

Filogenómica de elementos ultraconservados en genomas de cloroplasto: las fabáceas como caso de estudio

Phylogenomics of ultraconserved elements in chloroplast genomes: Fabaceae as a case study

Cristian Giovanni Díaz¹ y Juan Manuel Anzola Lagos²

Cómo citar:

Díaz, C. G., & Anzola Lagos, J. M. (2021). Filogenómica de elementos ultraconservados en genomas de cloroplasto: las fabáceas como caso de estudio. *Ingeciencia*, 6, 7-17.

¹ Estudiante de Biología de la Universidad Central y miembro del Semillero de Bioinformática y Biología Computacional. Correo: cdiazp5@ucentral.edu.co

² Profesor asociado de la Facultad de Ingeniería y Ciencias Básicas de la Universidad Central. Biólogo de la Universidad Nacional de Colombia y doctor en Ciencias Biológicas de Texas A&M University. Correo: janzolal@ucentral.edu.co

Resumen

En la última década, se han caracterizado regiones genómicas que se encuentran altamente conservadas entre dos o más taxones, denominadas *elementos ultraconservados*. Las *regiones flanqueantes* a estos elementos les brindan una variabilidad que permite usarlos como marcadores moleculares. Esta aproximación se ha explorado poco en plantas. El objetivo de este artículo es identificar elementos ultraconservados y regiones flanqueantes que puedan ser útiles en análisis filogenómicos. Como caso de estudio se seleccionaron las fabáceas (leguminosas) y fueron procesados 642 genomas por medio de un flujo de trabajo bioinformático que permitió localizar y extraer regiones ultraconservadas, así como regiones flanqueantes. Se identificaron 46 elementos ultraconservados, que en conjunto abarcan el 17% del genoma de referencia. Se caracterizaron 31 genes asociados a estas regiones y 18 regiones flanqueantes de interés. De las 18 regiones flanqueantes, una presenta potencial para emplearla como marcador en grupos de la familia *Fabaceae*.

Palabras clave: bioinformática, elementos ultraconservados, *fabaceae*, filogenómica, genómica, leguminosas.

Abstract

In the last decade, genomic regions that are highly conserved between two or more taxa, termed ultraconserved elements, have been characterized. The flanking regions of these provide variability that may allow their use as molecular markers. This approach has been seldom explored in plants and the present study aims to identify ultraconserved elements and flanking regions that might be useful in phylogenomic analyses. Fabaceae (legumes) were selected as a case study and 642 genomes were processed by means of a bioinformatics workflow that allowed the localization and extraction of ultraconserved regions as well as flanking regions. Forty-six ultraconserved elements were identified, which together comprise 17% of the reference genome. Thirty-one genes associated with these regions and 18 flanking regions of interest were characterized. Of the 18 flanking regions, one has potential for use as a marker in groups of the family *Fabaceae*.

Keywords: bioinformatics, fabaceae, genomics, legumes, phylogenomics, ultraconserved elements

Introducción

El cloroplasto constituye la unidad fundamental en la maquinaria energética de las plantas. Da lugar a procesos de suma importancia para la vida, además de cumplir un rol significativo en el crecimiento y desarrollo de estos organismos (Sakamoto & Takami, 2018). Su origen como organelo se le atribuye a un proceso de endosimbiosis debido a la presencia de su propio genoma. Al ser de carácter endosimbiótico, es semiautónomo, pues conserva gran parte de su maquinaria transcripcional y traduccional para la expresión de su información genética.

Con los avances en tecnología de secuenciación y en los métodos bioinformáticos se ha generado información sobre la estructura, evolución y funcionamiento del genoma cloroplástico. Por esto, hoy en día sabemos que el genoma es circular, altamente conservado y presenta una menor complejidad estructural en comparación con el genoma nuclear (Li *et al.*, 2020a). Su tamaño varía entre los 115 a 185 kb (Sorkheh *et al.*, 2017; Xu *et al.*, 2018), en el que se incluyen de 110 a 130 genes codificantes de proteínas (Wang *et al.*, 2020). Está compuesto por cuatro regiones: dos regiones pequeñas de repetición invertida (*inverted repeat region*: IRa e IRb), una región pequeña de copia única (*small single copy*: SSC) y una región grande de copia única (*large single copy*: LSC) (Xue *et al.*, 2019).

Debido a que el organelo se hereda por vía materna en gran parte las angiospermas, y a su baja tasa de mutación, el genoma cloroplástico ha sido el foco de atención en estudios filogenómicos (Lim *et al.*, 2020). Esto ha permitido la caracterización de marcadores moleculares basados en regiones de dicho genoma, con lo que se ha conseguido identificar y clasificar grupos en diferentes categorías taxonómicas. En las reconstrucciones filogenéticas se han empleado secuencias pertenecientes a regiones codificantes (*rbcl*, *matK*, *atpB*, *atpB-rbcL*, *ndhF*, *rpl16*, *atpB*), intrones y regiones intergénicas; no obstante, estas regiones suelen carecer de variación en especies que se encuentran estrechamente relacionadas, aún más en aquellas que han divergido recientemente (Dong *et al.*, 2012). Por consiguiente, se tiene la necesidad de encontrar regiones que brinden mayor eficiencia en la resolución taxonómica y, por lo tanto, que tengan mayor universalidad en las plantas.

En la última década, se han hallado regiones de ADN que poseen un alto grado de conservación ($\geq 80\%$ de identidad sobre ≥ 100 pb) entre dos o más taxones denominadas elementos ultraconservados. Estas poseen regiones flanqueantes en las que se aloja una variabilidad que incrementa con respecto a la distancia de la secuencia conservada (Faircloth *et al.*, 2012). Entre las ventajas que otorgan los elementos ultraconservados se encuentran la fácil alineación entre los taxones y la utilidad de la región flanqueante en análisis de genómica comparativa (Zhang *et al.*, 2019).

El uso de los elementos ultraconservados en análisis filogenómico se ha empleado en diversos taxones animales (Zhang *et al.*, 2019); sin embargo, no han sido explorados en profundidad en genomas vegetales, ya sea nuclear o cloroplástico. A la par, las preguntas por responder se basan principalmente en las escalas evolutivas y filogeográficas. En lo que nos concierne, este es el primer estudio que realiza una caracterización tanto de los elementos ultraconservados como de las regiones flanqueantes en plantas y genera una aproximación a dichas regiones como marcadores moleculares.

Métodos

Recuperación de secuencias

Se descargaron del Centro Nacional de Información Biotecnológica de Estados Unidos (NCBI) genomas completos de cloroplasto para *Embryophyta* (plantas terrestres). La ecuación de búsqueda fue “(txid3193[Organism:exp] AND 100000:400000000[Sequence Length] NOT UNVERIFIED [Title])” con el filtro “organelo: cloroplasto”. A las secuencias de los genomas se les hizo una asignación taxonómica por medio de *script in-house*.

El multifasta anotado constituye 13 253 genomas, distribuidos en 18 clases, 101 órdenes y 345 familias. El grupo de estudio (*Fabaceae*) se ve representado por 645 secuencias, distribuidas en 110 géneros y 485 especies.

Identificación de elementos ultraconservados

Se aplicó el flujo de trabajo diseñado por Faircloth (2017) al grupo estudio; de este se escogió un genoma base (*Inga leiocalycina*, ID KT428296) y como criterio de selección se tuvo en cuenta el nivel de anotación y ensamblaje.

Se generaron lecturas simuladas de 150 pb para los 642 genomas de *Fabaceae*, usando ART v20,160,605 (Huang *et al.*, 2012) y se mapearon al genoma de referencia empleando Stampy v1.0.31 (Lunter & Goodson, 2011). Las lecturas no asignadas se eliminaron con samtools v1.9 (Danecek *et al.*, 2021) y los archivos BAM resultantes se convirtieron a formato BED para ordenar y filtrar los datos utilizando Bedtools v2.28 (Quinlan & Hall, 2010). La lista resultante de loci representa las regiones potencialmente ultraconservadas en el grupo estudio.

El flujo de trabajo realizado se encuentra almacenado en el repositorio de GitHub (<https://bit.ly/3OZ1Cbv>).

Análisis de elementos ultraconservados

En Chlorobox (<https://chlorobox.mpimp-golm.mpg.de/>) se generó la anotación de los elementos ultraconservados para obtener información funcional. Con Geneious Prime se generó el mapeo de estos para identificar las regiones en las que se localizan y generar una medición indirecta de la conservación de posibles genes. En la medición se tuvo en cuenta la longitud de la región ultraconservada sobre la longitud del gen.

La visualización de los elementos ultraconservados identificados se realizó en Chloroplot (<https://irscope.shinyapps.io/Chloroplot/>).

Estimación de distancias en las regiones flanqueantes

La localización y extracción de las regiones flanqueantes a los elementos ultraconservados de interés se generó en Geneious Prime a partir de una búsqueda manual en el genoma de referencia. La extracción de las coincidencias de las regiones flanqueantes del genoma de referencia en los genomas del grupo estudio se hizo con BLAST+ (Camacho *et al.*, 2008) y luego se generó el alineamiento con Muscle (Edgar, 2004).

Por último, la estimación de distancia de los alineamientos sin *gaps* fue ejecutado en MEGA-X (Koichiro *et al.*, 2021) con el modelo de sustitución de Kimura 2-parámetros con delección completa de los *gaps*. Una vez computada la estimación de la distancia, se tomó en cuenta la media de estas para estimar la región con mayor variabilidad.

Resultados y discusión

Análisis de elementos ultraconservados

A partir de la anotación taxonómica se extrajo el grupo de estudio; para este, el criterio de selección fue escoger una de las familias con mayor número de representantes de los genomas descargados, que corresponde a la familia *Fabaceae*. En este estudio identificamos 46 elementos ultraconservados (figura 1) a partir de 642 genomas. Estas regiones constituyen 30,376 pb, que representan el 17% del genoma de referencia.

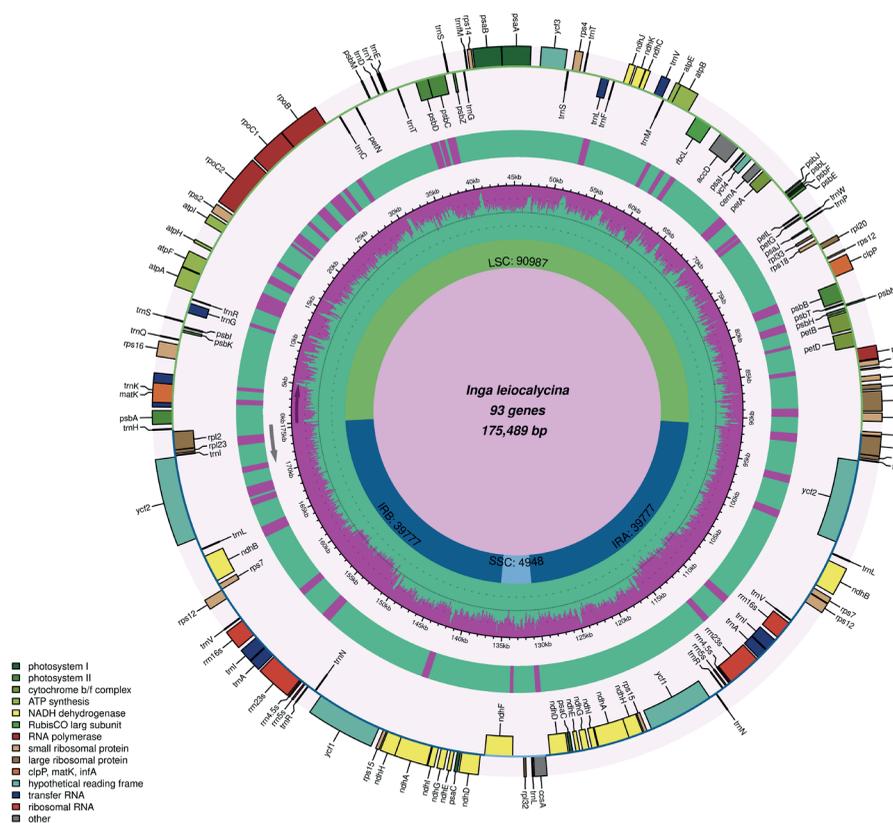


Figura 1. Mapa genómico del cloroplasto de *I. leiocalycina* y la distribución de los elementos ultraconservados (representados en el círculo flotante intermedio).

Fuente: elaboración propia.

El genoma del cloroplasto de *Inga leiocalycina* tiene un tamaño de 175 489 pb. Este posee la estructura de cuatro regiones genómicas, típicas del cloroplasto. El plastoma consta de 93 genes, incluidos 8 ribosomales, 37 tRNAs y 48 genes codificantes de proteínas. Identificamos 31 genes asociados a los elementos ultraconservados. Estas regiones abarcan desde el 30% hasta el 100% del tamaño de un gen (tabla 1).

Tabla 1. Resumen de los genes asociados a los elementos ultraconservados hallados en el genoma de *I. leiocalycina*

Categoría genes	Grupo de genes	Nombre del gen (IndC)
Genes relacionados con la fotosíntesis	Subunidades del fotosistema	<i>psbC</i> (0,45), <i>psbD</i> (0,84), <i>psbL</i> (1), <i>psbE</i> (0,85), <i>psbB</i> (0,41)
	Subunidades del complejo citocromo b/f	<i>petA</i> (0,76)
	Subunidad grande de la Rubisco	<i>rbcl</i> (0,47)
	Subunidades de la ATP sintasa	<i>atpA</i> (0,55), <i>atpF</i> (0,58), <i>atpI</i> (0,64), <i>atpB</i> (0,28), <i>atpH</i> (1)
	Subunidades de la NADH deshidrogenasa	<i>ndhF</i> (0,15), <i>ndhA</i> (0,21), <i>ndhB</i> (0,36)
Autorreplicación	ADN dependiente de ARN polimerasa	<i>rpoC2</i> (0,41), <i>rpoC1</i> (0,58), <i>rpoB</i> (0,48)
	Subunidad pequeña del ribosoma	<i>rps11</i> (0,96)
	Subunidad grande del ribosoma	<i>rps3</i> (0,66), <i>rpl2</i> (0,58), <i>rpl23</i> (0,89)
	Genes ARN ribosomal	<i>rrn23</i> (0,34), <i>rrn16</i> (0,35)
	Genes de ADN de transferencia	<i>trnG-UCC</i> (0,39), <i>trnL-UAA</i> (0,72), <i>trnI-GAU</i> (0,76)
Otros genes	Maturasa	<i>matK</i> (0,38)
	Gen de síntesis, citocromo tipo C	<i>ccsA</i> (0,47)
Genes con función desconocida	Marcos de lectura abiertos conservados	<i>ycf2</i> (0,33)

Nota: IndC: Índice de conservación. 0 = no presenta conservación; 1 = presenta conservación en la totalidad del gen, de acuerdo con los elementos ultraconservados hallados en 642 genomas de la familia *Fabaceae*.

Fuente: elaboración propia.

Se encontró que los genes *psbL* y *atpH* se hallan conservados en su totalidad, resultado que es consistente con lo citado por Steane (2005), quien menciona que estos se hallan altamente conservados entre varios taxones de plantas superiores. Otros genes por mencionar son *rps11*, *rpl23*, *psbE* y *psbD* debido a su alto nivel de conservación. Lo anterior cobra sentido al revisar los productos protéicos, que se encuentran involucrados en mecánicas importantes para el funcionamiento de la planta, ya sea a nivel fotosintético o en la autorreplicación (Wicke *et al.*, 2011).

Además, están los genes con un bajo nivel de conservación: de acuerdo con la métrica, encontramos a *ndhF*, *ndhA* y *ndhB*. Estos se encuentran relacionados con la maquinaria fotosintética, que hace parte del complejo NDH. A lo largo de la evolución del cloroplasto este complejo ha experimentado una alta variabilidad, que ha llevado a la pérdida de partes del complejo en algunos grupos, los cuales son necesarios en otros (Shikanai, 2016). *rpl2*, *rbcl* y *rpoC1* forman parte del grupo de genes que exhiben niveles de variabilidad intermedios; actualmente se emplean en estudios moleculares como códigos de barras de ADN.

Estimación de distancias en las regiones flanqueantes

De los 46 elementos ultraconservados identificados, se estudiaron particularmente 18 regiones flanqueantes. El criterio de selección se relacionó con el potencial de estos como marcadores moleculares en estudios filogenéticos. Para ello, se deben cumplir ciertos criterios como una alta variabilidad y una longitud del fragmento compatible con tecnologías de secuenciación actuales.

Las matrices generadas son una medida del grado de divergencia entre las secuencias; en promedio, presentan un valor de distancia de 0,06, lo cual indica que son altamente conservadas. No obstante, se destaca la extracción 0 que presentó valor de distancia de 0,13. Este fragmento se ubica en el gen codificante *matK* y su tamaño es de 863 pb (tabla 2).

Tabla 2. Distancias computadas de regiones flanqueantes

Extracción	Longitud	Dist.	Gen/ región	Extracción	Longitud	Dist.	Gen/ región
0	863	0,13	<i>matK</i>	9	604	0,05	<i>rpoB</i>
1	641	0,05	<i>atpA</i> – RI	10	549	0,06	<i>atpB</i> – RI
2	212	0,05	<i>atpF</i> – RI	11	737	0,06	<i>rbclL</i>
3	202	0,05	RI	12	157	0,07	<i>psbJ</i> – RI
4	716	0,08	<i>rps2</i> – <i>rpoC2</i>	13	906	0,04	<i>psbB</i> - <i>psbT</i>
5	603	0,09	<i>rpoC2</i>	14	814	0,05	<i>petB</i>
6	537	0,05	<i>rpoC2</i> - <i>rpoC1</i>	15	274	0,03	<i>rps3</i> - RI
7	279	0,07	<i>rpoC1</i>	16	869	0,04	<i>ycf2</i> (IRb)
8	417	0,05	<i>rpoB</i>	17	635	0,04	<i>ycf2</i> (IRa)

RI: región intergénica; dist.: distancia estimada

Nota: los rangos van de 0 a 1. Entre mayor sea el valor, mayor variabilidad

Fuente: elaboración propia.

Según estas estimaciones, nos encontramos que la mayoría de las regiones flanqueantes entre los elementos ultraconservados en el grupo de 642 genomas de *Fabaceae* carecen de la variabilidad suficiente para ser empleadas como marcadores moleculares. Esto se confirma al visualizar las topologías generadas y observar un alto número de nodos colapsados (figura 2). Sin embargo, la extracción 0 es un candidato por considerar en la identificación de miembros de la familia *Fabaceae*, debido a que la topología que generó no presentó nodos colapsados. Cabría evaluar su eficiencia en la tipificación, ya sea a nivel de género o especie.

Los genes cloroplásticos se encuentran bajo diferentes tipos de selección a lo largo del tiempo: pueden estar bajo selección positiva, y serían los responsables de adaptaciones funcionales, mientras otros se encuentran bajo selección negativa o purificadora, asociándose a funciones esenciales dentro de la maquinaria fotosintética o de autorreplicación (Bai *et al.*, 2021).

La primera conjetura sobre la alta conservación obtenida en las distancias se basa en genes como *rpoC2*, *rpoC1*, *rps2* y *rps3*, presentes en las regiones flanqueantes, que se encuentran bajo selección negativa. Esto refleja una posible presión selectiva por mantener las funciones de los genes, relacionadas con las subunidades del fotosistema o las subunidades ribosomales (Li *et al.*, 2020b). Asimismo, se ha de mencionar que generalmente los análisis filogenómicos no suelen abordar un número grande de géneros (para este caso, n = 110). Por esto, la alta conservación en las distancias de las regiones flanqueantes se puede atribuir a la presión selectiva purificadora dentro de la familia *Fabaceae*.

Otra posible conjetura por considerar sobre la alta conservación de las regiones flanqueantes obtenidas para el grupo es lo mencionado por Li *et al.* (2020b) y Shi *et al.* (2020), quienes plantean que las regiones codificantes se encuentran con alto nivel de conservación, en comparación con las regiones no codificantes y genes como *ycf1*, *ycf2*, *rpl23*, *rps3*, *rps18*, *accD*, *rpoC1*, *ycf4* y *rpl32*, poseen identidad relativamente baja, siendo inversamente proporcional a su conservación. Por lo tanto, aumentar el número de géneros —y a su vez de especies— incrementa la conservación de las regiones flanqueantes obtenidas por medio de esta aproximación.

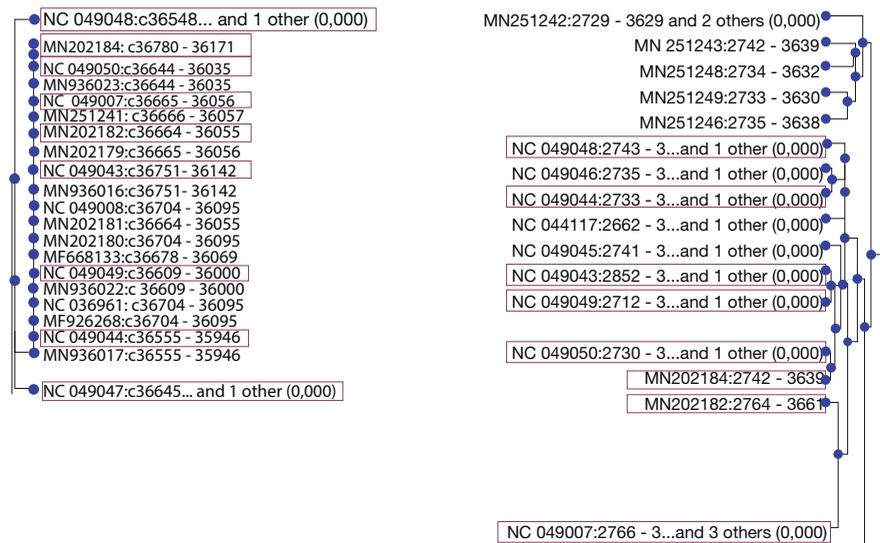


Figura 2. Topologías generadas a partir de regiones flanqueantes. Izquierda: nodo colapsado (extracción 9, Dist: 0,05). Derecha: nodo que presenta mayor estructura filogenética (extracción 0, Dist: 0,13).

Fuente: elaboración propia.

Conclusiones

El flujo de trabajo permitió identificar 46 regiones ultraconservadas entre 642 genomas, que se distribuyen principalmente en la región de copia única grande. Asociadas a estas 46 regiones, se seleccionaron e identificaron 18 regiones flanqueantes. La gran mayoría de ellas presenta altos niveles de conservación y topologías con ramas colapsadas en árboles

filogenéticos. Solo una de las regiones, con una media de distancia entre sus miembros de 0,13, generó un árbol filogenético con estructura sin colapsos, por lo que podría constituirse como posible marcador molecular para la familia *Fabaceae*. Se hace entonces necesario diseñar *primers* y estudios con el fin de evaluar esta región como potencial marcador molecular.

Agradecimientos

Agradezco al doctor Juan Manuel Anzola, por su apoyo y consejos en esta fase de mi proceso académico. Igualmente, a los compañeros del Semillero de Bioinformática y Biología Computacional por todas sus recomendaciones sobre este trabajo.

Referencias

- Bai, H. R., Oyebanji, O., Zhang, R., & Yi, T. S. (2021). Plastid phylogenomic insights into the evolution of subfamily Dialioideae (Leguminosae). *Plant Diversity*, 43(1), 27–34. <https://doi.org/10.1016/j.pld.2020.06.008>
- Dong, W., Liu, J., Yu, J., Wang, L., & Zhou, S. (2012). Highly Variable Chloroplast Markers for Evaluating Plant Phylogeny at Low Taxonomic Levels and for DNA Barcoding. *PLoS ONE*, 7(4), e35071. <https://doi.org/10.1371/journal.pone.0035071>
- Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5), 1792–1797. <https://doi.org/10.1093/nar/gkh340>
- Edwards, S. v. (2009). Is a new and general theory of molecular systematics emerging? *Evolution* (Vol. 63, Issue 1, pp. 1–19). <https://doi.org/10.1111/j.1558-5646.2008.00549.x>
- Faircloth, B. C. (2017). Identifying conserved genomic elements and designing universal bait sets to enrich them. *Methods in Ecology and Evolution*, 8(9), 1103–1112. <https://doi.org/10.1111/2041-210X.12754>
- Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61(5), 717–726. <https://doi.org/10.1093/sysbio/sys004>
- Koichiro Tamura, Glen Stecher, and Sudhir Kumar (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. *Molecular Biology and Evolution* 38:3022–3027
- Li, C., Zhao, Y., Xu, Z., Yang, G., Peng, J., & Peng, X. (2020a). Initial Characterization of the Chloroplast Genome of *Vicia sepium*, an Important Wild Resource Plant, and Related Inferences About Its Evolution. *Frontiers in Genetics*, 11. <https://doi.org/10.3389/fgene.2020.00073>

- Li, P., Lou, G., Cai, X., Zhang, B., Cheng, Y., & Wang, H. (2020b). Comparison of the complete plastomes and the phylogenetic analysis of *Paulownia* species. *Scientific Reports*, *10*(1), 1–9. <https://doi.org/10.1038/s41598-020-59204-y>
- Lim, L. W. K., Chung, H. H., & Hussain, H. (2020). Complete chloroplast genome sequencing of sago palm (*Metroxylon sagu* Rottb.): Molecular structures, comparative analysis and evolutionary significance. *Gene Reports*, *19*(March), 100662. <https://doi.org/10.1016/j.genrep.2020.100662>
- Lunter, G., & Goodson, M. (2011). Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research*, *21*(6), 936–939. <https://doi.org/10.1101/gr.111120.110>
- Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842. <https://doi.org/10.1093/bioinformatics/btq033>
- Sakamoto, W., & Takami, T. (2018). Chloroplast DNA dynamics: Copy number, quality control and degradation. *Plant and Cell Physiology* (Vol. 59, Issue 6, pp. 1120–1127). Oxford University Press. <https://doi.org/10.1093/pcp/pcy084>
- Shi, C., Han, K., Li, L., Seim, I., Lee, S. M. Y., Xu, X., Yang, H., Fan, G., & Liu, X. (2020). Complete Chloroplast Genomes of 14 Mangroves: Phylogenetic and Comparative Genomic Analyses. *BioMed Research International*, *2020*. <https://doi.org/10.1155/2020/8731857>
- Sorkheh, K., Dehkordi, M.K., Ercisli, S., Hegedus, A., & Halász J. (2017). Comparison of traditional and new generation DNA markers declares high genetic diversity and differentiated population structure of wild almond species. *Sci Rep*. Jul 20;7(1):5966. <https://doi.org/10.1038/s41598-017-06084-4>
- Steane, D. A. (2005). Complete nucleotide sequence of the chloroplast genome from the Tasmanian blue gum, *Eucalyptus globulus* (Myrtaceae). *DNA Research*, *12*(3), 215–220. <https://doi.org/10.1093/dnares/dsi006>
- Shikanai, T. (2016). Chloroplast NDH: A different enzyme with a structure similar to that of respiratory NADH dehydrogenase. *Biochimica et Biophysica Acta - Bioenergetics* (Vol. 1857, Issue 7, pp. 1015–1022). Elsevier B.V. <https://doi.org/10.1016/j.bbabi.2015.10.013>
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E.S., Fischer, A., Bock, R., & Greiner, S. (2017) GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research* *45*: W6–W11
- Wang, W., Yang, T., Wang, H. L., Li, Z. J., Ni, J. W., Su, S., & Xu, X. Q. (2020). Comparative and Phylogenetic Analyses of the Complete Chloroplast Genomes of Six Almond Species (*Prunus* spp. L.). *Scientific Reports*, *10*(1), 1–10. <https://doi.org/10.1038/s41598-020-67264-3>
- Wicke, S., Schneeweiss, G. M., dePamphilis, C. W., Müller, K. F., & Quandt, D. (2011). The evolution of the plastid chromosome in land plants: Gene content, gene order, gene function. *Plant Molecular Biology* (Vol. 76, Issues 3–5, pp. 273–297). <https://doi.org/10.1007/s11103-011-9762-4>
- Xu, X., Wen, J., Wang, W. et al. (2018). The complete chloroplast genome of the threatened *Prunus cerasoides*, a rare winter blooming cherry in the Himala-

- yan region. *Conservation Genet Resour* 10, 499–502. <https://doi.org/10.1007/s12686-017-0859-1>
- Xue, S., Shi, T., Luo, W., Ni, X., Iqbal, S., Ni, Z., Huang, X., Yao, D., Shen, Z., & Gao, Z. (2019). Comparative analysis of the complete chloroplast genome among *Prunus mume*, *P. armeniaca*, and *P. salicina*. *Horticulture Research*, 6(1), 1–13. <https://doi.org/10.1038/s41438-019-0171-1>.
- Zhang, Y. M., Williams, J. L., & Lucky, A. (2019). Understanding UCEs: A Comprehensive Primer on Using Ultraconserved Elements for Arthropod Phylogenomics. *Insect Systematics and Diversity*, 3(5). <https://doi.org/10.1093/isd/ixz016>