.348	20.803	20.803
<i>Triticum aestivum</i> ⁴	NC_002762	80.348	12.791	20.703	135.545
<i>Saccharum officinarum</i> ⁵	AE009947	83.048	12.544	22.795	141.182
<i>Phalaenopsis equestris</i> ⁶	JF719062	85.967	11.300	25.846	148.959
<i>Lemna minor</i> ⁷	DQ400350	89.906	13.603	31.223	165.955
<i>Nicotiana tabacum</i> ⁸	Z00044	86.686	18.573	25.342	155.943
<i>Beta vulgaris</i> ⁹	KR230391	83.111	17.745	24.433	149.722
<i>Solanum lycopersicum</i> ¹⁰	NC_007898	85.876	18.363	25.611	155.461
<i>Theobroma cacao</i>	HQ336404	89.396	20.188	25.510	160.604

Los tamaños correspondientes a cada genoma han sido tomados de las referencias indicadas mediante superíndices: ¹ Yang *et al.* (2010); ² Huang *et al.* (2013); ³ Maier *et al.* (1995); ⁴ Chang *et al.* (2005); ⁵ Asano *et al.* (2004); ⁶ Jheng *et al.* (2012); ⁷ Mardanov *et al.* (2008); ⁸ Yukawa *et al.* (2006); ⁹ Li *et al.* (2014); y ¹⁰ Daniell *et al.* (2006). Cuando no se dispuso de un artículo, los límites de la repetición invertida se determinaron mediante alineamientos realizados con BLAST 2 SEQUENCES.

5. Resultados

5.1. Ensamblaje *de novo* del genoma del cloroplasto del ajo

Con el propósito de contribuir al desarrollo de herramientas para la caracterización genómica de variedades de ajo, en este trabajo hemos llevado a cabo la secuenciación, ensamblaje *de novo* y anotación del genoma del cloroplasto. La secuenciación del genoma del cloroplasto ha sido realizada mediante la tecnología de secuenciación por síntesis desarrollada por Illumina, utilizando un equipo HiSeq 2500, empleando lecturas emparejadas (*paired-end reads*) de 101 nucleótidos. Secuenciamos una muestra de ADN total purificado a partir de tejido fresco de hoja de plantas jóvenes. En las muestras derivadas de tejido foliar, el ADN proveniente del genoma del cloroplasto está típicamente presente en un número de copias muy superior al de los genomas nuclear y mitocondrial, lo que facilita en gran medida la determinación de su secuencia.

La secuenciación rindió 84.919.628 lecturas (42.459.814 parejas de lecturas), que ensamblamos utilizando Velvet (versión 1.2.10), un programa que utiliza una estrategia basada en grafos de *de Bruijn* (Zerbino y Birney, 2008), con los siguientes parámetros: longitud de los *k*-meros: 51, longitud del inserto (*-ins_length*): 150, cobertura esperada (*-exp_cov*): 340 y cobertura mínima (*-cov_cutoff*): 173. En estas condiciones, el programa produjo un ensamblaje con 5.319 contigios, con una longitud acumulada de 503.236 nucleótidos. Estos contigios incorporan 10.186.160 lecturas (11,99% del total), que deben provenir de los genomas organulares o de la fracción repetitiva del genoma nuclear.

En el ensamblaje resultante, la longitud del contigio más largo fue de 82.086 nucleótidos. El valor de N50 del ensamblaje fue de 520 nucleótidos, lo que indica que la mitad de la secuencia ensamblada ($503.236/2=251.618$ nucleótidos) forma parte de contigios de longitud igual o superior a 520. El examen minucioso de los contigios más largos nos permitió identificar tres contigios que muy probablemente corresponden a las tres regiones características de los genomas cloroplásticos. Los contigios 59 y 62 presentaron longitudes de 18.142 y 82.086 nucleótidos, respectivamente, así como coberturas muy parecidas (370,79 y 377,80). El contigio 55 presentó una longitud de 26.490 nucleótidos y una cobertura de 787,79 (aproximadamente el doble de la observada para los contigios 59 y 62). Considerando los solapamientos entre las secuencias de estos contigios y los valores de cobertura, reconstruimos un genoma circular de 153.108 pares de bases (pb; descontando la longitud correspondiente a los solapamientos), donde el contigio 62 se corresponde con la región grande de copia única (LSC; *large single copy*), el contigio 59 se corresponde con la región pequeña de copia única (SSC; *small single copy*), y el contigio 55 –representado dos veces en el genoma circular– se corresponde con las repeticiones invertidas (IRa e IRb; *inverted repeats*).

Para evaluar la corrección del ensamblaje, alineamos las lecturas utilizando el programa Bowtie2 (Langmead *et al.*, 2009; Langmead y Salzberg, 2012), descartando todos los alineamientos no concordantes de las parejas de lecturas. El programa permitió alinear 1.244.644 lecturas (622.322 parejas de lecturas; 1,47% del total). El examen detallado del alineamiento permitió detectar una región a la que las lecturas no se alineaban correctamente. Corregimos la secuencia de dicha región manualmente utilizando la secuencia de las lecturas. Una vez corregida, el tamaño de la secuencia del genoma completo fue de 153.131 pb (Figura 1).

Con el fin de definir con precisión los límites de las regiones SSC, LSC, IRa e IRb, realizamos un alineamiento de la secuencia consigo misma utilizando el programa BLAST 2 SEQUENCES (Tatusova y Madden, 1999). Los alineamientos obtenidos indican que las longitudes de las regiones son las siguientes: 82.105 pb (LSC), 18.042 pb (SSC) y 26.492 pb (IRa e IRb). Los límites de estas regiones se muestran en el mapa de la Figura 1.

Tabla 2.- Frecuencias nucleotídicas del genoma del cloroplasto del ajo.

	G+C (%)	A (%)	C (%)	G (%)	T (%)	Longitud (pb)
Total	56.172 (36,68)	47.942 (31,31)	28.605 (18,68)	27.567 (18)	49.017 (32,01)	153.131
LSC	28.345 (34,52)	26.341 (32,08)	14.551 (17,72)	13.794 (16,80)	27.419 (33,4)	82.105
IRb	11.281 (42,58)	7.600 (28,69)	5.855 (22,1)	5.426 (20,48)	7.611 (28,73)	26.492
SSC	5.265 (29,18)	6.390 (35,42)	2.773 (15,37)	2.492 (13,81)	6.387 (35,4)	18.042
IRa	11.281 (42,58)	7.611 (28,73)	5.426 (20,48)	5.855 (22,1)	7.600 (28,69)	26.492

Se indica entre paréntesis el porcentaje de cada nucleótido calculado a partir de los valores recogidos en la Tabla.

Las regiones constituyentes del genoma del cloroplasto difieren no sólo en su tamaño sino también en su composición. La Tabla 2 recoge las frecuencias de cada nucleótido en las cuatro regiones del genoma. Los resultados obtenidos indican que las repeticiones invertidas son sustancialmente más ricas en G+C que las regiones de copia única. El menor contenido en G+C corresponde a la región SSC (29,18%) y el mayor a las repeticiones invertidas (42,58%).

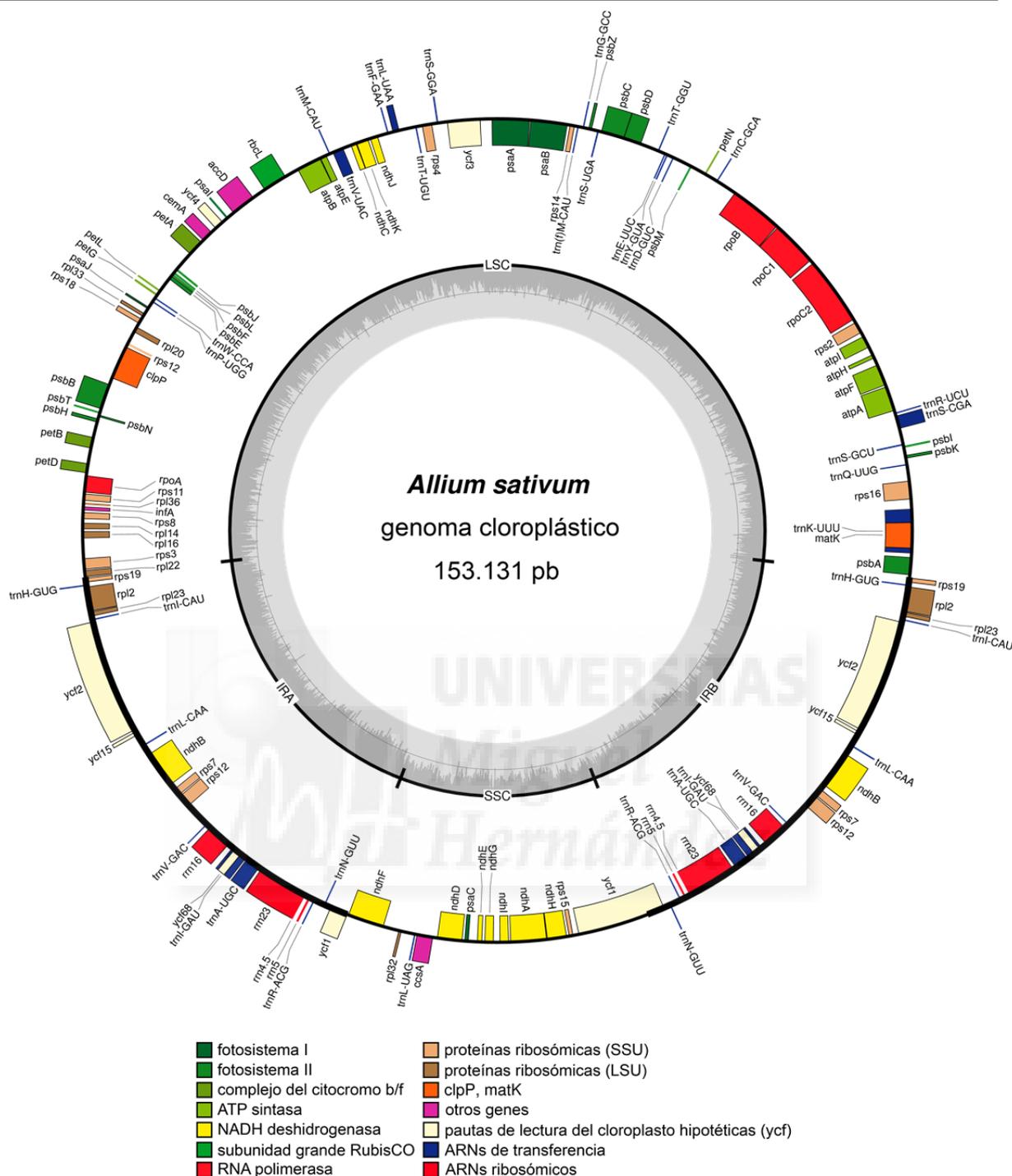


Figura 1.- Mapa del genoma completo del cloroplasto de *Allium sativum*. Los genes representados en el interior del círculo se transcriben en sentido de las agujas del reloj, mientras que los representados en el exterior se transcriben en sentido contrario. En el centro del círculo se representa la variación en el contenido en G+C del genoma.

5.2. Anotación del genoma

Para la identificación de los genes contenidos en el genoma del cloroplasto, hemos utilizado sistemáticamente una combinación de aproximaciones *ab initio* con otras basadas

en homología. La identificación de genes que codifican proteínas se realizó mediante la identificación de pautas de lectura abiertas (ORFs; *open reading frames*), a las que asignamos funciones mediante alineamientos realizados con BLASTP (Altschul *et al.*, 1997; 2010). Los genes correspondientes a moléculas de ARN no codificantes, como los ARN de transferencia (ARNt), se identificaron con ayuda de los programas tRNA-scan (Lowe y Eddy, 1997; Schattner *et al.*, 2005) y ARAGORN (Laslett y Canback, 2004). Los ARN ribosómicos (ARNr) se identificaron mediante la aplicación RNAmmer 1.2 Server (<http://www.cbs.dtu.dk/services/RNAmmer/>; Lagesen *et al.*, 2007).

Hemos determinado que el genoma del cloroplasto de *Allium sativum* contiene un total de 136 genes funcionales, a los que hay que añadir 6 pseudogenes (Tabla 3 y Figura 1). De estos 136 genes, 90 codifican proteínas implicadas en diversos procesos y los restantes corresponden a moléculas de ARN no codificantes. De éstos últimos, 38 corresponden a moléculas de ARNt y 8 a ARNr. A diferencia de los genes que residen en las regiones SSC y LSC, los genes de las repeticiones invertidas se encuentran presentes en dos copias. Por este motivo, el genoma del cloroplasto de *Allium sativum* contiene 94 genes de copia única (de los que 72 genes codifican proteínas y 22 corresponden a ARNt) y 21 genes presentes en 2 copias (de los que 9 codifican proteínas, 8 ARNt y 4 ARNr).

De los 136 genes identificados, 23 contienen intrones. De ellos, 14 corresponden a genes que codifican proteínas, 6 de ellos de copia única (*rpoC1*, *ycf3*, *ndhA*, *atpF*, *clpP* y *rps16*) y 4 presentes en dos copias (*rpl2*, *ndhB*, *ycf68* y *rps12*). Los 9 genes restantes corresponden a moléculas de ARNt, 5 de ellos de copia única (*trnE-UUC*, *trnK-UUU*, *trnL-UAA*, *trnS-CGA* y *trnV-UAC*) y 2 presentes en dos copias (*trnA-UGC* y *trnI-GAU*). Excepto los genes *ycf3* y *clpP*, que contienen dos intrones cada uno, todos los genes indicados poseen sólo un intrón.

En las secciones siguientes se describen sucintamente las distintas categorías funcionales a las que pertenecen los genes identificados en el genoma del cloroplasto.

5.2.1. Genes necesarios para la expresión génica del genoma del cloroplasto

Un grupo numeroso de los genes identificados desempeña funciones relacionadas con la expresión de los genes del genoma del cloroplasto, ya sea porque son necesarios para su transcripción o para la síntesis de proteínas. El genoma del cloroplasto contiene los genes *rpoA*, *rpoB*, *rpoC1* y *rpoC2* (Tabla 3), que codifican las subunidades α -, β -, β' - y β'' de la ARN polimerasa del cloroplasto codificada en el cloroplasto (PEP; *plastid-encoded plastid RNA polymerase*), respectivamente. Junto con la polimerasa del cloroplasto codificada en el núcleo (NEP; *nucleus-encoded plastid RNA polymerase*), PEP es la responsable de la transcripción de los genes que residen en el genoma del cloroplasto.

Tabla 3.- Genes identificados en la secuencia del genoma del cloroplasto del ajo.

Control de la expresión génica	ARN ribosómicos	<i>rrn4.5</i> (×2), <i>rrn5</i> (×2), <i>rrn16</i> (×2), <i>rrn23</i> (×2)
	ARN de transferencia	<i>trnA-UGC*</i> (×2), <i>trnC-GCA</i> , <i>trnD-GUC</i> , <i>trnE-UUC*</i> , <i>trnF-GAA</i> , <i>trn(f)M-CAU</i> , <i>trnG-GCC</i> , <i>trnH-GUG</i> (×2), <i>trnI-CAU</i> (×2), <i>trnI-GAU*</i> (×2), <i>trnK-UUU*</i> , <i>trnL-CAA</i> (×2), <i>trnL-UAG</i> , <i>trnL-UAA*</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> (×2), <i>trnP-UGG</i> , <i>trnQ-UUG</i> , <i>trnR-ACG</i> (×2), <i>trnR-UCU</i> , <i>trnS-GCU</i> , <i>trnS-CGA*</i> , <i>trnS-UGA</i> , <i>trnS-GGA</i> , <i>trnT-GGU</i> , <i>trnT-UGU</i> , <i>trnV-GAC</i> (×2), <i>trnV-UAC*</i> , <i>trnW-CCA</i> , <i>trnY-GUA</i>
	Subunidad pequeña del ribosoma	<i>rps3</i> , <i>rps4</i> , <i>rps7</i> (×2), <i>rps8</i> , <i>rps11</i> , <i>rps12</i> (transplicing) (×2), <i>rps14</i> , <i>rps15</i> , <i>rps18</i> , <i>rps19</i> (×2)
	Subunidad grande del ribosoma	<i>rpl2*</i> (×2), <i>rpl14</i> , <i>rpl16</i> , <i>rpl20</i> , <i>rpl22</i> , <i>rpl23</i> (×2), <i>rpl32</i> , <i>rpl33</i> , <i>rpl36</i>
	Subunidades de la ARN polimerasa	<i>rpoA</i> , <i>rpoB</i> , <i>rpoC1*</i> , <i>rpoC2</i>
Fotosíntesis	Subunidad grande de la RuBisCO	<i>rbcl</i>
	Fotosistema I	<i>psaA</i> , <i>psaB</i> , <i>psaC</i> , <i>psaI</i> , <i>psaJ</i> , <i>ycf3**</i> , <i>ycf4</i>
	Fotosistema II	<i>psbA</i> , <i>psbB</i> , <i>psbC</i> , <i>psbD</i> , <i>psbE</i> , <i>psbF</i> , <i>psbH</i> , <i>psbI</i> , <i>psbJ</i> , <i>psbK</i> , <i>psbL</i> , <i>psbM</i> , <i>psbN</i> , <i>psbT</i> , <i>psbZ</i>
	NADH deshidrogenasa	<i>ndhA*</i> , <i>ndhB*</i> (×2), <i>ndhC</i> , <i>ndhD</i> , <i>ndhE</i> , <i>ndhF</i> , <i>ndhG</i> , <i>ndhH</i> , <i>ndhI</i> , <i>ndhJ</i> , <i>ndhK</i>
	Complejo del citocromo b/f	<i>petA</i> , <i>petB</i> , <i>petD</i> , <i>petG</i> , <i>petL</i> , <i>petN</i>
	ATP sintasa	<i>atpA</i> , <i>atpB</i> , <i>atpE</i> , <i>atpF*</i> , <i>atpH</i> , <i>atpI</i>
Otros genes	Maturasa	<i>matK</i>
	Proteasa	<i>clpP**</i>
	Proteína de la envuelta de la membrana	<i>cemA</i>
	Subunidad de la acetil-CoA-carboxilasa	<i>accD</i>
	Gen de síntesis del citocromo tipo c	<i>ccsA</i>
	Pautas de lectura abierta (ORF) conservadas	<i>ycf1</i> , <i>ycf2</i> (×2), <i>ycf68*</i> (×2)
Pseudogenes	<i>rps2</i> , <i>ycf15</i> (×2), <i>rps16*</i> , <i>infA</i> , <i>ycf1</i>	

Uno y dos asteriscos indican genes con uno o dos intrones respectivamente. Los genes localizados en las regiones IR se indican con un (×2) después del nombre del gen.

La traducción de los genes transcritos por PEP y NEP es llevada a cabo en el estroma del cloroplasto por ribosomas de tipo bacteriano (70 S). El genoma del cloroplasto sólo contiene genes que codifican algunas de las proteínas ribosómicas que forman parte de las subunidades grande (subunidad 50 S; genes *rpl*) y pequeña (subunidad 30 S; genes *rps*) de estos ribosomas. De las 33 proteínas descritas en la subunidad 50 S del ribosoma de los cloroplastos de las plantas superiores (Yamaguchi y Subramanian, 2000), el genoma del cloroplasto codifica únicamente 9 proteínas distintas (genes *rpl2*, *rpl14*, *rpl16*, *rpl20*, *rpl22*, *rpl23*, *rpl32*, *rpl33* y *rpl36*; Tabla 3). Las restantes proteínas necesarias para el ensamblaje de la subunidad 50 S se sintetizan en el citoplasma a partir de ARNm derivados de genes nucleares, desde donde deben transportarse al cloroplasto. De las 25 proteínas descritas en la subunidad 30 S del ribosoma (Yamaguchi *et al.*, 2000), el genoma del cloroplasto de *Allium sativum* sólo codifica 10 de ellas (genes *rps3*, *rps4*, *rps7*, *rps8*, *rps11*, *rps12*, *rps14*, *rps15*, *rps18* y *rps19*). Si bien es posible identificar secuencias relacionadas con dos genes *rps* adicionales, *rps2* y *rps16*, éstos parecen haberse pseudogenizado.

Los genes de los ARN ribosómicos *rrn5*, *rrn4.5*, *rrn23* y *rrn16* residen en la repetición invertida, por lo que todos ellos se encuentran en dos copias.

Para la traducción de los genes del cloroplasto, es necesario que exista un complemento mínimo de moléculas de ARNt, capaces de reconocer los 61 codones que corresponden a aminoácidos. Debido al fenómeno denominado “tambaleo” (*wobbling*), por el que se consideran válidos ciertos pares de bases atípicos entre la tercera base de un codón y la primera base del anticodón (como por ejemplo, los pares de bases G-U y U-G), se considera que un complemento de 32 ARNt diferentes deberían permitir la lectura de todos los codones (Alkatib *et al.*, 2012b).

Utilizando varias herramientas bioinformáticas y búsquedas BLAST, hemos identificado 38 genes que corresponden a ARNt (genes *trn*), 22 de ellos de copia única y 8 de ellos presentes en 2 copias (Tabla 3). Conforme a lo explicado en el párrafo anterior, un complemento de 30 ARNt parece, *a priori*, insuficiente para traducir los 61 codones. La complementariedad entre residuos de G y U proporciona una explicación al reconocimiento de un buen número de codones no complementarios a los anticodones de los ARNt identificados (se indica entre paréntesis el ARNt encontrado): UUU (*trnF*-GAA), UCU (*trnS*-GGA), UAU (*trnY*-GUA-GGA), UGU (*trnC*-GCA), CUG (*trnL*-UAG), CCG (*trnP*-UGG), CAU (*trnH*-GUG), CAG (*trnQ*-UUG), AUU (*trnI*-GAU), ACU (*trnT*-GGU), ACG (*trnT*-UGU), AAU (*trnN*-GUU), AAG (*trnK*-UUU), AGU (*trnS*-GCU), AGG (*trnR*-UCU), GUU (*trnV*-GAC), GUG (*trnV*-UAC), GCG (*trnA*-UGC), GAU (*trnD*-GUC), GAG (*trnE*-UUC), GGU (*trnG*-GGC) y GGG (*trnG*-UCC).

Tabla 4.- Anticodones presentes en los ARNt del genoma del cloroplasto del ajo.

		Segunda base			
		U	C	A	G
Primera base	U	UUU Phe	UCU Ser	UAU Tyr	UGU Cys
		UUC Phe <i>trnF-GAA</i>	UCC Ser <i>trnS-GGA</i>	UAC Tyr <i>trnY-GUA</i>	UGC Cys <i>trnC-GCA</i>
		UUA Leu <i>trnL-UAA</i>	UCA Ser <i>trnS-UGA</i>	UAA -	UGA -
		UUG Leu <i>trnL-CAA</i> ×2	UCG Ser <i>trnS-CGA</i>	UAG -	UGG Trp <i>trnW-CCA</i>
	C	CUU Leu	CCU Pro	CAU His	CGU Arg <i>trnR-ACG</i> ×2
		CUC Leu	CCC Pro	CAC His <i>trnH-GUG</i> ×2	CGC Arg
		CUA Leu <i>trnL-UAG</i>	CCA Pro <i>trnP-UGG</i>	CAA Gln <i>trnQ-UUG</i>	CGA Arg
		CUG Leu	CCG Pro	CAG Gln	CGG Arg
	A	AUU Ile	ACU Thr	AAU Asn	AGU Ser
		AUC Ile <i>trnI-GAU</i> ×2	ACC Thr <i>trnT-GGU</i>	AAC Asn <i>trnN-GUU</i> ×2	AGC Ser <i>trnS-GCU</i>
		AUA Ile <i>trnI-CAU</i> ×2	ACA Thr <i>trnT-UGU</i>	AAA Lys <i>trnK-UUU</i>	AGA Arg <i>trnR-UCU</i>
		AUG Met <i>trnM-CAU</i> , <i>trn(f)M-CAU</i>	ACG Thr	AAG Lys	AGG Arg
	G	GUU Val	GCU Ala	GAU Asp	GGU Gly
		GUC Val <i>trnV-GAC</i> ×2	GCC Ala	GAC Asp <i>trnD-GUC</i>	GGC Gly <i>trnG-GCC</i>
		GUA Val <i>trnV-UAC</i>	GCA Ala <i>trnA-UGC</i> ×2	GAA Glu <i>trnE-UUC</i>	GGA Gly <i>trnG-UCC</i>
		GUG Val	GCG Ala	GAG Glu	GGG Gly

Tres de las 16 secciones de la Tabla 4 (CUN, CCN y GCN) están representadas, respectivamente, por un único anticodón. La lectura de los codones CUU, CUC, CCU, CCC, GCU y GCC podría ser realizada gracias al fenómeno denominado “supertambaleo” (*superwobbling*; Alkatib *et al.*, 2012b), por el que un anticodón cuya primera base es U permite reconocer cuatro codones distintos. De este modo, los codones CUN podrían ser leídos por *trnL-UAG*, CCN por *trnP-UGG*, y GCN por *trnA-UGC*. Únicamente tres de los 61 codones (CGC, CGA y CGG, todos ellos correspondientes al aminoácido arginina) parecen carecer de moléculas de ARNt que posibiliten su lectura. Esta circunstancia se discute más ampliamente en la Discusión.

Tres genes distintos corresponden a moléculas de ARNt que portan el mismo anticodón CAU pero desempeñan diferentes funciones. Dos de ellas, *trnM-CAU* y *trn(f)M-CAU* reconocen el mismo codón AUG, correspondientes a los aminoácidos metionina y formilmetionina. Las diferencias en la secuencia y estructura secundaria de estos ARNt, sin embargo, determinan que el *trn(f)M-CAU* se utilice únicamente para iniciar la síntesis de proteínas. Se ha descrito que la lisinación de la C en el anticodón de *trnI-CAU* permite la lectura del codón AUA (correspondiente al aminoácido isoleucina), pero no del codón AUG,

evitando así que el “tambaleo” del anticodón UAU interfiera con la lectura de los codones correspondientes a metionina (Alkatib *et al.*, 2012a). La Figura 2 ilustra las diferencias en la secuencia y estructura secundaria existentes entre los diferentes ARNt para metionina, formilmetionina e isoleucina identificados en el genoma del cloroplasto de *Allium sativum*.

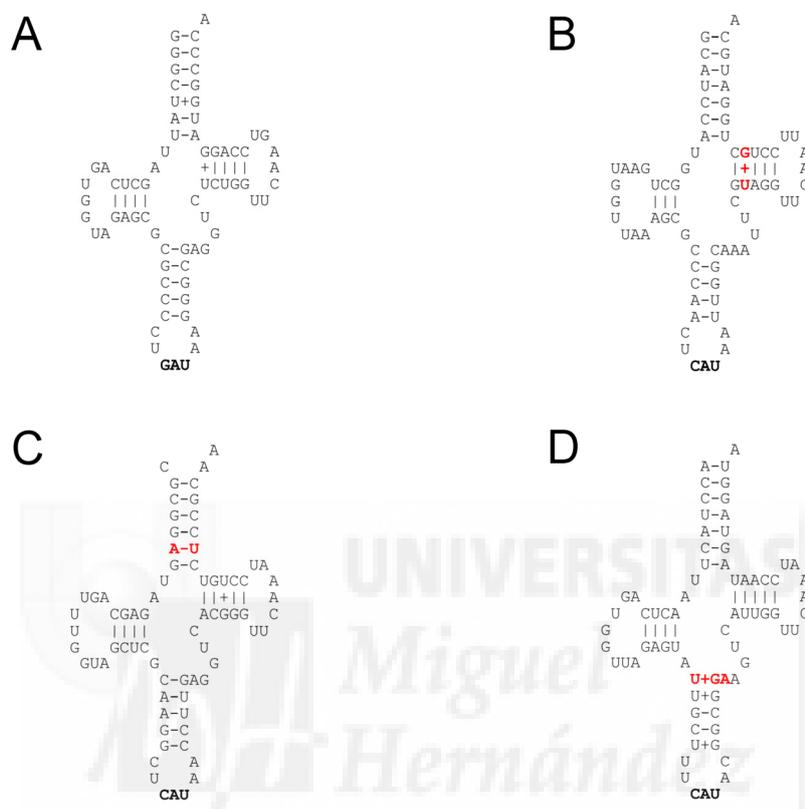


Figura 2.- Estructura secundaria de cuatro moléculas de ARNt del cloroplasto de *Allium sativum* para los aminoácidos metionina, formilmetionina e isoleucina. (A) trnI-GAU. (B) trnI-CAU. (C) trn(f)M-CAU. (D) trnM-CAU. Figura adaptada del artículo de Alkatib *et al.* (2012a). Se destacan en rojo las diferencias entre las secuencias de los ARNt del artículo y los del ajo. Se destacan en negrita los anticodones de cada ARNt. Se indica mediante un signo + la formación de pares de bases no canónicos entre G y U.

5.2.2. Genes implicados en la fotosíntesis

Otro grupo numeroso de genes codifica proteínas que desempeñan funciones relacionadas con la fotosíntesis (Tabla 3). Entre estos genes destacan algunos que codifican subunidades de los fotosistemas I y II. El genoma de *Allium sativum* contiene genes para cinco subunidades del fotosistema I (de un total de 15 descritas en las plantas superiores; Jensen *et al.*, 2007): PsaA, PsaB, PsaC, PsaI y PsaJ, exactamente las mismas que se encuentran codificadas en el genoma del cloroplasto de la dicotiledónea *Arabidopsis thaliana*. Los genes *ycf3* e *ycf4*, codifican proteínas conservadas que, si bien no forman

parte del fotosistema I, participan en su biogénesis en el alga *Chlamydomonas reinhardtii* (Boudreau *et al.*, 1997).

También hemos anotado genes que codifican 15 subunidades del fotosistema II, 11 subunidades del complejo NADH deshidrogenasa, 6 subunidades del complejo del citocromo b/f, 6 subunidades del complejo ATP sintasa, y 1 subunidad de la ribulosa 1,5-bisfosfato carboxilasa-oxigenasa (RuBisCO). Todos estos genes están presentes en una sola copia, excepto el gen *ndhB*, que reside en las repeticiones invertidas (Tabla 3).

5.2.3. Otros genes

El genoma del cloroplasto contiene varios genes que codifican proteínas con funciones diversas: *matK*, *clpP*, *cemA*, *accD* y *ccsA*. Además de estos genes, es posible identificar tres pautas de lectura abiertas conservadas, denominadas *ycf1*, *ycf2* e *ycf68*, cuya función precisa se desconoce (la denominación *ycf* deriva de *hypothetical conserved frame*). Estos tres genes residen en la región repetida y, en consecuencia, están presentes en dos copias. Además de estos genes, hemos clasificado 6 secuencias como pseudogenes, a juzgar por la presencia de mutaciones y/o cambios en la pauta de lectura observados en su secuencia. Cuatro de estos pseudogenes están presentes en una sola copia y uno está presente en dos copias.

5.3. Análisis filogenético

Hemos utilizado la secuencia obtenida para determinar la posición filogenética de *Allium sativum*. Para ello, hemos seleccionado las secuencias de aminoácidos de los productos de 17 genes (*rps4*, *rps8*, *rps11*, *rps15*, *rps19*, *rpl22*, *psbA*, *psbB*, *psbD*, *petA*, *petB*, *petD*, *petN*, *rpoA*, *ccsA*, *cemA* y *rbcL*), que fueron seleccionados por hallarse presentes en las 14 especies que hemos incorporado a nuestro estudio, 10 monocotiledóneas y 4 dicotiledóneas. Entre las monocotiledóneas, hemos incluido al ajo (*Allium sativum*), la cebolla (*Allium cepa*), la palmera datilera (*Phoenix dactylifera*), el cocotero (*Cocos nucifera*), el trigo (*Triticum aestivum*), el arroz (*Oryza sativa*), el maíz (*Zea mays*), la caña de azúcar (*Saccharum* sp.), una orquídea (*Phalaenopsis equestris*) y la lenteja de agua (*Lemna minor*). Entre las dicotiledóneas, hemos incluido al tabaco (*Nicotiana tabacum*), el tomate (*Solanum lycopersicum*), la remolacha (*Beta vulgaris*) y el cacao (*Theobroma cacao*).

Hemos realizado alineamientos múltiples con las secuencias de cada una de las 17 proteínas por separado con el programa T-Coffee, que utiliza un algoritmo que maximiza la consistencia del alineamiento múltiple producido (Notredame *et al.*, 2000; Di Tommaso *et al.*, 2011). Los alineamientos resultantes se guardaron en el formato estándar del programa ClustalW, y se concatenaron en una única secuencia para su estudio conjunto. El análisis

filogenético fue llevado a cabo con ayuda del programa MEGA7, que implementa diversos algoritmos de reconstrucción filogenética (Kumar *et al.*, 2016). En este trabajo, hemos utilizado dos métodos diferentes para inferir la historia evolutiva de las 14 especies consideradas: uno basado en caracteres (método de máxima parsimonia; *maximum parsimony*) y otro basado en distancias (el método de unión al vecino; *neighbor-joining*). El método de máxima parsimonia busca el árbol que explica la evolución de las secuencias con un menor número de mutaciones. El método de unión al vecino, por su parte, es un algoritmo de agrupamiento que asume aditividad de las distancias determinadas entre las secuencias analizadas.

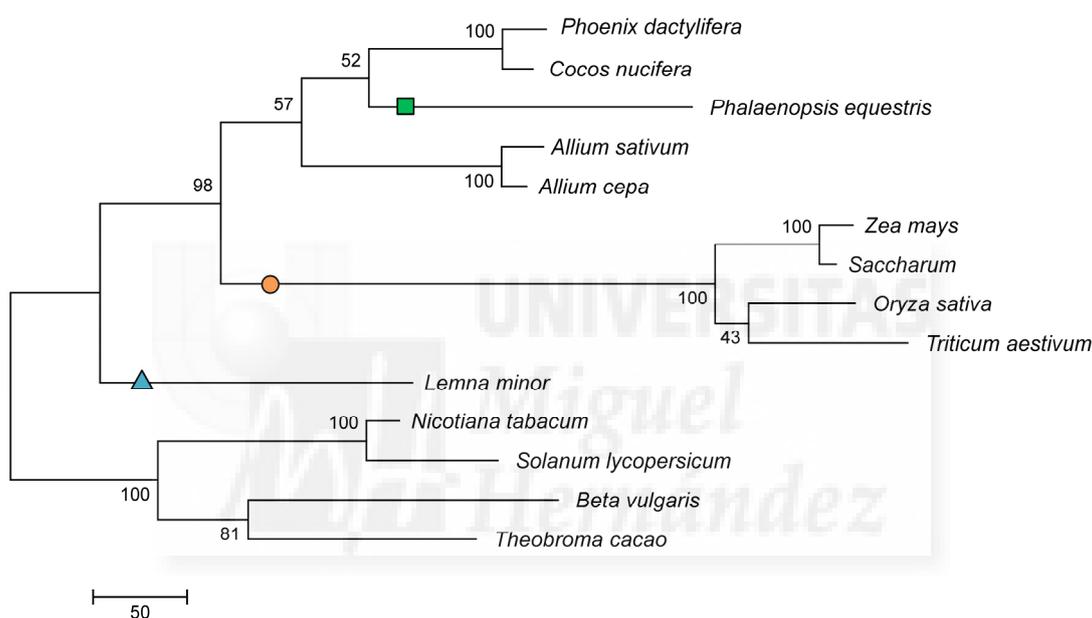


Figura 3.- Historia evolutiva de 14 especies de plantas inferida mediante el método de máxima parsimonia. Se muestra el primer árbol de los dos más parsimoniosos, de longitud 1.894. Las ramas han sido etiquetadas con el porcentaje de árboles (pseudorréplicas) que presentan una rama que agrupa a los mismos taxones, según una prueba de *bootstrap* con 1.000 réplicas (Felsenstein, 1985). Los árboles más parsimoniosos se identificaron mediante el algoritmo de ramificación y acotación Max-mini, que garantiza la identificación de todas las topologías óptimas. La longitud de las ramas es proporcional al número de sustituciones ocurridas. El conjunto final de datos comprende información para 3.903 posiciones. El árbol ha sido enraizado seleccionando como grupo externo el clado formado por las dicotiledóneas. Los símbolos indican las ramas en las que se produjo la pérdida de los genes *ndhF* y *ndhH* (cuadrado verde), *ycf1* y *ycf2* (círculo naranja), y *trnH* y *trnN* (triángulo azul).

Los árboles filogenéticos obtenidos se muestran en la Figura 3 (máxima parsimonia) y en la Figura 4 (unión al vecino). Conforme a lo esperado, la especie más próxima al ajo es la cebolla, ya que ambas pertenecen al mismo género (*Allium*). Ambas especies forman parte de un mismo clado junto a *Phalaenopsis equestris* (una orquídea) y las palmeras

Phoenix dactylifera y *Cocos nucifera*. Los cuatro miembros del orden Poales (*Zea mays*, *Saccharum* sp., *Oryza sativa* y *Triticum aestivum*) se agrupan en un clado distinto del anterior. En ambos árboles, la lenteja de agua ocupa una posición basal respecto de las restantes monocotiledóneas incluidas en el estudio.

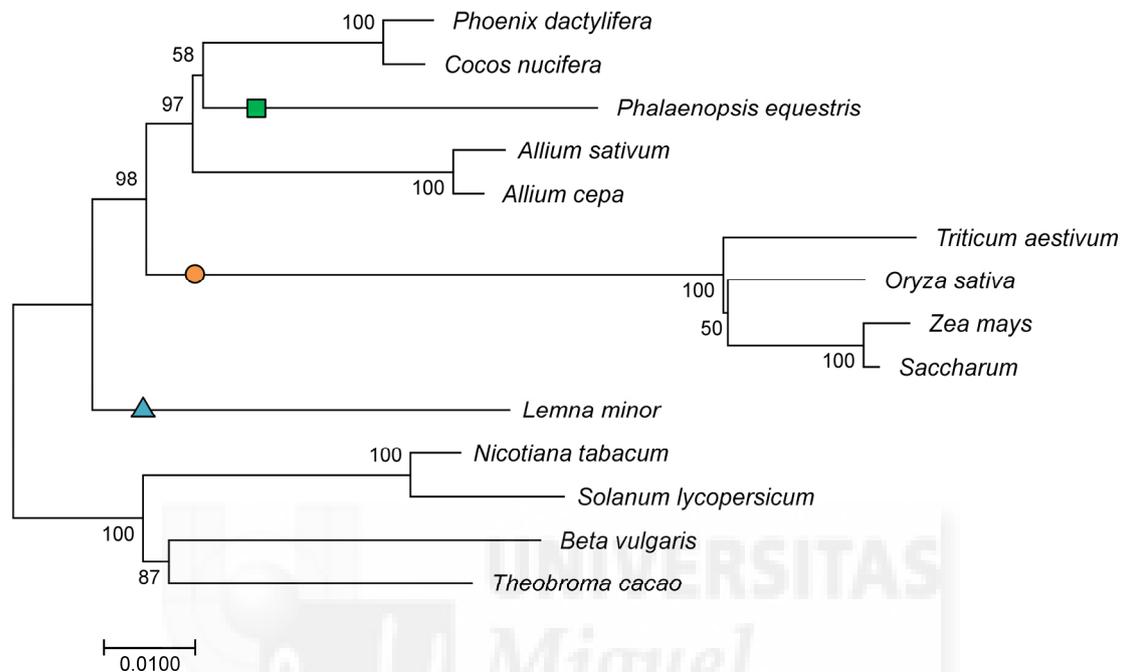


Figura 4.- Historia evolutiva de 14 especies de plantas inferida mediante el método de unión al vecino (Saitou y Nei, 1987). Se representa el árbol óptimo, cuyas ramas presentan una longitud total de 0,44100792. Las ramas del árbol han sido etiquetadas con el porcentaje de árboles (pseudorréplicas) que presentan una rama que agrupa a los mismos taxones, según una prueba de *bootstrap* con 1.000 réplicas (Felsenstein, 1985). El árbol ha sido dibujado a escala, con ramas de longitud proporcional a las distancias evolutivas empleadas para inferir el árbol filogenético. Dichas distancias evolutivas han sido corregidas conforme a la distribución de Poisson (Zuckerandl y Pauling, 1965), y se expresan en unidades de número de sustituciones por sitio. Todas las posiciones ambiguas de cada par de secuencias se eliminaron para el análisis. El conjunto final de datos comprende información para 3.903 posiciones. El árbol ha sido enraizado seleccionando como grupo externo el clado formado por las cuatro dicotiledóneas. Los símbolos indican las ramas en las que se produjo la pérdida de los genes *ndhF* y *ndhH* (cuadrado verde), *ycf1* e *ycf2* (círculo naranja), y *trnH* y *trnN* (triángulo azul).

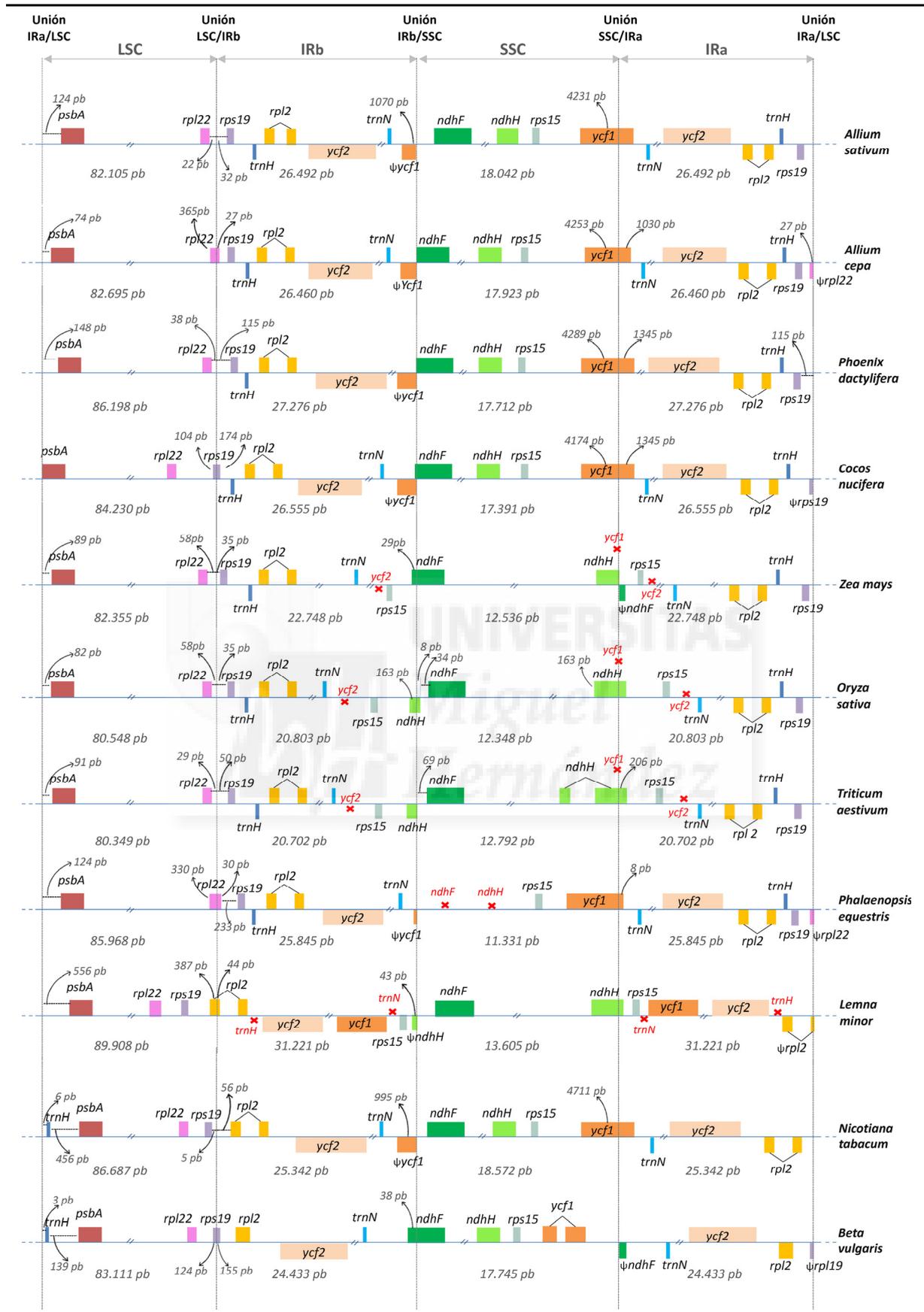
Durante la evolución de las angiospermas, se han producido variaciones espontáneas en las secuencias que forman parte de las repeticiones invertidas (IRa e IRb) (Goulding *et al.*, 1996). Pese a que el genoma del cloroplasto ha experimentado pocos cambios a lo largo de la evolución de las plantas superiores, durante la misma se han producido inserciones y deleciones que han causado desplazamientos de los límites entre las regiones de copia única y las repeticiones invertidas. La incorporación de un gen

completo al interior de la región repetida conlleva un incremento en su dosis. Por su parte, cuando un gen se dispone sobre el límite entre una región de copia única y una de las repeticiones, una de sus copias suele pseudogenizarse. Estos cambios en la posición de los genes han sido utilizados previamente como marcadores para elucidar la historia evolutiva de los genomas cloroplásticos. En la Figura 5 se han representado gráficamente los genes que ocupan las posiciones más próximas a puntos de unión entre las regiones de copia única (SSC y LSC) y las dos repeticiones invertidas (IRa e IRb) en 11 especies distintas (9 monocotiledóneas y 2 dicotiledóneas)

Como ilustra la Figura 5, las cinco especies del clado que contiene a *Allium sativum* presentan varias características comunes, que las distinguen de las Poales. Entre estas características diferenciales destacan: (1) la presencia de una copia completa del gen *ycf1* en el sitio de unión SSC/IRa, (2) una copia parcial del mismo gen (Ψ *ycf1*) en la región IRb adyacente al sitio de unión IRb/SSC, (3) la presencia del gen *rps15* en la región SSC, y (4) la presencia de una copia completa del gen *ycf2* en las regiones IRa e IRb. Todas estas características son compartidas, al menos, por algunas dicotiledóneas, como el tabaco (*Nicotiana tabacum*), lo que sugiere que ya estaban presentes en el genoma del cloroplasto del ancestro común a todas las angiospermas.

El genoma de *Phalaenopsis equestris* presenta, a su vez, algunas diferencias con los restantes miembros del clado. La diferencia más obvia es la pérdida completa de los genes *ndhF* y *ndhH*, que parece ser una característica común en el linaje de las orquídeas. El clado de las Poales ha experimentado la pérdida completa de los genes *ycf1* e *ycf2*.

Figura 5.- (Véase la página siguiente) Comparación de las uniones entre las repeticiones invertidas (IRa, IRb) y las secuencias de copia única (SSC, LSC) en 11 especies de plantas (9 monocotiledóneas y 2 dicotiledóneas). Se indican los tamaños de cada una de las regiones los tamaños de los genes relevantes y su distancia a las uniones. Los pseudogenes se denotan con el símbolo ψ . Las pérdidas completas de genes se indican mediante cruces rojas con el nombre del gen en la posición donde normalmente se encuentran. Los genes han sido representados a un lado u otro de la línea horizontal dependiendo de si se transcriben a partir de una hebra o la otra.



6. Discusión

En este Trabajo de Fin de Grado, hemos llevado a cabo la caracterización del genoma completo del cloroplasto de una planta cultivada, el ajo (*Allium sativum*). Este genoma ha resultado ser muy parecido en tamaño y contenido en genes al genoma del cloroplasto de la cebolla (*Allium cepa*), que ha sido descrito recientemente (Kim *et al.*, 2015b). Nuestro trabajo se enmarca dentro de una línea de investigación más amplia desarrollada en el laboratorio del tutor, que tiene por objetivo el desarrollo de herramientas para el análisis genético y genómico de esta especie. En el marco de esta investigación, en el laboratorio también se está trabajando actualmente en el ensamblaje *de novo* y anotación del genoma mitocondrial y el transcriptoma de diversos tejidos. La disponibilidad de la secuencia de ambos genomas organulares representa un primer punto de partida para futuros estudios del genoma nuclear.

Nuestros resultados indican que el genoma del cloroplasto de *Allium sativum* presenta características comunes con los genomas de otros grupos de plantas monocotiledóneas pertenecientes al mismo clado, como la cebolla y las palmeras. Sin embargo, los genomas de estas especies difieren del de la orquídea *Phalaenopsis equestris*, que también pertenece al mismo clado, ya que este último ha perdido numerosos genes (en concreto los genes *ndhA*, *ndhF* y *ndhH*) presumiblemente como resultado de su transferencia al genoma nuclear (Chang *et al.*, 2006). Por otro lado, los genomas de los cloroplastos de estas monocotiledóneas son, a su vez, claramente distinguibles de los genomas de los cloroplastos de las gramíneas (Poales), que comparten entre sí numerosas características y forman un clado distinto dentro del orden monocotiledóneas. Así mismo hemos observado un gran parecido estructural entre los genomas cloroplásticos de las monocotiledóneas *Allium sativum*, *Allium cepa*, *Phoenix dactylifera* y *Cocos nucifera* y el de la dicotiledónea *Nicotiana tabacum*. Esta observación indica que el orden y contenido en genes de estas especies deriva del de su ancestro común.

El genoma del cloroplasto del ajo contiene muchos genes implicados en la expresión de sus genes. Entre estos genes destaca un complemento completo de subunidades de la ARN polimerasa codificada en el cloroplasto (PEP), así como numerosos genes que codifican proteínas de las subunidades grande (50 S) y pequeña (30 S) de los ribosomas 70 S de los cloroplastos. Hemos prestado especial atención al conjunto de genes de los ARN de transferencia (ARNt). Los anticodones de los ARNt transcritos a partir del genoma del cloroplasto permiten leer directamente 30 de los 61 codones codificantes. Las reglas del “tambaleo” (*wobbling*) explican la lectura de 22 codones adicionales. Los codones de los tipos CUN, CCN y GCN podrían ser leídos por los ARNt existentes gracias al fenómeno denominado “supertambaleo” (*superwobbling*). En la sección de resultados dejamos

pendiente la discusión del mecanismo que permitiría la lectura de los codones CGC, CGA y CGG (correspondientes al aminoácido arginina), que no pueden ser leídos por el trnR-ACG. Como puede apreciarse, la tercera base de estos codones (C, A y G, respectivamente) no es complementaria de la primera base del anticodón (A), lo que impide el reconocimiento de los mismos por este ARNt. Como mecanismo alternativo que posibilitaría la lectura de estos codones, proponemos la hipótesis de que la A del anticodón podría ser editada a inosina (I). Dado que la inosina permite la formación de pares de bases con estos nucleótidos, el trnR-ICG resultante permitiría la traducción correcta de los tres codones huérfanos. De ser correcta esta hipótesis, el genoma del cloroplasto de *Allium sativum* contendría todos los ARNt necesarios para la traducción de cualquiera de sus genes.

Coincidiendo con la finalización de este trabajo, se ha publicado recientemente un breve artículo que describe sucintamente la secuencia del genoma del cloroplasto de una variedad de ajo originaria de Uzbekistán (Filyushin *et al.*, 2016). Nuestros resultados difieren en algunos aspectos de los presentados en dicho artículo. En concreto, hemos identificado un mayor número de genes en la secuencia del genoma del cloroplasto. Dado que hemos identificado el mismo número de pseudogenes (6), genes de ARNt (38) y ARNr (8), esta diferencia se debe al mayor número de genes que codifican proteínas que hemos identificado (90 en nuestro trabajo frente a 82 en el suyo). Apreciamos también algunas diferencias en lo referido a la identificación de genes que contienen intrones (mayor en nuestro caso) y a la descripción de los genes duplicados. A modo de ejemplo, el trabajo de Filyushin *et al.* (2016) considera que el gen *trnM-CAU* se encuentra duplicado, cuando en realidad las dos copias difieren en su secuencia, lo que presumiblemente les permite distinguir la metionina de la formilmetionina.

Las pequeñas diferencias existentes en el tamaño de los genomas cloroplásticos de dos variedades de ajo (153.131 pb en nuestro trabajo y 153.372 pb en el trabajo de Filyushin *et al.*, 2016) y cinco de cebolla (153.529, 153.440 y 153.568 bp, descritas en Kim *et al.*, 2015b; 153.355 y 153.538, descritas en Von Kohn *et al.*, 2013) reflejan la existencia de polimorfismos intraespecíficos e intragenéricos que podrían ser utilizados para genotipar las distintas variedades. Si bien la propagación del ajo se realiza habitualmente de forma vegetativa, mediante la multiplicación de sus bulbos, en la actualidad se están llevando a cabo esfuerzos para restaurar la fertilidad de sus flores, un objetivo factible gracias a la existencia de algunas variedades cultivadas que no han perdido completamente la capacidad de reproducirse mediante semillas. Esta posibilidad abre un horizonte prometedor para la mejora genética de la especie y el desarrollo de nuevas variedades híbridas. En la cebolla, los métodos actuales de producción de semilla híbrida se basan en la disponibilidad de un selecto grupo de variedades que manifiestan esterilidad de origen citoplasmático. Como en la cebolla (Von Kohn *et al.*, 2013; Kim *et al.*, 2015b), el conocimiento de la

secuencia de los genomas del cloroplasto y la mitocondria del ajo debería permitir el desarrollo de marcadores moleculares con los que caracterizar rápidamente e identificar con precisión los genotipos asociados a la esterilidad.



7. Conclusiones y proyección futura

Este trabajo de Fin de Grado surge como continuación y complemento a las diversas líneas de investigación que se están llevando a cabo en relación a los ajos.

Hemos puesto a punto un método para la purificación de ADN total de *Allium sativum* a partir de tejidos foliares, con calidad suficiente para la realización de experimentos de secuenciación masivamente paralela.

Hemos preparado una muestra de ADN total a partir de tejidos de la variedad “Spring blanco”, que ha sido secuenciada mediante la tecnología de secuenciación por síntesis de Illumina, en un equipo HiSeq 2500.

Para el ensamblaje *de novo* de las secuencias obtenidas, hemos utilizado el programa Velvet, que nos ha permitido identificar tres contigios de gran longitud, que se corresponden de manera precisa con las tres regiones del genoma del cloroplasto: la región LSC (grande de copia única), la región SSC (pequeña de copia única) y la secuencia de la repetición invertida (IR).

El tamaño del genoma del cloroplasto del ajo es 153.131 pb, ligeramente más pequeño que el de la cebolla (153.538 pb). El tamaño de las regiones LSC, SSC e IR (presente en dos copias) es, respectivamente, de 82.105, 18.042 y 26.492 pb.

Hemos realizado la anotación del genoma del cloroplasto mediante una combinación de métodos *ab initio* y basados en homología. La anotación nos ha permitido determinar que el genoma contiene 136 genes funcionales y 6 pseudogenes. Algunos de estos genes, los que residen en las repeticiones invertidas, se encuentran presentes en dos copias.

Como otros genomas de cloroplastos, el genoma del cloroplasto del ajo es particularmente rico en genes implicados en la expresión de sus genes y en genes relacionados con la función fotosintética.

Hemos determinado que el genoma del cloroplasto contiene un complemento completo de moléculas de ARNt, que debe permitir la traducción de todos los codones de sus transcritos.

De manera inminente, planeamos depositar la secuencia obtenida en la base de datos GenBank, para que quede a disposición de la comunidad científica. Planeamos redactar un manuscrito en el que se describan los resultados que hemos obtenido.

En el laboratorio del tutor, estamos llevando a cabo la secuenciación, ensamblaje *de novo* y anotación del genoma mitocondrial de esta especie. La secuencia de ambos genomas orgánulares representa un primer paso hacia el desarrollo de herramientas moleculares para el estudio de esta especie.

8. Bibliografía

- Alkatib, S., Fleischmann, T.T., Scharff, L.B., y Bock, R. (2012a). Evolutionary constraints on the plastid tRNA set decoding methionine and isoleucine. *Nucleic Acids Res.* **40**: 6713-6724.
- Alkatib, S., Scharff, L.B., Rogalski, M., Fleischmann, T.T., Matthes, A., Seeger, S., Schöttler, M.A., Ruf, S., y Bock, R. (2012b). The contributions of wobbling and superwobbling to the reading of the genetic code. *PLoS Genet.* **8**: e1003076.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., y Lipman D.J. (2010). Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**: 403-410.
- Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., y Lipman D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**: 3389-3402.
- Asano, T., Tsudzuki, T., Takahashi, S., Shimada, H., y Kadowaki K. (2004). Complete nucleotide sequence of the sugarcane (*Saccharum officinarum*) chloroplast genome: a comparative analysis of four monocot chloroplast genomes. *DNA Res.* **11**: 93-99.
- Austin, J.R., Frost, E., Vidi, P-A., Kessler, F., y Staehelin A. (2006). Plastoglobules are lipoprotein subcompartments of the chloroplast that are permanently coupled to thylakoid membranes and contain biosynthetic enzymes. *Plant Cell* **18**: 1693-1703.
- Bausher, M.G., Singh, N.D., Mozoru, J., Lee, S-B., Jansen, R.K., y Daniell, H. (2006). The complete chloroplast genome sequence of *Citrus sinensis* (L.) Osbeck var. 'Ridge Pineapple': organization and phylogenetic relationships to other angiosperms. *BMC Plant Biol.* **6**: 21.
- Benkeblia, N. (2003). Antimicrobial activity of essential oil extracts of various onions (*Allium cepa*) and garlic (*Allium sativum*). *Lebensm.-Wiss. -Technol.* **37**: 263-268.
- Borlinghaus, J., Albrecht, F., Gruhlke, M.C., Nwachukwu, I.D., Slusarenko, A.J. (2014). Allicin: Chemistry and Biological Properties. *Molecules* **19**: 12591-12618.
- Boudreau, E., Takahashi, Y., Lemieux, C., Turmel, M., y Rochaix, J.D. (1997) The chloroplast *ycf3* and *ycf4* open reading frames of *Chlamydomonas reinhardtii* are required for the accumulation of the photosystem I complex. *EMBO J.* **16**: 6095-6104.
- Bungard, R.A. (2004). Photosynthetic evolution in parasitic plants: insight from the chloroplast genome. *Bioessays* **26**: 235-247.
- Buso, G.S., Paiva, M.R., Torres, A.C., Resende, F.V., Ferreira, M.A., Buso, J.A., y Dusi, A.N. (2008). Genetic diversity studies of Brazilian garlic cultivars and quality control of garlic-clover production. *Genet. Mol. Res.* **7**: 534-541.
- Cattolico, R.A. (1985). Chloroplast biosystematics: chloroplast DNA as a molecular probe. *Biosystems* **18**: 299-306.
- Chand, S.K., Nanda, S., Rout, E., y Joshi, R.K. (2015). Mining, characterization and validation of EST derived microsatellites from the transcriptome database of *Allium sativum* L. *Bioinformatics* **11**: 145.
- Chang, C-C., Lin, H-C., Lin, I-P., Chow, T-Y., Chen, H-H., Chen, W-H., Cheng, C-H., Lin, C-Y., Liu, S-M., Chang, C-C., y Chaw S-M. (2006). The chloroplast genome of *Phalaenopsis*

- aphrodite* (Orchidaceae): comparative analysis of evolutionary rate with that of grasses and its phylogenetic implications. *Mol. Biol. Evol.* **23**: 279-291.
- Chumley, T.W., Palmer, J.D., Mower, J.P., Fourcade, H.M., Calie, P.J., Boore, J.L., y Jansen, R.K. (2006). The complete chloroplast genome sequence of *Pelargoniumxhortorum*: organization and evolution of the largest and most highly rearranged chloroplast genome of land plants. *Mol. Biol. Evol.* **23**: 2175-2190.
- Clegg, M.T. (1993). Chloroplast gene sequences and the study of plant evolution. *Proc. Natl. Acad. Sci. USA* **90**: 363-367.
- Cunha, C.P., Hoogerheide, E.S., Zucchim, M.I., Monteiro, M. y Pinheiro, J.B. (2012). New microsatellite markers for garlic, *Allium sativum* (Alliaceae). *Am. J. Bot.* **99**: e17-19.
- Curtis, S.E., y Clegg M.T. (1984). Molecular evolution of chloroplast DNA sequences. *Mol. Biol. Evol.* **1**: 291-301.
- Daniell, H., Lee, S-B., Grevich, J., Saski, C., Quesada-Vargas, T., Guda, C., Tomkins, J., y Jansen, R.K. (2006). Complete chloroplast genome sequences of *Solanum bulbocastanum*, *Solanum lycopersicum* and comparative analyses with other Solanaceae genomes. *Theor. Appl. Genet.* **112**: 1503-1518.
- De Pamphilis, C.W., y Palmer, .JD. (1990). Loss of photosynthetic and chlororespiratory genes from the plastid genome of a parasitic flowering plant. *Nature* **348**: 337-339.
- Di Tommaso, P., Moretti, S., Xenarios, I., Orobítg, M., Montanyola, A., Chang, J.M., Taly, J.F., y Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res.* **39**: W13–W17.
- Do, H.D.K., Kim, J.S., y Kim, J.H. (2013). Comparative genomics of four liliales families inferred from the complete chloroplast genome sequence of *Veratrum patulum* O. Loes. (Melanthiaceae). *Gene* **530**: 229-235.
- Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783-791.
- Filyushin, M.A., Beletsky, A.V., Mazur, A.M., y Kochieva, E.Z. (2016). The complete plastid genome sequence of garlic *Allium sativum* L. *Mitochondrial DNA Part B: Resources* **1**: 831-832.
- Funk, H., Berg, S., Krupinska, K., Maier, U., y Krause, K. (2007). Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biol.* **7**: 45.
- García-Lampasona, S., Asprelli, P., y Burba, J.L. (2012). Genetic analysis of a garlic (*Allium sativum* L.) germplasm collection from Argentina. *Scientia Horticulturae* **138**: 183-189.
- Goulding, S., Wolfe, K., Olmstead, R., y Morden, C. (1996). Ebb and flow of the chloroplast inverted repeat. *Mol. Gen. Genet.* **252**: 195-206.
- Gray, M.W. (1999). Evolution of organellar genomes. *Curr. Opin. Genet. Dev.* **9**: 678-687.

- Guo, X.Y., Ruan, S.L., Hu, W.M., Ca, D.G., y Fan, L.J. (2008). Chloroplast DNA insertions into the nuclear genome of rice: the genes, sites and ages of insertion involved. *Funct. Integr. Genom.* **8**: 101-108.
- Howe, C.J., Barbrook, A.C., Koumandou, V.L., Nisbet, R.E.R., Symington, H.A., y Wightman, T.F. (2003). Evolution of the chloroplast genome. *Phil. Trans. R. Soc. Lond. B.* **358**: 99-107.
- Huang, Y.Y., Matzke, A.J., y Matzke, M. (2013). Complete sequence and comparative analysis of the chloroplast genome of coconut palm (*Cocos nucifera*). *PLoS One* **8**: e74736.
- Ingvarsson, P.K., Ribstein, S., y Taylor, D.R. (2003). Molecular evolution of insertions and deletion in the chloroplast genome of silene. *Mol. Biol. Evol.* **20**: 1737-1740.
- Ipek, M., Ipek, A., Almquist, S.G., y Simon, P.W. (2004). Demonstration of linkage and development of the first low-density genetic map of garlic, based on AFLP markers. *Theoretical Applied Genet.* **110**: 228-236.
- Jensen, P.E., Bassi, R., Boekema, E.J., Dekker, J.P., Jansson, S., Leister, D., Robinson, C., y Scheller, H.V. (2007). Structure, function and regulation of plant photosystem I. *Biochim. Biophys. Acta* **1767**: 335-352.
- Jheng, C.F., Chen, T.C., Lin, J.Y., Wu, W.L., y Chang, C.C. (2012). The comparative chloroplast genomic analysis of photosynthetic orchids and developing DNA markers to distinguish *Phalaenopsis* orchids. *Plant Sci.* **190**: 62-73.
- Jo, M.H., Ham, I.K., Moe, K., Kwon, S.-W., Lu, F.-H., Park, Y.-J., Kim, W.S., Kim, M.K., Kim, T., y Lee, E.M. (2012). Classification of genetic variation in garlic (*Allium sativum* L.) using SSR markers. *Australian J. Crop Sci.* **6**: 625-631.
- Kamenetsky, R., Faigenboim, A., Shemesh Mayer, E., Ben Michael, T., Gershberg, C., Kimhi, S., Esqira, I., Rohkin Shalom, S., Eshel, D., Rabinowitch, H.D., y Sherman, A. (2015). Integrated transcriptome catalogue and organ-specific profiling of gene expression in fertile garlic (*Allium sativum* L.). *BMC Genomics* **16**: 12.
- Kim, D.W., Jung, T.S., Nam, S.H., Kwon, H.R., Kim, A., Chae, S.H., Choi, S.H., Kim, D.W., Kim, R.N., y Park, H.S. (2009). GarlicESTdb: an online database and mining tool for garlic EST sequences. *BMC Plant Biol.* **9**: 61.
- Kim, S., Kim, M.S., Kim, Y.M., Yeom, S.I., Cheong, K., Kim, K.T., Jeon, J., Kim, S., Kim, D.S., Sohn, S.H., Lee, Y.H., y Choi, D. (2015a). Integrative structural annotation of de novo RNA-Seq provides an accurate reference gene set of the enormous genome of the onion (*Allium cepa* L.). *DNA Res.* **22**: 19-27.
- Kim, S., Park, J.Y., y Yang, T.J. (2015b). Comparative analysis of complete chloroplast genome sequences of a normal male-fertile cytoplasm and two different cytoplasms conferring cytoplasmic male sterility in onion. *J. Hortic. Sci. Biotechnol.* **90**: 459-468.
- Koonin, E.V. (2001). Computational genomics. *Curr. Biol.* **11**: R155-158.
- Kumar, S., Stecher, G., Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**: 1870-1874.

- Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.H., Rognes, T., y Ussery, D.W. (2007). RNAMmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* **35**: 3100-3108.
- Langmead, B., Trapnell, C., Pop, M., y Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25.
- Langmead, B., y Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357-359.
- Laslett, D., y Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucl. Acids Res.* **32**: 11-16.
- Leister, D. (2003). Chloroplast research in the genomic age. *Trends Genet.* **19**: 47-56.
- Li, G., Ma, X., Deng, L., Zhao, X., Wei, Y., Gao, Z., Jia, J., Xu, J., y Sun, C. (2015). Fresh Garlic Extract Enhances the Antimicrobial Activities of Antibiotics on Resistant Strains *in Vitro*. *Jundishapur J Microbiol.* **8**: e14814.
- Li, H., Cao, H., Cai, Y-F., Wang, J-H., Qu, S-P., y Huang, X-Q. (2014). The complete chloroplast genome sequence of sugar beet (*Beta vulgaris ssp. vulgaris*). *Mit. DNA* **25**: 209-211.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., y 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
- Lohse, M., Drechsel, O., Kahlau, S., y Bock, R. (2013). OrganellarGenome DRAW – a suite of tools for generating physical maps of plastid and mitochondrial genomes and visualizing expression data sets. *Nucleic Acids Res.* **41**: W575-W581.
- Lohse, M., Drechsel, O., y Bock, R. (2007). OrganellarGenomeDRAW (OGDRAW): A tool for the easy generation of high-quality custom graphical maps of plastid and mitochondrial genomes. *Curr. Genet.* **52**: 267-274.
- Lowe, T.M., y Eddy, S.R. (1997). tRNAscan-SE: A Program program for Improved improved Detection detection of Transfer transfer RNA Genes genes in Genomic genomic Sequencessequence. *Nucleic Acids Res.* **25**: 955-964.
- Maier, R.M., Neckermann, K., Igloi, G.L., y Kössel, H. (1995). Complete sequence of maize chloroplast genome: Gene content, hotspots of divergence and the tuning of genetic information by transcript editing. *J. Mol. Biol.* **251**: 614-628.
- Mardanov, A.V., Ravin, N.V., Kuznetsov, B.B., Samigullin, T.H., Antonov, A.S., Kolganova, T.V., y Skyabin, K.G. (2008). Complete sequence of the duckweed (*Lemna minor*) chloroplast genome: structural organization and phylogenetic relationships to other angiosperms. *J. Mol. Evol.* **66**: 555-564.
- Margulis, L. (1975). Symbiotic theory of the origin of eukaryotic organelles; criteria for proof. *Symp. Soc. Exp. Biol.* **29**: 21-38.
- Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M., y Kowallik, K.V. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* **393**: 162-165.

- Matsuo, M., Ito, Y., Yamauchi, R., y Obokata, J. (2005). The rice nuclear genome continuously integrates, shuffles, and eliminates the chloroplast genome to cause chloroplast-nuclear DNA flux. *Plant Cell* **17**: 665-675.
- Mes, T.H.M., Fritsch, R.M., Pollner, S., y Bachmann, K. (1999). Evolution of the chloroplast genome and polymorphic ITS regions in *Allium* subg. *Melanocrommyum*. *Genome* **42**: 237-247.
- Milne, I., Stephen, G., Bayer, M., Cock, P.J.A., Pritchard, L., Cardle, L., Shaw, P.D., y Marshall, D. (2013). Using Tablet for visual exploration of second-generation sequencing data. *Brief. Bioinform.* **14**: 193-202.
- Nakazono, M., y Hira, A. (1993). Identification of the entire set of transferred chloroplast DNA sequences in the mitochondrial genome of rice. *Mol. Gen. Genet.* **236**: 341-346.
- Nei, M., y Kumar, S. (2000). Molecular Evolution and Phylogenetics. *Oxford University Press*, New York.
- Notredame, C., Higgins, D.G., y Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**: 205-17.
- Ohiwa, T. (1977). Response of *Spirogyra* chloroplast to local illumination. *Planta* **136**: 7-11.
- Palmer, J.D. (1990). Contrasting modes and tempos of genome evolution in land plant organelles. *Trends Genet.* **6**: 115-120.
- Palmer, J.D. (1991). Plastid chromosomes: structure and evolution. En: *Molecular biology of plastids* (Bogorad L., editor). Academic Press.
- Peña-Iglesias, A. (1988). El ajo: virosis, fisiopatías y selección clonal y sanitaria. I Parte teórico-descriptiva. *Bol. San. Veg. Plagas* **14**: 461-483.
- Pires, J.C., Maureira, I.J., Givnish, T.J., Sytsma, K.J., Seberg, O., Petersen, G., Davis, J.I., Stevenson, D.W., Rudall, P.J., Fay, M.F., y Chase, M.W. (2006). Phylogeny, genome size, and chromosome evolution of the Asparagales. *Aliso* **22**: 287-304.
- Pyke, K.A. (1999). Plastid division and development. *Plant Cell* **11**: 549-556.
- Rochaix, J.D. (1997). Chloroplast reverse genetics: new insights into the function of plastid genes. *Trends Plant Sci.* **2**: 419-425.
- Rouze, P., Pavy, N., y Rombauts, S. (1999). Genome annotation: which tools do we have for it? *Curr. Opin. Plant Biol.* **2**: 90-95.
- Sagan, L. (1967). On the origin of mitosing cells. *J. Theoret. Biol.* **14**: 225-274.
- Saitou, N., y Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406-425.
- Savard, L., Li, P., Strauss, H., Chase, M. W., Michaud, M., y Bousquet, J. (1994). Chloroplast and nuclear gene sequences indicate Late Pennsylvanian time for the last common ancestor of extant seed plants. *Proc. Natl. Acad. Sci. USA* **91**: 5163-5167.
- Schattner, P., Brooks, A.N., y Lowe, T.M. (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* **33**: W686-689.

- Shalom, S.R., Gillett, D., Zemach, H., Kimhi, S., Forer, I., Zutahy, Y., Tam, Y., Teper-Bamnlker, P., Kamenetsky, R., y Eshel, D. (2015). Storage temperature controls the timing of garlic bulb formation via shoot apical meristem termination. *Planta* **242**: 951-962.
- Shemesh-Mayer, E., Ben-Michael, T., Rotem, N., Rabinowitch, H.D., Doron-Faigenboim, A., Kosmala, A., Perlikowski, D., Sherman, A., y Kamenetsky, R. (2015). Garlic (*Allium sativum* L.) fertility: transcriptome and proteome analyses provide insight into flower and pollen development. *Front. Plant Sci.* **6**: 271.
- Shemesh-Mayer, E., Winiarczyk, K., Blaszczyk, L., Kosmala, A., Rabinowitch, H.D., y Kamenetsky, R. (2013). Male gametogenesis and sterility in garlic (*Allium sativum* L.): barriers on the way to fertilization and seed production. *Planta* **237**: 103-120.
- Sheppard, A.E., y Timmis, J.N. (2009) Instability of plastid DNA in the nuclear genome. *PLoS Genet.* **55**: e1000323.
- Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B.Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J. Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H., y Sugiura, M. (1986). The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO J.* **5**: 2043-2049.
- Stein, L. (2001) Genome annotation: from sequence to biology. *Nat. Rev. Genet.* **2**: 493-503.
- Stern, D.B., y Astwood, J.D. (1986). Tripartite mitochondrial genome of spinach: physical structure, mitochondrial gene mapping, and locations of transposed chloroplast DNA sequences. *Nucl. Acids Res.* **14**: 5651-5666.
- Sugiura, M. (1992). The chloroplast genome. *Plant Mol. Biol.* **19**: 149-168.
- Sun, X., Zhou, S., Meng, F., y Liu, S. (2012). De novo assembly and characterization of the garlic (*Allium sativum*) bud transcriptome by Illumina sequencing. *Plant Cell Reports* **31**: 1823-1828.
- Sun, X.D., Ma, G.Q., Cheng, B., Li, H., y Liu, S.Q. (2013). Identification of differentially expressed genes in shoot apex of garlic (*Allium sativum* L.) using Illumina sequencing. *Journal of Plant Studies* **2**: 136-148.
- Sung, S.Y., Sin, L.T., Tee, T.T., Bee, S.T., y Rahmat, A.R. (2014). Effects of *Allium sativum* essence oil as antimicrobial agent for food packaging plastic film. *Innov. Food Sci. Emerg. Tech.* **26**: 406-414.
- Tanksley, S.D., y Pichersky, E. (1988). Organization and evolution of sequences in the plant nuclear genome. Gottlieb L.D., Jain S.K., editors. *Plant Evolutionary Biology. Chapman and Hall.* p: 55-83.
- Tatusova, T.A., y Madden, T.L. (1999). BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* **174**: 247-250.
- Timmis, J.N., Ayliffe, M.A., Huang, C.Y., y Martin, W. (2004) Endosymbiotic gene transfer: organelle genomes forge eukaryotic chromosomes. *Nat. Rev. Genet.* **55**: 123-135.

- Tsukazaki, H., Yaguchi, S., Sato, S., Hirakawa, H., Katayose, Y., Kanamori, H., Kurita, K., Itoh, T., Kumagai, M., Mizuno, S., Hamada, M., Fukuoka, H., Yamashita, K., McCallum, J.A., Shigyo, M., y Wako, T. (2015). Development of transcriptome shotgun assembly-derived markers in bunching onion (*Allium fistulosum*). *Molecular Breeding* **35**: 1-11.
- Von Kohn, C., Kielkowska, A., y Havey M.J. (2013). Sequencing and annotation of the chloroplast DNAs and identification of polymorphisms distinguishing normal male-fertile and male-sterile cytoplasms of onion. *Genome* **56**: 737-42.
- Weeden, N.F. (1981). Genetic and biochemical implications of the endosymbiotic origin of the chloroplast. *J. Mol. Evol.* **17**: 133-139.
- Whatley, J.M., (1981). Chloroplast evolution—ancient and modern. *Ann. N.Y. Acad. Sci.* **361**: 154-165.
- Wicke, S., Schneeweiss, G., de Pamphilis, C., Müller, K., y Quandt, D. (2011). The evolution of the plastid chromosome in land plants: gene content, gene order, gene function. *Plant Mol. Biol.* **76**: 273-297.
- Wolfe, K.H., Morden, C.W., y Palmer, J.D. (1992). Function and evolution of a minimal plastid genome from a nonphotosynthetic parasitic plant. *Proc. Natl. Acad. Sci. USA* **89**:10648-10652
- Yamaguchi, K., von Von Knoblauch, K., y Subramanian, A.R. (2000). The plastid ribosomal proteins. Identification of all the proteins in the 30 S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* **275**: 28455-28465.
- Yamaguchi, K., y Subramanian, A.R. (2000). The plastid ribosomal proteins. Identification of all the proteins in the 50 S subunit of an organelle ribosome (chloroplast). *J. Biol. Chem.* **275**: 28466-28482.
- Yang, M., Zhang, X., Liu, G., Yin, Y., Chen, K., Yun, Q., Zhao, D., Al-Mssallem, I.S., y Yu., J. (2010). The complete chloroplast genome sequence of date palm (*Phoenix dactylifera dactylifera dactylifera* L.). *PLoS One* **5**: e12762.
- Yukawa, M., Tsudzuki, T., y Sugiura, M. (2006). The chloroplast genome of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*: complete sequencing confirms that the *Nicotiana sylvestris* progenitor is the maternal genome donor of *Nicotiana tabacum*. *Mol. Genet. Genomics.* **275**: 367-373.
- Zerbino, D.R., y Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**: 821-829.
- Zewdie, Y., Havey M.J., Prince, J.P., y Jenderek, M.M. (2005). The first genetic linkages among expressed regions of the garlic genome. *J. American Soc. Horticultural Science* **130**: 569-574.
- Zhou, S.M., Chen, L.M., Liu, S.Q., Wang, X.F., y Sun, X.D. (2015). De Novo Assembly and Annotation of the Chinese Chive (*Allium tuberosum* Rottler ex Spr.) Transcriptome Using the Illumina Platform. *PLoS One* **10**: e0133312.
- Zimorski, V., Ku, C., Martin, W.F., y Gould, S.B. (2014)). Endosymbiotic theory for organelle origins. *Curr. Opin. Microbiol.* **22**: 38-48.

Zuckerlandl, E., y Pauling, L. (1965). Evolutionary divergence and convergence in proteins. Edited in *Evolving Genes and Proteins* by V. Bryson and H.J. Vogel, pp. 97-166. *Academic Press*, New York.

