



Ciencia Nicolaita 86

ISSN: 2007-7068



Universidad
Michoacana
de San Nicolás
de Hidalgo

Evaluación bioinformática de datos Hyb-Seq para tres especies de *Salvia* subgénero *Calosphace* (Lamiaceae)

Bioinformatic evaluation of Hyb-Seq data for three species of *Salvia* subgenus *Calosphace* (Lamiaceae)

Daniel Simbrón Romero, Carlos Alonso Maya Lastra y Sabina I. Lara Cabrera.*

Para citar este artículo: Daniel Simbrón Romero, Carlos Alonso Maya Lastra y Sabina I. Lara Cabrera*, 2022. Evaluación bioinformática de datos Hyb-Seq para tres especies de *Salvia* subgénero *Calosphace* (Lamiaceae). Ciencia Nicolaita no. 86, 69-83. DOI: <https://doi.org/10.35830/cn.vi86.689>



Historial del artículo:

Recibido: 30 de agosto de 2022

Aceptado: 5 de noviembre de 2022

Publicado en línea: diciembre de 2022



Ver material suplementario



Correspondencia de autor: sabina.lara@umich.mx



Términos y condiciones de uso: <https://www.cic.cn.umich.mx/cn/about/privacy>



Envíe su manuscrito a esta revista: <https://www.cic.cn.umich.mx/cn/about/submissions>

Evaluación bioinformática de datos Hyb-Seq para tres especies de *Salvia* subgénero *Calosphace* (Lamiaceae)

Bioinformatic evaluation of Hyb-Seq data for three species of *Salvia* subgenus *Calosphace* (Lamiaceae)

Daniel Simbrón Romero,¹ Carlos Alonso Maya Lastra² y Sabina I. Lara Cabrera.^{1*}

¹Facultad de Biología, Universidad Michoacana de San Nicolás de Hidalgo, México.

²Department of Ecology, Evolution and Environmental Biology, Columbia University, 1200 Amsterdam Ave., New York, USA.

Resumen

El subgénero *Calosphace* es el más diverso del género *Salvia*; su distribución abarca el continente americano, siendo México y Centroamérica las regiones de mayor diversidad. Estudios filogenéticos del grupo con secuenciación Sanger para marcadores de núcleo y cloroplasto, han resuelto las relaciones profundas para *Salvia*; sin embargo, en las relaciones superficiales, frecuentemente se observan politomías. Con los avances tecnológicos e informáticos, se han desarrollado técnicas de secuenciación masiva (NGS del inglés Next Generation Sequencing) que, al aplicarlas para inferencias filogenéticas, han resultado en filogenias más robustas a niveles superficiales. En este estudio se analizaron 30 especies del subgénero *Calosphace* con el protocolo NGS Hyb-Seq para poner a prueba la monofilia de tres especies representadas por dos muestras cada una de distinta procedencia y determinar su posición filogenética. Se amplificaron 98 genes nucleares y con el pipeline HybPhylo-Maker se generó el mapeo, alineamiento y filtrado de los genes amplificados. Finalmente, se construyeron árboles de genes y árboles de especies con FastTree y ASTRAL, respectivamente. La inferencia filogenética resuelve cinco clados principales, dos de los cuales son coincidentes a secciones *sensu* Epling. De las tres especies en las que se evalúa su monofilia, *S. hispanica* y *S. purpurea* se recuperan como monofiléticas, con altos valores de probabilidad posterior local; sin embargo, *S. polystachia* muestra no ser monofilética. Adicionalmente, se muestran dos alineamientos MAUVE con pocos reordenamientos en los pares de muestras de *S. hispanica* y *S. purpurea* y reacomodos importantes para el par de *S. polystachia*. Datos obtenidos con Hyb-seq, pueden ayudar a resolver relaciones filogenéticas en niveles superficiales de la filogenia de *Salvia* y, a su vez, nos permite identificar qué especies requieren de otros estudios complementarios para esclarecer sus relaciones evolutivas.

Palabras clave: filogenia, secuenciación masiva, ASTRAL, *Calosphace*



Abstract

Salvia subgenus *Calosphace* is the most diverse of the five subgenera. Of wide distribution in America, being Mexico and Central America the most diverse areas. Phylogenetic studies utilizing Sanger sequencing for both nuclear and chloroplast markers, have solved the deep relationships but shallow relationships generally remain unsolved with polytomies. Recent technological and bioinformatic advances resulted in Next Generation Sequencing (NGS) which when applied to phylogenetic inferences result in robust hypotheses at shallow levels. Here we evaluate 30 species in *Calosphace* through next gen Hyb-Seq to assess monophyly of three species' with two samples each from distinct provenances and define their phylogenetic position. 98 nuclear genes were amplified, mapping, alignment and gene filtering were performed in HybPhyloMaker pipeline. Gene trees and species trees were built in FastTree y ASTRAL respectively. Five main clades are resolved, two of which agree to sections *sensu* Epling. *Salvia hispanica* y *S. purpurea* are both monophyletic species with high local posterior probability, however *S. polystachia* is not. Additionally, two MAUVE analyses were performed, few rearrangements are shown between *S. hispanica* and *S. purpurea* paired samples, non the less important rearrangements are shown for *S. polystachia*. Hyb-seq data aid in solving phylogenetic relationships at shallow level in *Salvia* phylogeny, and allows us to identify species which need to be subjected to additional and more thorough analyses to better clarify species relationships.

Key words: Phylogeny, High throughput sequencing, ASTRAL, *Calosphace*

Introducción

El género *Salvia* L. es uno de los más diversos en el mundo (Mabberly, 1997), contando con cerca de 1,000 especies y una distribución casi global (Drew *et al.*, 2017). *Salvia* pertenece a la familia Lamiaceae y se clasifica en cinco subgéneros: *Audibertia* (Benth.) Epling ex. Walker, *Calosphace* (Benth.) Epling, *Leonia* Benth, *Salvia* y *Sclarea* Benth. (Will y Classen-Bockhoff, 2017). *Calosphace* es el subgénero más diverso de *Salvia* (Epling, 1939) con aproximadamente 600 especies distribuidas en América, siendo México la zona de mayor diversidad con 295 especies (González-Gallegos *et al.*, 2020).

La sistemática y clasificación de los seres vivos, ha ido cambiando a lo largo del tiempo, habiendo dominado por muchas décadas el estudio de la morfología, anatomía, química y distribución (Montero-Castro *et al.*, 2016). Para *Calosphace*, la revisión taxonómica más completa es la de Epling (1939) y subsecuentes contribuciones (Epling 1939, 1940, 1941, 1944, 1947, 1951; Epling y Mathias 1957; Epling y Játiva 1963, 1966), resultando en 102 secciones, clasificadas con base en caracteres morfológicos del cuerpo vegetativo, reproductivo y la distribución geográfica.

Las últimas cuatro décadas han visto un cambio significativo en cómo se hace sistemática y qué fuentes de evidencia se emplean para proponer relaciones filogenéticas. La secuenciación Sanger (Sanger *et al.*,

1977) fue el método predominante hasta mediados de los años 2000 (Barrett *et al.*, 2016), con este método se obtienen secuencias de 500 a 3 000 pares de bases (pb) y han permitido resolver relaciones filogenéticas profundas, principalmente en niveles superiores en la jerarquía clasificatoria, como orden, familia, géneros, e incluso a nivel de especies; sin embargo, en muchos casos, no resuelve y provee baja resolución para niveles superficiales, particularmente en grupos de radiación evolutiva reciente (Saarela *et al.*, 2015).

Este ha sido el caso del subgénero *Calosphace*, donde se resolvieron muchas relaciones filogenéticas profundas; sin embargo, las relaciones superficiales no fueron totalmente resueltas (Fragoso-Martínez *et al.*, 2018; Jenks *et al.*, 2013; Walker, 2006), particularmente en el llamado “núcleo *Calosphace*” (Walker, 2006) donde se encuentra la mayoría de las especies del subgénero y las de mayor plasticidad fenotípica.

El avance tecnológico y bioinformático ha permitido el desarrollo de nuevos métodos y protocolos que revolucionan las ciencias biológicas (Barrett *et al.*, 2016) mediante una serie de técnicas de secuenciación de nueva generación, permitiendo obtener lecturas de secuencias, abarcando gran parte de genomas, e incluso genomas completos (Barrett *et al.*, 2016). Conforme se van optimizando los protocolos en grupos modelo como *Arabidopsis thaliana* (L.) Heynh, se

empiezan a emplear en organismos no-modelo, permitiendo abordar preguntas evolutivas complejas como eventos de poliploidía, hibridación, introgresión y coalescencia estocástica (Brassac y Blattner, 2015; Eriksson *et al.*, 2018; Uribe-Convers *et al.*, 2016).

A la fecha, se han reportado cinco estudios de NGS en *Calosphace*, dos con el protocolo de Enriquecimiento Híbrido Anclado o AHE, por sus siglas en inglés (Lemmon *et al.*, 2012), donde obtuvieron propuestas filogenéticas robustas con altos valores de soporte (Fragoso-Martínez *et al.*, 2017; Kriebel *et al.*, 2019). También se reporta el protocolo Hyb-Seq (Weitemier *et al.*, 2014), con muestreo del genoma de cloroplasto y cistrones ribosomales para tres secciones *sensu* Epling cercanas filogenéticamente (Olvera-Mendoza *et al.*, 2020) y para 98 genes nucleares en 75 especies (Lara-Cabrera *et al.*, 2021; Pérez-García, 2019).

Hyb-Seq es un protocolo que amplifica selectivamente regiones exónicas con base en el diseño de sondas específicas. En este estudio, se evalúan los productos de Hyb-Seq para 33 muestras de *Calosphace* pertenecientes a un clado monofilético (Lara-Cabrera *et al.*, 2021; Pérez-García, 2019), con el propósito de evaluar la monofilia de tres especies interesantes y útiles.

De las tres especies, la más famosa es *Salvia hispanica* L., la chía que tiene cerca de 5 000 años de haberse domesticado, llegando a tener gran importancia para las culturas precolombinas que la utilizaban como pseudocereal, para extracción de aceite y en herbolaria (Sosa-Baldivia *et al.*, 2018). *Salvia polystachia* Ort. es una especie con extensa distribución en

Mesoamérica y de alta plasticidad fenotípica, lo que ocasionalmente hace difícil su identificación, posee compuestos químicos útiles con potencial médico (Maldonado y Ortega 2000; Calzada *et al.*, 2010), además, en propuestas filogenéticas (Walker, 2006) se ha observado que no es monofilética. Finalmente, *Salvia purpurea* Cav. es una especie muy llamativa que comienza a utilizarse en horticultura que, además, como *S. polystachia*, posee elementos químicos valiosos con potencial médico (Cuevas-Morales *et al.*, 2022; Velazco-Lezama *et al.*, 2021) (Figura 1).

Metodología

Se muestrearon 30 especies de *Salvia* subgénero *Calosphace* en un clado monofilético (Lara-Cabrera *et al.*, 2021; Pérez-García, 2019) con *Salvia leucantha* Cav. como grupo externo con base en ambas filogenias y tres especies representadas por dos muestras provenientes de localidades distintas, *S. hispanica* [85] de Oaxaca (Tenorio, P & Dieringer, D. 10685. México, Oaxaca, Teotitlán de Flores Magón, Cerro Verde, 23 Km al NE de Teotitlán, carr. Teotitlán-Huautla. 12/10/1985) y *S. hispanica* [19] de Guerrero (González, s.n. México. Guerrero. Mpio. Leonardo Bravo. La Laguna, 6 km adelante del poblado. 17°43' 54"N, 99°44'40"W, 1997), *S. polystachia* [163] de Morelos (Lara-Cabrera, S. 163. México. Morelos, 5 km al S de Tetela del Volcán camino a Temoac. 18° 50' 55"N, 98° 44' 16"W, 2030m. 11/15/2009), y *S. polystachia* [065] de Michoacán (Bedolla-García, B. 65. México. Michoa-



Figura 1. De izquierda a derecha. *Salvia hispanica* (Foto: Yocupitzia Ramírez), *Salvia polystachia* (Foto: S.I. Lara-Cabrera) y *Salvia purpurea* (Foto: S.I. Lara-Cabrera).

cán. En los alrededores de Tlalpujahua. Mpio. Tlalpujahua. 19° 48' 05" N 100° 11' 40" W, 2633m. 11/14/2009) y *S. purpurea* [103] de Chiapas (Bedolla-García, B. 103. México. Chiapas. A las afueras de Siltepec. Mpio. Siltepec. 15° 26' 57"N, 92° 15' 58"W. 2874m. 12/28/2009) y *S. purpurea* [156] de Morelos (Lara-Cabrera, S. 156. México. Morelos. Carretera Tepoztlán San Juan Tlacotenco, 19°.00.08N; 99°.06.99W. 10/31/2009). Las condiciones de extracción, diseño de sondas, amplificación y procesamiento de Hyb-seq, se pueden consultar en Lara-Cabrera y colaboradores (2021).

El procesamiento y análisis de las secuencias se realizó usando distintos programas incluidos en el pipeline HybPhyloMaker v.1.6.4 (Fér y Schmickl, 2018), como se menciona a continuación. Con base en los 98 genes amplificados con las sondas, se generó la pseudoreferencia (alineamiento de secuencias de las sondas diseñadas para amplificar los genes deseados), contra la que se mapearon las lecturas (Lara-Cabrera *et al.*, 2021) y se eliminaron los duplicados en las secuencias usando BWA (Li y Durbin, 2009). Las secuencias se separaron con base en la pseudoreferencia en genes nucleares independientes y se utilizó MAFFT v.7.429 (Katoh y Standley, 2013) para alinearlos. Para la reconstrucción filogenética de los árboles de genes se utilizó FastTree-2.1.11 (Price *et al.*, 2010). Estos árboles de genes fueron usados para reconstruir el árbol de especies en ASTRAL 5.6.3 (Mirarab *et al.*, 2014). Los nodos con valores de soporte inferior al 70 % de probabilidad posterior local, fueron colapsados. Finalmente, el árbol fue editado en TreeGraph (Stöver y Müller, 2010) y Photoshop (Adobe Inc. 2019).

Con base en el árbol de especies, se seleccionaron dos clados para realizar análisis de MAUVE (Darling *et al.*, 2004) y evaluar si hay reacomodos genómicos entre las tres especies representadas por dos muestras usando la opción Progressive MAUVE con los valores predeterminados del programa. Los alineamientos genómicos MAUVE se han utilizado para alinear genomas completos de mitocondria (Lee *et al.*, 2019), cloroplasto (Wang *et al.*, 2016) y núcleo (ADN ribosomal) (Olvera-Mendoza *et al.*, 2019; Hegarthy *et al.*, 2019) por mencionar algunos, para comparar las anotaciones a nivel genes y determinar si hay homología y reacomodos, ya sean en orden u orientación de genes entre varios genomas (Armstrong *et al.*, 2018).

Resultados

El mapeo en BWA presentó promedio de lecturas de 1,614,216 pares de bases (pb), en un rango de 300,000 a 3,000,000 pb; este amplio intervalo de longitud de las lecturas apunta a baja homogeneidad en el mapeo y/o captura de las sondas (Tabla. 1). En el caso de especies con dos muestras, se obtuvieron diferencias en longitud de lecturas pareadas de 841,172 pb en el par *S. hispanica*; 1,123,299 pb en el par *S. polystachia*; y 1,826,728 pb en el par *S. purpurea*.

Se alinearon 96 genes en MAFFT (Apéndice A), con promedio de longitud de secuencia de 1,237.21 pb e intervalo de 154 pb - 3,322 pb. De los 98 genes amplificados, se tuvieron que quitar dos de análisis subsecuentes al no presentarse en *S. leucantha*.

El árbol filogenético de especies (Figura 2) presenta 5 clados bien definidos, con altos valores de probabilidad posterior local, incluyendo a las secciones *Curtiflorae* y *Lavanduloidae sensu* Epling y otros 3 clados con especies clasificadas en distintas secciones. Las muestras pareadas de *S. hispanica* y *S. purpurea*, se observan como monofiléticas y con altos valores de probabilidad posterior local, pero *S. polystachia* no se recupera como especie monofilética.

El primer análisis de MAUVE incluye parte del clado D, con probabilidad posterior local de 0.97 con *Salvia rhyacophila* (Fernald) Epling como especie hermana de *S. hispanica*, este alineamiento presenta un solo reordenamiento y diferencia en longitud del alineamiento de ca. 30,000 pb. (Figura 3). El segundo análisis MAUVE incluye los clados E2.2.1. y E2.2.2; presenta importantes diferencias en longitud y reordenamientos, particularmente en los primeros 50,000 pb, sobre todo en *S. polystachia* [065]; menos reordenamientos entre *S. polystachia* [163] y *S. purpurea* [156] y más entre *S. polystachia* [065] y *S. tiliifolia* (Figura 4)

Tabla 1

Resultado del mapeo con BWA para las 33 muestras analizadas, indicando el porcentaje de lecturas pareadas, número de lecturas y lecturas pareadas. Se resaltan las muestras con mayor y menor cantidad de lecturas.

Especies	Porcentaje de lecturas mapeadas	Número de lecturas mapeadas	Número de lecturas pareadas
<i>Salvia areolata</i> Epling	40.333	1,762,389	2,104,618
<i>Salvia brachyodonta</i> Briq.	36.357	1,551,950	2,031,227
<i>Salvia chiapensis</i> Fernald	31.296	648,248	983,441
<i>Salvia connivens</i> Epling	40.962	1,588,767	1,687,323
<i>Salvia curtiflora</i> Epling	40.846	1,956,676	2,097,640
<i>Salvia curviflora</i> Benth	29.565	436,539	572,554
<i>Salvia decora</i> Epling	33.85	737,425	861,407
<i>Salvia farinacea</i> Benth.	34.679	1,510,414	1,923,431
<i>Salvia filipes</i> Benth.	38.018	2,066,606	2,384,762
<i>Salvia greggii</i> A. Gray	36.604	1,715,118	2,060,304
<i>Salvia helianthemifolia</i> Benth.	42.39	2,241,950	2,331,029
<i>Salvia tiliifolia</i> Vahl.	33.45	1,201,132	1,687,650
<i>Salvia hispanica</i> L. [19]	35.147	889,091	1,175,011
<i>Salvia hispanica</i> L. [85]	31.226	271,532	333,839
<i>Salvia iodantha</i> Fernald	42.859	1,706,012	1,734,365
<i>Salvia lavanduloides</i> Kunth	30.813	969,661	1,359,272
<i>Salvia leucantha</i> Cav.	42.081	2,815,039	2,946,595
<i>Salvia longispicata</i> M.Martens & Galeotti	26.9	569,094	860,396
<i>Salvia longistyla</i> Benth.	38.947	2,108,733	2,383,302
<i>Salvia mexicana</i> L.	41.206	1,957,267	2,077,869
<i>Salvia nervata</i> M.Martens & Galeotti	35.551	1,100,385	1,297,246
<i>Salvia occidua</i> Epling	26.673	457,891	735,646
<i>Salvia perblanda</i> Epling	38.548	1,027,185	1,124,198
<i>Salvia plurispicata</i> Epling	42.149	1,759,249	1,785,830
<i>Salvia polystachia</i> Ort. [065]	39.443	1,488,092	1,613,417
<i>Salvia polystachia</i> Ort. [163]	42.878	2,701,015	2,736,716
<i>Salvia purepecha</i> Bedolla & Lara Cabrera	30.374	938,863	1,237,170
<i>Salvia purpurea</i> Cav. [103]	37.429	1,532,314	1,741,100
<i>Salvia purpurea</i> Cav. [156]	41.367	1,861,188	1,912,355

Especies	Porcentaje de lecturas mapeadas	Número de lecturas mapeadas	Número de lecturas pareadas
<i>Salvia rhyacophila</i> (Fernald) Epling	26.986	290,255	488,674
<i>Salvia roscida</i> Fernald	23.372	427,666	711,865
<i>Salvia tiliifolia</i> Vahl.	35.046	1,653,494	2,033,111
<i>Salvia tonaticensis</i> Ramamoorthy ex Lara Cabrera, Bedolla & Zamudio	38.012	1,972,024	2,255,818

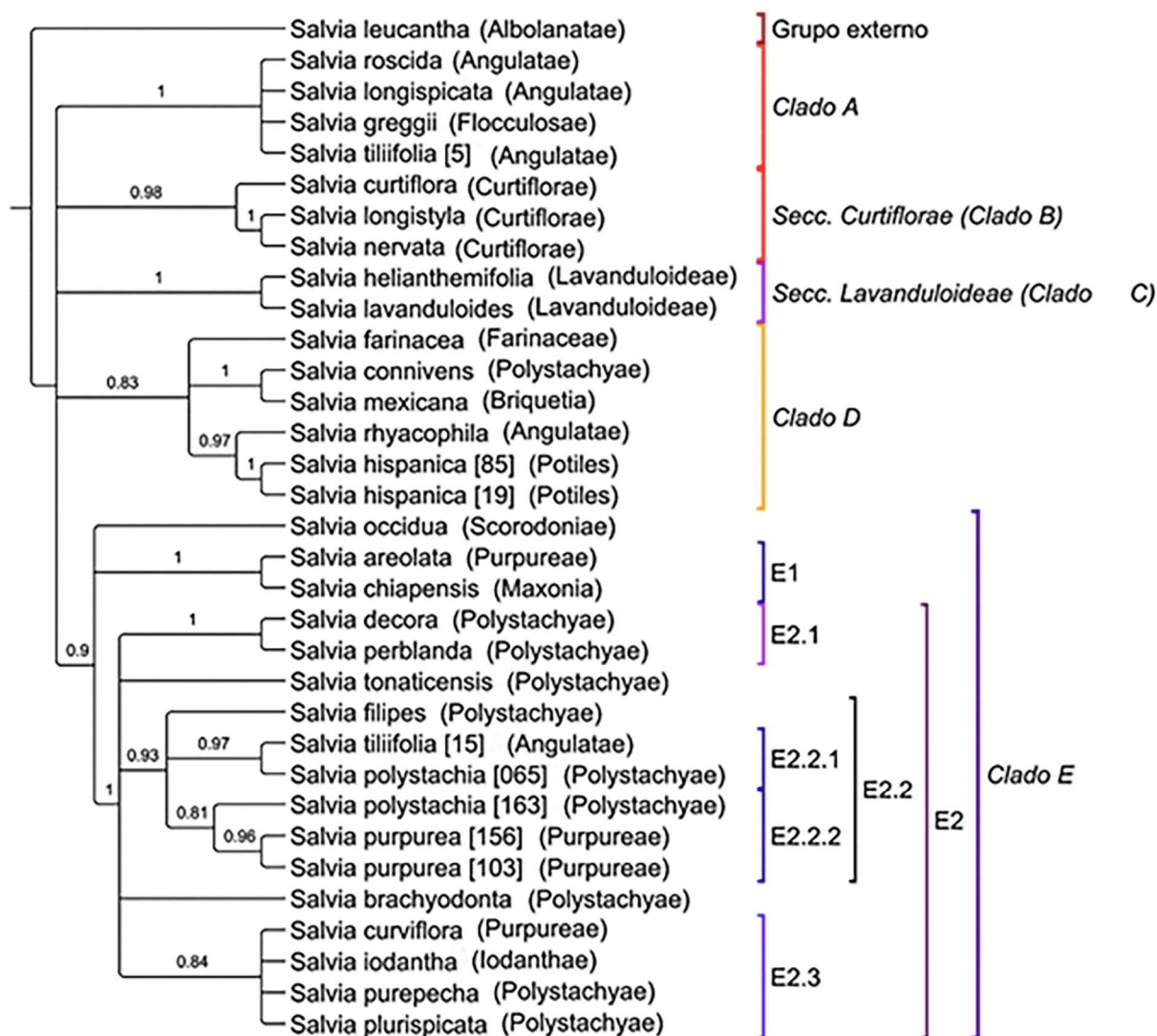


Figura 2. Árbol de especies reconstruido en ASTRAL. Los diferentes clados están indicados con corchetes de distintos colores. Los valores numéricos sobre las ramas del árbol señalan los valores de probabilidad posterior local.

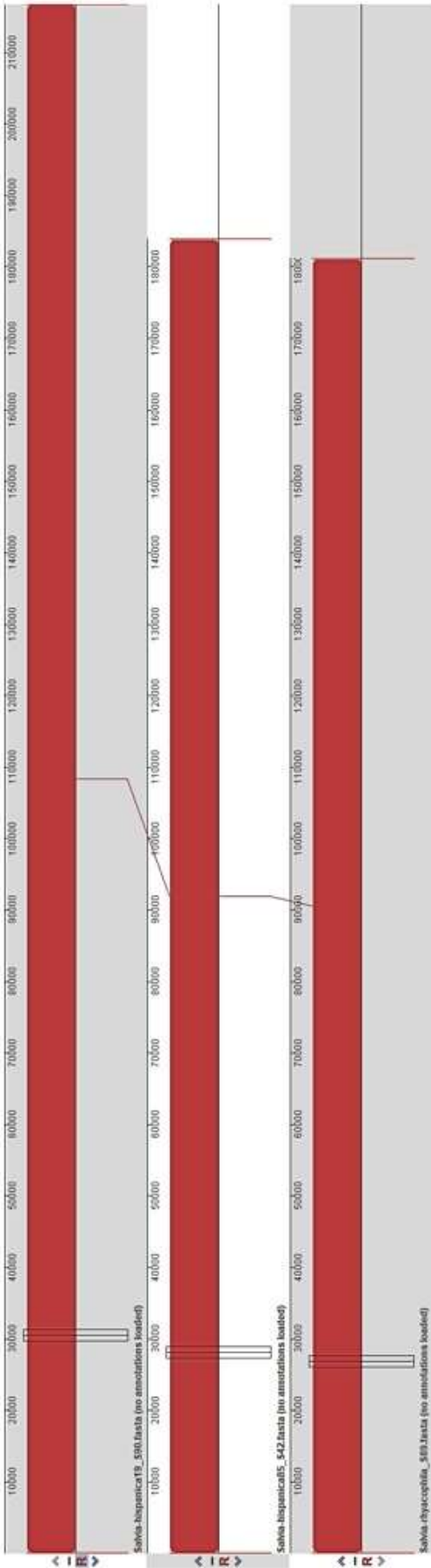


Figura 3. Análisis de reordenamiento de secuencias empleando MAUVE para el renglón inferior al superior *S. rhyacophila*, *S. hispanica* [85] y *S. hispanica* [19]. La línea negra perpendicular indica el fragmento/posición

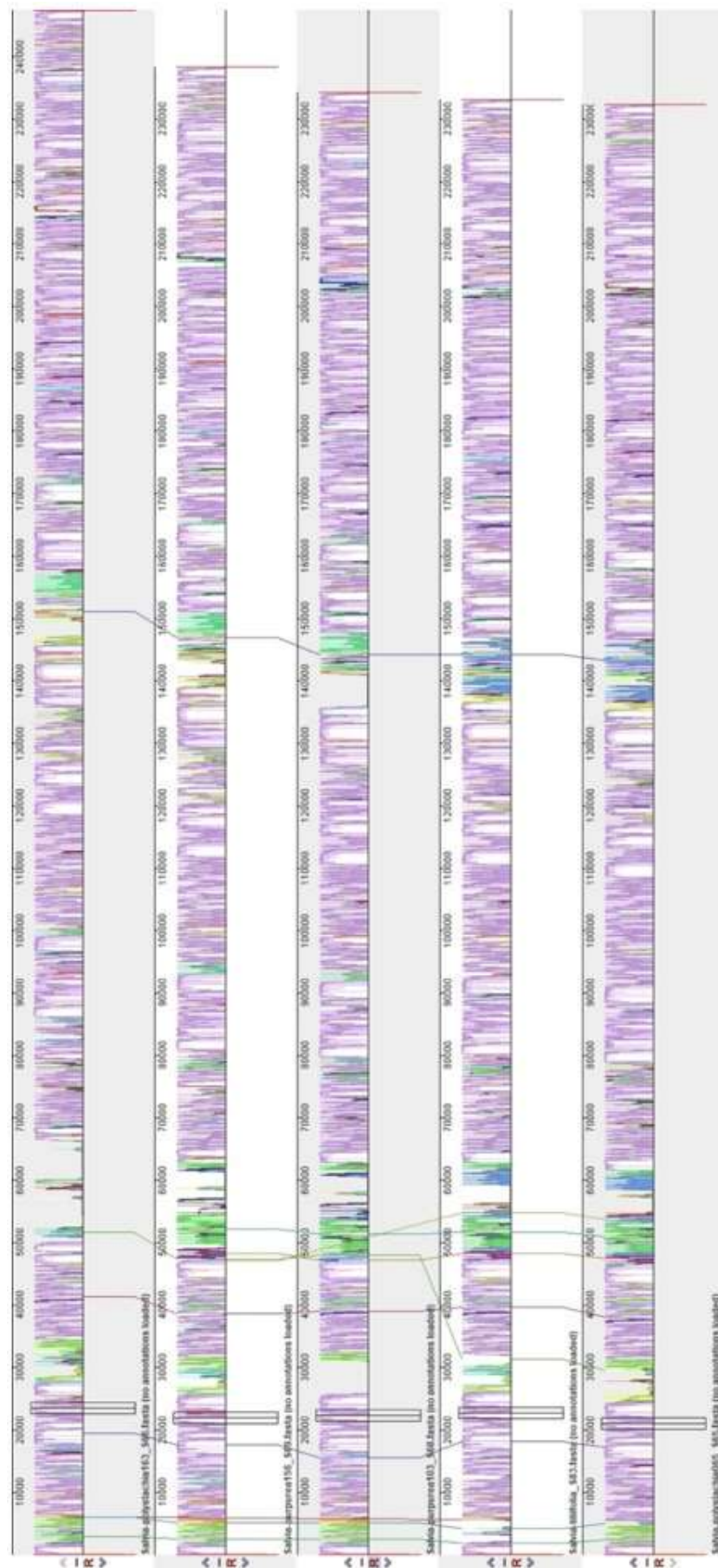


Figura 4. Análisis de reordenamiento de secuencias empleando MAUVE para cinco muestras de *Salvia* ubicadas en los clados E.2.2.1 y E.2.2.2 del árbol ASTRAL. Los renglones, de abajo arriba, muestran a *S. polystachia* [65], *S. tiliifolia* [83], *S. purpurea* [156] y *S. polystachia* [163]. Las secciones en la secuencia en diferente color indican áreas con diferencias y reordenamientos en las secuencias.

Discusión

El árbol de especies de ASTRAL, con base en 96 genes nucleares, está bien resuelto al mantener las secciones *sensu* Epling *Curtiflorae* y *Lavaduloideae* y con clados soportados por altos valores de probabilidad posterior local y un clado con integrantes principalmente de las secciones *Polystachyae* y *Purpureae*. Sin embargo, al representar este un muestreo pequeño, comparado con la cantidad de especies ubicadas en el “núcleo *Calosphace*”, creemos que en futuros análisis será beneficioso que se incorporen otras especies a estos clados.

En este estudio, hacemos un primer acercamiento para determinar la monofilia de tres especies representadas por dos individuos procedentes de poblaciones geográficamente distantes; si bien se trata de un muestreo representativo de los rangos de distribución, es la primera vez que se hace datos NGS, mostrando que la técnica empleada es robusta y amena para utilizarse en muestreo más amplios para determinar la monofilia de especies. Confirmamos la monofilia de *S. hispanica* y *S. purpurea*, concordando con estudios previos (Lara-Cabrera *et al.*, 2021; Pérez-García, 2019). *S. hispanica*, se considera hoy en día un superalimento (Sosa-Baldivia *et al.*, 2018), por lo que determinar sus relaciones filogenéticas, monofilia y grupos hermanos, es de relevancia para futuros esfuerzos de mejoramiento genético, siendo *S. rhyacophila* de la sección Angulatae su especie hermana. La estrecha relación entre las especies *S. polystachia* y *S. purpurea*, fue observada desde el estudio de Walker (2006) y recuperada también en el estudio de Jenks y colaboradores (2013). Por su parte, en este estudio, nuevamente se presenta la cercanía filogenética entre estas dos especies, con una de las muestras de *S. polystachia* perteneciendo al grupo hermano de las *S. purpurea*. *Salvia polystachia* no se muestra como monofilética, pero *S. purpurea* es monofilética y hermana de la primera.

Las relaciones inferidas en el árbol ASTRAL de especies, adicionalmente se correlacionan con los alineamientos en MAUVE para las tres especies representadas por dos muestras. Estudios de alineamientos genómicos por MAUVE, buscan nucleótidos homólogos entre dos o más genomas, pero en el caso de especies cercanamente relacionadas, han presentado

problemas para distinguir mutaciones locales (inserción-delección) de mutaciones a mayor escala (ganancia y pérdida de segmentos genómicos grandes producto de eventos de recombinación). Aquí utilizamos progressive MAUVE que mejora las comparaciones genómicas entre muestras de especies cercanamente emparentadas, que pudieron pasar por eventos de recombinación, reacomodos, duplicación de segmentos, así como por pérdida y ganancia de regiones (Darling *et al.*, 2019). Los pocos reacomodos genómicos para las muestras de las dos especies monofiléticas y extensos reacomodos en *S. polystachia* y su cercana relación con *S. purpurea*, indican que en *S. polystachia* pueden haber ocurrido eventos de recombinación complejos, probablemente involucradas otras especies de *Salvia* y que se requerirá utilizar de otras herramientas, como estudios filogeográficos para definir una *S. polystachia* monofilética.

Los resultados aquí obtenidos, demuestran que el protocolo Hyb-Seq es una buena opción para análisis filogenéticos, al proveer resolución de clados a nivel profundo y superficial en la mayoría de los casos, dada la gran cantidad de datos obtenidos, observándose altos valores de soporte; asimismo, se recuperan *S. hispanica* y *S. purpurea* como especies monofiléticas. Sin embargo, *S. polystachia* no es monofilética, por lo que Hyb-Seq también ayuda a identificar especies de relaciones complejas dentro del “núcleo *Calosphace*” que, por lo tanto, deberán seguir siendo estudiadas.

Agradecimientos

Al Proyecto 8.16 de la Coordinación de la Investigación Científica otorgado a SILC; y a los revisores por sus comentarios.

Bibliografía

- Armstrong, J., Fiddes, I.T., Diekhans, M. y Paten, B., 2019, Whole-Genome Alignment and Comparative Annotation: *Annual Review of Animal Biosciences*, 15, 41-64. PMID: 30379572; PMCID: PMC6450745. doi: 10.1146/annurev-animal-020518-115005
- Barrett, C.F., Bacon, C.D., Antonelli, A., Cano, Á. y Hofmann, T., 2016, An introduction to plant phylogenomics with a focus on palms: *Botanical Journal of the Linnean Society*, 182(2), 234-255. ISSN: 0024-4074 (print) 1095-8339 (web), <https://doi.org/10.1111/boj.12399>

- Brassac, J. y Blattner, F.R., 2015, Species-Level Phylogeny and Polyploid Relationships in *Hordeum* (Poaceae) Inferred by Next-Generation Sequencing and In Silico Cloning of Multiple Nuclear Loci: *Systematic Biology*, 64(5), 792-808. ISSN: 1063-5157 (print) 1076-836X (web). <https://doi.org/10.1093/sysbio/syv035>
- Calzada, F., Yepez-Mulia, L., Tapia-Contreras, A., Bautista, E., Maldonado, E. y Ortega, A., 2010, Evaluation of the antiprotozoal activity of neo-clerodane type diterpenes from *Salvia polystachya* against *Entamoeba histolytica* and *Giardia lamblia*: *Phytotherapy Research: An International Journal Devoted to Pharmacological and Toxicological Evaluation of Natural Product Derivatives*, 24(5), 662-665. <https://doi.org/10.1002/ptr.2938>
- Cuevas-Morales, C., Zavala-Ocampo, L.M., Miguel-Chávez, S., González-Trujano, M.E., Basurto-Pena, F.A., Muñoz-Ocoteo, V. y Aguirre-Hernández, E., 2022, Pharmacological evaluation of the antinociceptive activity and phytochemical analysis of active extracts of *Salvia purpurea* Cav.: *Botanical Sciences*, 100(2), 383-396. ISSN 2007-4476 (online) ISSN 2007-4298 (print). <https://doi.org/10.17129/botsci.3013>
- Darling, A.C.E., Mau, B., Blattner, F.R. y Perna, N.T., 2004, Mauve: Multiple Alignment of Conserved Genomic Sequence with Rearrangements: *Genome Research*, 14(7), 1394-1403. ISSN: 1081-9051 (print) 1549-5469 (web). <https://doi.org/10.1101/gr.2289704>
- Drew, B.T., González-Gallegos, J.G., Xiang, C.L., Kriebel, R., Drummond, C.P., Walked, J.B. y Sytsma, K.J., 2017, *Salvia* united: The greatest good for the greatest number: *TAXON*, 66(1), 133-145. ISSN: 1196-8175 (web). <https://doi.org/https://doi.org/10.12705/661.7>
- Epling, C., 1939, A revision of *Salvia* subgenus *Calosphace*: *Repert. Spec. Nov. Regni Veg. Beih*, 110. ISSN: 0233-187X (print).
- Epling, C., 1940, Supplementary Notes on American Labiatae: *Bulletin of the Torrey Botanical Club*, 67(6), 509-534. ISSN: 0040-9618 (web). <https://doi.org/10.2307/2480972>
- Epling, C., 1941, Supplementary Notes on American Labiatae-II: *Bulletin of the Torrey Botanical Club*, 68(8), 552-568. ISSN: 0040-9618 (web). <https://doi.org/10.2307/2481456>
- Epling, C., 1944, Supplementary Notes on American Labiatae-III: *Bulletin of the Torrey Botanical Club*, 71(5), 484-497. ISSN: 0040-9618 (web). <https://doi.org/10.2307/2481241>
- Epling, C., 1947, Supplementary Notes on American Labiatae-IV: *Bulletin of the Torrey Botanical Club*, 74(6), 512-518. ISSN: 0040-9618 (web). <https://doi.org/10.2307/2481876>
- Epling, C., 1951, Supplementary Notes on American Labiatae-V: *Brittonia*, 7(3), 129-142. ISSN: 0007-196X (print). <https://doi.org/10.2307/2804702>
- Epling, C. y Játiva, C., 1963, Supplementary notes on American Labiatae-VIII: *Brittonia* 15(4), 366-376. ISSN: 0007-196X (print). <https://doi.org/10.2307/2805381>
- Epling, C. y Játiva, C., 1966, Supplementary notes on American Labiatae-IX: *Brittonia* 18(3), 255-265. ISSN: 0007-196X (print). <https://doi.org/10.2307/2805366>
- Epling, C. y Mathias, M.E., 1957, Supplementary notes on American Labiatae-VI: *Brittonia*, 8(4), 297-313. ISSN: 0007-196X (print). <https://doi.org/10.2307/2804980>
- Eriksson, J.S., de Sousa, F., Bertrand, Y.J.K., Antonelli, A., Oxelman, B. y Pfeil, B.E., 2018, Allele phasing is critical to revealing a shared allopolyploid origin of *Medicago arborea* and *M. strasseri* (Fabaceae): *BMC Evolutionary Biology*, 18(9). ISSN: 1471-2105 (web). <https://doi.org/10.1186/s12862-018-1127-z>
- Fér, T. y Schmickl, R.E., 2018, HybPhyloMaker: Target Enrichment Data Analysis From Raw Reads to Species Trees: *Evolutionary Bioinformatics*, 14. ISSN: 1176-9343 (web). <https://doi.org/10.1177/1176934317742613>
- Fragoso-Martínez, I., Martínez-Gordillo, M., Salazar, G.A., Sazatornil, F., Jenks, A.A., García Peña, M. del R., Barrera-Aveleida, G., Benitez-Vieyra, S., Magallón, S., Cornejo-Tenorio, G. y Granados Mendoza, C., 2018, Phylogeny of the Neotropical sages (*Salvia* subg. *Calosphace*; Lamiaceae) and insights into pollinator and area shifts: *Plant Systematics and Evolution*, 304(1), 43-55. ISSN: 0378-2697 (print) 1610-6110 (web). <https://doi.org/10.1007/s00606-017-1445-4>
- Fragoso-Martínez, I., Salazar, G.A., Martínez-Gordillo, M., Magallón, S., Sánchez-Reyes, L., Moriarty Lemmon, E., Lemmon, A.R., Sazatornil, F. y Granados Mendoza, C., 2017, A pilot study applying the plant Anchored Hybrid Enrichment method to New World sages (*Salvia* subgenus *Calosphace*; Lamiaceae): *Molecular Phylogenetics and Evolution*, 117, 124-134. ISSN: 1055-7903 (print). <https://doi.org/https://doi.org/10.1016/j.ympev.2017.02.006>
- González-Gallegos, J.G., Bedolla-García, B.Y., Cornejo-Tenorio, G., Fernández-Alonso, J.L., Fragoso-Martínez, I.,

- García-Peña, M. del R., Harley, R.M., Klitgaard, B., Martínez-Gordillo, M.J., Wood, J.R.I., Zamudio, S., Zona, S. y Xifreda, C.C., 2020, Richness and Distribution of *Salvia* Subg. *Calosphace* (Lamiaceae): *International Journal of Plant Sciences*, 181(8), 831-856. ISSN: 1058-5893 (print) 1537-5315 (web).
<https://doi.org/10.1086/709133>
- Heraghty, S.D., Sutton, J.M., Pimsler, M.L., Fierst, J.L., Strange, J.P., y Lozier, J.D., 2020, *De novo* genome assemblies for three North American bumble bee species: *Bombus bifarius*, *Bombus vancoverensis*, and *Bombus vosnesenskii*: *G3. Genes, Genomes, Genetics*, 10(8), 2585-2592. ISSN 2160-1836 (web).
<https://doi.org/10.1534/g3.120.401437>
- Jenks, A.A., Walker, J.B. y Kim, S.C., 2013, Phylogeny of New World *Salvia* subgenus *Calosphace* (Lamiaceae) based on cpDNA (*psbA-trnH*) and nrDNA (ITS) sequence data: *Journal of Plant Research*, 126(4), 483-496. ISSN: 0918-9440 (print) 1618-0860 (web).
<https://doi.org/10.1007/s10265-012-0543-1>
- Katoh, K. y Standley, D.M., 2013, MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability: *Molecular Biology and Evolution*, 30(4), 772-780. ISSN: 1537-1719 (web).
<https://doi.org/10.1093/molbev/mst010>
- Kriebel, R., Drew, B.T., Drummond, C.P., González-Gallegos, J.G., Celep, F., Mahdjoub, M.M., Rose, J.P., Xiang, C., Hu, G., Walker, J.B., Lemmon, E.M., Lemmon, A.R. y Sytsma, K.J., 2019, Tracking temporal shifts in area, biomes, and pollinators in the radiation of *Salvia* (sages) across continents: leveraging anchored hybrid enrichment and targeted sequence data: *American Journal of Botany*, 106(4), 573-597. ISSN: 1537-2197 (web). <https://doi.org/10.1002/ajb2.1268>
- Lara-Cabrera, S.I., Pérez-García, M. de la L., Maya-Lastra, C.A., Montero-Castro, J.C., Godden, G.T., Cibrian-Jaramillo, A., Fisher, A.E. y Porter, J.M., 2021, Phylogenomics of *Salvia* L. subgenus *Calosphace* (Lamiaceae): *Frontiers in Plant Science*, 12, 725900. ISSN: 1664-462X (web).
<https://doi.org/10.3389/fpls.2021.725900>
- Lee, Y., Kwak, H., Shin, J., Kim, S.C., Kim, T., y Park, J.K., 2019, A mitochondrial genome phylogeny of Mytilidae (Bivalvia: Mytilida): *Molecular Phylogenetics and Evolution*, 139, 106533. ISSN 1055-7903 (web).
<https://doi.org/10.1016/j.ympev.2019.106533>
- Lemmon, A.R., Emme, S.A. y Lemmon, E.M., 2012, Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics: *Systematic Biology* 61(5), 727-744. ISSN: 1063-5157 (print) 1076-876X (web).
<https://doi.org/10.1093/sysbio/sys049>
- Li, H. y Durbin, R., 2009, Fast and accurate short read alignment with Burrows–Wheeler transform: *Bioinformatics*, 25(14), 1754-1760. ISSN: 1367-4803 (print) 1460-2059 (web). <https://doi.org/10.1093/bioinformatics/btp324>
- Mabberley, D.J., 1997, The plant-book. A portable dictionary of the vascular plants: *Cambridge University Press*, 2a Ed. ISBN 0-521-41421-0.
- Maldonado, E. y Ortega, A., 2000, Polystachynes A–E, five cis-neo-clerodane diterpenoids from *Salvia polystachya*: *Phytochemistry*, 53(1), pp.103-109.
[https://doi.org/10.1016/S0031-9422\(99\)00466-5](https://doi.org/10.1016/S0031-9422(99)00466-5)
- Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S. y Warnow, T., 2014, ASTRAL: genome-scale coalescent-based species tree estimation: *Bioinformatics*, 30(17), i541-i548. ISSN: 1367-4803 (print) 1460-2059 (web). <https://doi.org/10.1093/bioinformatics/btu462>
- Olvera-Mendoza, E.I., Godden, G.T., Montero-Castro, J.C., Porter, J.M. y Lara-Cabrera, S.I., 2020, Chloroplast and nuclear ribosomal cistron phylogenomics in a group of closely related sections in *Salvia* subg. *Calosphace*: *Brazilian Journal of Botany*, 43(1), 177-191. ISSN: 1806-9959 (web). <https://doi.org/10.1007/s40415-019-00572-9>
- Pérez-García, M. de la L., 2019, Filogenómica de *Salvia* subgénero *Calosphace* (Lamiaceae): Msc thesis, *Universidad Michoacana de San Nicolás de Hidalgo*. http://biotecavir-tual.dgb.umich.mx:8083/xmlui/handle/DGB_UMICH/1808
- Price, M.N., Dehal, P.S. y Arkin, A.P., 2010, FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments: *PLOS ONE*, 5(3), e9490. ISSN: 1932-6203 (web). <https://doi.org/10.1371/journal.pone.0009490>
- Saarela, J., Wysocki, W., Barrett, C., Soreng, R., Davis, J., Clark, L., Kelchner, S., Pires, J., Edger, P., Mayfield-Jones, D. y Duvall, M., 2015, Plastid phylogenomics of the cool-season grass subfamily: Clarification of relationships among early-diverging tribes: *AoB PLANTS*, 7. ISSN: 2041-2851 (web).
<https://doi.org/10.1093/aobpla/plv046>

- Sanger, F., Nicklen, S. y Coulson, A.R., 1977, DNA sequencing with chain-terminating inhibitors: *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463-5467. ISSN: 0027-8424 (print) 1091-6490 (web).
<https://doi.org/10.1073/pnas.74.12.5463>
- Sosa-Baldivia, A., Ruiz-Ibarra, G., de la Torre, R.R.R., López, R.R., y López, A.M., 2018, The chia (*Salvia hispanica*): past, present and future of an ancient Mexican crop.: *Australian Journal of Crop Science*, 12(10), 1626-1632. ISSN: 1835-2693 (print).
- Stöver, B.C. y Müller, K.F., 2010, TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses: *BMC Bioinformatics*, 11(1), 7. ISSN: 1471-2105 (web). <https://doi.org/10.1186/1471-2105-11-7>
- Uribe-Convers, Settles, M.L. y Tank, D.C., 2016, Data from: A phylogenomic approach based on PCR target enrichment and high throughput sequencing: resolving the diversity within the South American species of *Bartsia* l. (Orobanchaceae). *PLOS ONE*, Dryad Digital Repository. ISSN: 1932-6203 (web).
<https://doi.org/doi:10.5061/dryad.fh592>
- Velasco-Lezama, R., Quintero-Reyes, M. y Tapia-Aguilar, R., 2021, Evaluation of the antibacterial activity of *Salvia purpurea* Cav.: *Pharmacology on line*, 3, 413-420. ISSN: 1827-8620. <http://pharmacologyonline.silae.it>
- Walker, J.B., 2006, A preliminary molecular phylogenetic analysis of *Salvia* subgenus *Calosphace*: Ph.D thesis, *University of Wisconsin*.
<http://ci.nii.ac.jp/naid/10029563726/en/>
- Wang, L., Wuyun, T.N., Du, H., Wang, D. y Cao, D., 2016, Complete chloroplast genome sequences of *Eucommia ulmoides*: genome structure and evolution: *Tree Genetics & Genomes*, 12(1), 1-15. ISSN 1614-2950 (web).
<https://doi.org/10.1007/s11295-016-0970-6>
- Weitemier, K., Straub, S.C.K., Cronn, R.C., Fishbein, M., Schmickl, R., McDonnell, A. y Liston, A., 2014, Hyb-Seq: Combining target enrichment and genome skimming for plant phylogenomics: *Applications in Plant Sciences*, 2(9), 1400042. ISSN: 2168-0450 (web).
<https://doi.org/10.3732/apps.1400042>
- Will, M. y Claßen-Bockhoff, R., 2017, Time to split *Salvia sl* (Lamiaceae)—new insights from Old World *Salvia* phylogeny: *Molecular Phylogenetics and Evolution*, 109, 33-58. ISSN: 1055-7903 (print).
<https://doi.org/10.1016/j.ympev.2016.12.041>

Apéndice A

Longitud del alineamiento de 96 genes anotados en referencia a *Arabidopsis thaliana* (AT) para los cinco cromosomas de la especie, indicando el número de muestras que se amplificaron con las sondas diseñadas para el gen; longitud del alineamiento, número de sitios variables y sitios parsimoniosamente informativos, además, porcentaje de bases AT y GC. En rojo se indican los dos genes que se eliminaron de los análisis subsecuentes por carecer de *S. leucantha*.

Locus	# de taxa	Longitud alineamiento (pb)	# de sitios variables	Sitios de información parsimoniosa	% Adenina - Timina	% Guanina - Citosina
AT1G05120	33	2700	179	106	57.6	42.4
AT1G05350	33	154	9	2	61.3	38.7
AT1G07010	33	1075	60	27	54.4	45.6
AT1G14300	33	815	53	22	56.9	43.1
AT1G14810	33	651	61	45	44.9	55.1
AT1G16970	33	1935	85	34	58.1	41.9
AT1G28340	33	1538	138	85	55.2	44.8
AT1G43860	33	849	51	33	58.5	41.5
AT1G51940	33	1980	174	84	53.7	46.3
AT1G53280	33	1087	124	66	50.8	49.2
AT1G62750	33	2129	109	42	54.1	45.9
AT1G64550	33	1439	89	51	53.6	46.4
AT2G04305	25	1359	132	85	51.8	48.2

Locus	# de taxa	Longitud alineamiento (pb)	# de sitios variables	Sitios de información parsimoniosa	% Adenina - Timina	% Guanina - Citosina
AT2G15230	33	584	29	14	57	43
AT2G18710	33	1292	112	67	56.3	43.7
AT2G19940	33	357	39	19	55.2	44.8
AT2G36895	33	738	50	22	49	51
AT2G37500	33	447	43	24	55.2	44.8
AT3G01720	33	2508	175	104	55.1	44.9
AT3G04260	33	2749	183	89	56.2	43.8
AT3G04480	32	2236	94	32	59.7	40.3
AT3G05350	33	1338	109	52	58.7	41.3
AT3G09180	33	727	27	11	45.9	54.1
AT3G10230	27	1402	117	62	51.6	48.4
AT3G11830	33	1683	132	72	56.6	43.4
AT3G12290	33	908	85	43	53.2	46.8
AT3G17040	33	1860	166	78	52.7	47.3
AT3G17810	33	1149	122	85	51.3	48.7
AT3G17940	33	1090	120	68	53.1	46.9
AT3G20790	33	392	59	33	51.7	48.3
AT3G22590	33	1148	73	33	47.1	52.9
AT3G23620	33	957	78	54	56.8	43.2
AT3G25660	33	1958	124	72	53.3	46.7
AT3G25900	32	1001	67	34	51.7	48.3
AT3G27530	33	2756	167	76	55.9	44.1
AT3G29010	33	874	43	19	59	41
AT3G43540	31	1064	63	30	52.2	47.8
AT3G45300	33	951	48	24	55.4	44.6
AT3G47610	33	1093	94	47	52.8	47.2
AT3G51050	33	1655	151	102	51.8	48.2
AT3G52190	33	795	78	49	58.2	41.8
AT3G52640	33	2072	138	59	56.8	43.2
AT3G55070	33	1091	112	72	55.1	44.9
AT3G55260	33	1686	158	80	53.3	46.7
AT3G55500	27	383	36	22	34.3	65.7
AT3G56460	33	1111	134	76	49.4	50.6
AT3G57790	33	1503	106	55	51.2	48.8
AT3G58460	33	175	13	6	50.6	49.4
AT3G58690	33	1191	105	51	47.3	52.7
AT3G59040	33	1798	116	59	54.8	45.2
AT3G59380	33	1065	94	51	52.2	47.8
AT3G60830	33	1088	88	55	55.3	44.7
AT3G60850	33	1975	233	146	52.2	47.8
AT3G66658	33	1230	92	52	54.9	45.1
AT4G00090	33	645	61	27	55.4	44.6
AT4G00740	33	1324	147	94	52.1	47.9



Locus	# de taxa	Longitud alineamiento (pb)	# de sitios variables	Sitios de información parsimoniosa	% Adenina - Timina	% Guanina - Citosina
AT4G01880	28	1411	70	19	56.3	43.7
AT4G02990	33	1573	240	131	54.8	45.2
AT4G09750	33	969	64	23	47.1	52.9
AT4G19490	33	3322	234	107	53.1	46.9
AT4G19860	33	907	82	49	53.2	46.8
AT4G20130	33	1795	175	73	55.4	44.6
AT4G29490	33	590	48	26	59.3	40.7
AT4G29830	33	1037	72	22	40.5	59.5
AT4G30310	33	1077	78	38	54.8	45.2
AT4G30510	33	1161	50	24	59	41
AT4G31990	33	960	73	34	55.4	44.6
AT4G35850	33	510	45	20	55.5	44.5
AT4G35870	33	2625	244	138	52.5	47.5
AT4G37040	33	575	44	23	53.2	46.8
AT4G38460	33	1085	84	34	44.9	55.1
AT5G04420	33	606	48	24	56.7	43.3
AT5G05200	33	1079	92	39	56.3	43.7
AT5G05660	33	2581	261	129	46.1	53.9
AT5G08100	31	774	36	16	46.3	53.7
AT5G08170	33	527	54	39	55.7	44.3
AT5G09860	33	272	17	11	60.6	39.4
AT5G13030	33	1408	98	39	55.5	44.5
AT5G13520	33	2064	189	110	47.9	52.1
AT5G13650	33	1309	107	60	56.3	43.7
AT5G14250	33	832	71	35	53.8	46.2
AT5G14720	33	2025	90	37	56.3	43.7
AT5G17530	33	578	48	24	56.5	43.5
AT5G18070	33	1679	209	132	55.2	44.8
AT5G30510	33	704	53	30	53.4	46.6
AT5G42310	33	1494	138	83	50.1	49.9
AT5G42480	33	1051	136	84	53.6	46.4
AT5G43600	33	575	76	53	54	46
AT5G50390	33	1900	141	76	57.1	42.9
AT5G56580	33	1062	57	28	57.9	42.1
AT5G57030	33	855	91	51	57.2	42.8
AT5G57655	33	809	60	30	55.5	44.5
AT5G61530	33	1133	72	37	54.5	45.5
AT5G62530	32	352	47	27	56.2	43.8
AT5G63610	33	1418	89	46	54.1	45.9
AT5G63890	33	541	48	17	54.6	45.4
AT5G64370	33	1004	134	88	53.8	46.2
AT5G65720	33	1258	142	81	47.4	52.6