

## ARTÍCULO CIENTÍFICO

### Phylogeny of lulo, tree tomato and their wild relatives

## Filogenia de lulo, tomate de árbol y sus parientes silvestres

Oscar C. Bedoya-Reina<sup>1</sup>, Luz Stella Barrero<sup>2</sup>

#### ABSTRACT

Among the Solanaceae family of plants, the monophyletic genus *Solanum* is the most diverse one.

Among the *Solanum* genus, *S. quitoense* Lam. (lulo) and *S. betaceum* (tree tomato) represent some of the most promising species to become premium crops for local and export markets. *S. quitoense* and its wild relative *S. hirtum* belong to the *Leptostemomum* clade and *S. betaceum* and its wild relative *S. unilobum* belong to the *Cyphomandra* clade of the *Solanum* genus.

Previous studies have reported that *Leptostemomum* and *Cyphomandra* form a sister clade to the *Potato* clade (which groups *S. lycopersicum* and *S. tuberosum* among others). An alternative hypothesis explains that these three clades form a polytomy inside the *Solanum* genus. In order to clarify the relationships among the species of the *Solanum* genus and to contribute in understanding its divergence process, we conducted phylogenetic analyses with COSII markers. Results obtained support that *Leptostemomum* and *Cyphomandra* form a subclade which is sister to the *Potato* clade. In addition, our results suggest that divergence of *S. hirtum* and *S. quitoense* happened 1 million years ago approximately, and that of *S. unilobum* and *S. betaceum* occurred about 450 thousand years ago. Both of these periods are coincident with global climate changes which have been associated with the diversification processes of several Andean clades.

**Keywords:** *Cyphomandra betacea*, biogeography, Solanaceae, clade, *Solanum*.

#### RESUMEN

Dentro de la familia Solanaceae, el género monofilético *Solanum* es el más grande y diverso. Dentro de este género, *S. quitoense* Lam. (lulo) y *S. betaceum* Cav. (tomate de árbol) representan dos de las especies más importantes de Solanáceas neotropicales para convertirse en cultivos prominentes en mercados locales y de exportación. *S. quitoense* y su pariente silvestre *S. hirtum* pertenecen al clado *Leptostemomum* del género *Solanum*, mientras que *S. betaceum* y su pariente silvestre *S. unilobum* pertenecen al clado *Cyphomandra* del mismo género. En estudios filogenéticos previos se ha reportado que el clado *Leptostemomum* y el *Cyphomandra* podrían formar un clado hermano al de *Papa* (que agrupa a *S. lycopersicum* y *S. tuberosum* entre otras especies); alternativamente, estos tres clados formarían una politomía al interior del género. Con el propósito de clarificar las relaciones entre los clados del género *Solanum* y contribuir con el entendimiento de su proceso de divergencia, el presente estudio utilizó secuencias de marcadores COSII para conducir análisis filogenéticos y de datación. Los resultados obtenidos soportan la hipótesis que indica que *Leptostemomum* y *Cyphomandra* forman un subclado hermano al clado *Papa*. Adicionalmente, éstos sugieren que la divergencia entre los pares *S. hirtum* y *S. quitoense*, y *S. unilobum* y *S. betaceum* ocurrió hace aproximadamente un millón y 450 mil años respectivamente, periodos que coinciden con eventos de cambio climático a los que se les ha atribuido la diversificación de varios clados andinos.

**Palabras clave:** *Cyphomandra betacea*, biogeografía, Solanaceae, clado, *Solanum*.

#### INTRODUCCIÓN

LA FAMILIA SOLANACEAE comprende entre 9.000 y 10.000 especies distribuidas globalmente (Bernatzky y Tanksley, 1986a; Bohs, 2007; D'Arcy, 1991; Lobo, 2006); dentro de ésta, el género monofilético *Solanum* es el más grande y diverso, agrupando entre 1.000 y 2.000 especies ampliamente distribuidas desde selvas lluviosas hasta desiertos (Bohs, 2007; Chung *et al.*, 2006; Heiser y Anderson, 1999; Manoko *et al.*, 2008). Las especies del género *Solanum* revisten gran importancia económica, algunas de ellas hacen parte de la dieta tradicional en diferentes partes del mundo (incluyendo tomate y papa) y otras son apreciadas por su potencial como frutales novedosos (Bohs, 2007; Heiser, 1985; Heiser y Anderson, 1999; Knapp, 2002).

Radicado: 15 de julio de 2008  
Aprobado: 6 de octubre de 2009

<sup>1</sup> Candidato a Ph.D en Bioinformática y Genómica. Laboratorio de Genética Molecular Vegetal, Centro de Biotecnología y Bioindustria (CBB), Corpoica. Mosquera, Cundinamarca. oscarbed@gmail.com

<sup>2</sup> Ph.D en Genética vegetal. Laboratorio de Genética Molecular Vegetal. Corpoica, Mosquera, Cundinamarca. lbarrero@corpoica.org.co

Dentro del género *Solanum*, *S. quitoense* Lam. (lulo) y *S. betaceum* Cav. (tomate de árbol) representan dos de las especies más importantes de Solanáceas neotropicales para convertirse en cultivos prominentes en mercados locales y de exportación (Heiser y Anderson, 1999; Lobo, 2006). *S. quitoense* pertenece a la sección *Lasiocarpa* del clado *Leptostemonum* que agrupa cerca de 13 especies tropicales principalmente distribuidas en Colombia, Ecuador y Venezuela (Bohs, 2004; Heiser, 1972; Heiser, 1985; Miller y Diggle, 2003). Es una planta alta (2 metros) con desarrollo de coloración purpúrea pronunciada en el tallo y las hojas, y frutos pequeños y medianos de color naranja cuyo sabor ha sido descrito como una mezcla de naranja, piña y tomate (Bohs, 2004; Heiser, 1972; Heiser, 1985; Heiser, 1989; Heiser y Anderson, 1999; Miller y Diggle, 2003). Dos variedades botánicas de esta especie han sido reconocidas: *quitoense*, forma sin espinas, y *septentrionale*, forma con espinas encontrada en Panamá, Costa Rica y la parte central de Colombia (Heiser, 1972). El lulo es una planta de cultivo que, en tanto tiene características de una silvestre (alogamia, presencia de antocianinas y alto número de semillas, entre otras), es considerada una especie en proceso de domesticación (Lobo *et al.*, 2007). El tomate de árbol es un árbol relativamente pequeño (2-7 metros) nativo de los Andes, con frutos ovoides entre 4 y 10 centímetros de longitud que se consume crudo, en jugos y mermeladas (Heiser, 1972). Pertenece a la sección *Pachyphylla* del clado *Cyphomandra* (*Solanum*) que agrupa 46 especies, se cultiva alrededor del mundo en áreas subtropicales y tiene poblaciones silvestres en el sur de Bolivia y el norte de Argentina, sitios considerados su centro de origen (Heiser y Anderson, 1999; Kou *et al.*, 2009; Lobo *et al.*, 2000).

Los parientes silvestres de *S. quitoense* y *S. betaceum* son una fuente de características deseables que pueden ser fácilmente explotadas en mejoramiento genético (Heiser, 1985; Lobo *et al.*, 2007). Así, el desarrollo de híbridos entre las especies cultivadas y sus relativos silvestres ha sido una estrategia exitosamente utilizada para introducir resistencia a patógenos, una barrera común para la explotación comercial de lulo y tomate de árbol (Benítez *et al.*, 1991; Heiser, 1989; Lobo *et al.*, 2007). Los cruces entre *S. hirtum* (sección *Lasiocarpa* del clado *Leptostemonum*) y *S. quitoense* han resultado en variedades con sabores agradables, parecidos o mejores que los del lulo y resistencia al nematodo de la pudrición de la raíz; también se han desarrollado híbridos entre *S. unilobum* (sección *Pachyphylla* del clado *Cyphomandra*) y *S. betaceum* con resistencia mejorada a la antracnosis (Lobo *et al.*, 2000).

Aunque la taxonomía del género *Solanum* es compleja por su tamaño y diversidad morfológica, las aproximaciones

moleculares han sido útiles para elucidar su estructura filogenética. Con el uso de datos moleculares del gen *ndhF* de cloroplasto y las regiones nucleares *ITS* y *waxy*, al menos doce clados mayores se han encontrado bien soportados dentro del género (*Leptostemonum*, *Dulcamaroides*, *Morelloide*, *Normania*, *Archaeosolanum*, *Africanas sin espinas*, *Papa*, *Regmandra*, *Thelopodium*, *Brevantherum*, *Geminata* y *Cyphomandra*) (Weese y Bohs, 2007). Varios de estos clados conforman grupos infragenéricos reconocidos previamente con base en su similitud morfológica; sin embargo, otros representan nuevos agrupamientos no sugeridos previamente (Bohs, 2004; Weese y Bohs, 2007). A pesar de los estudios moleculares conducidos en cuanto al género, aún existen incongruencias en las relaciones de estos clados, mayormente debidas a la baja resolución dada por algunos genes (Weese y Bohs, 2007).

El clado *Cyphomandra* (sensu Weese y Bohs 2007) es uno, con buen soporte, que agrupa especies neotropicales en tres secciones: sección *Pachyphylla* (Dunal) Dunal, sección *Cyphomandropsis* Bitter y sección *Glaucophyllum*. A. Child. Todos los taxones del clado son arbustos leñosos o árboles pequeños y la mayoría tienen poros en las anteras relativamente pequeños que no se abren en rendijas longitudinales (Bohs, 2007). Las sinapomorfias más consistentes del grupo son la presencia de cromosomas muy largos, y grandes cantidades de ADN nuclear que han sido encontradas en todas las especies del clado (Stevens, 2001 y en adelante; Bohs, 2007). Anteriormente, la sección *Pachyphylla* era formalmente reconocida como el género *Cyphomandra* Sendt., cuyas especies se encuentran distribuidas desde México hasta el norte de Argentina y el sureste de Brasil (Stevens, 2001 y en adelante; Bohs, 2004; Bohs, 2007). La sinapomorfia que define a la sección *Pachyphylla* es la presencia de anteras conectivas alargadas que funcionan al menos en algunas especies como osmóforos florales para atraer abejas macho (Gracie, 1993). Análisis moleculares (Bohs, 2007) indican que el género *Cyphomandra* se encuentra anidado dentro del género *Solanum*, y por ello todas las especies de *Cyphomandra* fueron transferidas al género *Solanum* (Bohs, 2004; Bohs, 2007).

De otra parte, el clado *Leptostemonum* (sensu Bohs 2004) agrupa especies con espinas y tricomas estrellados en dos secciones: sección *Lasiocarpa* Dunal y sección *Acanthophora* Dunal (Bohs, 2004). La sección *Lasiocarpa* une cerca de 12 especies de arbustos perennes o pequeños árboles con centro de distribución en el noroeste suramericano (Colombia, Venezuela, Ecuador y Perú) cuya distribución se extiende hacia América Central y el noreste suramericano (Stevens, 2001 y en adelante; Bohs, 1994; 2004; 2007). Las características morfológicas que definen la sección incluyen unidades simpodiales

difoliadas, hojas largas onduladas inflorescencias no ramificadas, corolas y tricomas estrellados con rayos laterales reducidos (Bohs, 1994; 2004; 2007). A pesar la gran cantidad de datos moleculares y morfológicos acumulados, los especialistas no han alcanzado un consenso respecto a las relaciones filogenéticas de los taxones en este clado, debido especialmente a los conflictos de topologías de diferentes tipos de datos y la baja resolución en algunos clados (Bohs, 1994; 2004; 2007).

Para mejorar la comprensión y promover el acceso a la diversidad genética de los parientes silvestres de lulo y tomate de árbol, el objetivo del presente estudio fue conducir análisis filogenéticos de seis especies del género *Solanum*: *S. lycopersicum*, *S. tuberosum*, y *S. quitoense* y *S. betaceum*, con sus parientes silvestres *S. hirtum* y *S. unilobum*, respectivamente. Con este estudio se espera contribuir a la clarificación de las relaciones filogenéticas de los clados que conforman el género y obtener una idea de los procesos biogeográficos que les dieron origen. Este estudio utiliza marcadores COSII (Wu *et al.*, 2009a; 2009b), un nuevo tipo de herramienta molecular en astéridas compuesta de secuencias de ADN de genes ortólogos que han sido transferidos exitosamente a *S. quitoense*, *S. betaceum* y parientes silvestres (Pratt *et al.*, 2008) y que ofrecen una imagen más amplia de todo el genoma, en comparación con la obtenida por el análisis de secuencias nucleares y de cloroplasto (Clegg, 1993; Graham y Olmstead, 2000; Palmer *et al.*, 1988).

## MATERIALES Y MÉTODOS

### Material vegetal y aislamiento de ADN

Para el presente estudio, se incluyeron muestras de uno a dos individuos de una accesión por cada especie estudiada; éstas fueron obtenidas del banco de germoplasma de Colombia a cargo de Corpoica (C.I. La Selva, Rionegro, Antioquia, Colombia) y de la Universidad de Cornell (Ithaca, Estados Unidos). Se incluyeron dos individuos de la accesión 04T142 de *S. quitoense* (Universidad de Cornell, Estados Unidos) donados por el profesor C. Heiser de la Universidad de Indiana, Estados Unidos, un individuo de la accesión ILS 505 de *S. betaceum* (Corpoica) procedente del departamento de Antioquia, dos individuos de la accesión ILS 3234 de *S. hirtum* (Corpoica) de Aguas Calientes, Santander, y un individuo de la accesión ILS 3332 de *S. unilobum* (Corpoica) de Santa Cruz, Bolivia.

El ADN genómico total fue aislado de hojas jóvenes usando el procedimiento descrito por Bernatzky y Tanksley (1986b). Para ello, aproximadamente 25 gramos de hojas jóvenes fueron homogenizados en un mortero con 150 mililitros de tampón de extracción (100 mM Tris pH

8,0; 0,35 M Sorbitol; 5 mM EDTA y 1% B-mercaptoetanol), filtrados y centrifugados a 700 g durante 15 minutos. El producto fue resuspendido en tampón de extracción, ajustado a 1% de CTAB, 1 M NaCl, 25 mM EDTA, y luego calentado a 60 °C durante 20 minutos. Para precipitar el ADN, fue lavado una vez con una solución cloroformo/etanol (24:1 v/v) y su fase acuosa fue mezclada con dos tercios de su volumen de isopropanol. Posteriormente, el ADN extraído se lavó con una solución 76% etanol, 10 mM acetato de amonio pH 7,0, y se diluyó a una concentración final de 2 ng/μl.

### Amplificación de marcadores COSII y secuenciación

Para el presente estudio, se secuenciaron seis marcadores COSII (At2g47580, At3g07100, At3g52120, At3g63190, At3g63490, At4g34350) buscando muestrear diferentes regiones genómicas. Esta selección se sustentó en la posición de los marcadores en el mapa EXPEN 2000 de tomate (Mueller *et al.*, 2005), tomando en cuenta que se ha observado una sintenia alta en las especies del género *Solanum* (Huang *et al.*, 2005; Wu *et al.*, 2009a; 2009b).

Para amplificar y secuenciar los marcadores, se obtuvieron cebadores universales a partir de la base de datos de la Red Genómica de las Solanaceae (SGN) (Mueller *et al.*, 2005). La amplificación por PCR del ADN genómico se condujo utilizando los cebadores seleccionados, en un volumen de reacción de 20 microlitros con aproximadamente 20 nanogramos de ADN genómico, 0,2 μg de ambos cebadores, 0,125 mM de cada uno de los cuatro dNTP, 0,25 unidades de Taq polimerasa y 2,5 mM de MgMCl<sub>2</sub>. Las condiciones de amplificación fueron 3 minutos a 94 °C y 30 ciclos de 30 segundos a 94 °C, 30 segundos a 55 °C, y 2 minutos a 72 °C, seguidos por un paso de extensión final de 5 minutos a 72 °C. Los productos amplificados de bandas sencillas fueron limpiados con Exo-SAP-IT (Amersham Pharmacia Biotech Inc., Piscataway, NJ, EE.UU.) y secuenciados con el cebador directo en un solo paso.

### Procesamiento de cromatogramas y edición de secuencias

Los cromatogramas fueron procesados con el grupo de programas desarrollado por Ewing y Green (1998); el llamado de bases, asignación de calidad y corte de secuencias se realizó con PHRED 0.020425.c, el ensamble de secuencias se condujo con PHRAP y la edición posterior se realizó en Consed 17.0 (Gordon *et al.*, 1998). Aquellos marcadores secuenciados para más de un individuo por especie fueron ensamblados en una sola secuencia.

Las secuencias de ortólogos de los marcadores analizados de *Solanum lycopersicum* (tomate), *Solanum tuberosum*

(papa), *Capsicum annuum* (pimentón), *Coffea canephora* (café) y *Arabidopsis thaliana* fueron obtenidas de la base de datos de COSII de la SGN ([ftp://sgn.cornell.edu/cosii/ortholog\\_sequences/](ftp://sgn.cornell.edu/cosii/ortholog_sequences/)) (Mueller *et al.*, 2005). Por su identidad, las secuencias del marcador At3g52120 para papa y tomate fueron completadas con las secuencias de unigenes SGN-E601714 y SGN-E247720 respectivamente; de igual manera la secuencia de café para el marcador At4g34350 fue complementada con la del unigen SGN-U352133, disponibles todas en la misma base de datos.

La totalidad de las secuencias de cada marcador fueron unidas en matrices y posteriormente alineadas utilizando Dialign 2.2.1 (Morgenstern, 2004). Las matrices alineadas fueron editadas y corregidas manualmente, eliminando las regiones no presentes en las secuencias de referencia (aquellas descargadas de la base de datos de COSII del SGN) (figura 1). Adicionalmente, se construyó una matriz combinada, que reunía la totalidad de las secuencias editadas y corregidas manualmente. Dado que las secuencias de referencia (*Arabidopsis thaliana*, *Capsicum annuum*, *Coffea canephora*, *Solanum betaceum*, *Solanum hirtum*, *Solanum lycopersicum*, *Solanum quitoense*, *Solanum tuberosum* y *Solanum unilobum*) pertenecen a unigenes ensamblados con un marcador de secuencia expresada o EST (acrónimo del inglés expressed sequence tag), se asumió que todas las regiones removidas no eran exones y podrían interferir en los análisis posteriores.

### Análisis de datos

Se realizaron análisis filogenéticos con las matrices alineadas bajo los criterios de máxima parsimonia (MP) y máxima verosimilitud (ML) en PAUP\* 4b10 (Swofford, 2004), eligiendo como grupo externo a *Arabidopsis thaliana* (Brassicales). Para los análisis de máxima parsimonia se realizaron búsquedas exhaustivas con el algoritmo "branch and bound" soportadas con 1000 repeticiones con el análisis de remuestreo (método *bootstrap* -BS). El análisis de máxima verosimilitud se condujo con la matriz de los genes unidos, utilizando el modelo TIM+G con parámetros seleccionados por el criterio de información de Akaike en ModelTest 3.7 (Posada y Crandall, 1998); y fue soportado por 100 réplicas con el análisis de remuestreo. Clados con soporte de análisis de remuestreo superior a 75 se asumieron como bueno, y aquéllos con soporte inferior a 75, como bajo.

La datación de la filogenia se realizó en r8s 1.71 (Sander-son, 2006) utilizando como base la topología reconstruida por máxima verosimilitud y las tasas de sustituciones obtenidas con ModelTest. Para calibrar el reloj molecular, se utilizaron tres puntos diferentes: la fecha de divergencia del orden Brassicales (85-90 millones de años), la del

orden Gentianales (83-89 millones de años) y la del orden Solanales (82-86 millones de años) (Wikstrom *et al.*, 2001). Para probar si la matriz era ultramétrica (y podía usarse directamente como reloj molecular) se condujo un test de probabilidades (likelihood ratio test -LRT), usando los valores de verosimilitud de las topologías obtenidas al forzar el reloj molecular y al no hacerlo. Dado que las diferencias del LRT fueron significativas ( $P < 0,05$ ), el método utilizado para datación fue LF con reloj molecular local y tres parámetros diferentes: uno para *Arabidopsis thaliana*, otro para café y otro para las Solanaceae (*C. annuum*, *S. betaceum*, *S. hirtum*, *S. lycopersicum*, *S. quitoense*, *S. tuberosum* y *S. unilobum*). La validez de la datación se comprobó realizando validaciones de datos cruzados que en todos los casos confirmaron la robustez del análisis realizado.

## RESULTADOS

### Análisis filogenéticos

Secuencias genómicas de seis marcadores COSII se obtuvieron para cuatro especies de dos clados del género *Solanum*: *S. hirtum* y *S. quitoense* (clado *Leptostemonum*); y *S. betaceum* y *S. unilobum* (clado *Cyphomandra*); adicionalmente, secuencias de los mismos marcadores para dos especies del clado *Papa* (*S. lycopersicum* y *S. tuberosum*), *Capsicum annuum*, *Coffea canephora* y *Arabidopsis thaliana* se obtuvieron de bases de datos del SGN (ver Materiales y métodos) (Mueller *et al.*, 2005). Para cada marcador se construyeron matrices que fueron alineadas, corregidas y editadas, con cuya unión se creó una matriz combinada (figura 1).

Con las matrices de los genes alineados se condujeron análisis cladísticos por máxima parsimonia, mediante una búsqueda exhaustiva soportada por 1000 réplicas de *bootstrap* (figura 2). Respecto a las topologías obtenidas con cada uno de los marcadores, At2g47580 define dos clados: el primero es integrado por *C. canephora* de divergencia basal soportado con un valor de 100 en el análisis de remuestreo, y el segundo está conformado por las especies de la familia Solanaceae, con un valor de 77 con este mismo análisis. Este último, en su parte basal presenta un politomía integrada de las especies del clado *Papa*, *C. annuum* y un subclado que con bajo soporte agrupa a las especies esperadas de los clados *Cyphomandra* y *Leptostemonum*. La topología obtenida con el marcador At3g07100 define los mismos grupos que el gen At2g47580, sólo que no tiene poder para discriminar los clados *Cyphomandra* y *Leptostemonum* cuyas especies agrupa en una politomía bien soportada. La topología reconstruida con el gen At3g52120 es totalmente dicotomizada; el clado que agrupa *C. canephora* y aquel que agrupa las especies de la familia Solanaceae tienen una divergencia basal bien soportada. Dentro del clado que agrupa a las especies de la

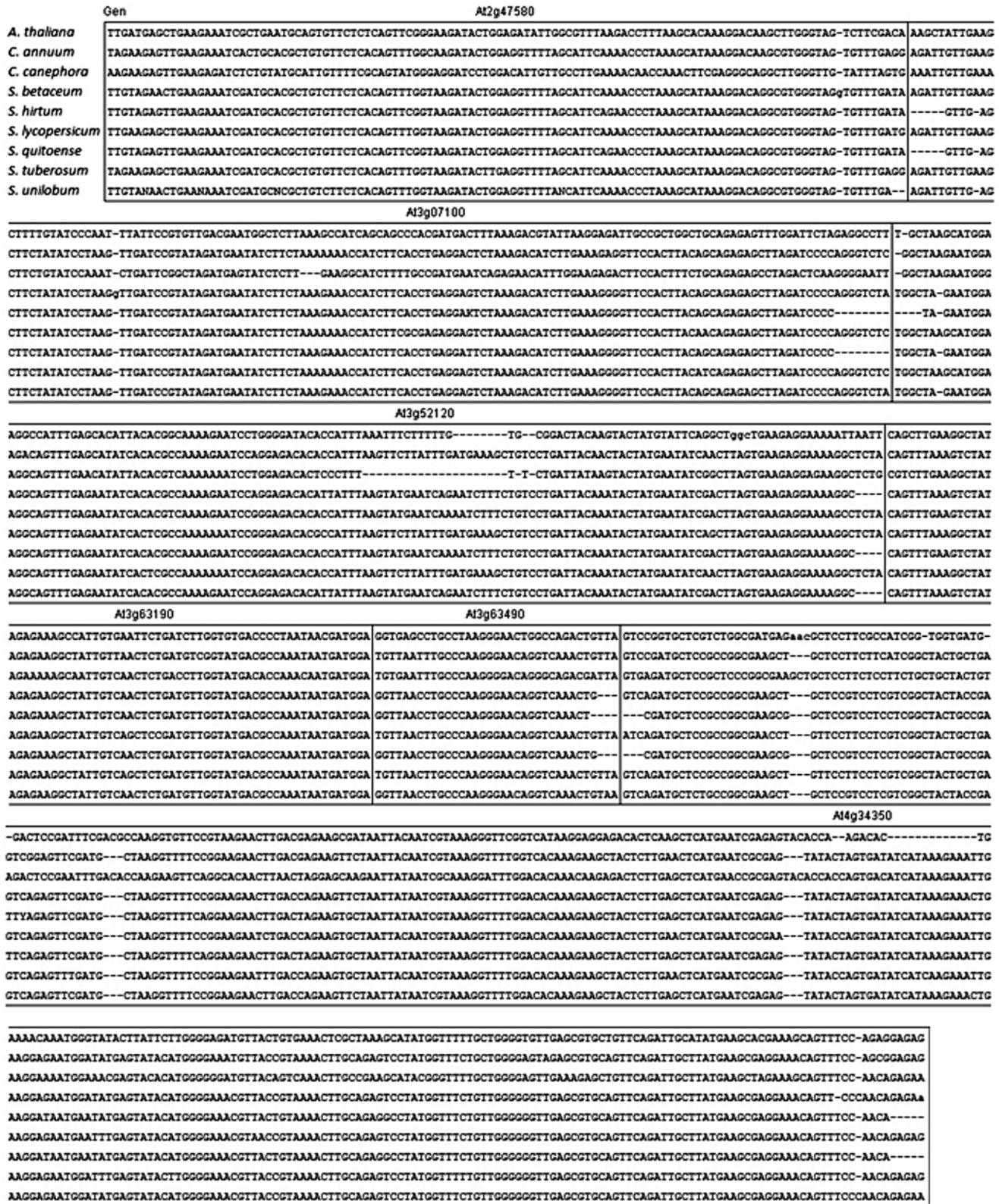
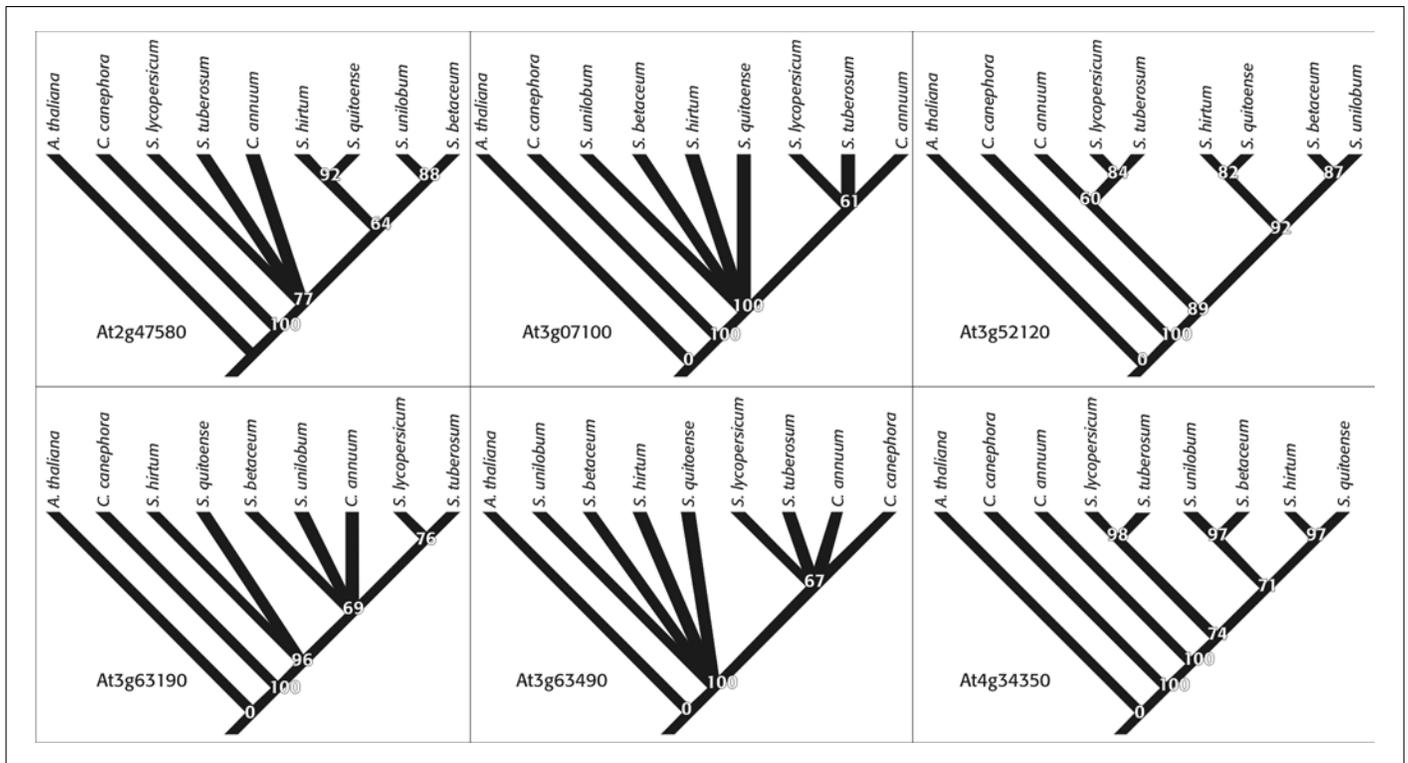


Figura 1. Alineamiento de seis marcadores COSII utilizados en el presente estudio. Cada sección corresponde a la matriz alineada de las regiones codificantes parciales de cada gen. La totalidad de las matrices alineadas se integró en una matriz combinada para realizar análisis por máxima parsimonia y máxima verosimilitud y para datar las filogenias



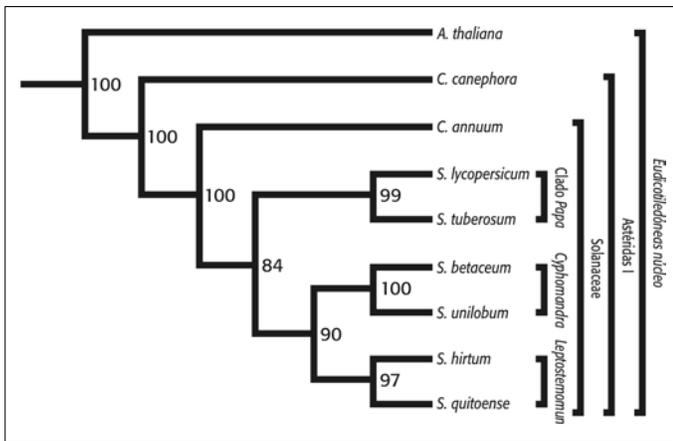
**Figura 2.** Consenso de la mayoría de los árboles más parsimoniosos obtenidos para cada marcador COSII. Los soportes del análisis de remuestreo se realizaron con 1000 réplicas y búsquedas exhaustivas. Los nodos con valores inferiores a 50% se colapsaron en politomías

familia Solanaceae, se distinguen dos subclados: uno que agrupa el clado *Papa* (que agrupa a su vez a *S. lycopersicum* y *S. tuberosum*) y *C. annuum* con bajo soporte, y otro que agrupa los clados *Leptostemonum* (que agrupa a su vez a *S. hirtum* y *S. quitoense*) y *Cyphomandra* (que a su vez agrupa a *S. betaceum* y *S. unilobum*) con un buen soporte.

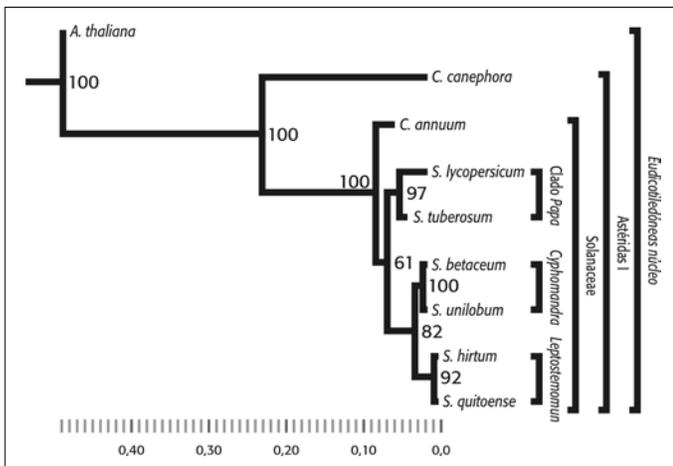
El gen At3g63190 define al clado *C. canephora* y agrupa las especies de la familia Solanaceae en otro clado con buen soporte, en el cual hay tres subclados unidos en una politomía basal bien soportada: el primero integra a *S. hirtum*, el segundo a *S. quitoense* y el tercero agrupa tres especies (*S. betaceum*, *S. unilobum* y *C. annuum*) y el clado *Papa* (*S. lycopersicum* y *S. tuberosum*) con un bajo soporte. La topología reconstruida con el gen At3g63490 agrupa en una politomía con un buen soporte cuatro especies (*S. unilobum*, *S. betaceum*, *S. hirtum* y *S. quitoense*) y un clado conformado por las especies restantes. Este último tiene una politomía con un soporte bajo que agrupa a *S. lycopersicum*, *S. tuberosum*, *C. annuum* y *C. canephora*. Finalmente, al igual que con el marcador At3g52120, la topología reconstruida con el gen At4g34350 está completamente dicotomizada. En la parte basal del cladograma se desprende *C. canephora* y un subclado que agrupa las especies restantes; este último está conformado por una dicotomía conformada por *C. annuum* y un clado que agrupa a las especies del género *Solanum*. En éste se agrupan con un buen soporte *S. lycopersicum* y *S.*

*tuberosum* en un clado, y en otro los clados *Cyphomandra* (agrupando *S. unilobum* y *S. betaceum*) y *Leptostemonum* (agrupando a su vez a *S. hirtum* y *S. quitoense*).

Con la matriz de datos combinados, se realizaron análisis de máxima parsimonia (figura 3) y máxima verosimilitud (figura 4), soportados en el primer caso con 1000 réplicas con el análisis de remuestreo y en el segundo con 100. Las topologías resultantes del análisis por los dos criterios son las mismas, totalmente dicotomizadas, agrupando las especies en los clados previamente descritos en la literatura. En la parte más basal del filograma y del cladograma reconstruidos se unen el grupo externo (*Arabidopsis thaliana*) y un clado que agrupa las restantes especies (todas astéridas I) con un buen soporte. En el siguiente nodo de las topologías, se agrupan *C. canephora* (especie del orden Gentianales) y el clado de las especies pertenecientes a la familia Solanaceae (análisis de remuestreo en MP y ML = 100). Este último clado está conformado a su vez por *C. annuum* y un subclado de las especies del género *Solanum* (análisis de remuestreo en MP y ML = 100). En el estudio presente, *S. lycopersicum* y *S. tuberosum* se agrupan en un clado (*Papa*), al igual que *S. betaceum* y *S. unilobum* (*Cyphomandra*), y *S. hirtum* y *S. quitoense* (*Leptostemonum*); el clado *Cyphomandra* se agrupa con el clado *Leptostemonum* (análisis de remuestreo en MP = 90, en ML = 82) en un grupo que se une con el clado *Papa* (BS en MP = 84, en ML = 61).



**Figura 3.** Cladograma más parsimonioso obtenido para la matriz combinada de los marcadores COSII. Los soportes del análisis de remuestreo se realizaron con 1000 réplicas y búsquedas exhaustivas. Las asignaciones de los nombres de los clados externos siguen las nomenclaturas asignadas por el Grupo de Filogenia de las Angiospermas (APG) (Stevens, 2001 y en adelante), y los clados dentro del género *Solanum* siguen las denominaciones propuestas por Bohs (2007). Los clados *Cyphomandra* y *Leptostemomun* forman un subclado hermano al clado *Papa* dentro de la familia Solanaceae



**Figura 4.** Filograma obtenido bajo el criterio de máxima verosimilitud para la matriz combinada de los marcadores COSII. Los soportes del análisis de remuestreo se realizaron con 100 réplicas y búsquedas heurísticas. Las asignaciones de los nombres de los clados externos siguen las nomenclaturas asignadas por el Grupo de Filogenia de las Angiospermas (APG) (Stevens, 2001 y en adelante), y los clados dentro del género *Solanum* siguen las denominaciones propuestas por Bohs (2007). Los clados *Cyphomandra* y *Leptostemomun* forman un subclado hermano al clado *Papa* dentro de la familia Solanaceae

Respecto a las distancias observadas en el análisis de máxima verosimilitud, se observa que las parejas de especies *S. betaceum* y *S. unilobum*, y *S. hirtum* y *S. quitoense* poseen distancias cercanas a cero desde el nodo más cercano. De otra parte, el clado *Cyphomandra* posee una distancia más corta al nodo más cercano respecto al clado *Leptostemomun*. El subclado, que las especies de estos dos

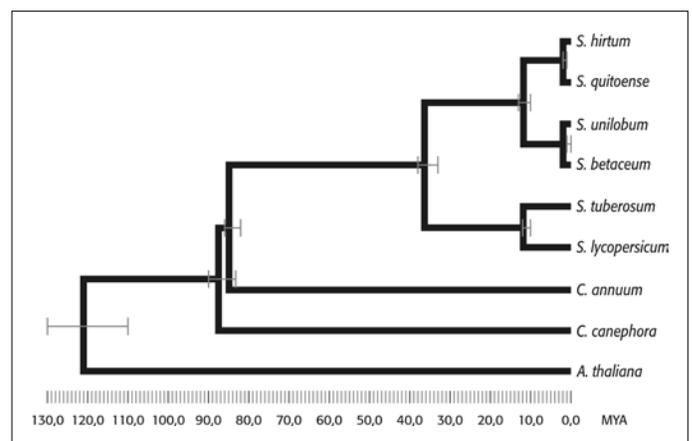
últimos clados forman, tiene una longitud mayor desde el nodo que comparte con el clado *Papa* que aquella del clado *Papa* al mismo nodo. Más alejadas las especies *C. canephora* y *A. thaliana* muestran una diferencia de longitud notoria en comparación con aquellas del género *Solanum*.

### Datación de la filogenia reconstruida

Para conocer las fechas de divergencia de las especies de los clados estudiados, se condujo una datación de la topología reconstruida previamente con la matriz combinada (figura 5). Se determinó que la divergencia entre el grupo de los clados *Cyphomandra* y *Leptostemomun* ocurrió entre 10,57 y 11,98 millones de años, rango cercano al determinado para el clado *Papa* que se sugiere ocurrió entre 10,5 y 11,92 millones de años. Los resultados del estudio sugieren que la divergencia entre las especies del clado *Leptostemomun* ocurrió recientemente, entre 970 mil años y 1,09 millones de años, y aún más recientemente, la divergencia entre la especies del clado *Cyphomandra* habría ocurrido hace 430 mil años a 480 mil años.

### Análisis de resultados

Las topologías construidas con los genes individuales muestran para la mayoría de casos, clados no definidos que son resueltos en los análisis de la matriz combinada. El marcador At2g47580 pareciera tener poder para discriminar las especies de reciente divergencia, ya que define el clado *Cyphomandra* y el *Leptostemomun*; sin embargo, no podría diferenciar al género *Solanum* ya que incluye en una politomía basal una especie del género hermano *Capsicum* (Bohs, 1991; 2002; 2004). De otra parte, el marcador At3g07100 podría discriminar los clados de divergencia



**Figura 5.** Datación de la filogenia de las especies estudiadas con el método LF y relojes moleculares locales. Los puntos de calibración fueron las fechas de divergencia de *A. thaliana* (85-90 millones de años), *C. canephora* (83-89 millones de años) y el orden Solanales (82-86 millones de años). La divergencia entre *S. hirtum* y *S. quitoense* se estima que ocurrió hace 970 mil años ó 1,09 millones de años, y la divergencia entre *S. betaceum* y *S. unilobum* hace 430 mil ó 480 mil años

basal, incluyendo *C. canephora* y el género *Solanum*, pero no podría discriminar las diferencias entre los clados *Cyphomandra* y *Leptostemomun* y en adición agrupa el clado *Papa* con *C. annuum*. La topología obtenida con el marcador At3g52120 está completamente dicotomizada y define los clados *Papa*, *Cyphomandra* y *Leptostemomun*, agrupando los dos últimos en un subclado; sin embargo, este gen une a *C. annuum* con el clado *Papa* en su parte basal. El marcador At3g63190 agrupa en su cladograma a las especies del clado *Papa*, pero a éste lo agrupa en una politomía con *C. annuum* y las especies del clado *Cyphomandra*; en la misma topología, el clado que agrupa a todas estas especies se une en una politomía en su parte basal con las dos especies del clado *Leptostemomun*. El gen At3g63490 no tiene poder para discriminar a ninguno de los clados del género *Solanum* y agrupa a *C. canephora*, *C. annuum*, *S. tuberosum* y *S. lycopersicum* en un grupo en una politomía basal con las especies restantes. Finalmente, el marcador At4g34350 diferencia la totalidad de los clados de la misma forma como lo hace la matriz de todos los genes unidos. En esta topología se agrupan los clados *Cyphomandra* y *Leptostemomun* en un subclado que se une con el clado *Papa*, en su parte basal, el clado que agrupa todas estas especies se une con *C. annuum* y éste a su vez con *C. canephora*.

Como producto de la edición, para filtrar las regiones no codificantes de las secuencias genómicas, la longitud de las matrices de cada marcador es inferior a la de aquellas normalmente utilizadas para hacer análisis filogenéticos (como *ITS*, *trnL*, *ndhF*, *rpl16* entre otros (Acevedo-Rosas *et al.*, 2004; Chaw *et al.*, 1997; Downie *et al.*, 2000; Graham y Olmstead, 2000; Sang *et al.*, 1997). Sin embargo, los índices de consistencia (CI) y de retención (RI) para todas las topologías construidas fueron altos ( $> 0,878$  y  $> 0,720$  respectivamente), mientras que los índices de homoplasia fueron bajos ( $< 0,125$ ) (tabla 1). De tal forma, a pesar del tamaño de las secuencias, los datos muestran ser suficientemente robustos para hacer inferencias elaboradas sobre ellos. En comparación con

otros marcadores, los COSII utilizados muestran dar una buena resolución para definir las topologías de las solanáceas. Al respecto, las reconstrucciones filogenéticas hechas por D'Arcy (1991) con el gen *ndhF* (aproximadamente 2000 pares de bases) y por Weese y Bohs (2007) (cerca de 2100 pares de bases) con el interespeciador *trnT-F*, no pudo dar resolución para los clados del género *Solanum*, agrupando los clados *Papa*, *Leptostemomun* y *Cyphomandra* en una politomía. Sin embargo, la topología construida con el gen At4g34350 (de 327 pares de bases) y aquella producto de la unión de todos los genes (de 823 pares de bases), resuelven completamente y con buen soporte las relaciones entre estos clados.

Los cladogramas obtenidos mediante análisis por máxima parsimonia y máxima verosimilitud de la matriz de datos combinada, definen con un buen soporte clados a diferentes niveles de manera consistente a lo consensuado por el Grupo de Filogenia de Angiospermas (APG) (Stevens, 2001 y en adelante). En primer lugar se agrupan, como es esperado, todas las especies en el clado astéridas I, exceptuando al grupo externo *Arabidopsis thaliana*. El siguiente clado incluye todas las especies de la familia Solanaceae, y excluye a *C. canephora*, perteneciente a la familia Rubiaceae del orden Gentianales. El siguiente clado agrupa todas las especies del género *Solanum* y excluye a *C. annuum*, especie perteneciente al género *Capsicum* que ha sido reportado en varios estudios como taxón hermano del género *Solanum* (D'Arcy, 1991; Olmstead *et al.*, 1999; Weese y Bohs, 2007). El siguiente clado agrupa las especies *S. lycopersicum* y *S. tuberosum* con un buen soporte, este mismo clado ha sido denominado *Papa* (*Potato*) en estudios previos (Weese y Bohs, 2007). Este clado se une con otro que agrupa los reconocidos como *Cyphomandra* (integrado por *S. betaceum* y *S. unilobum*) y *Leptostemomun* (integrado por *S. hirtum* y *S. quitoense*), siendo consistentes con las reconstrucciones hechas por Weese y Bohs (2007) para las especies del género *Solanum* con los genes *ndhF* y *waxy*. Nuestros resultados siguieron

**Tabla 1.** Comparación de los grupos de datos utilizados en el presente estudio

	At2g47580	At3g07100	At3g52120	At3g63190	At3g63490	At4g34350	Todos
Número de pasos en árbol(es) más parsimonioso(s)	68	87	67	24	16	172	439
Número de árboles más parsimoniosos	4	111	1	2	26	1	1
Número de caracteres (longitud de secuencia)	115	139	140	67	35	327	823
Número de caracteres informativos	16	17	25	6	5	64	133
Índice de consistencia	0,897	0,989	0,910	0,958	0,875	0,878	0,902
Índice de retención	0,720	0,947	0,875	0,917	0,778	0,772	0,790
Índice de consistencia reescalado	0,646	0,936	0,797	0,878	0,681	0,678	0,713
Índice de homoplasia	0,103	0,011	0,090	0,042	0,125	0,122	0,098

la necesidad de conducir estudios taxonómicos posteriores para buscar caracteres homólogos que diferencien el subclado *Leptostemomun* + *Cyphomandra* para luego asignarle una jerarquía taxonómica.

Para la datación realizada se utilizó una metodología que permite otorgar diferentes tasas de mutaciones a diferentes clados, esta mostró producir no sólo resultados robustos sino también consistentes con al menos un reporte previo: las fechas obtenidas para la divergencia de papa (*S. tuberosum*) y tomate (*S. lycopersicum*) son cercanas a las sugeridas previamente por Alba y colaboradores (2000) (aproximadamente 10 millones de años). La fecha de divergencia entre las especies del clado *Cyphomandra*, *S. betaceum* y *S. unilobum* muestra ser muy reciente (de 430 mil a 480 mil años), este tiempo corto de divergencia limitaría el número acumulado de mutaciones así como las barreras genéticas para su cruzamiento, razón por la cual se obtienen híbridos vigorosos entre ellos (Lobo *et al.*, 2000). El tiempo calculado para este evento se ubica entre el período glacial de Kansan (de 0,30 a 0,45 millones de años) y el interglacial aptoniano (0,45 a 0,64 millones de años) (Cossios *et al.*, 2009), lapso asociado con eventos de aislamiento y posterior expansión/divergencia en el sur de los Andes, producidos por cambios climáticos (Cossios *et al.*, 2009). Los resultados obtenidos en la presente investigación indican que la divergencia de *S. hirtum* y *S. quitoense* habría ocurrido en un período comprendido entre 970 mil y 1,09 millones de años, en el período glacial prepastoniano (0,80 a 1,30 millones de años), en el que ocurrieron las más extensivas glaciaciones en los Andes (Manuel, 2004) y múltiples eventos de diversificación de grandes clados de animales suramericanos (Cossios *et al.*, 2009; Manuel, 2004).

La clarificación de las relaciones filogenéticas entre los clados del género *Solanum* permitirá conducir con precisión futuros estudios de detección de genes ortólogos para plantas de este género. Estos estudios permitirán trasladar el conocimiento que se tiene de genes de interés agronómico (i.e. resistencia biótica y abiótica, relacionados a condiciones organolépticas o características fisio-

lógicas) en especies del género ampliamente estudiadas (como papa y tomate), a otras menos conocidas, nativas y prominentes. En experimentos previos (Jeanneau *et al.*, 2002) se ha probado de forma exitosa la búsqueda de genes ortólogos para mejoramiento vegetal, en este caso, para mejorar la resistencia a las sequías en maíz.

## CONCLUSIONES

Los marcadores COSII secuenciados mostraron ser robustos para resolver completamente las relaciones entre todos los clados estudiados.

Mediante análisis filogenéticos se confirmaron tres clados dentro del género *Solanum*: *Papa*, *Cyphomandra* y *Leptostemomun*, los dos últimos forman a su vez un subclado.

Se determinó que la fecha de divergencia de *S. hirtum* y *S. quitoense* habría ocurrido hace 970 mil a 1,09 millones de años, mientras que la de *S. unilobum* y *S. betaceum* habría ocurrido hace 430 mil a 480 mil años. Ambas fechas coinciden con períodos de cambios climáticos a los que se les ha atribuido la diversificación de varios clados andinos.

La clarificación de las relaciones filogenéticas dentro de los clados del género *Solanum*, permitirá a futuro aplicar el conocimiento de genes de interés agronómico en solanáceas ampliamente estudiadas (como tomate y papa), a otras nativas de Colombia prominentes como lulo y tomate de árbol.

## AGRADECIMIENTOS

Los autores expresan sus agradecimientos a S.D. Tanskley y F. Wu por la información suministrada y el soporte para el trabajo con los marcadores COSII. También agradecen a L. Bohs y M. Lobo por proveer el material vegetal para el presente estudio. La secuenciación para este estudio fue financiada por Colciencias y la NSF, la financiación restante fue suministrada por Corpoica.

## REFERENCIAS

- Acevedo-Rosas R, Cameron K, Sosa V, Pell S. 2004. A molecular phylogenetic study of *Graptopetalum* (Crassulaceae) based on ETS, ITS, RPL16, and TRNL-F nucleotide sequences 91(7):1099-1104.
- Alba R, Kelmenson PM, Cordonnier-Pratt M-M, Pratt LH. 2000. The phytochrome gene family in tomato and the rapid differential evolution of this family in Angiosperms. *Molecular Biology and Evolution* 17(3):362-373.
- Benítez A, Valencia J, Estrada EI, Baena D. 1991. Phenotypic characterization of certain accessions in the lulo (*Solanum quitoense*) germplasm bank. En: Hawkes J, Lester RN, Nee M, Estrada N (eds.) *Solanaceae III: Taxonomy, Chemistry, Evolution*. Royal Botanic Gardens Kew and Linnean Society of London, London. pp. 437-450.
- Bernatzky R, Tanksley SD. 1986a. Genetics of actin-related sequences in tomato. *Theoretical and Applied Genetics* 72:314-321.
- Bernatzky R, Tanksley SD. 1986b. Toward a saturated linkage map in tomato based on isozymes and random cDNA sequences. *Genetics* 112:887-898.
- Bohs L. 1994. *Cyphomandra* (Solanaceae) *En Flora Neotropica* vol. 63 New York Botanical Garden, p. 1-175.
- Bohs L. 1991. Crossing studies in *Cyphomandra* (Solanaceae) and their systematic and evolutionary significance. *American Journal of Botany* 78(12):10.
- Bohs L. 2004. A chloroplast DNA phylogeny of *Solanum* section *Lasiocarpa*. *Systematic Botany* 29(1):177-187.
- Bohs L. 2007. Phylogeny of the *Cyphomandra* clade of the genus *Solanum* (Solanaceae) based on ITS sequence data. *Taxon* 56(4):1012-1026.
- Chaw SM, Zharkikh A, Sung HM, Lau TC, Li WH. 1997. Molecular phylogeny of extant gymnosperms and seed plant evolution: analysis of nuclear 18S rRNA sequences. *Molecular Biology and Evolution* 14(1):56-68.
- Chesser RT. 2000. Evolution in the High Andes: The phylogenetics of *Muscisaxicola* ground-tyrants. *Molecular Phylogenetics and Evolution* 15(3):369-380.
- Chung HJ, Jung JD, Park HW, Kim JH, Cha HW, Min SR, Jeong WJ, Liu JR. 2006. The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. *Plant Cell Reports* 25(12):1369-1379.
- Clegg MT. 1993. Chloroplast gene sequences and the study of plant evolution. *PNAS* 90:363-367.
- Cossios D, Lucherini M, Ruiz-García M, Angers B. 2009. Influence of ancient glacial periods on the Andean fauna: the case of the pampas cat (*Leopardus colocolo*). *BMC Evolutionary Biology* 9(1):68.
- D'Arcy WG. 1991. The Solanaceae since 1976, with a review of its biogeography. En: Hawkes JG, Lester RN, Nee M, Estrada E (eds.) *Solanaceae III: Taxonomy, Chemistry, Evolution*. Royal Botanic Garden, Kew, pp. 75-137.
- Downie SR, Katz-Downie DS y Watson MF. 2000. A phylogeny of the flowering plant family Apiaceae based on chloroplast DNA rpl16 and rpoC1 intron sequences: towards a suprageneric classification of subfamily Apioideae. *American Journal of Botany* 87(2):273-292.
- Ewing B, Green P. 1998. Base-Calling of Automated Sequencer Traces Using Phred. II. Error Probabilities. *Genome Research* 8(3):186-194.
- Gordon D, Abajian C, Green P. 1998. Consed: A Graphical Tool for Sequence Finishing. *Genome Research* 8(3):195-202.
- Gracie C. 1993. Pollination of *Cyphomandra endopogon* var. *endopogon* (Solanaceae) by *Eufriesea* spp. (Euglossini) in French Guiana. *Brittonia* 45(1):39-46.
- Graham SW and Olmstead RG. 2000. Utility of 17 chloroplast genes for inferring the phylogeny of the basal angiosperms. *American Journal of Botany* 87:1712-1730.
- Heiser C. 1972. The Relationships of the Naranjilla, *Solanum quitoense*. *Biotropica* 4(2):77-84.
- Heiser C. 1985. Ethnobotany of the Naranjilla (*Solanum quitoense*) and its relatives. *Economic Botany* 39(1): 4-11.
- Heiser C. 1989. Artificial Hybrids in *Solanum* sect. *Lasiocarpa*. *Systematic Botany* 14(1):3-6.
- Heiser C, Anderson G. 1999. "New" Solanums. En: Perspectives on new crops and new uses. Janick J. ASHS Press, Alexandria, VA. pp. 379-384.
- Huang S, van der Vossen EA, Kuang H, Vleeshouwers V, Zhang N, Borm T, van Eck HJ, Baker B, Jacobsen E, Visser R. 2005. Comparative genomics enabled the isolation of the *R3a* late blight resistance gene in potato. *The Plant Journal* 42(2): 251-261.
- Jeanneau M, Gerentes D, Foueillassar X, Zivy M, Vidal J, Toppan A, Perez P. 2002. Improvement of drought tolerance in maize: towards the functional validation of the *Zm-Asr1* gene and increase of water use efficiency by over-expressing C4-PEPC. *Biochimie* 84(11):1127-1135.
- Knapp S. 2002. Tobacco to tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae. *Journal of Experimental Botany* 53(377):2001-2022.
- Kou MC, Yen JH, Hong JT, Wang CL, Lin CW, Wu MJ. 2009. *Cyphomandra betacea* Sendt. phenolics protect LDL from oxidation and PC12 cells from oxidative stress. *Lwt-Food Science and Technology* 42(2):458-463.
- Lobo M, Medina CI, Cardona M. 2000. Resistencia de campo a la antracnosis de los frutos (*Colletotrichum gloeosporioides*) en tomate de árbol (*Cyphomandra* (Solanum) *betacea* (betaceum) Cav. Sendt.). *Revista Facultad Nacional de Agronomía, Medellín* 53(2):1129-1141.
- Lobo M. 2006. Recursos genéticos y mejoramiento de frutales andinos: una visión conceptual. *Revista Corpoica – Ciencia y Tecnología Agropecuaria* 7(2):40-54.
- Lobo M, Medina CI, Delgado OA, Bermeo A. 2007. Morphological variability of the Colombian collection of lulo (*Solanum quitoense* Lam.) and related *Lasiocarpa* section species. *Revista Facultad Nacional de Agronomía, Medellín* 60(2):3939-3964.
- Manoko MLK, Van den Berg RG, Feron RMC, Van der Weerden GM, Mariani C. 2008. Genetic diversity of the African hexaploid species *Solanum scabrum* Mill. and *Solanum nigrum* L. (Solanaceae). *Genetic Resources and Crop Evolution* 55(3):409-418.
- Manuel N. 2004. The implications of Tertiary and Quaternary sea level rise events for avian distribution patterns in the lowlands of northern South America. *Global Ecology & Biogeography* 13(2):149-161.
- Miller JS, Diggle PK. 2003. Diversification of andromonoecy in *Solanum* section *Lasiocarpa* (Solanaceae): The roles of phenotypic plasticity and architecture. *American Journal of Botany* 90(5):707-715.
- Morgenstern B. 2004. DIALIGN: multiple DNA and protein sequence alignment at BiBiServ. *Nucleic Acids Research* 32(suppl. 2): W33-36.
- Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD. 2005. The SOL Genomics Network. A Comparative Resource for Solanaceae Biology and Beyond. *Plant Physiology* 138(3):1310-1317.
- Olmstead RG, Sweere JA, Spangler RE, Bohs L, Palmer JD. 1999. Phylogeny and provisional classification of the Solanaceae based on chloroplast DNA. En: M Nee, DE Symon, RN Lester, JP Jessop. *Solanaceae IV*. Royal Botanic Gardens, Kew. pp. 111-137.

- Palmer JD, Jansen RK, Michaels HJ, Chase MW and Manhart JR. 1998. Chloroplast DNA Variation and Plant Phylogeny. *Annals of the Missouri Botanical Garden* 75(4):1180-1206.
- Posada D, Crandall KA. 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14(9):817-818.
- Pratt RC, Francis DA, Barrero LS. 2008. Genomics of Tropical Solanaceous species: Established and emerging crops. En: Moore PH, Ming R. (eds.) *Plant Genetics, Genomics: Crops and Models*. Springer, New York, NY, EU. pp. 453-467.
- Sanderson M. 2006. r8s, <http://loco.biosci.arizona.edu/r8s/>. University of California, Davis, Davis, California, EEUU.
- Sang T, Crawford DJ y Stuessy TF. 1997. Chloroplast DNA phylogeny, reticulate evolution, and biogeography of *Paeonia* (Paeoniaceae) *American Journal of Botany* 84(9):1120-1136.
- Stevens PF. 2001 y en adelante. Angiosperm Phylogeny Website. Version 9, June 2008 [and more or less continuously updated since].
- Swofford D. 2004. Paup 4.0 for Macintosh: Phylogenetic Analysis Using Parsimony (Software and User's Book for Macintosh). Sinauer Associates, Incorporated.
- Weese TL, Bohs L. 2007. A Three-Gene Phylogeny of the genus *Solanum* (Solanaceae). *Systematic Botany* 32: 445-463.
- Wikstrom N, Savolainen V, Chase MW. 2001. Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society B*. 268:2211-2220.
- Wu F, Eannetta NT, Xu Y, Durrett R, Mazourek M, Jahn MM, Tanksley SD. 2009a. A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. *Theoretical and Applied Genetics* 118(7):1279-1293.
- Wu F, Eannetta NT, Xu Y, Tanksley SD. 2009b. A detailed synteny map of the eggplant genome based on conserved ortholog set II (COSII) markers. *Theoretical and Applied Genetics* 118(5):927-935.