



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

---

---

FACULTAD DE ESTUDIOS SUPERIORES IZTACALA

**Análisis de los cambios estructurales del  
genoma completo de cloroplasto de especies  
del orden Caryophyllales**

T E S I S

Que para obtener el grado académico de:

Licenciado de Biología

P R E S E N T A:

Lucio Omar Estévez Rico

DIRECTORA DE TESIS:  
Dra. Sofía Solórzano Lujano



Estado de México, México, 2021



Universidad Nacional  
Autónoma de México



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

## **Agradecimientos Académicos**

A los integrantes del jurado revisor, al Dr. Ángel Salvador Arias Montes, a la M. en C. Delil Andrea Chincoya Martínez, a la Dra. Patricia Dávila Aranda y a la Dra. Sandra Luz Gómez Acevedo por sus excelentes comentarios y sugerencias que mejoraron este trabajo. A la Dra. Sofía Solórzano Lujano, directora de tesis que desde el primer momento mostró interés en el desarrollo del proyecto y siempre estuvo al tanto del progreso de este y de mi aprendizaje.

Esta tesis forma parte del proyecto de investigación PAPIIT-UNAM IN228619 “Caracterización de la estructura del genoma completo de cloroplasto de *Mammillaria* y sus especies cercanas”. Gracias a la Universidad Nacional Autónoma de México, porque me brindó el apoyo para mi formación académica y profesional desde mi ingreso a nivel preparatoria hasta la finalización de la carrera.

## **Agradecimientos Personales**

A mis padres, Lucio Estévez Martínez y María Magdalena Rico Esqueda: cuyo apoyo, cariño, comprensión y paciencia durante todos estos años ha servido para convertirme en una persona dedicada y fuerte para el porvenir.

A la familia Rico: que siempre estuvo interesada en mi desarrollo académico y personal, brindando un apoyo moral y humano siempre que pudieron.

A mis mejores amigos; Abraham, Ernesto, Javier y Lucio: con quienes he construido una amistad y confianza inquebrantable y que han sido parte de etapas importantes de mi vida; incluyendo ésta.

A los amigos y compañeros del grupo 04: con quienes compartí los 4 años de carrera y que, tanto en las aulas de la Facultad como en las prácticas de campo, siempre hicieron del aprendizaje una experiencia más amena.

## Índice

|   |    |
|---|----|
| 1. Resumen  | 5  |
| 2. Antecedentes                                   | 8  |
| 3. Objetivos                                      | 14 |
| Objetivo General                                  | 14 |
| Objetivos Particulares                            | 14 |
| 4. Métodos  | 15 |
| Recopilación y curado de las secuencias in silico | 15 |
| Alineamiento estructural.                         | 16 |
| Matriz de caracteres                              | 16 |
| 5. Resultados                                     | 16 |
| 6. Discusión                                      | 27 |
| 7. Conclusiones                                   | 29 |
| 8. Referencias                                    | 30 |
| 9. Anexos   | 38 |

## 1. Resumen

El orden Caryophyllales agrupa a ca. 11,555 especies que se distribuyen en 749 géneros y 40 familias. Este orden forma parte del grupo monofilético de las eudicotiledóneas de acuerdo con la filogenia de la APG III. Los caracteres diagnósticos que unifican a los miembros de este orden son la nula formación de micorrizas, su gineceo unilocular con placentación central, un polen tipo colpado y la presencia de betalainas o antocianinas. En este orden se incluyen especies con hábitos de vida y adaptaciones morfológicas contrastantes, entre las que destacan los cactus (Cactaceae), las piedras vivas (Aizoaceae) y las plantas carnívoras (Droseraceae). También se agrupan especies que sirven de alimento a la especie humana (Amaranthaceae, Euphorbiaceae, Portulacaceae). Actualmente, Caryophyllales cuenta con 21 secuencias completas de genoma mitocondrial, 423 secuencias completas del genoma de cloroplasto y 325 secuencias nucleares. En particular, la comparación de los genomas completos de cloroplasto permite identificar cambios en el número e identidad de los genes que lo componen y en cómo éstos se organizan dentro del genoma, lo que comúnmente se refiere a la estructura del genoma de cloroplasto. A partir de estas comparaciones, entre algas y plantas terrestres, y dentro de grupos taxonómicos particulares, se han identificado cambios estructurales tales como la pérdida o ganancia de genes, rearrreglos génicos y la formación o pérdida de inversos repetidos (IRs); mismos que se interpretan como pasos y procesos evolutivos. De acuerdo con los antecedentes, los genomas de cloroplasto de las especies de Caryophyllales están seccionados en cuatro partes: una región grande (LSC), una región pequeña (SSC) y dos inversos repetidos (IRs), pero también se ha registrado la pérdida de los IRs en algunos cactus columnares (Cactaceae). Por ello, en esta tesis se espera que la tendencia sea que las especies del orden Caryophyllales tengan en su genoma de cloroplasto una estructura seccionada en cuatro regiones. El objetivo general de esta tesis fue identificar los cambios estructurales del genoma completo de cloroplasto en especies del orden Caryophyllales para determinar si estos cambios son evolutivamente informativos. Los objetivos particulares fueron caracterizar la composición génica y la longitud de los IRs y construir una filogenia para las 70 especies estudiadas a partir de la estructura del genoma completo de cloroplasto. Para cumplir estos objetivos, se usaron las 70 secuencias *in silico* del genoma completo de cloroplasto, que representan 16 familias y 40 géneros diferentes, debido a que eran las secuencias disponibles al inicio de esta tesis. Las 70 secuencias son de acceso libre y fueron previamente ensambladas en sus estudios originales, sin embargo, 53 de ellas se publicaron sin su anotación por ello, previo a los análisis, estas secuencias se anotaron en la plataforma gratuita de CHLOROBX. Las 70 secuencias se revisaron en su ensamblado y en

su anotación y fueron alineadas estructuralmente con MAUVE. Para determinar los cambios entre las secuencias estudiadas, se caracterizaron tanto la identidad como el número de genes del genoma completo, enfatizando en la extensión y composición de los IRs. Con el fin de determinar las relaciones filogenéticas de las especies estudiadas, se construyó una matriz basada en caracteres estructurales, para ello, los nombres de los genes y su ubicación relativa se asignaron a categorías numéricas. Esta matriz integró un total de 106 caracteres y se analizó en el programa Mesquite. Los resultados principales de esta tesis confirman que en Caryophyllales el genoma completo de cloroplasto se organiza en cuatro secciones típicas de angiospermas ya que, de las 70 especies aquí estudiadas, sólo dos especies de cactus columnares (*Carnegiea gigantea* y *Lophocereus schottii*) carecen de IRs. Por tanto, es una novedad evolutiva la pérdida de IRs para todo el orden y para la familia (Cactaceae). En esta familia, los IRs se registran en las nueve especies de cactus de los géneros *Mammillaria*, *Opuntia* y *Rhipsalis* para las que se tienen secuenciados los genomas de cloroplasto. En el caso de las plantas carnívoras (Droseraceae), en particular *Dionaea muscipula*, presenta un acortamiento de sus IRs a < 2.1 kbp mientras que, en las otras tres especies analizadas de distintos géneros, su longitud varía de 23 a 27 kbp. Este acortamiento también se registró en las especies *Mammillaria albiflora* y *M. pectinifera* cuyos IRs tienen una longitud de 1.3 kbp y 1.5 kbp, respectivamente, en tanto que en otras cinco especies del mismo género su longitud varía de 14 a 28 kbp. Estas comparaciones permiten establecer que la diversidad estructural del genoma de cloroplasto se presenta incluso dentro de un mismo género, no solo a nivel de familia y de orden. Sin embargo, en esta tesis no se pudieron identificar las fuerzas y mecanismos evolutivos que dirigen dichos cambios estructurales. Los resultados muestran que el genoma completo de cloroplasto, en el orden Caryophyllales, tiende a una composición numérica y de identidad de genes altamente conservada, independientemente de si hay o no la formación de IRs. Entre las 70 especies, los genes que codifican variaron de 89 (*Dionaea muscipula*, Droseraceae) a 113 (*Amaranthus caudatus*, Amaranthaceae), los ARN de transferencia de 30 (*Carnegiea gigantea*, Cactaceae) a 36 (*Fagopyrum esculentum*, Polygonaceae), y el total de las especies estudiadas contienen cuatro ARN ribosomales (*rrn4.5*, *rrn5*, *rrn16* y *rrn23*) y cuatro genes de la ARN polimerasa (*rpoA*, *rpoB*, *rpoC1* y *rpoC2*). El análisis filogenético basado en la estructura del cloroplasto verificó los resultados de las filogenias previas, en las que se separan las llamadas Caryophyllales centrales de las no centrales. A partir de esta filogenia se concluye que la estructura del genoma completo de cloroplasto representa información filogenética que debe analizarse en estudios futuros para entender cabalmente las fuerzas evolutivas subyacentes a esos cambios. La filogenia de esta

tesis también permite concluir que la pérdida y acortamientos de los IRs documentados en Cactaceae y Droseraceae son eventos evolutivos que han ocurrido independientemente. Esta tesis confirma que la pérdida completa de los genes de la familia *ndh*, o su pseudogenización, al igual que la de otros genes, como los *ycf* en Cactaceae, son cambios estructurales que repercuten en el acortamiento relativo del tamaño total del genoma del cloroplasto. También se concluye que el análisis de los cambios estructurales debe profundizarse en estudios que relacionen estructura del genoma con su función en la fotosíntesis.



## 2. Antecedentes

Las cianobacterias son procariotes y por tanto carecen de cloroplastos, pero aun así contienen como genoma una molécula de ADN circular (Kaneko y Tabata, 1997). Actualmente, prácticamente todos los análisis de secuencias de ADN sugieren un ancestro cianobacteriano de los cloroplastos de las algas y de las plantas terrestres (Raven y Allen, 2003). Las cianobacterias, tales como *Oxyphotobacteria* y *Melainabacteria*, tuvieron un papel fundamental en la formación de la biósfera terrestre debido a su capacidad fotosintética. La fotosíntesis transforma el dióxido de carbono en oxígeno, proceso que consolidó las formas de vida ancestrales que en su momento prosperaron en el planeta Tierra (Soo et al. 2017). El origen posible de las cianobacterias se remonta al Proterozoico Inferior, hace unos 2,500-1,700 millones de años, con la aparición de una atmósfera con contenido de oxígeno muy cercano al de la actualidad. Esto favoreció la evolución de microorganismos anaerobios (Shestakov y Karbysheva, 2017). El biólogo y botánico ruso, Konstantín Merezhkovski (1905) presentó oficialmente la teoría de la simbiogénesis donde manifiesta que la selección natural no se produce mediante mutaciones aleatorias sino por la incorporación de simbioses (Brummitt y Powell, 1992). Posteriormente, esta teoría sería retomada y modificada por Lynn Margulis (1967), para convertirse en la Teoría de la Endosimbiosis Seriada.

En la Teoría de la Endosimbiosis Seriada se describe la existencia de sucesivas simbiosis hasta la aparición de las células eucariotas tal como las conocemos actualmente. En esta teoría, se propone una primera simbiosis que se produjo al fusionarse una bacteria nadadora (del tipo espiroqueta) con otra que utilizaba el azufre y el calor como fuente de energía, originando un organismo con las características de ambas. Este sería el primer eucarionte ancestral de todos los organismos pluricelulares actuales. La segunda simbiosis se propone que surgió entre este eucarionte anaerobio y una bacteria aerobia capaz de realizar la respiración celular. Así surgieron las células eucariotas con mitocondrias que darían lugar a hongos y animales. La tercera simbiosis se originó entre estos últimos organismos aerobios y las cianobacterias, mismas que aportaron la capacidad de obtener energía a partir de materia inorgánica mediante la fotosíntesis. Así surgieron los cianoplastos en glaucofitas, rhodoplastos en algas rojas, y los cloroplastos en algas verdes y plantas terrestres (Margulis, 2002 y Kleine et al. 2009). Las secuencias de ADN de cloroplasto analizadas en una perspectiva filogenética han apoyado la evidente cercanía entre los cloroplastos de plantas como derivados directos de las cianobacterias (Douglas y Raven, 2003).

La función principal de los cloroplastos es la conversión de energía solar en carbohidratos, proceso conocido como fotosíntesis y que se conforma de dos partes: las reacciones de luz que ocurren en el sistema de membranas del tilacoide del cloroplasto y que producen ATP y NADPH, y las reacciones de carbono independientes de la luz que usan los productos anteriores para fijar el CO<sub>2</sub> atmosférico en moléculas orgánicas (Johnson, 2016). La fotosíntesis se traduce en una reacción química simple:  $6\text{CO}_2 + 6\text{H}_2\text{O} \rightarrow \text{C}_6\text{H}_{12}\text{O}_6 + 6\text{O}_2$  (Whittingham, 1952). También en los cloroplastos, se producen aminoácidos, nucleótidos, lípidos, fitohormonas y vitaminas (Alberts et al. 2002) además de una gran cantidad de metabolitos esenciales a las respuestas de la planta ante el calor, la sequía, la salinidad e incluso la presencia de algunos patógenos (Bobik y Burch-Smith, 2015).

Los cloroplastos se encuentran localizados en todo el citoplasma de las células vegetales de las plantas terrestres en un número que varía desde uno, en algas unicelulares, hasta 100 por célula en plantas como *Arabidopsis thaliana* y *Triticum aestivum* (Ono et al. 1995, Trojan y Gabrys, 1996 y Wardley et al. 1984). Gracias a la fotosíntesis se pueden sostener las complejas relaciones tróficas de todas las formas de vida del planeta. Sin embargo, algunas especies de algas del género *Polytomella*, y plantas del género *Epifagus* y *Monotropa*, presentan cloroplastos que no son funcionales debido a la retención de los pseudogenes para la mayoría de los genes relacionados a la fotosíntesis (Smith y Lee, 2014, Abbate y Campbell, 2013, Logacheva et al. 2016). Además, existen organismos parcialmente heterótrofos, como es el caso de las plantas carnívoras, en donde algunos nutrientes provienen directamente de sus presas mientras que otros son procesados mediante cloroplastos funcionales (Poppinga et al. 2013).

Eventualmente, la fotosíntesis primitiva en medios completamente acuáticos permitió la colonización de la superficie terrestre hasta el surgimiento de las plantas superiores en este medio, lo que fue un evento de suma importancia en la evolución de las plantas terrestres, y de todas las formas de vida (Kenrick y Crane, 1997). A partir de ahí, se originaron los dos grandes grupos de plantas conocidos como angiospermas y gimnospermas hace aproximadamente 180 millones de años (Soltis et al. 2008). Cerca del 40% de las gimnospermas están representadas por un diverso grupo de especies de coníferas distribuidas en ocho familias, 70 géneros y más de 630 especies (Farjon, 2001). Los estudios de genómica comparada de los cloroplastos en estas plantas han revelado elementos transponibles, secuencias repetidas y duplicación de genes como características comunes (Morse et al. 2009).

En plantas terrestres se ha documentado una mayor frecuencia de cloroplastos con una molécula de ADN de doble cadena el cual se hereda a partir de los óvulos (Sugiura, 1989). Una de las tendencias evolutivas identificadas es que el genoma completo del cloroplasto (ADNcp, de aquí en adelante) de las plantas terrestres tiene el tamaño más pequeño en relación con los genomas mitocondrial y nuclear de la misma especie (Alexeyev et al. 2004, Greilhuber et al. 2005) pero también con el de sus ancestros cianobacterianos (Gray & Doolittle, 1982). En plantas terrestres el ADNcp varía de 100 a 200 kbp (Beverley, 2011) y en algunos casos puede alcanzar hasta 217,000 bp de longitud total con un número de genes que varían de 29 a 147 (De las Rivas et al. 2012). Actualmente, se conoce que la maquinaria fotosintética requiere muchas más funciones genéticas que aquellas que están especificadas en el ADNcp, por tanto, se asume que muchas de estas funciones se transfirieron al genoma nuclear (Rochaix, 2004) o mitocondrial (Brandvain y Wade, 2009). A pesar de esas transferencias o migraciones de genes desde el cloroplasto a otros genomas su contenido genético está altamente conservado (Gantt et al. 1991). Este contenido genómico conservado y los cambios relativos lentos de sustitución de nucleótidos en genes codificantes hacen que el ADNcp sea un modelo base para el estudio de la historia evolutiva en plantas terrestres (Clegg, 1993).

En particular, en el grupo de las angiospermas terrestres es una tendencia que su ADNcp esté organizado en cuatro secciones por ello se dice que tiene una estructura cuatripartita (Downie y Palmer, 1994). Dos de ellas se llaman regiones únicas (SC, por sus siglas en inglés "single copy"), una de éstas tiene un mayor número de genes que se refleja en una mayor longitud y por ello se nombra como región única grande (LSC, por sus siglas en inglés) y a la otra región única pequeña (SSC, por sus siglas en inglés); las otras dos regiones se llaman inversos repetidos (IRs). En la mayoría de las plantas terrestres la LSC mide de 80 a 90 kbp y está compuesta por cerca de 60 genes codificantes y 21 ARNs de transferencia (ARNt); por su parte la SSC suele tener una longitud entre 16 a 27 kbp con un promedio de 11 genes codificantes mientras que los IRs, principalmente compuestos de 7 ARNt y 4 ARNr (Li et al. 2018 y Mower y Vickrey, 2018), oscilan entre los cinco y los 27 kbp de longitud. Sin embargo, dentro de cada especie es usual que los dos IRs sean idénticos en longitud y en número de genes, e incluso en secuencias idénticas de ADN, lo único que lo diferencia es la orientación (un IR está en sentido río abajo respecto al otro) (Zhu et al. 2016). Sin embargo, en la familia Cactaceae, para las especies *Mammillaria zephyranthoides* (Solórzano et al. 2019) y *Rhipsalis baccifera* (Oulo et al. 2020) se documentaron IRs divergentes, es decir, que difieren en la composición de sus genes.

Dicha composición generalmente incluye un conjunto de cuatro genes ribosomales (*rrn4.5*, *rrn5*, *rrn16* y *rrn23*) y siete ARNt (*trnA-UGC*, *trnL-CAA*, *trnI-GAU*, *trnM-CAU*, *trnN-GUU*, *trnR-ACG* y *trnV-GAC*) a los cuales se les ha atribuido la estabilidad estructural del ADNcp debido a su baja tasa de sustitución de nucleótidos (Zheng et al. 2016).

Los datos actuales muestran que es un patrón que las angiospermas terrestres tengan una estructura cuatripartita en su ADNcp, aunque pocas especies no poseen los IRs. Esta carencia de IRs se ha interpretado como una pérdida evolutiva (Sanderson et al. 2015) puesto que su estado ancestral fue la estructura cuatripartita. Los tres grandes cambios estructurales en la organización del ADNcp de plantas terrestres que se han señalado son la pérdida de los IRs, la pérdida completa de familias de genes y la pseudogenización (Oldenburg y Bendich, 2015). Estos tres cambios estructurales han sido registrados en distintas especies de los miembros de algunas familias, pero no para el total de sus miembros. Por ejemplo, en la familia Cactaceae, la pérdida de los IRs, junto con la pérdida de familias completas de los genes de deshidrogenasas (*ndh*) fue documentado en el cactus columnar *Carnegiea gigantea* (Sanderson et al. 2015), en tanto que el proceso de la pseudogenización de las familias de genes *ndh* y *ycf*, fue documentada en los cactus globosos pequeños del género *Mammillaria*, pero no la pérdida de los IRs (Solórzano et al. 2019). Sin embargo, también se ha documentado la pérdida de IRs en diversos linajes de angiospermas tales como Leguminosae (Palmer y Thompson, 1982), Geraniaceae (Ruhlman et al. 2017) y Arecaceae (Choi et al. 2019) cuya pérdida se ha interpretado como resultado de un solo evento de mutación ancestral que fue heredado a todos los descendientes. Estos resultados sugieren que el estudio de las secuencias de ADN y de la organización estructural del genoma de cloroplasto (entiéndase como identidad, número, arreglo y orientación de los genes) tienen un uso potencial en los análisis filogenéticos de los taxa (Jansen et al. 2007).

En particular, el orden Caryophyllales, cuenta con cerca de 11,555 especies agrupadas en 749 géneros y 40 familias (APG, 2016). Los miembros de este orden se distribuyen en todos los continentes, ocupando ambientes terrestres y acuáticos. Algunas de las familias de Caryophyllales se destacan por sus adaptaciones a las condiciones ambientales variables, tales como la sequía/calor (Cactaceae), el frío extremo (Montiaceae) o la alta salinidad (Amaranthaceae) (Hernández-Ledesma et al. 2015). En otras familias, hay especies que se destacan por su grado comercial y ornamental (Caryophyllaceae, Droseraceae, Portulacaceae, Polygonaceae) (López-Nieves et al. 2018).

De acuerdo con el análisis filogenético basado en secuencias del gen *matK* y el intrón *trnK*, el orden Caryophyllales se divide en dos clados principales: las Caryophyllales centrales y las Caryophyllales no centrales (Meimberg et al. 2000 y Meimberg et al. 2001). Las Caryophyllales centrales agrupan 19 familias que se caracterizan por tener nodos uniloculares, tallos a menudo con anillos céntricos de xilema y floema, tubos de tamiz de floema con plastidios con un anillo periférico de filamentos proteicos y un cristal proteico central. Además, todos los miembros de este grupo poseen betalaínas y perdieron el intrón en el gen *rpl2*; características que los unifica. La mayoría de las familias de este grupo presentan morfologías múltiples de su polen, lo que indica que la estructura de éste es variable en este clado (Eckardt, 1976 y Norwicke, 1975). En cambio, las Caryophyllales no centrales corresponden a aquellos miembros anteriormente colocados en el orden Polygonales que describieron Judd et al. (2002) y que han sido recientemente identificadas correctamente a través de las filogenias moleculares basadas en las secuencias de cloroplasto del *rbcL*, *atpB* y *matK*, así como en las secuencias de ADN ribosomal nuclear 18S (Cuénoud et al. 2002).

A partir de la comparación del genoma completo de cloroplasto de 19 familias de Caryophyllales se documentaron rearrreglos de genes (cambios en la ubicación relativa) significativos en las familias Amaranthaceae, Caryophyllaceae, Chenopodiaceae, Droseraceae y Cactaceae (Ali, 2019). Actualmente, se encuentran disponibles 423 secuencias de genoma completo de cloroplasto que representan un total de 33 familias de Caryophyllales. En esta tesis se eligieron 70 de estas secuencias debido a que, al inicio de la investigación, eran las únicas que estaban completas y verificadas en el GenBank.

Por otra parte, los estudios filogenéticos tradicionales en plantas han utilizado un número de una a 15 secuencias de ADNcp como marcadores moleculares para investigar relaciones filogenéticas. Entre ellos destacan secuencias de genes codificantes (e.g. *atpB*, *matK*, *ndhF*, *psbB*, *rbcL*, *rpoC2*, *rps4*, *rpl16* y *rpoC1*) y no codificantes (i.e. intrones y espaciadores intergénicos) (Patwardhan et al. 2014). Sin embargo, es frecuente que este bajo número de secuencias de ADN no resuelvan las relaciones filogenéticas (Crawley y Hilu, 2012) entre las especies de Caryophyllales. Por ello se recomienda incluir otros marcadores, tales como los genes codificantes de las unidades ribosomales (ARNr) ubicadas en el núcleo como son la 16S, 5S y 28S, o de la mitocondria como subunidades del citocromo oxidasa (CO), la subunidad 12S

del ARNr o el citocromo b para tratar de resolver dichas relaciones filogenéticas. Por otro lado, las secuencias conservadas en los IRs muestran que Caryophyllales es un grupo monofilético (Cuénoud et al. 2002 y Brockington, 2009), aunque dentro del mismo se han identificado aparentes relaciones parafiléticas, como es el caso de los miembros del género *Mammillaria* (Cactaceae) (Butterworth y Wallace, 2004), algunos géneros de la familia Molluginaceae (Fior et al. 2006; Harbaugh et al. 2010) y ciertas especies del género *Salicornia* (Amaranthaceae) (Steffen et al. 2015).

### **3. Objetivos**

#### **Objetivo General**

Analizar la variación estructural del genoma completo de cloroplasto de 70 especies del orden Caryophyllales con la finalidad de identificar cambios estructurales y asignarlos como posibles eventos evolutivos en una filogenia de estas especies.

#### **Objetivos Particulares**

Comparar los arreglos estructurales de los genomas completos de cloroplasto de 70 especies del orden Caryophyllales para identificar si las diferencias encontradas tienen una posible interpretación evolutiva.

Obtener las relaciones filogenéticas entre las 70 especies del orden Caryophyllales con la finalidad de comprobar si la variación estructural en el genoma de cloroplasto tiene información filogenética relevante.

## 4. Métodos

### Recopilación y curado de las secuencias *in silico*

A partir de los genomas publicados en artículos y aquellos directamente disponibles en la base de datos digitales del GenBank (<https://www.ncbi.nlm.nih.gov/genome/>), se obtuvieron 70 ADNcp disponibles de miembros del orden Caryophyllales. Estas especies representan a 16 familias y a 40 géneros distintos. Del total de estas familias, las que suman más genomas completos de cloroplasto son Amaranthaceae (14 especies), Caryophyllaceae (18), Cactaceae (11) y Polygonaceae (7).

Las secuencias de ADNcp de las 70 especies se obtuvieron en formato FASTA para sistematizar la información. Para cada una de las 70 especies estudiadas se documentaron las características de tamaño total del ADNcp (bp), el número total de genes y la ganancia y pérdida de genes en las regiones de IRs, SSC y LSC. También se documentaron las figuras de los genomas ensamblados y anotados previamente publicados para construir una base digital. Del total de los 70 ADNcp localizados, 53 de ellos fueron publicados como genomas ensamblados, pero no fueron anotados; por ello en esta tesis se obtuvo su anotación. La anotación se hizo en la plataforma CHLOROBX de libre acceso y que cuenta con la herramienta GeSeq (<http://chlorobox.mpimp-golm.mpg.de/geseq.html>). En esta plataforma de anotación, el GeSeq usa los algoritmos basados en los programas ARAGORN y tRNAscan-SE (Laslett y Canback, 2004), los cuales sirvieron para la anotación de los ARNt, las regiones codificantes y los intrones, y por convención las secuencias de ADN localizadas entre dos secuencias anotadas se llaman espaciadores intergénicos (IGs) pero no se anotan. La detección de los bloques de genes que se encuentran repetidos en dos copias, pero en orientación invertida (IRs) que se encuentran típicamente en los ADNcp, fueron automáticamente detectadas por el anotador. En esta misma plataforma se generó un archivo de salida de GeSeq necesario para su visualización con OGDRAW (Tillich et al. 2017). La filogenia de Caryophyllales (Yao et al. 2019) se usó como referencia para detectar los cambios en las estructuras de los genomas de cloroplasto y si éstos tienen una tendencia evolutiva.



### **Alineamiento estructural.**

Para visualizar las diferencias y similitudes estructurales en el total de 70 genomas ensamblados, se usó el programa MAUVE 2.4.0 (Darling et al. 2009). En este programa se utilizó la herramienta progressiveMauve para obtener el alineamiento estructural para los miembros de una familia y del total de las especies del orden. El resultado que arroja este alineamiento permite visualizar los rearrreglos de genes de todo el genoma.

### **Matriz de caracteres**

Los 70 genomas de cloroplasto ya se encontraban ensamblados y sirvieron como soporte para la elaboración de una matriz de caracteres en el programa Mesquite (<https://www.mesquiteproject.org/home.html>). Dicho software es modular y extensible para la biología evolutiva pues su énfasis está en el análisis filogenético, lo cual permite el manejo y procesamiento de una gran cantidad de datos. Así, la matriz contó con 70 especies y 106 caracteres que corresponden a cada gen (o secuencias anotadas) del ADNcp de estas especies. La matriz fue codificada del siguiente modo: 1) carácter de presencia de un cierto gen: (símbolo X) en el ADNcp, este carácter con dos estados de carácter: ausente (0), presente (1). 2) carácter de ubicación (localización) relativa del cierto gen X en el ADNcp, con los siguientes estados de carácter: localizado en IR (1), localizado en SSC (2) y localizado en LSC (3). En el caso de las dos especies sin estructura cuadripartita, donde no se le puede asignar una región estructural a sus genes, éstos últimos se consideraron como caracteres binarios tomando en cuenta su ausencia (0) o presencia (1) en el genoma. Una vez que la matriz se completó, se elaboró una filogenia para observar de manera gráfica la relación entre las especies con base en su composición genómica.

## **5. Resultados**

El análisis de los 70 genomas completos (Tabla 1) mostró que 68 de éstos tienen una estructura cuadripartita; las dos especies que carecen de IRs corresponden a la familia Cactaceae y ambos son cactus columnares (*Carnegiea gigantea* y *Lophocereus schottii*). De las 68 especies con estructura cuadripartita, 64 de ellas mostraron sus dos IRs y las regiones únicas, LSC y SSC, con composición y número de genes idénticos (Tabla 2), las otras cuatro especies, todas ellas de la familia Droseraceae, presentan un aparente rearrreglo de genes (Tabla3).

Tabla 1. Total de las 70 especies estudiadas del orden Caryophyllales. Para cada especie se incluye su familia, su registro en la base digital de datos *GenBank* (NCBI), el tamaño total del genoma de cloroplasto en pares de bases (bp), el número total de genes para el genoma completo, para la región única grande (LSC), para la región única pequeña (SSC) y para los inversos repetidos (IRs). Para esta última región se duplica el número total de genes (2X). Se señalan con asterisco (\*) aquellas especies que carecen de IRs.

| Familia          | Género                     | Especie                | Registro en NCBI | Tamaño (bp) | # total de genes | LSC | SSC | IR (2X) |
|------------------|----------------------------|------------------------|------------------|-------------|------------------|-----|-----|---------|
| 1. Aizoaceae     | 1. <i>Mesembryanthemum</i> | <i>crystallinum</i>    | NC_029049        | 153,722     | 132              | 85  | 13  | 34      |
|                  | 2. <i>Tetragonia</i>       | <i>tetragonioides</i>  | NC_036991        | 149,506     | 129              | 82  | 13  | 34      |
| 2. Amaranthaceae | 1. <i>Alternanthera</i>    | <i>philoxeroides</i>   | NC_042798        | 152,255     | 128              | 82  | 12  | 34      |
|                  | 2. <i>Amaranthus</i>       | <i>caudatus</i>        | NC_040143        | 150,523     | 130              | 83  | 13  | 34      |
|                  |                            | <i>hypochondriacus</i> | NC_030770        | 150,518     | 113              | 74  | 11  | 28      |
| 3. Cactaceae     | 1. <i>Carnegiea</i>        | <i>gigantea</i>        | NC_027618        | 113,064     | 127              | *   | *   | *       |
|                  | 2. <i>Lophocereus</i>      | <i>schottii</i>        | NC_041727        | 113,204     | 109              | *   | *   | *       |
|                  | 3. <i>Mammillaria</i>      | <i>albiflora</i>       | NC_517610        | 110,789     | 114              | 87  | 26  | 4       |
|                  |                            | <i>crucigera</i>       | NC_517613        | 115,505     | 119              | 75  | 24  | 20      |
|                  |                            | <i>huitzilopochtli</i> | NC_517612        | 115,886     | 119              | 75  | 24  | 20      |
|                  |                            | <i>pectinifera</i>     | NC_519716        | 108,756     | 109              | 87  | 26  | 4       |
|                  |                            | <i>solisioides</i>     | NC_518341        | 115,356     | 117              | 75  | 24  | 20      |

|                 |                           |                              |           |         |     |    |    |    |
|-----------------|---------------------------|------------------------------|-----------|---------|-----|----|----|----|
|                 |                           | <i>supertexta</i>            | NC_508963 | 116,175 | 117 | 75 | 24 | 20 |
|                 |                           | <i>zephyranthoides</i>       | NC_517611 | 107,343 | 131 | 82 | 11 | 38 |
|                 | 4. <i>Opuntia</i>         | <i>quimilo</i>               | NC_114084 | 150,374 | 131 | 92 | 5  | 34 |
|                 | 5. <i>Rhipsalis</i>       | <i>baccifera</i>             | NC_821847 | 122,333 | 110 | 85 | 19 | 6  |
| 4.              |                           |                              |           |         |     |    |    |    |
| Caryophyllaceae | 1. <i>Agrostemma</i>      | <i>githago</i>               | NC_023357 | 151,733 | 128 | 83 | 13 | 32 |
|                 | 2. <i>Colobanthus</i>     | <i>apetalus</i> <sup>G</sup> | NC_036424 | 151,228 | 128 | 82 | 12 | 34 |
|                 |                           | <i>quitensis</i>             | NC_028080 | 151,276 | 125 | 85 | 12 | 28 |
|                 | 3. <i>Dianthus</i>        | <i>caryophyllus</i>          | NC_039650 | 147,604 | 125 | 84 | 13 | 28 |
|                 | 4. <i>Dysphania</i>       | <i>ambrosioides</i>          | NC_041201 | 151,689 | 127 | 82 | 13 | 32 |
|                 |                           | <i>botrys</i>                | NC_042166 | 152,022 | 127 | 84 | 13 | 30 |
|                 |                           | <i>pumilio</i>               | NC_041159 | 151,962 | 127 | 82 | 13 | 32 |
|                 | 5. <i>Gymnocarpos</i>     | <i>przewalskii</i>           | NC_036812 | 150,636 | 128 | 85 | 13 | 30 |
|                 | 6. <i>Haloxylon</i>       | <i>ammodendron</i>           | NC_027668 | 151,572 | 135 | 84 | 13 | 38 |
|                 |                           | <i>persicum</i>              | NC_027669 | 151,586 | 135 | 84 | 13 | 38 |
|                 | 7. <i>Lychnis</i>         | <i>wilfordii</i>             | NC_035225 | 152,320 | 128 | 82 | 12 | 34 |
|                 | 8. <i>Pseudostellaria</i> | <i>heterophylla</i>          | NC_044183 | 149,765 | 126 | 83 | 13 | 30 |
|                 |                           | <i>longipedicellata</i>      | NC_039454 | 149,626 | 128 | 85 | 13 | 30 |
|                 |                           | <i>okamotoi</i>              | NC_039974 | 149,653 | 126 | 83 | 13 | 30 |
|                 |                           | <i>palibiniana</i>           | NC_041166 | 149,668 | 126 | 83 | 13 | 30 |

|                |                      |                           |           |         |     |    |    |    |
|----------------|----------------------|---------------------------|-----------|---------|-----|----|----|----|
|                | 9. <i>Salicornia</i> | <i>brachiata</i>          | NC_027224 | 153,324 | 131 | 84 | 13 | 34 |
|                |                      | <i>bigelovii</i>          | NC_027226 | 153,076 | 132 | 85 | 13 | 34 |
|                |                      | <i>europaeae</i>          | NC_027225 | 153,232 | 129 | 82 | 13 | 34 |
|                | 10. <i>Silene</i>    | <i>capitata</i>           | NC_035226 | 150,224 | 130 | 83 | 13 | 34 |
|                |                      | <i>chalcedonica</i>       | NC_023359 | 148,081 | 125 | 84 | 13 | 28 |
|                |                      | <i>conica</i>             | NC_016729 | 147,208 | 126 | 84 | 12 | 30 |
|                |                      | <i>conoidea</i>           | NC_023358 | 147,896 | 125 | 83 | 12 | 30 |
|                |                      | <i>latifolia</i>          | NC_016730 | 151,736 | 126 | 83 | 13 | 30 |
|                |                      | <i>noctiflora</i>         | NC_016728 | 151,639 | 127 | 84 | 11 | 32 |
|                |                      | <i>paradoxa</i>           | NC_023360 | 151,632 | 126 | 83 | 13 | 30 |
|                |                      | <i>vulgaris</i>           | NC_016727 | 151,583 | 124 | 81 | 13 | 30 |
|                | 11. <i>Suaeda</i>    | <i>japonica</i>           | NC_042675 | 152,109 | 131 | 84 | 13 | 34 |
|                |                      | <i>malacosperma</i>       | NC_139180 | 151,989 | 131 | 84 | 13 | 34 |
| 5. Droseraceae | 1. <i>Aldovandra</i> | <i>vesiculosa</i>         | NC_035416 | 141,568 | 128 | 70 | 6  | 52 |
|                | 2. <i>Dionaea</i>    | <i>muscipula</i>          | NC_035147 | 117,589 | 105 | 52 | 49 | 4  |
|                | 3. <i>Drosera</i>    | <i>regia</i>              | NC_035415 | 136,810 | 117 | 79 | 6  | 32 |
|                |                      | <i>rotundifolia</i>       | NC_029770 | 192,912 | 143 | 54 | 21 | 68 |
| 6. Montiaceae  | 1. <i>Cistanthe</i>  | <i>longiscapa</i>         | NC_035410 | 156,830 | 133 | 96 | 13 | 24 |
|                | 2. <i>Nepenthes</i>  | <i>ventricosa x alata</i> | NC_044185 | 156,637 | 146 | 89 | 13 | 44 |

|                         |                        |                      |           |         |     |    |    |    |
|-------------------------|------------------------|----------------------|-----------|---------|-----|----|----|----|
| 7.Nyctaginaceae         | 1. <i>Acleisanthes</i> | <i>obtusa</i>        | NC_041416 | 154,004 | 133 | 84 | 13 | 36 |
|                         | 2. <i>Nyctaginia</i>   | <i>capitata</i>      | NC_041415 | 152,995 | 133 | 84 | 13 | 36 |
| 8.Petiveraceae          | 1. <i>Monococcus</i>   | <i>echinophorus</i>  | NC_041414 | 155,099 | 133 | 84 | 13 | 36 |
|                         | 2. <i>Petiveria</i>    | <i>alliaceae</i>     | NC_041417 | 154,947 | 133 | 84 | 13 | 36 |
|                         | 3. <i>Segueira</i>     | <i>aculeata</i>      | NC_041418 | 155,849 | 133 | 84 | 13 | 36 |
| 9.Phytolaccaceae        | 1. <i>Phytolacca</i>   | <i>insularis</i>     | NC_141113 | 156,419 | 132 | 84 | 12 | 36 |
| 10.Plumbaginaceae       | 1. <i>Plumbago</i>     | <i>auriculata</i>    | NC_041245 | 168,765 | 134 | 84 | 12 | 38 |
| 11.Portulacaceae        | 1. <i>Portulaca</i>    | <i>oleracea</i>      | NC_063236 | 156,533 | 132 | 83 | 13 | 36 |
| 12.Polygonaceae         | 1. <i>Fagopyrum</i>    | <i>dibotrys</i>      | NC_037705 | 159,320 | 130 | 84 | 12 | 34 |
|                         |                        | <i>esculentum</i>    | NC_010776 | 159,599 | 132 | 84 | 12 | 36 |
|                         |                        | <i>luojishanense</i> | NC_037706 | 159,265 | 132 | 84 | 12 | 36 |
|                         |                        | <i>tataricum</i>     | NC_027161 | 159,272 | 131 | 84 | 12 | 36 |
|                         | 2. <i>Rheum</i>        | <i>palmatum</i>      | NC_027728 | 161,541 | 134 | 86 | 12 | 36 |
|                         | 3. <i>Rumex</i>        | <i>acetosa</i>       | NC_042390 | 160,269 | 132 | 84 | 12 | 36 |
|                         | 4. <i>Oxyria</i>       | <i>sinensis</i>      | NC_032031 | 160,404 | 132 | 84 | 12 | 36 |
| 13.Stegnosperma<br>ceae | 1. <i>Stegnosperma</i> | <i>halimifolium</i>  | NC_041235 | 155,485 | 133 | 84 | 13 | 36 |
| 14.Talinaceae           | 1. <i>Talinum</i>      | <i>paniculatum</i>   | NC_037748 | 159,929 | 133 | 84 | 13 | 36 |

Los resultados mostraron que los IRs de las 64 especies conservadas mantienen siete ARNt (*trnA-UGC*, *trnL-CAA*, *trnL-GAU*, *trnM-CAU*, *trnN-GUU*, *trnR-ACG* y *trnV-GAC*) y cuatro ARNr (*rrn4.5*, *rrn5*, *rrn16* y *rrn23*). La comparación de estos 64 genomas mostró que en todas ellas la SSC posee los genes asociados a la Nicotinamida-Adenina-Dinucleótido (NADH) mientras que su composición genómica cuenta con un número de genes codificantes que van desde los 89, para el caso de *Dianthus caryophyllus*, hasta 113 para *Amaranthus caudatus*. En cuanto a los ARNt se reporta un rango que oscila entre los 30 en *Colobanthus apetalus* hasta 36 en *Fagopyrum esculentum*.

Por otro lado, el total de especies que son 70 cuentan con cuatro genes de la ARN polimerasa (*rpoA*, *rpoB*, *rpoC1* y *rpoC2*) y el gen *rbcL* que codifica para la subunidad grande de la ribulosa 1-5, bifosfato (RuBisCO) y otros genes que corresponden a las subunidades respectivas del complejo citocromo b6-f; y a la mayoría de las subunidades de deshidrogenasas (NADH), que usualmente están en la SSC.

Tabla 2. Identidad génica de las regiones de IRs, SSC y LSC de las 64 especies que mantienen su composición y número de genes conservados.

| Región                      | Identidad del gen   | Gen   |
|-----------------------------|---|---|
| IRs                         | Marcos de lectura hipotéticos ( <i>ycf1</i> )   | fragmento de <i>ycf1</i> , <i>ycf2</i> completo             |
|                             | Deshidrogenasas (NADH)  | <i>ndhB</i>   |
|                             | Proteínas de la subunidad grande del ribosoma ( <i>rpl</i> )  | <i>rpl2</i> , <i>rpl23</i>                                  |
|                             | Proteínas de la subunidad pequeña del ribosoma ( <i>rps</i> )   | <i>rps7</i> , <i>rps12</i> , <i>rps19</i>                   |
|                             | ARN ribosomal (ARNr)  | <i>rrn4.5</i> , <i>rrn5</i> , <i>rrn16</i> , <i>rrn23</i> , |
| ARN de transferencia (ARNt) | <i>trnA-UGC</i> , <i>trnL-CAA</i> , <i>trnL-GAU</i> , <i>trnM-CAU</i> , <i>trnN-GUU</i> , <i>trnR-ACG</i> , <i>trnV-GAC</i> |   |

|  |  |  |
|--|--|--|
| SSC  | Marcos de lectura hipotéticos (ycf)                              | fragmento de <i>ycf1</i>   |
|  | Deshidrogenasas (NADH)   | <i>ndhA, ndhB, ndhD, ndhE, ndhF, ndhG, ndhH, ndhI</i>                                      |
|  | Proteínas de la subunidad grande del ribosoma (rpl)              | <i>rpl32</i>   |
|  | Proteínas de la subunidad pequeña del ribosoma (rps)             | <i>rps15</i>   |
|  | Fotosistema I (psa)  | <i>psaC</i>  |
| ARN de transferencia (ARNt)                          | <i>trnI-UAG</i>  |  |
| LSC  | Ribulosa 1,5-bifosfato   | <i>rbcL</i>  |
|  | Marcos de lectura hipotéticos (ycf)                              | <i>ycf3, ycf4</i>  |
|  | Complejo ATP sintasa   | <i>atpA, atpB, atpE, atpF, atpH, atpI</i>  |
|  | Deshidrogenasas (NADH)   | <i>ndhC, ndhJ, ndhK</i>  |
|  | Complejo citocromo b6f (pet)                                     | <i>petA, petB, petD, petG, petL, petN</i>  |
|  | Fotosistema I (psa)  | <i>psaA, psaB, psaI, psaJ</i>  |
|  | Fotosistema II (psb)   | <i>psbA, psbB, psbC, psbD, psbE, psbF, psbI, psbJ, psbK, psbL, psbM, psbT, psb30, psbZ</i> |
|  | ARN polimerasa (rpo)   | <i>rpoA, rpoB, rpoC1, rpoC2</i>  |
|  | Proteínas de la subunidad grande del ribosoma (rpl)              | <i>rpl4, rpl16, rpl20, rpl22, rpl33, rpl36</i>   |
| Proteínas de la subunidad pequeña del ribosoma (rps) | <i>rps2, rps3, rps4, rps8, rps11, rps12, rps14, rps16, rps18</i> |  |

|  |                             |   |
|--|-----------------------------|---|
|  | ARN de transferencia (ARNt) | <i>trnC-GCA, trnD-GUC, trnE-UUC, trnG-GCC, trnF-GAA, trnH-GUG, trnM-CAU, trnP-UGC, trnQ-UUG, trnR-UCU, trnS-CGA, trnS-GCU, trnS-UGA, trnT-GCU, trnW-CGA, trnY-GUA</i> |
|--|-----------------------------|---|

Por su parte, las cuatro especies de la familia Droseraceae difieren en la composición génica de su ADNcp respecto a aquellas presentadas en la Tabla 2 ya que presentan ganancia o pérdida de genes e incluso hasta subunidades completas. En consecuencia, *Aldrovanda vesiculosa*, *Dionaea muscipula*, *Drosera regia* y *Drosera rotundifolia* presentan un importante rearrreglo estructural en su genoma (Tabla 3).

Tabla 3. Identidad génica de las regiones de IRs, SSC y LSC de las cuatro especies de la familia Droseraceae.

| Región | Identidad del gen                                    | Gen                                   |
|--------|--|---------------------------------------|
| IR     | Proteínas de la subunidad grande del ribosoma (rpl)  | <i>rpl14, rpl19, rpl22, rpl36</i>     |
|        | Proteínas de la subunidad pequeña del ribosoma (rps) | <i>rps2, rps3, rps8, rps16, rps23</i> |
|        | ARN de transferencia (ARNt)                          | <i>trnH-GUG</i>                       |
|        | Fotosistema II (psb)                                 | <i>psbA, psbD</i>                     |



|     |  |   |
|-----|--|---|
| SSC | Marcos de lectura hipotéticos (ycf)                  | <i>ycf2, ycf3</i>   |
|     | Proteínas de la subunidad grande del ribosoma (rpl)  | <i>rpl2, rpl20, rpl23, rpl36</i>  |
|     | Proteínas de la subunidad pequeña del ribosoma (rps) | <i>rps3, rps4, rps8, rps11, rps12, rps14, rps16, rps19, rps22</i>           |
|     | Fotosistema I (psa)                                  | <i>psaA, psaB</i>   |
|     | ARN de transferencia (ARNt)                          | <i>trnS-GCA, trnM-CAU, trnD-GUC, trnY-GUA, trnE-UUC, trnN-GUU, trnR-ACG</i> |
|     | Complejo citocromo b6f (pet)                         | <i>petB, petN</i>   |
|     | Fotosistema II (psb)                                 | <i>psbB, psbH, psbM, psbT</i>   |
| LSC | Deshidrogenasas (NADH)                               | <i>ndhB</i>   |
|     | Proteínas de la subunidad grande del ribosoma (rpl)  | <i>rpl32</i>  |
|     | ARN de transferencia (ARNt)                          | <i>trnF-GCG, trnK-UUU, trnL-CAA, trnL-GUU, trnM-GUG</i>                     |

Los resultados de la anotación de las 53 especies que aquí se elaboró, se presentan junto con los genomas de las 17 especies previamente anotados en otros estudios (Anexo 1) al igual que los alineamientos comparativos generados en MAUVE 2.4.0 (Anexo 2).

La filogenia resultante de la matriz de caracteres (Fig. 1) se elaboró tomando en cuenta al orden Ericales como grupo externo ya que también forma parte del grupo de las Astéridas al igual que Caryophyllales. En dicha filogenia, se observan 7 principales agrupamientos : I: *Dysphania botrys*, *D. pumilio*, *Fagopyrum dibotrys*, *F. esculentum*, *F. luojishanense*, *F. tataricum*, *Gymnocarpos przewalskii*, *Haloxylon ammodendron*, *H. persicum*, *Chenopodium album*, *C. quinoa*, *C. ficifolium*, *Cistanthe longiscapa*, *Colobanthus apetalus*, *C. quitensis*, *Alternanthera philoxeroides*, *Acleisanthes obtusa*, *Agrostemma githago*, *Dianthus caryophyllus*, *Amaranthus caudatus* y *A. hypochondriacus* (genes *ycf1* y *ycf2* en SSC y LSC respectivamente), II: *Talinum paniculatum*, *Tetragonia tetragonioides*, *Stegnosperma halimifolium*, *Suaeda japonica*, *S. malacosperma*, *Seguieria aculeata*, *Silene capitata*, *S. paradoxa*, *S. vulgaris*, *S. chalcedonica*, *S. conica*, *S. conoidea*, *S. latifolia* y *S. noctiflora* (pérdida de los genes *infA* y *rpoC1*), III: *Portulaca oleraceae*, *Pseudostellaria heterophylla*, *P. longipedicellata*, *P. okamotoi*, *P. palibiniana*, *Rheum palmatum*, *R. acetosa*, *Salicornia bigelovii*, *S. brachiata* y *S. europeae* (pérdida de tres tRNAs en la LSC y dos en los IRs), IV: *Oxyria sinensis*, *Petiveria alliaceae*, *Phytolacca insularis*, *Plumbago auriculata*, *Mesembryanthemum crystallinum* y *Monococcus echinophorus* (grupo externo), V: *Opuntia quimilo*, *Arbutus unedo*, *Rhododendron datianginense* y *Vaccinum bracteatum* (grupo externo), VI: *Mammillaria zephyranthoides*, *Dionaea muscipala*, *Aldovandra vesiculosa*, *Drosera regia*, *D. rotundifolia* y *Nepenthes ventricosa* (gen *rpl32* ausente), VII: *Rhipsalis baccifera*, *Mammillaria albiflora*, *M. pectinifera*, *Carnegia gigantea*, *Lophocereus schottii*, *Lychnis wilfordii*, *Mammillaria crucigera*, *M. huitzilpochtli*, *M. solisioides* y *M. supertexta* (pérdida de cuatro genes *ndh* en la LSC y SSC).

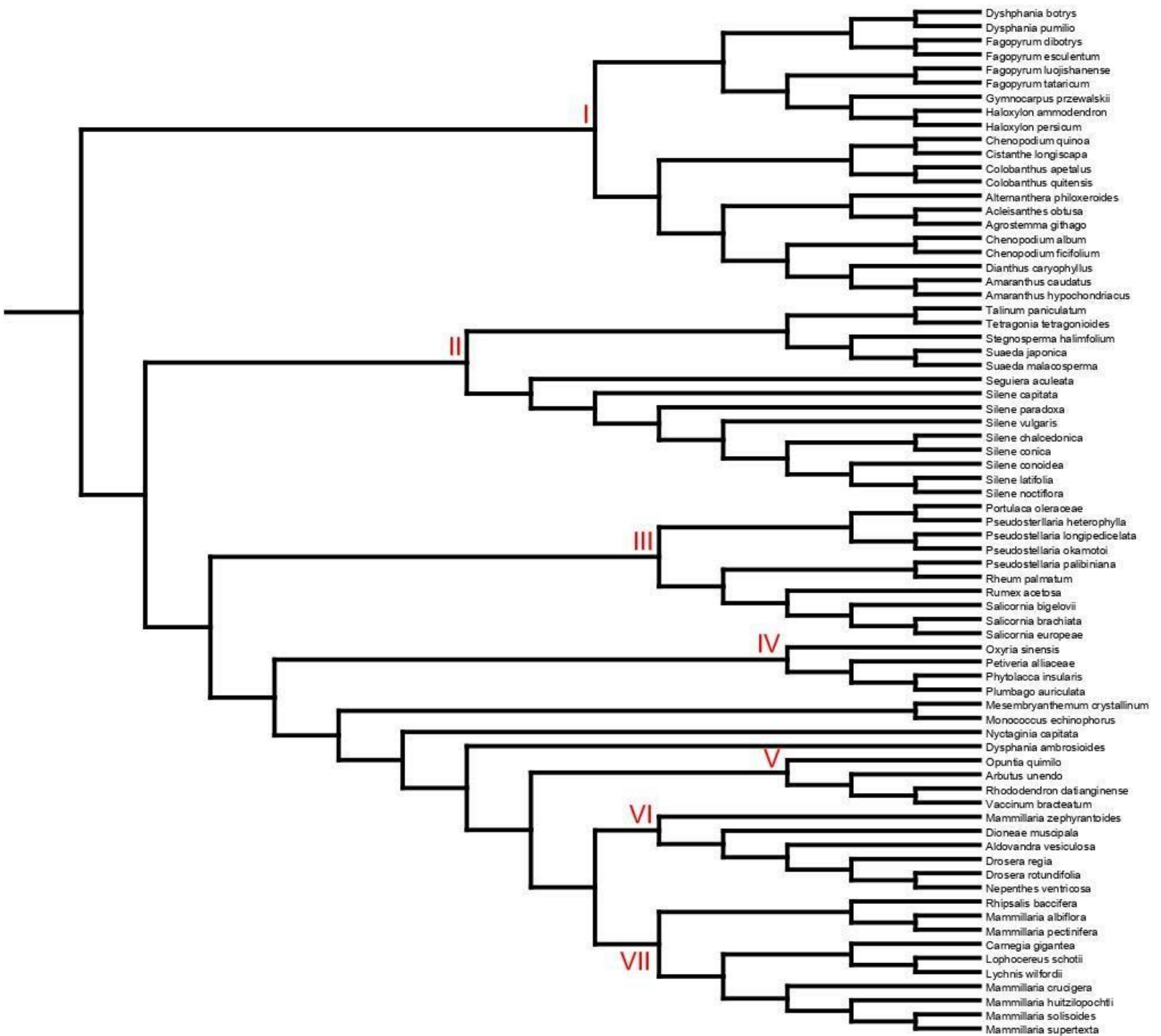


Fig. 1. Filogenia de Caryophyllales generada por medio del análisis SPR rearranger por método de parsimonia. Se identifican 7 principales grupos: I (genes *ycf1* y *ycf2* en SSC y LSC respectivamente), II (pérdida de los genes *infA* y *rpoC1*), III (pérdida de tres tRNAs en la LSC y dos en los IRs), IV (pérdida de genes *accD* y *ndhF*), V (grupo externo), VI (gen *rp132* ausente) y VII (pérdida de 4 genes *ndh* en la LSC y SSC).

De las 68 especies con IR, se encontró que los genes ribosomales *rrn4.5*, *rrn5*, *rrn16* y *rrn23* al igual que los genes de transferencia *trnN-GUU*, *trnL-CAA*, *trnR-ACG* y *trnV-GAC* están frecuentemente insertados en dicha región, aunque pueden encontrarse en la LSC como el caso de las especies *D. muscipala* y *T. tetragonioides* o en la región SSC de las especies *M. albiflora*, *M. crucigera*, *M. huitzilopochtli*, *M. pectinifera*, *M. solisioides*, *M. supertexta* y *R. baccifera*.

## 6. Discusión

A pesar de que en esta tesis sólo se analizaron 70 genomas de las especies del orden Caryophyllales, se encontró que en este grupo existe una tendencia a formar inversos repetidos en su estructura del genoma de cloroplasto. Las únicas especies que carecen de estos inversos repetidos son los cactus de crecimiento columnar *C. gigantea* y *L. schottii*. El origen de Cactaceae se remonta al Eoceno Tardío, hace 30-35 millones de años (Hernández-Hernández et al. 2014, Magallón et al. 2015), época geológica donde se experimentó una caída de CO<sub>2</sub>; condición que originó a este linaje altamente eficiente en el uso de dicho gas. Por otro lado, se propone que la pérdida de los inversos repetidos se llevó a cabo hace cerca de 23 millones de años durante el Mioceno Tardío (Mauseth, 1990, Hershkovitz y Zimmer, 1997 y Hernández-Hernández, 2011) donde las condiciones áridas del ambiente favorecieron la diversificación de Cactaceae debido a la alta prevalencia de hibridación y poliploidización (Machado, 2008 y Majure et al. 2012). A partir de los análisis de los genomas completos de 70 especies del orden Caryophyllales, se indica que quizá la pérdida de los IRs es algo inusual, ya que solo se documentó en dos especies de cactus columnares, *C. gigantea* y *L. schottii*, pertenecientes a la subtribu Pachycereinae cuya divergencia es aún más reciente, hace 4-6 millones de años (Hernández-Hernández et al. 2011, 2014). Es por eso que se necesitan más estudios para aseverar si en las más de 11,000 especies de este orden, los inversos repetidos se perdieron tal y como sucede con las cactáceas columnares. Por otro lado, los análisis estructurales enfocados a identificar los procesos mediante los cuales se llevó a cabo la pérdida de dichas regiones no forman parte de los enfoques principales de esta tesis. Es posible que, en un futuro, con la disponibilidad de más secuencias y con la integración de relacionar la estructura y composición del ADNcp con las funciones a la cual están ligados, se podría entender cabalmente las consecuencias de los rearrreglos estructurales del genoma de cloroplasto. Por otro lado, respecto a la composición de genes, Gantt et al

(1991) propusieron que el contenido génico del ADNcp está altamente conservado en plantas terrestres debido a la transferencia de las funciones del cloroplasto al núcleo. El presente trabajo identificó una composición común de las regiones IRs, SSC y LSC en 64 de las 70 especies analizadas; las seis especies restantes presentan pérdidas o ganancias en sus regiones que van de uno a cinco genes. Lo anterior posiblemente se debe al mantenimiento de cuatro genes ribosomales (*rps4.5*, *rps5*, *rps16* y *rps23*) y siete ARNt (*trnA-UGC*, *trnL-CAA*, *trnI-GAU*, *trnM-CAU*, *trnN-GUU*, *trnR-ACG* y *trnV-GAC*) en los inversos repetidos, mismos a los que Zheng et al (2016) les atribuyeron la estabilidad de la estructura del ADNcp debido a la baja tasa de sustitución de nucleótidos. Con relación al número de genes codificantes y ARNt, las SSC y LSC en 64 especies (Tabla 2) están altamente conservadas de acuerdo con los valores reportados en los trabajos de Li et al. (2018) y Mower y Vickrey (2018).

La comparación de los resultados obtenidos en este trabajo con respecto a la filogenia publicada por Brockington et al (2009) que establece una separación entre las Caryophyllales no centrales y las centrales, apoya la idea de que las regiones de IRs y las regiones únicas son variables para el primer grupo y son conservadas para el segundo. En la filogenia de Brockington et al (2009), las llamadas Caryophyllales no centrales forman un grupo monofilético con dos subclados: uno con Plumbaginaceae y Polygonaceae como parientes de Frankeniaceae y Tamaricaceae. El segundo contiene a los taxa carnívoros y sus parientes cercanos. Los resultados de esta tesis muestran que la variación estructural del ADNcp de las cuatro especies de la familia Droseraceae es concordante con la separación de este segundo clado con el primero. En cuanto a las Caryophyllales centrales, éstas forman un grupo fuertemente soportado por las familias Amaranthaceae, Cactaceae, Caryophyllaceae y Chenopodiaceae. En la filogenia resultante, las cactáceas globosas (*Mammillarias*) tienen relación con las especies de las familias anteriormente mencionadas debido a la pérdida de algunos genes *ndh* en sus regiones LSC y SSC. En contraste, no se muestra una concordancia para Cactaceae únicamente, lo cual puede sugerir procesos evolutivos internos dentro de esta familia que deberán ser analizados en estudios futuros. Esto se puede deber a que en esta familia se registra la ausencia de IR, pero solo a las especies de crecimiento columnar (*C. gigantea* y *L. schottii*) (Sanderson et al. 2015) pero no para el resto de las especies analizadas.

Respecto al tamaño total del ADNcp, el genoma más pequeño de las especies de Caryophyllales analizadas fue el de *C. gigantea* (113 kbp) y el más grande el de *D.*

*rotundifolia* (192 kbp). En particular, la reducción del ADNcp de *C. gigantea* conlleva la pérdida de las regiones IR que involucran nueve de los 11 genes *ndh* fundamentales para el complejo tilacoide que actúa como válvula de alimentación para ajustar el nivel redox de los transportadores de electrones fotosintéticos (Martin y Sabater, 2010). Sin embargo, en esta especie se mantienen dos subunidades de *ndh* (*ndhB* y *ndhD*), las cuales parecen haber perdido su funcionalidad (Sanderson et al. 2015). Una posible explicación a la pérdida de estas regiones IR, es que sea un subproducto de la racionalización total del genoma de los plastidios (Wu et al. 2009), lo que explicaría la fuerte asociación entre la pérdida de los genes *ndh* y los IRs. Sin embargo, para establecer las causas específicas de la pérdida de cualquier componente del ADNcp, es necesario realizar estudios bioquímicos y moleculares para combinar los resultados generados en este estudio.

Los resultados muestran por tanto que el alargamiento de los IRs ocurrió en cuatro especies a expensas de genes que están típicamente en la LSC y en dos especies en la región SSC. Por otro lado, los grandes cambios en la composición de genes no se registran en la región LSC ya que 68 especies de Caryophyllales mantienen 80 genes de los que suelen encontrarse en angiospermas.

Por tanto, de acuerdo con las estructuras identificadas en este estudio, se concluye que la pérdida de IRs para la familia Cactaceae representa un paso evolutivo que no es disperso o bien frecuente para el resto de los miembros del orden Caryophyllales.

## **7. Conclusiones**

Los estudios de genómica comparada del genoma completo de cloroplasto permitió identificar tendencias en la organización estructural de este genoma en el orden Caryophyllales. A partir de los resultados, se concluye que en Caryophyllales la tendencia de la organización estructural del genoma de cloroplasto es en cuatro secciones: LSC, SSC y dos IRs. La pérdida de los IRs es una novedad evolutiva identificada solo en cactus columnares (Cactaceae) pertenecientes a la subtribu Pachycereinae, lo cual posiblemente se originó hace 4-6 millones de años.

Es posible que el ambiente sea el promotor de cambios estructurales en el genoma de cloroplasto, ya que los rearrreglos génicos más profundos se identificaron en las especies

de hábito carnívoro (Droseraceae) y en los cactus columnares (Cactaceae) cuya especiación se atribuye a la aridificación global ocurrida en el Mioceno tardío hace aproximadamente 10 millones de años. Sin embargo, los resultados de esta tesis no permiten apoyar esta idea, la que deberá de ser analizada en estudios futuros.

Aunque faltan más estudios con mayor inclusión taxonómica, la filogenia elaborada en esta tesis con datos estructurales recuperó los grandes agrupamientos filogenéticos obtenidos previamente en otros estudios con Caryophyllales con secuencias de ADN de cloroplasto y de otros genomas, lo que indica que la estructura del genoma de cloroplasto si revela información filogenética.

## 8. Referencias

Abbate, A. y Campbell, J. (2013). Parasitic Beechdrops (*Epifagus virginiana*): A possible ant-pollinated plant. *Southeastern Naturalist*, 12(3):661-665.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. y Walter, P. (2002). Chloroplast and Photosynthesis (en) *Molecular Biology of the Cell*. 4ta Ed. Garland Science, New York. 250-342.

Ali, M. (2019). Comparative chloroplast genomic analysis revealed extensive genomic rearrangements in some core and non-core Caryophyllales. *Plant Taxon*, 26(1):107-116.

Angiosperm Phylogeny Group. (2016). An Update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants. *Botanical Journal of the Linneal Society*, 181(1):399-436.

Asaff, S., Khan, AL., Khan, MA., Shahzad, R., Lubna y Kang, SM. (2018). Complete chloroplast genome sequence and comparative analysis of loblolly pine (*Pinus taeda* L.) with related species. *Public Library of Science ONE*, 13(3):1-29.

Bendich, A. (2004). Circular chloroplast chromosomes: the grand illusion. *The Plant Cell*, 16(7):1661-1666.

Beverley, R. (2011). Chloroplast genomes of photosynthetic eukaryotes. *The Plant Journal*, 66(1):34-44.

Blazier, J., Ruhlman, T., Weng, M., Rehman, S., Sabir, J. y Jansen, R. (2016). Divergence of RNA polymerase alpha subunits in angiosperms plastid genomes is mediated by genomic rearrangement. *Nature*, 6(1):1-15.

Bobik, K. y Burch-Smith, TM. (2015). Chloroplast signaling within, between and beyond cells. *Frontiers in Plant Science*, 6(1):781.

Brandvain, Y. y Wade, M. (2009). The Functional transfer of genes to the mitochondria. The Effects of Selection, Mutation, Population Size and Rate-Self Fertilization. *Genetics*, 182(4):1129-1139.

Brockington, S., Alexandre, R., Ramdial, J., Moore, M., Crawley, S. y Dzinga, A. (2009). Phylogeny of the Caryophyllales sensu lato: revisiting hypotheses on pollination biology and perianth differentiation in the core Caryophyllales. *International Journal of Plant Science*, 170(5): 627-643.

Brummitt, R. y Powell, C. (1992). Authors of Plant Names. *Royal Botanic Gardens*, 21(1):131-145.

Choi, I., Jansen, R. y Ruhlman, T. (2019). Lost and found: return of the inverted repeat in the legume clade defined by its absence. *Genome Biology and Evolution*, 11(4):1321-1333.

Clegg, M. T. (1993). Chloroplast gene sequences and the study of plant evolution. *Proceedings of the National Academy of Sciences*, 90(2):363–367.

Cuenoud, P., Savolainen, V., Chatrou, L., Powell, M., Grayer, R. y Chase, M. (2002). Molecular Phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK sequences. *American Journal of Botany*, 89(1): 132-144.

Crawley, S. y Hilu, K. (2012). Impact of missing data, gene choice and taxon sampling on phylogenetic reconstruction: the Caryophyllales (angiosperms). *Plant Systematics and Evolution*, 298(2):297-312.

Darling, A., Mau, B. y Perna, N. (2009). Progressive Mauve: Multiple alignment of genomes with gene flux and rearrangement. *Public Library of Science ONE*, 5(6):e11147.



De las Rivas, J., Lozano, J. y Ortiz, A. (2012). Comparative Analysis of Chloroplast Genomes: Functional Annotation, Genome-Based Phylogeny, and Deduced Evolutionary Patterns. *Genome Research*, 12(4):567-583.

Douglas, A.E. y Raven, J.A. (2003). Genomes at the interface between bacteria and organelles. *Philosophical Transactions of the Royal Society: Biological Sciences*, 358(1429):5-18.

Downie, S. y Palmer, J. (1994). A chloroplast DNA phylogeny of the Caryophyllales based on structural and inverted repeats restriction site variation. *Systematic Botany*, 2(19):135-252.

Eckardt, T. (1976). Classical morphological features of centrosperous families. *Plant Systematics and Evolution*, 1(126): 5-25.

Farjon, A. (2001). World Checklist and Bibliography of Conifers. *Royal Botanic Gardens*, 1(287):1-76.

Gantt, J. S., Baldauf, S. L., Calie, P. J., Weeden, N. F. & Palmer, J. D. (1991). Transfer of rpl22 to the nucleus greatly preceded its loss from the chloroplast and involved the gain of an intron. *The European Molecular Biology Organization Journal*, 10(10):3073–3078.

Hernández-Hernández, T., Brown, J., Schlumpberger, B., Eguiarte, L. y Magallón, S. (2014). Beyond aridification: multiple explanations for the elevated diversification of cacti in the New World Succulent Biome. *New Phytologist*, 1(202):1382-1397.

Hernández-Ledesma, P., Berendsohn, W., Borsch, T., Mering, S., Akhiani, H., Arias, S., Castañeda-Noa, I., Eggli, U., Eriksson, R., Flores-Olvera, H., Fuentes-Bazán, S., Kadereit, G., Klak, C., Korotkova, N., Nyffeler, R., Ocampo, G., Ochoterena, H., Oxelman, B., Sánchez, R., Schlumpberger, B. y Uotila, P. (2015). A taxonomic backbone for the global synthesis of species diversity in the angiosperm order Caryophyllales. *Willdenowia*, 45(3): 281-383.

Hershkovitz, M. y Zimmer, E. (1997). On the evolutionary origins of the cacti. *Taxon*, 46(2): 217-232.

- Jansen, R.K., Cai, Z., Raubeson, L.A., Daniell, H., Leebans, J. y Muller, K.F. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperm and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences*, 104(49): 19369-19374.
- Judd, W. S., Campbell, C. S., Kellogg, E. A. and P. F. Stevens. (2002). *Plant systematics: a phylogenetic approach*. Sinauer Associates, Inc., Sunderland, Massachusetts.
- Johnson, M. (2016). Photosynthesis. *Essays in Biochemistry*, 60(3):255-273.
- Kaneko, T. y Tabata, S. (1997). Complete genome structure of the unicellular cyanobacterium *Synechocystis* sp. PCC6803. *Plant Cell Physiology*, 38(11):1171-1176.
- Kenrick, P. y Crane, P. (1997). The origin and early evolution of plants on Land. *Nature*, 389(6646):33-39.
- Kleine, T., Maier, U. y Leister, D. (2009). DNA transfer from organelles to the nucleus: the idiosyncratic genetics of endosymbiosis. *Annual Review of Plant Biology*, 60(1):115-138.
- Kohler, M., Reginato, M., Teixeira de Souza, T. y Majure, L. (2020). Insights into chloroplast genome evolution across Opuntioideae (Cactaceae) reveals robust yet sometimes conflicting phylogenetic topologies. *Frontiers in Plant Science*, 11(1):729.
- Von Kohn, C.; Kielkowska, A.; Havey, M.J. 2013. Sequencing and annotation of the chloroplast DNAs and identification of polymorphisms distinguishing normal male-fertile and male-sterile cytoplasms of onion. *Genome*, 56(12):737-742.
- Laslett, D. y Canback, B. (2004). ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, 32(1):11-16.
- Li, Y., Zhang, J., Li, L., Gao, L., Xu, J. y Yang, M. (2018). Structural and comparative analysis of the complete chloroplast genome of *Pyrus hopeiensis* “wild plants with a tiny population” and three other *Pyrus* species. *International Journal of Molecular Sciences*, 19(10): 32-62.
- Logacheva, M., Schelkunov, M., Shtratnikova, V., Matveeva, M. y Penin, A. (2016). Comparative analysis of the plastid genome of nonphotosynthetic Ericaceae and their photosynthetic relatives. *Scientific Reports*, 6(1):30-42.

Landrum, J. (2002). Four succulent families and 40 million years of evolution and adaptation to xeric environments: what can stem and leaf anatomical characters tell us about their phylogeny?. *Taxon*, 51(3):463-473.

López-Nieves, S., Yang, Y., Timoneda, A., Wang, M., Feng, T., Smith, S., Brockington, S. y Maeda, H. (2018). Relaxation of tyrosine pathway regulation underlines the evolution of betalain pigmentation in Caryophyllales. *New Phytologist*, 217(2):896-908.

Machado, M. (2008). What is the role of hybridization in the evolution of the Cactaceae? *Bradleya*, 43(26):1-18.

Magallón, S., Gómez-Acevedo, S., Sánchez-Reyes, L. y Hernández-Hernández, T. (2015). A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, 3(207):437-453.

Majure, L., Puente, R., Griffith, M., Judd, W., Soltis, P. y Soltis, D. (2012). Phylogeny of *Opuntia* s.s (Cactaceae): clade declination, geographic origins, and reticulate evolution. *American Journal of Botany*, 1(99):847-864.

Margulis, L. (2002). Planeta Simbiótico: Un nuevo punto de vista sobre la evolución. Victoria Laporta Gonzalo (Trad.). Madrid: Editorial Debate.

Martin, M. y Sabater, B. (2010). Plastid *ndh* genes in plant evolution. *Plant Physiology and Biochemistry*, 48(8):636-645.

Mauseth, J. (1990). Continental drift, climate, and the evolution of cacti. *Cactus & Succulent Journal*, 62(6):302-308.

Meimberg, H., Dittrich, P., Bringmann, G., Schlauer, J. y Heubl, G. (2000). Molecular phylogeny of Caryophyllidae s.l. based on *matK* sequences with special emphasis on carnivorous taxa. *Plant Biology*, 2(2):218-228.

Meimberg, H., Witsuba, A., Dittrich, P. y Heubl, G. (2001). Molecular phylogeny of the Nepenthaceae based on cladistic analysis of plastid *trnK* intron sequences. *Plant Biology*, 3(2):164-175.

Morse, A., Peterson, D., Faridi, M., Smith, K., Magbanua, Z., García, S., Kubisiak, T., Amerson, H., Carlson, J., Nelson, C. y Davis, J. (2009). Evolution of Genome Size and Complexity in *Pinus*. *Public Library of Science ONE*, 4(2):e4332.

Mower, J. y Vickrey, T. (2018). Advances in Botanical Research Plastid Genome Evolution; Chaw, S., Jansen, R. (Eds). *Academic Press of Elsevier*, 85(9): 263-292.

Nowicke, J. W. 1975. Pollen morphology in the order Centrospermae. *Grana* 15(1):51-77.

Oldenburg, DJ y Bendich, AJ. (2015). DNA maintenance in plastids and mitochondria of plants. *Frontiers in Plant Science*, 6(1):883.

Ono, K., Hashimoto, H. y Katoh, S. (1995). Changes in the number and size of chloroplasts during sequence of primary leaves of wheat grown under different conditions. *Plant and Cell Physiology*, 36(1): 9-17.

Oulo, M., Yang, J., Dong, X., Wang, V., Mkala, E., Munyao, J., Onjolo, V., Cheptoo, P., Hu, G. y Wang, Q. (2020). Complete chloroplast genome of *Rhipsalis baccifera*, the only cactus with natural distribution in the old World: genome rearrangement, intron gain and loss, and implications for Phylogenetic studies. *Plants*, 9(8):979.

Palmer, J. y Thompson, W. (1982). Chloroplast DNA rearrangements are more frequent when a large inverted repeat sequence is lost. *Cell*, 29(2):537-550.

Patwardhan, A., Ray, S. y Roy, A. (2014). Molecular markers in phylogenetic studies-A Review. *Journal of Phylogenetic & Evolutionary Biology*, 2(2):100-131.

Poppinga, S., Hartmeyer, S. y Masselter, T. (2013). Trap diversity and evolution in the family Droseraceae. *Plant signaling & behavior*, 8(7):e24685.

Raven, J. y Allen, J. (2003). Genomics and chloroplast evolution: what did cyanobacteria do for plants?. *Genome Biology*, 4(1):209-215.

Rochaix, J. (2004). Genetics of the Biogenesis and Dynamics of the Photosynthetic Machinery in Eukaryotes. *The Plant Cell Online*, 16(7):1650-1660.

Ruhlman, T., Zhang, J., Blazier, J., Sabir, J. y Jansen, R. (2017). Recombination-dependant Replication and gene conversion homogenize repeat sequences and diversify plastid genome structure. *American Journal of Botany*, 104(4):559-572.

Sanderson, M., Coppetti, D., Burquez, A., Bustamante, E., Eguiarte, L., Kumar, S., Lee, H., McMahon, M., Steele, K., Wing, R., Yang, T. y Zwicky, D. (2015). Exceptional reduction of the plastid genome of saguaro cactus (*Carnegieia gigantea*): Loss of the *ndh* gene suite and inverted repeat. *American Journal of Botany*, 102(7):1115-1127.

Shestakov, S.V. y Karbysheva, E.A. (2017). The Origin and Evolution of Cyanobacteria. *Biology Bulletin Reviews*, 7(4):259-272.

Smith, D. y Lee, R. (2014). A plastid without a genome: evidence from the nonphotosynthetic green algal genus *Polytomella*. *Plant Physiology*, 164(4):1812-1819.

Solórzano, S., Chincoya, D., Sánchez-Flores, A., Estrada, K., Díaz-Velázquez, C., González-Rodríguez, A., Vaca-Paniagua, F., Dávila, P y Arias, S. (2019). *De novo* assembly discovered novel structures in genome of plastids and revealed divergent inverted repeats in *Mammillaria* (Cactaceae, Caryophyllales). *Plants*, 8(10):392.

Soltis, D., Albert, V., Leebens-Mack, J. y Bell, C. (2008). Polyploidy and angiosperm diversification. *American Journal of Botany*, 96(1):336-348.

Soo, R., Hemp, J., Parks, D., Fischer, W. y Hugenholtz, P. (2017). On the origins of oxygenic photosynthesis and aerobic respiration in Cyanobacteria. *Science*, 355(6332):1436-1440.

Straub, S., Parks, M., Weitemier, K., Fishbein, M., Cronn, R. y Liston, A. (2012). Navigating the tip of the genomic iceberg: next-generation Sequencing for plant Systematics. *American Journal of Botany*, (99):349-364.

Sugiura, M. (1989). The Chloroplast Chromosome in Land Plants. *Annual Review of Cell Biology*, 5(1):51-70.

Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht, E., Fischer, A. y Greiner, S. (2017). GeSeq- versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, 45(1):6-11.

Trojan, A. y Gabrys, H. (1996). Chloroplast distribution in *Arabidopsis thaliana* (L.) depends on light conditions during growth. *Plant Physiology*, 111(2):419-425.

Wardley, T., Bhalla, P. y Dalling, M. (1984). Changes in the number and composition of chloroplasts during senescence of mesophyll cells of attached and detached primary leaves of wheat (*Triticum aestivum* L. ) *Plant Physiology*, 75(2):421-424.

Whittingham, C. (1952). The Chemical Mechanism of Photosynthesis. *Botanical Review*, 18(4): 245-290.

Yao, G., Jin, J., Li, H., Yang, J., Mandala, V., Croley, M., Mostow, R., Douglas, N., Chase, M., Christenhusz, M., Soltis, D., Soltis, P., Smith, P., Brockington, S., Moore, M., Yi, T. y Li, D. (2019). Plastid phylogenomic insights into the evolution of Caryophyllales. *Molecular Phylogenetics and Evolution*, 134(1): 74-86.

Zheng, W., Chen, J. y Hao, Z. (2016). Comparative analysis of the chloroplast genomic information of *Cunninghamia lanceolata* (Lamb.) Hook with sibling species from the genera *Cyrtomeria*. *International Journal of Molecular Sciences*, 17(7):1084.

Zhu, A., Guo, W., Gupta, S., Fan, W. y Mower, J. (2016). Evolutionary dynamics of the plastid inverted repeat: the effects of expansión, contraction and loss of substitution rates. *New Phytologist*, 209(4): 1747-1756.

## 9. Anexos

### Anexo 1. Anotaciones de los 70 genomas completos de cloroplasto.

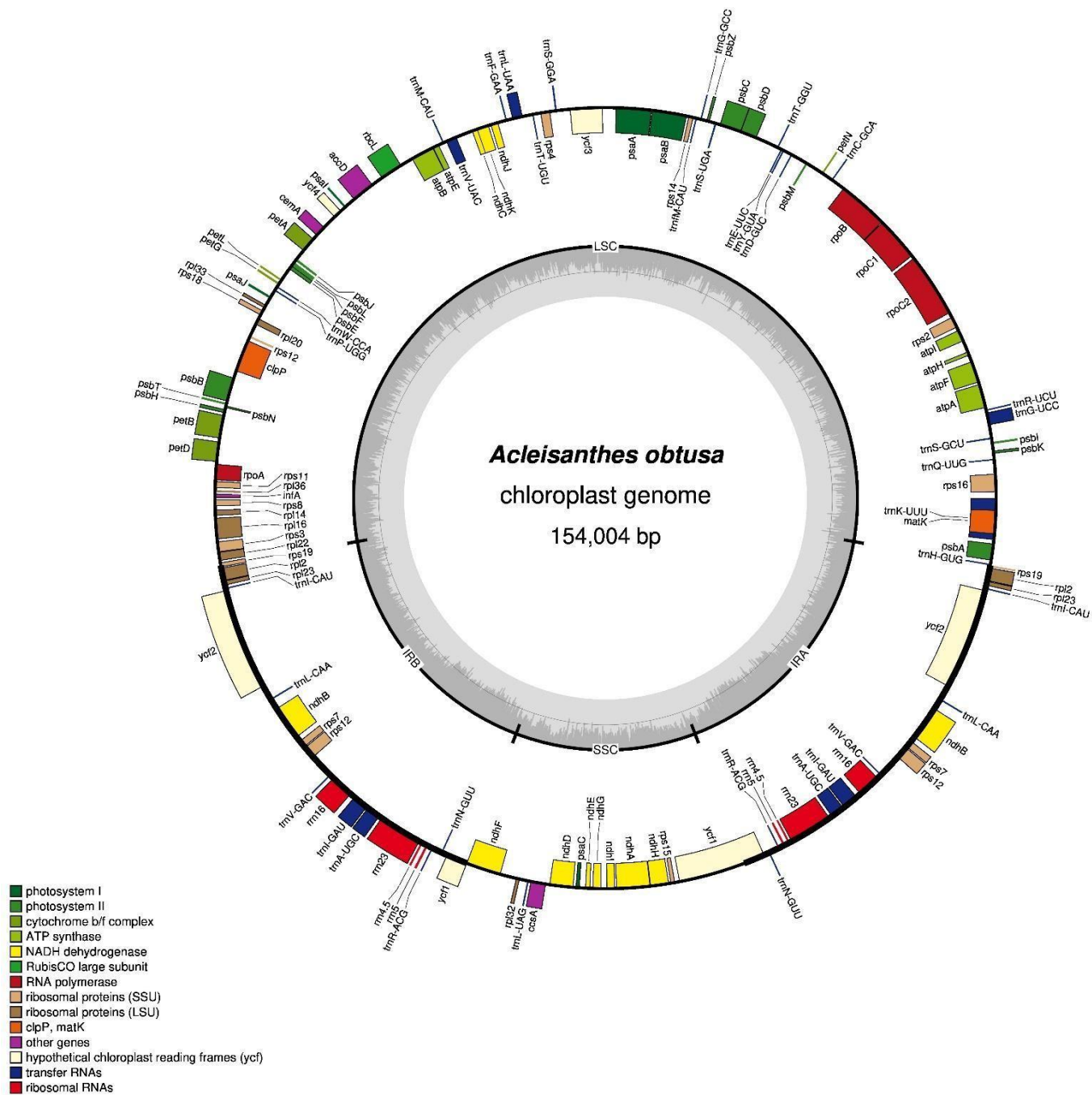


Figura 1. Estructura cuatripartita de *Acleisanthes obtusa* (Generada).

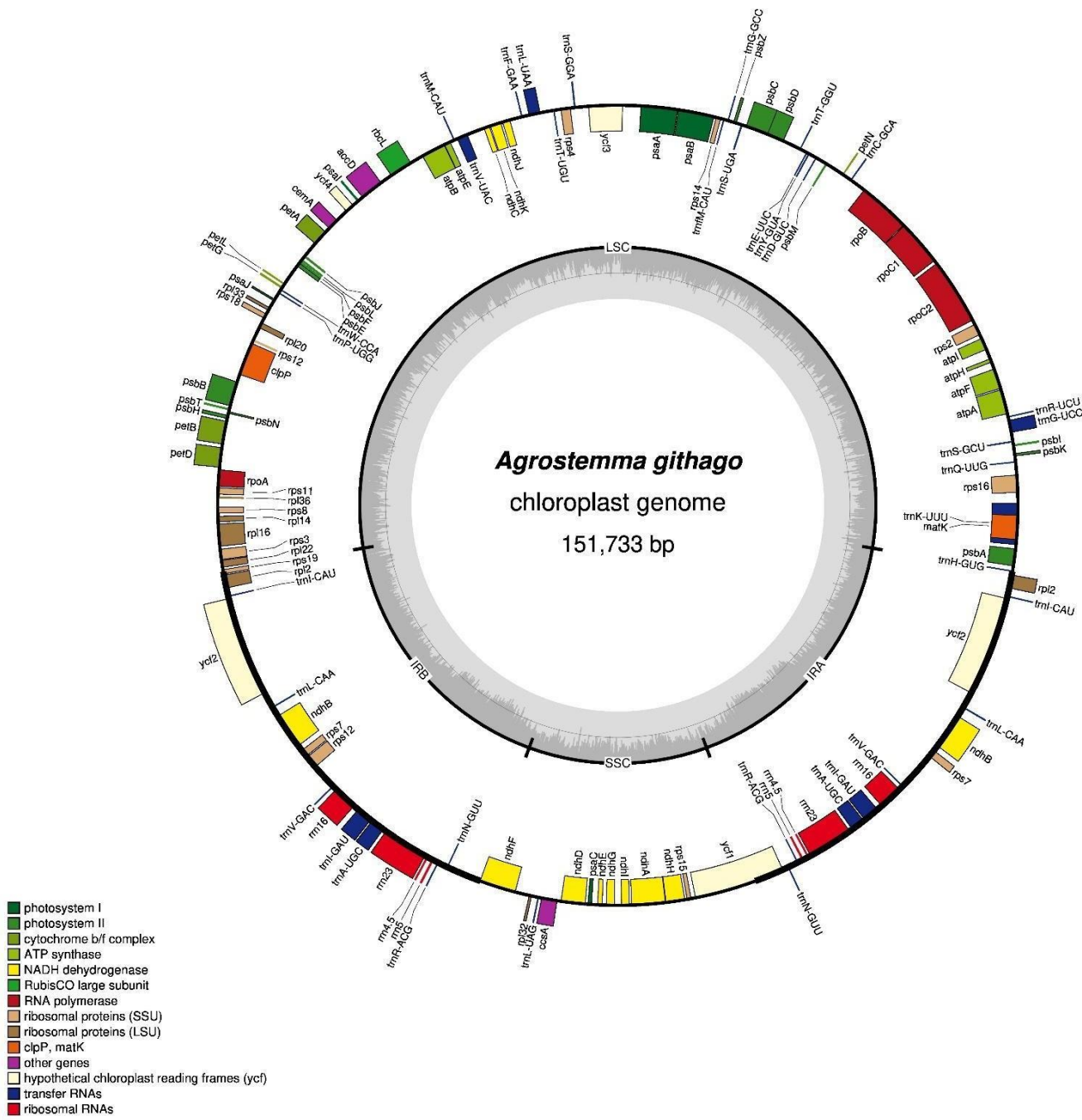


Figura 2. Estructura cuadripartita de *Agrostemma githago* (Generada).



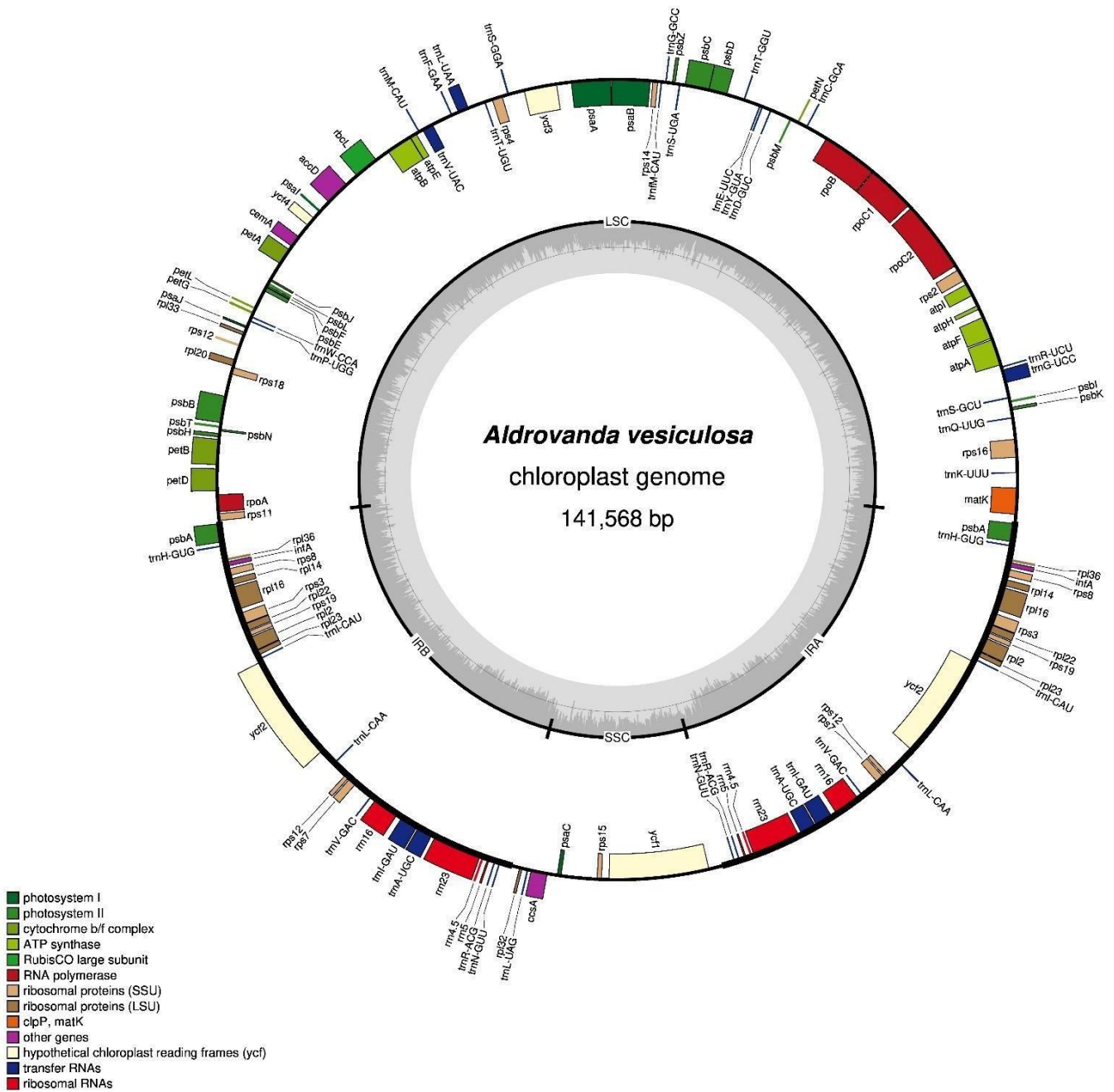


Figura 3. Estructura cuatripartita de *Aldrovandra vesiculosa* (Generada).

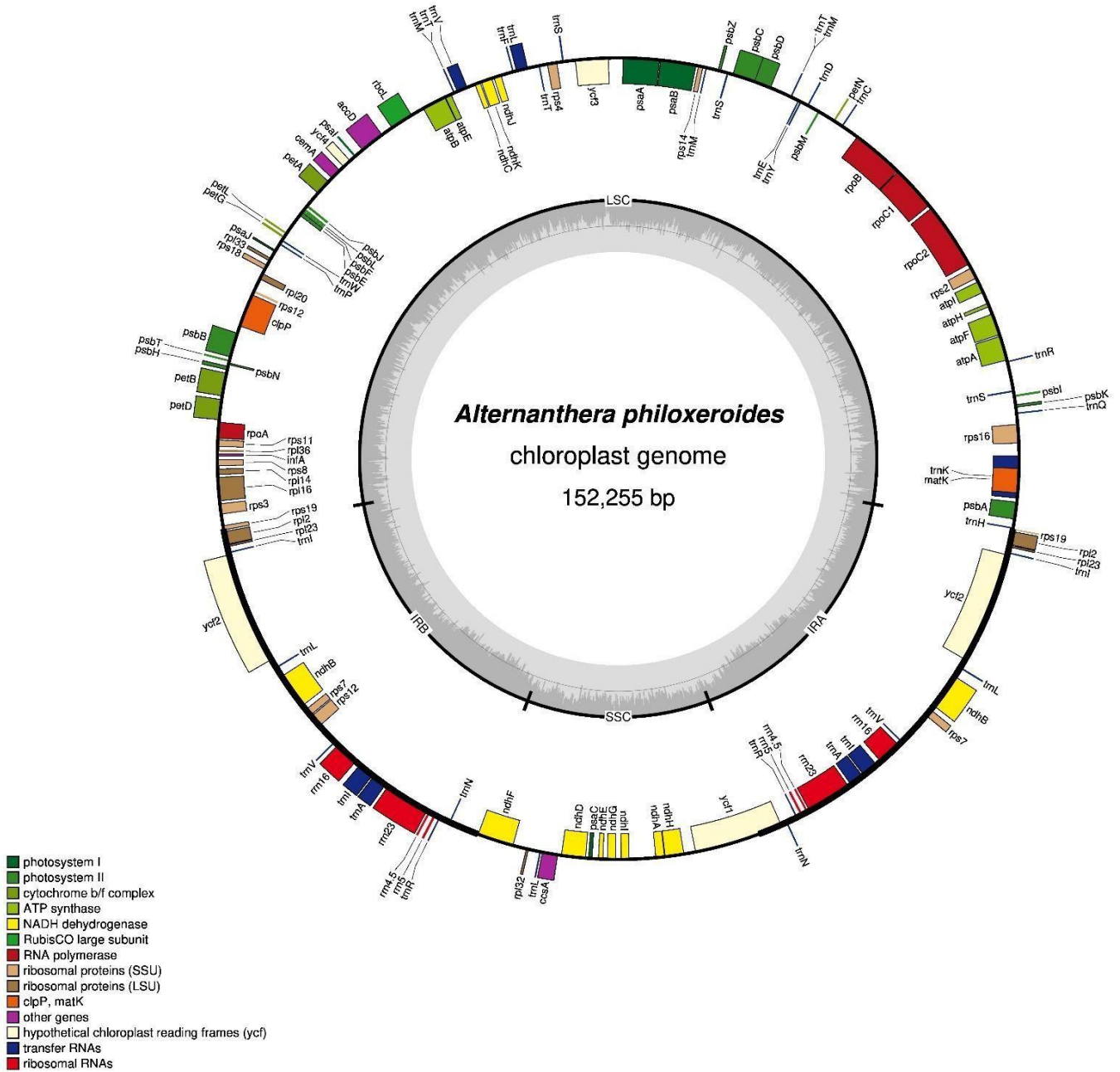


Figura 4. Estructura cuadripartita de *Alternanthera philoxeroides* (Generada).

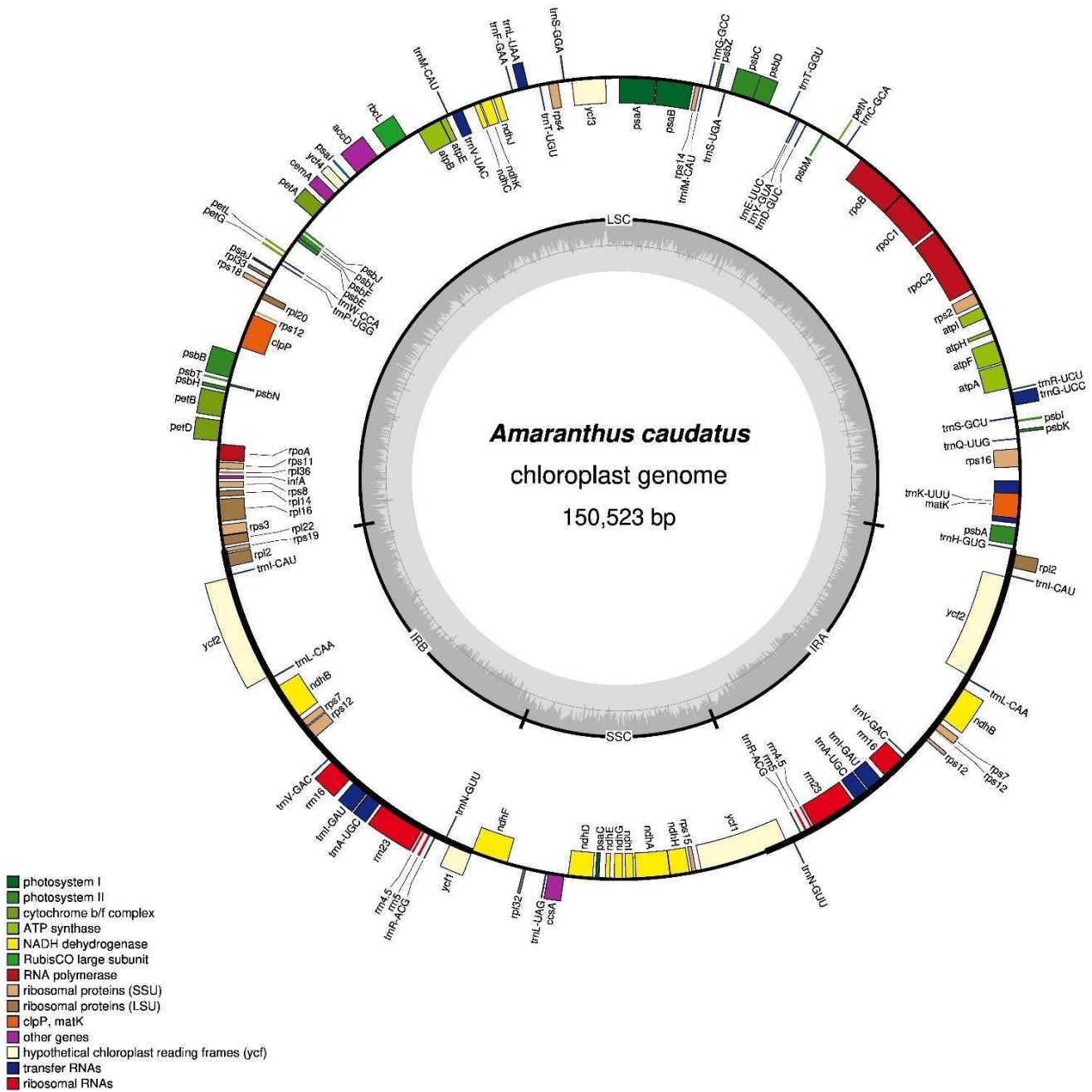


Figura 5. Estructura cuatripartita de *Amaranthus caudatus* (Generada).

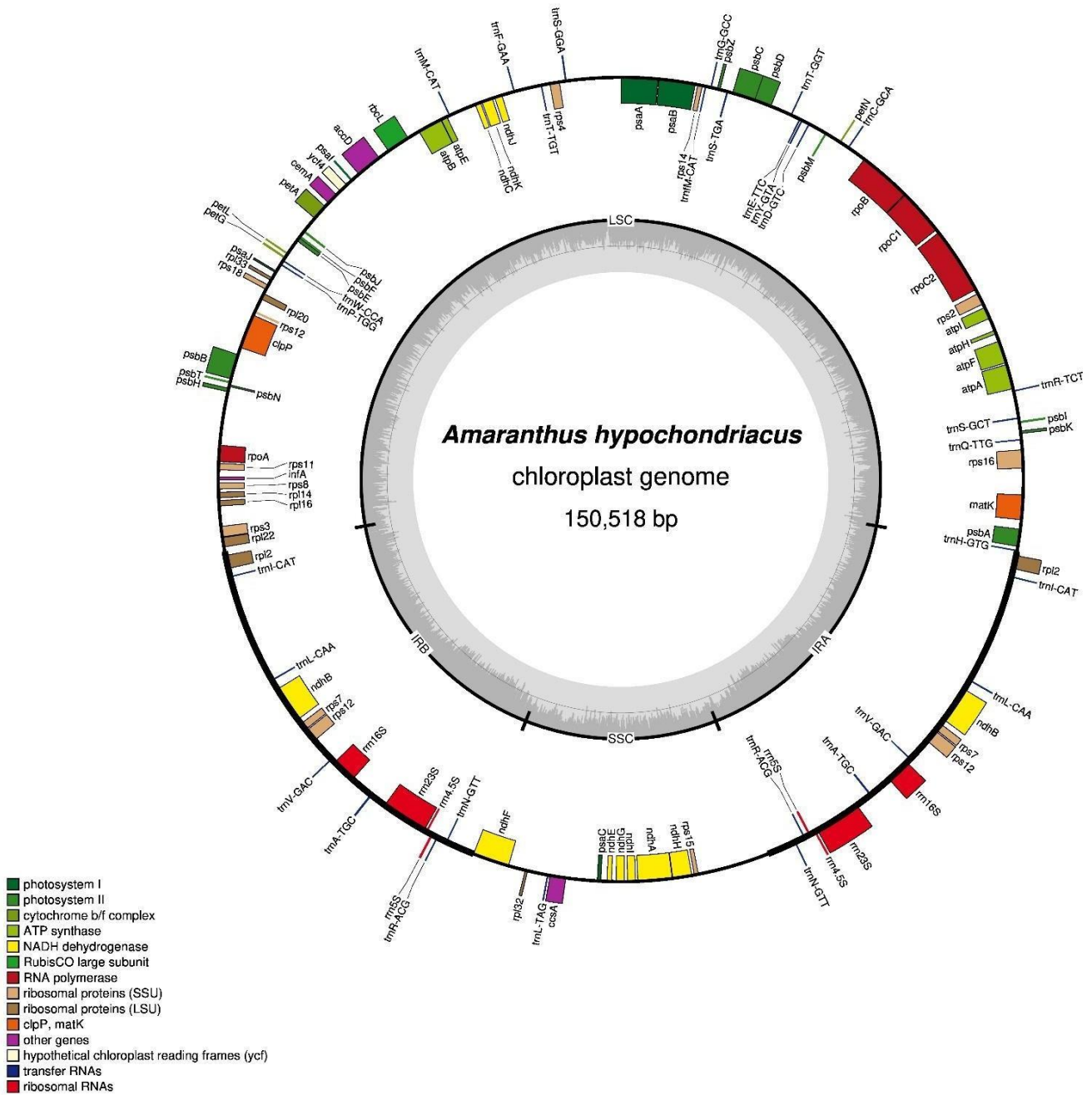


Figura 6. Estructura cuatripartita de *Amaranthus hypochondriacus* (Generada).

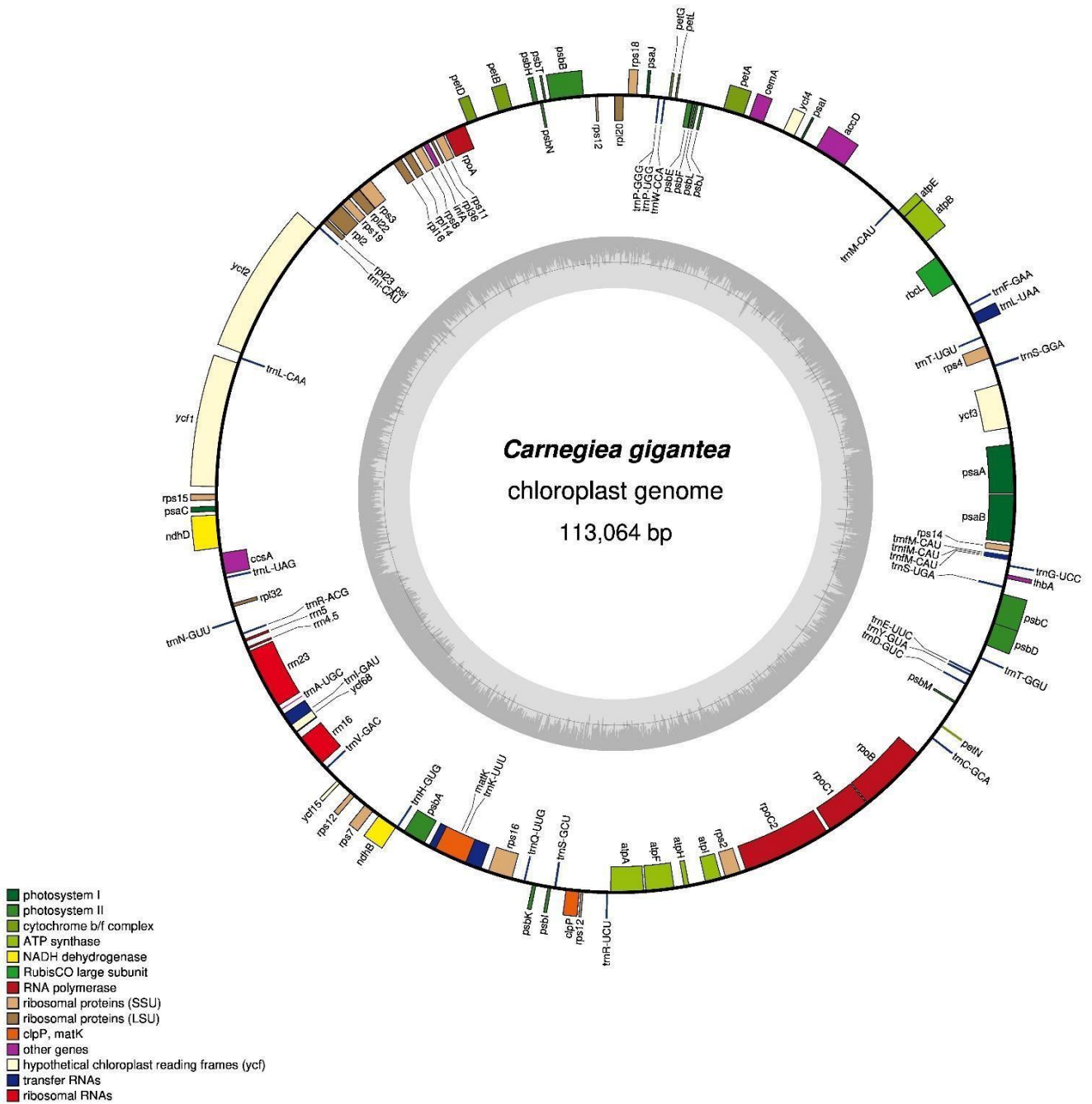


Figura 7. Genoma completo de cloroplasto de *Carnegieia gigantea* (Generada).

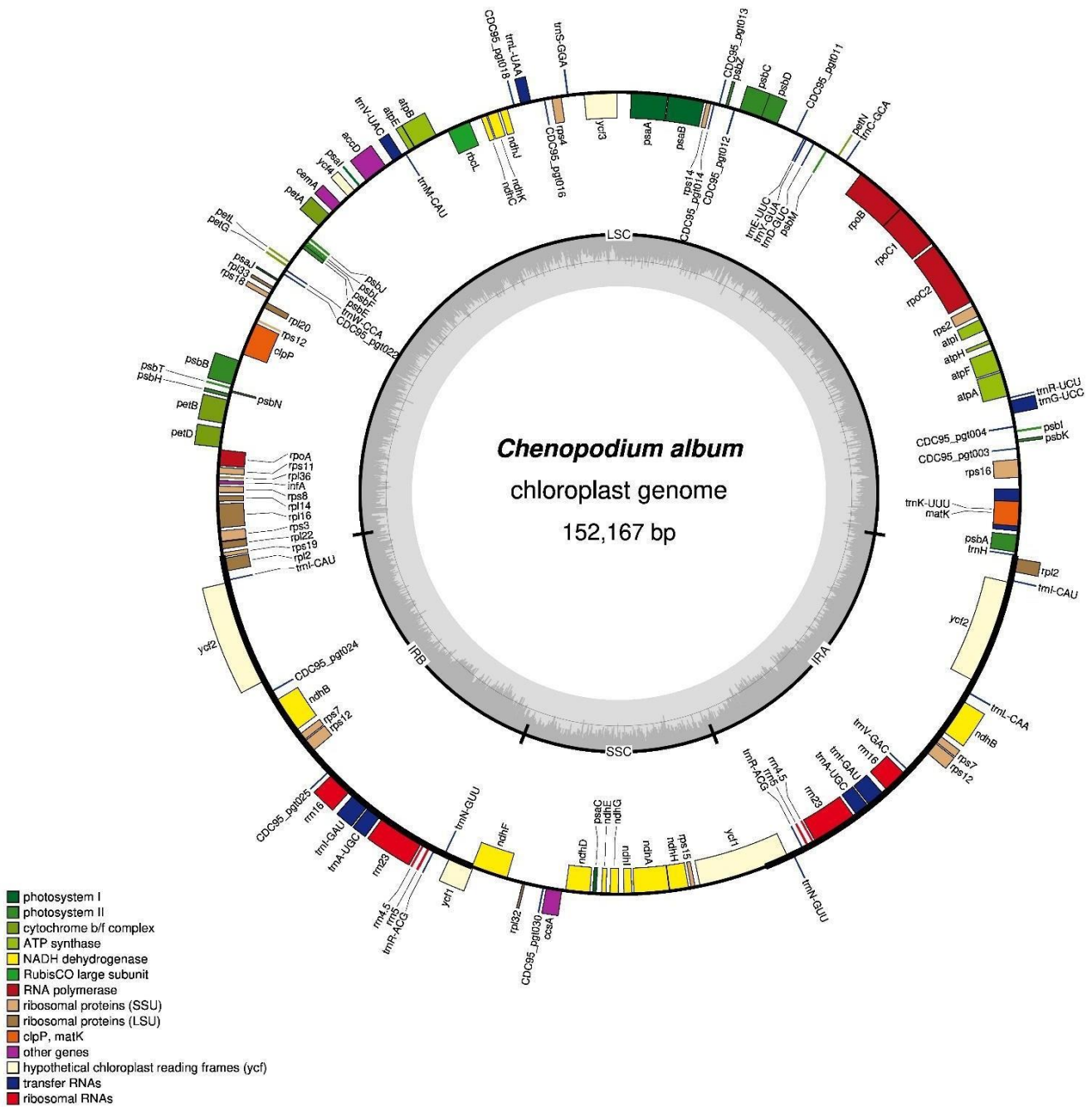


Figura 8. Estructura cuatripartita de *Chenopodium album* (Generada).

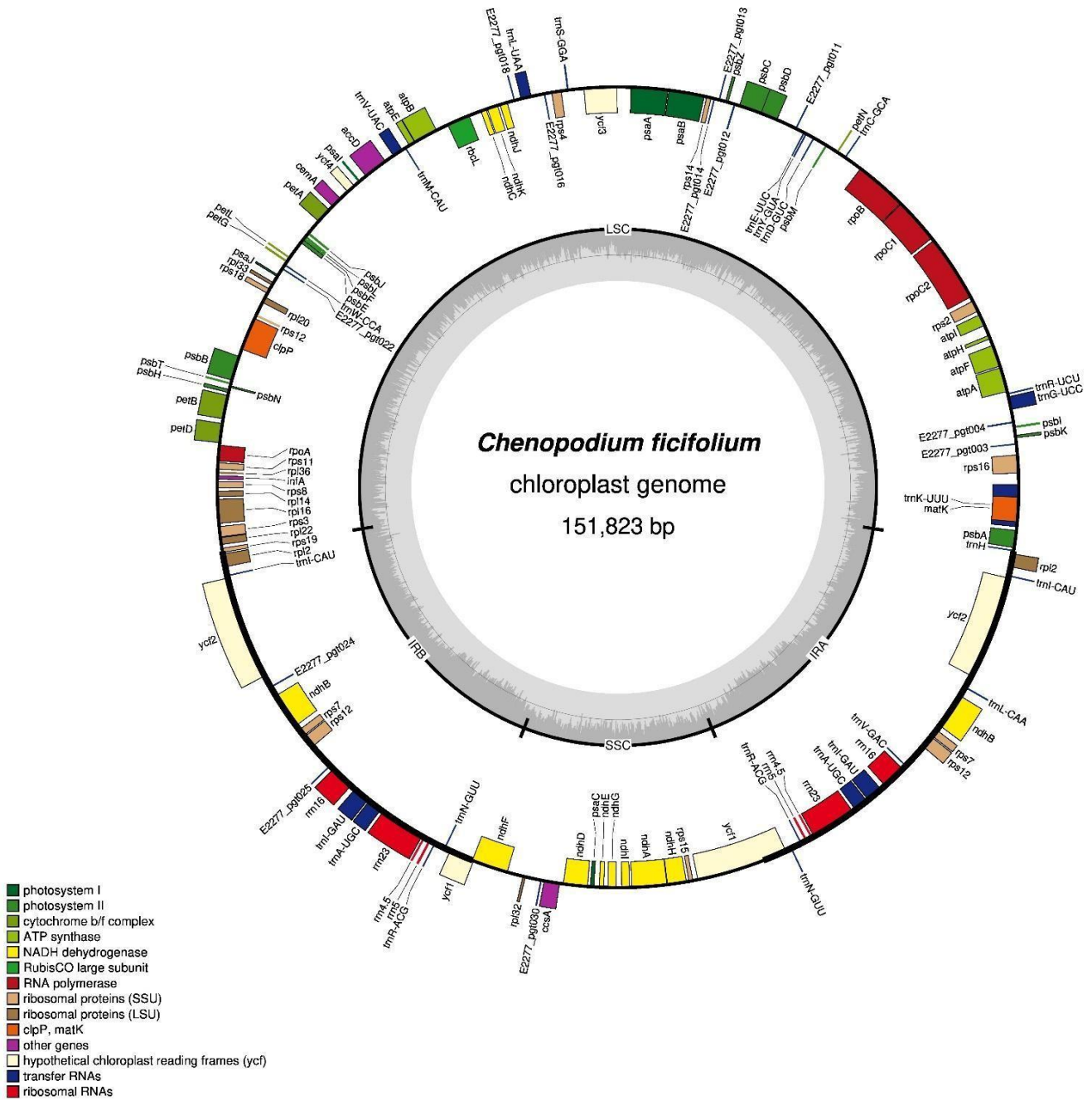


Figura 9. Estructura cuatripartita de *Chenopodium ficifolium* (Generada).

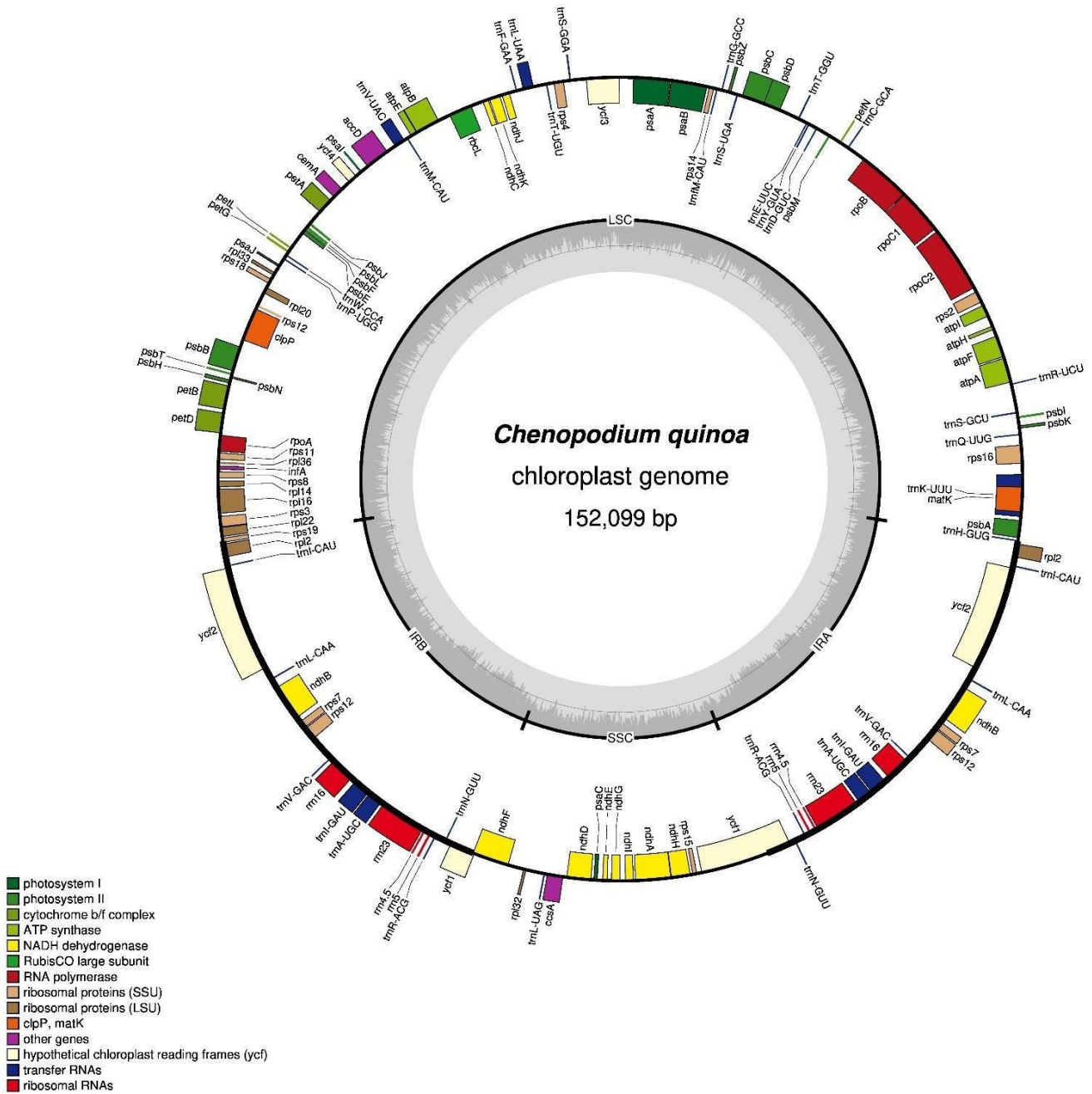


Figura 10. Estructura cuatripartita de *Chenopodium quinoa* (Generada).



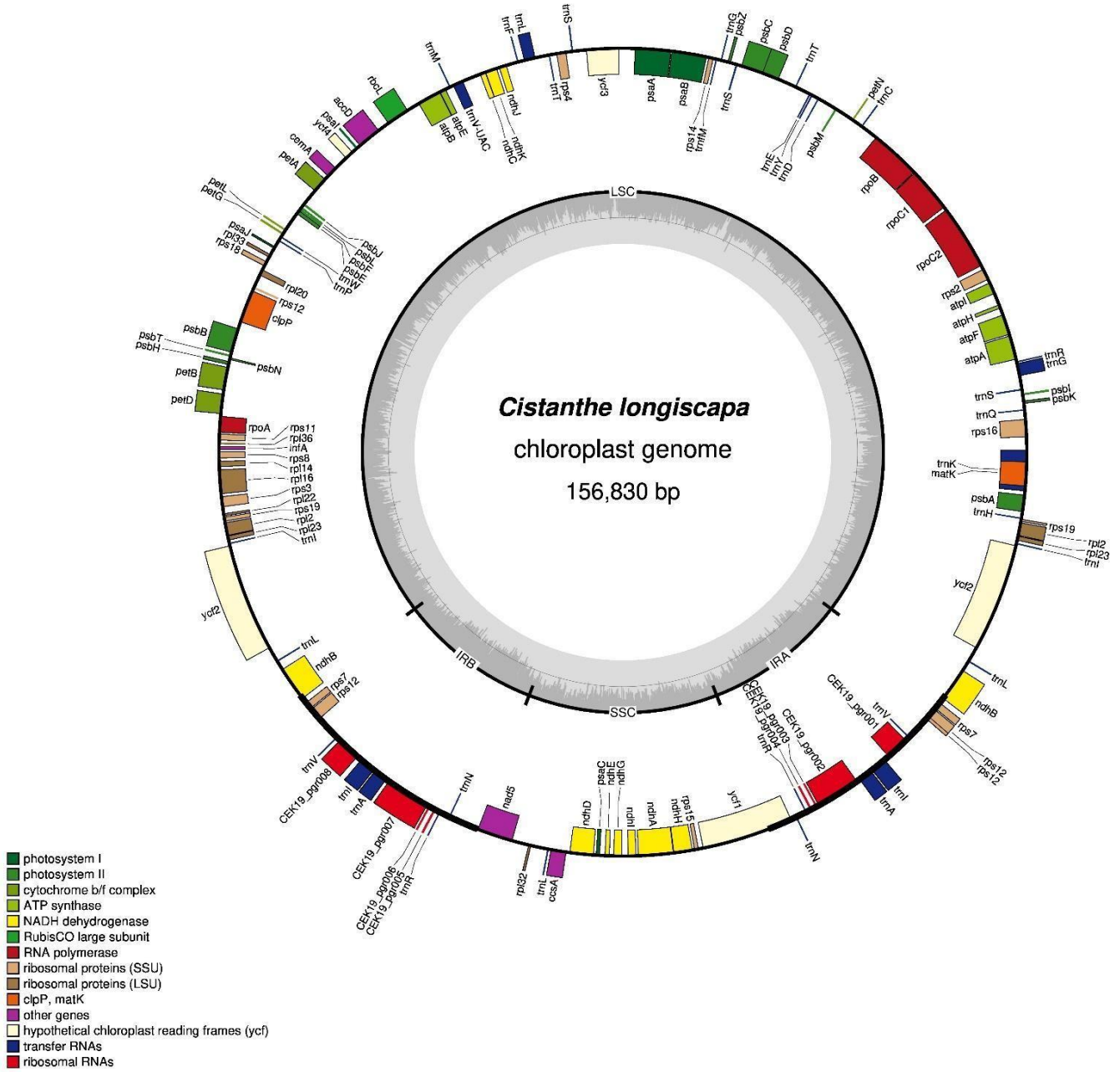


Figura 11. Estructura cuatripartita de *Cistanthe longiscapa* (Generada).

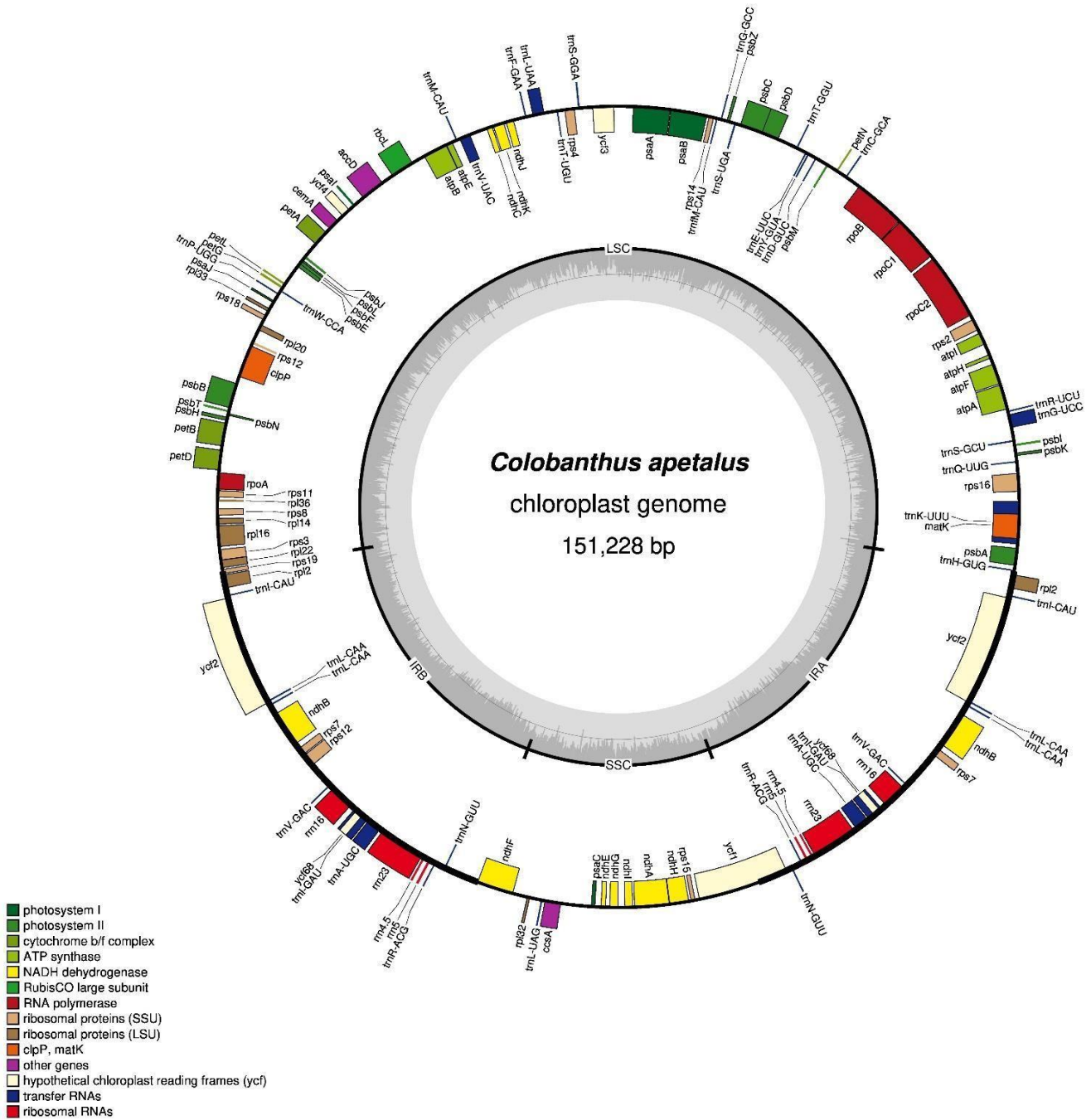


Figura 12. Estructura cuatripartita de *Colobanthus apetalus* (Generada).

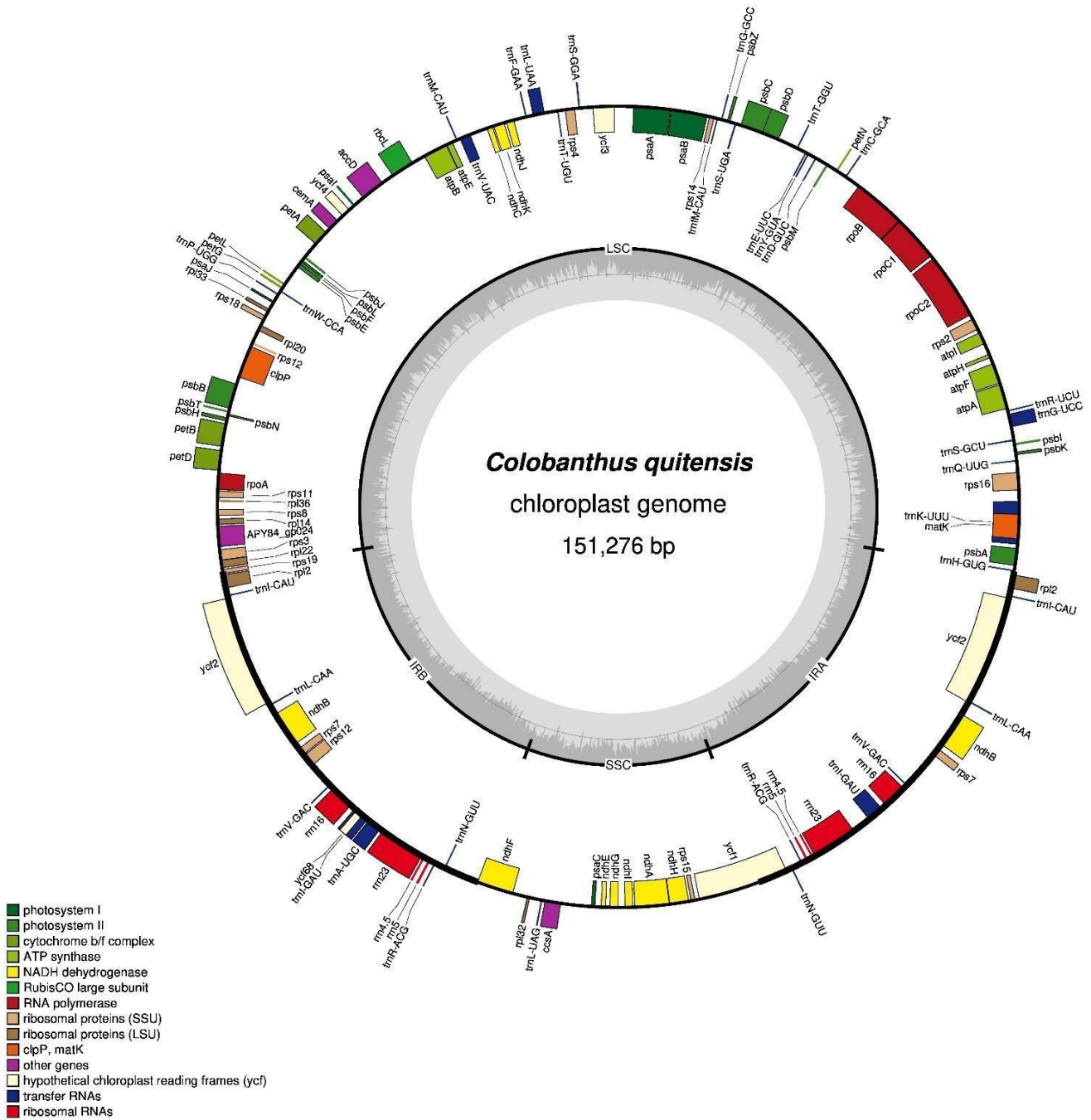


Figura 13. Estructura cuatripartita de *Colobanthus quitensis* (Generada).

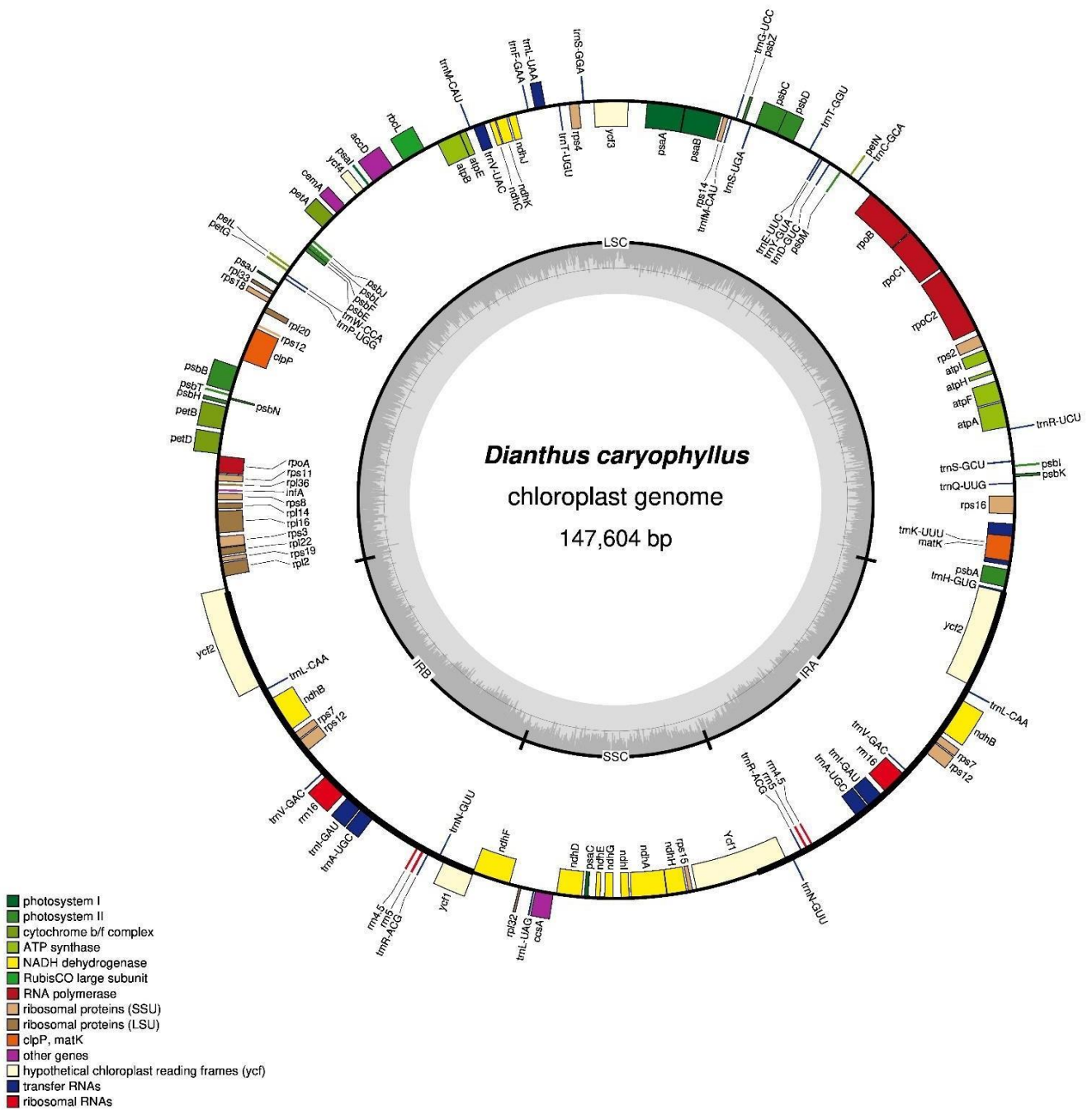


Figura 14. Estructura cuatripartita de *Dianthus caryophyllus* (Generada).

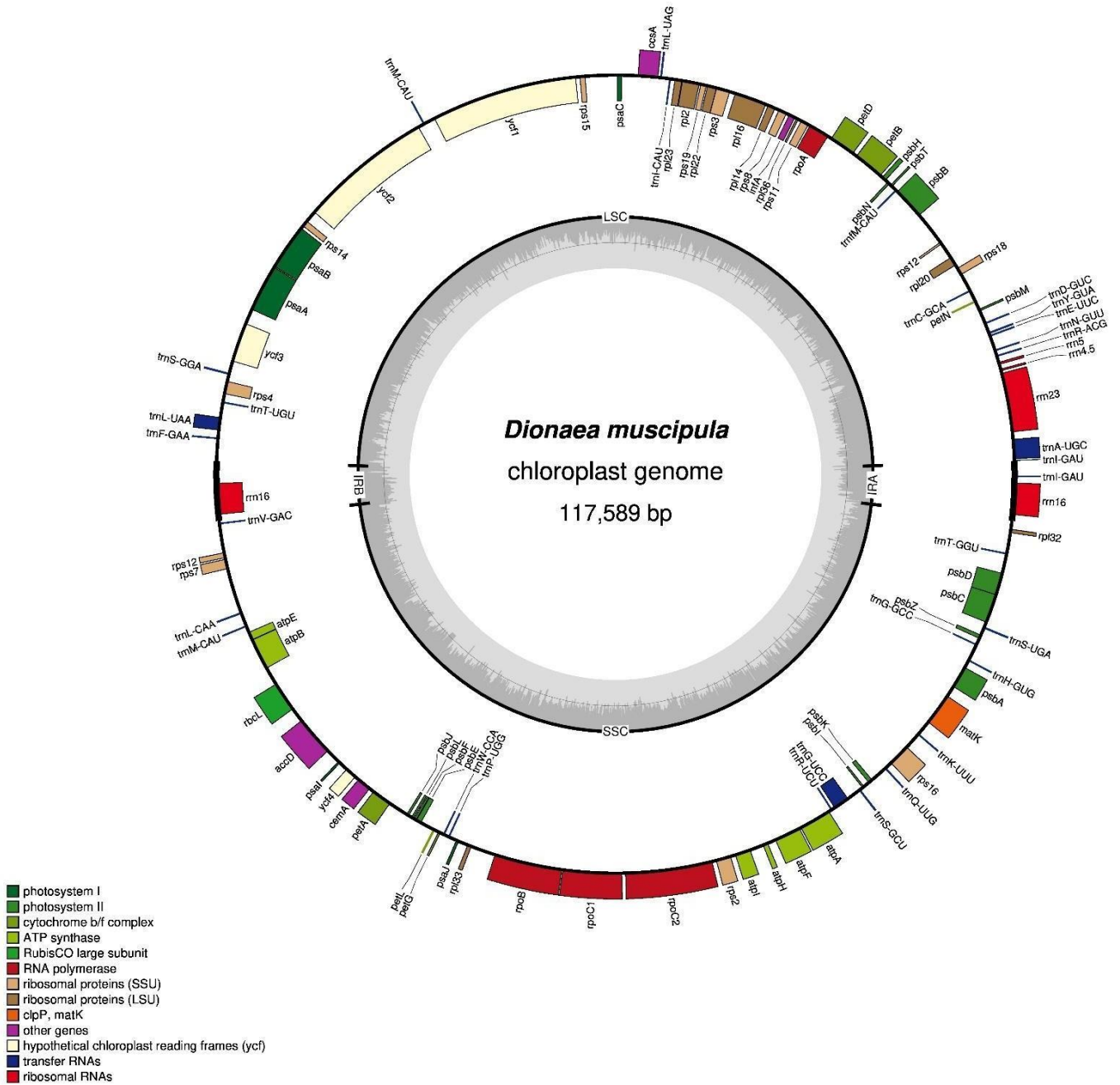


Figura 15. Estructura cuadripartita de *Dionaea muscipala* (Generada).

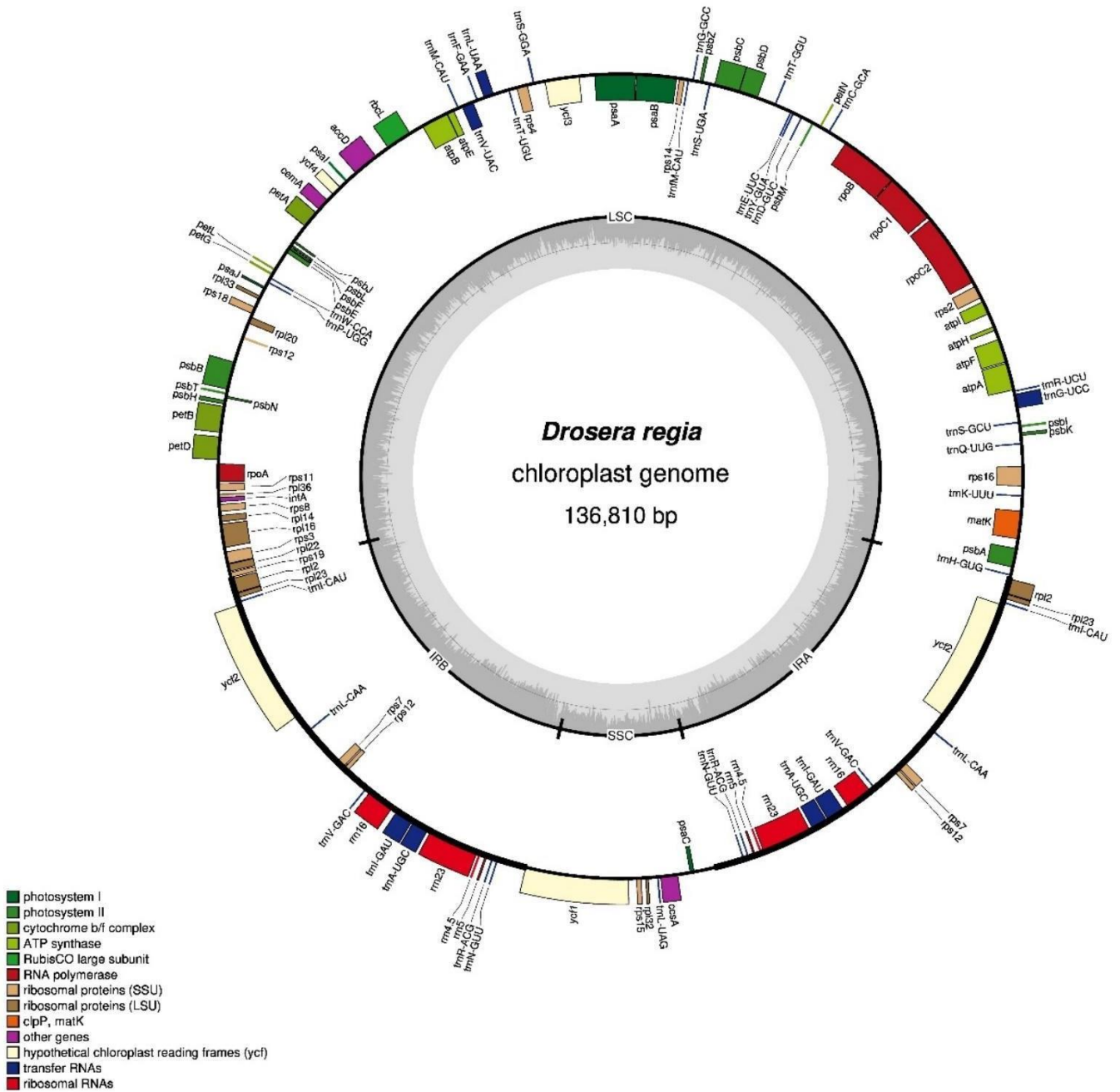


Figura 16. Estructura cuadripartita de *Drosera regia* (Generada).

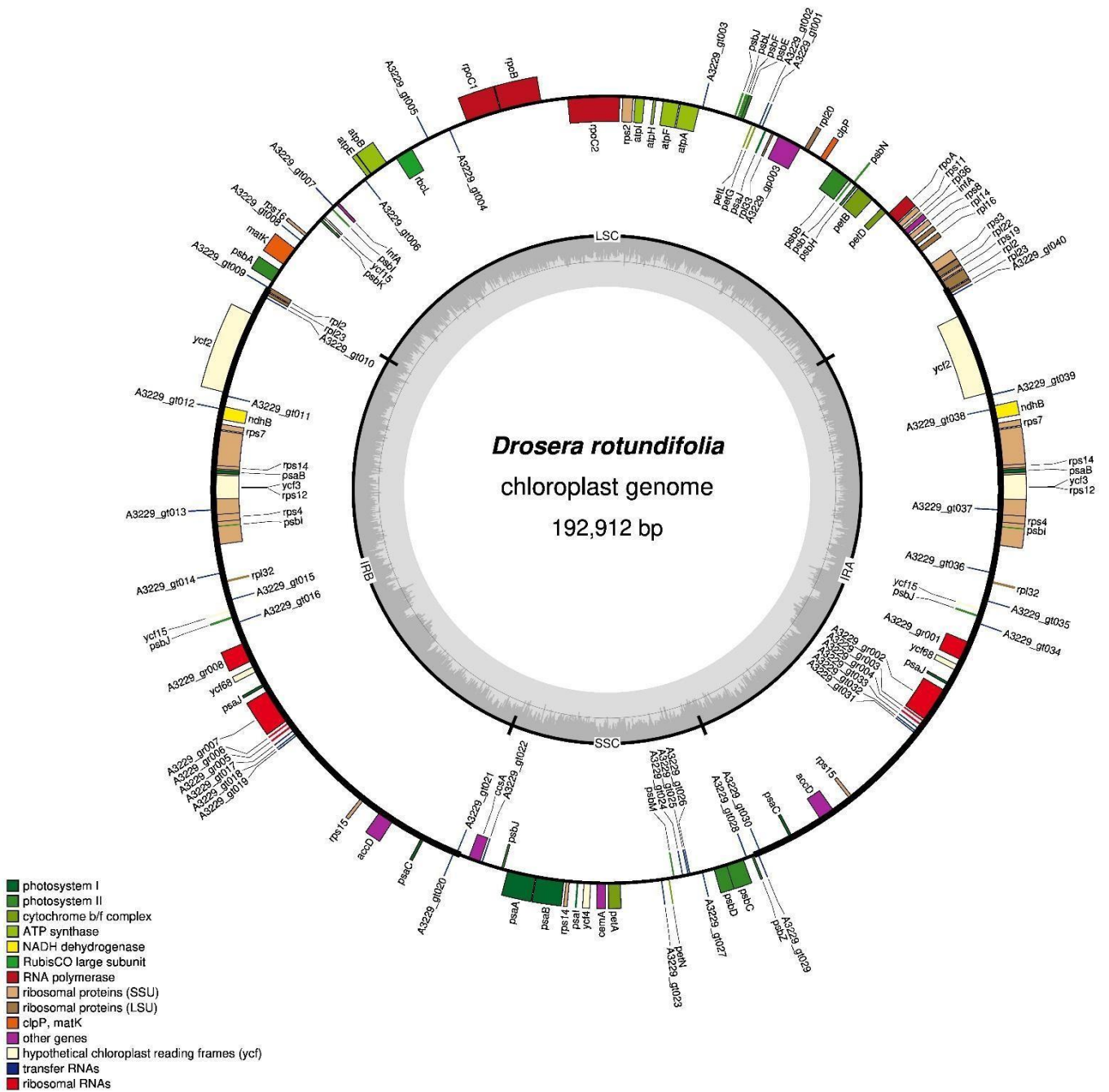


Figura 17. Estructura cuatripartita de *Drosera rotundifolia* (Generada).

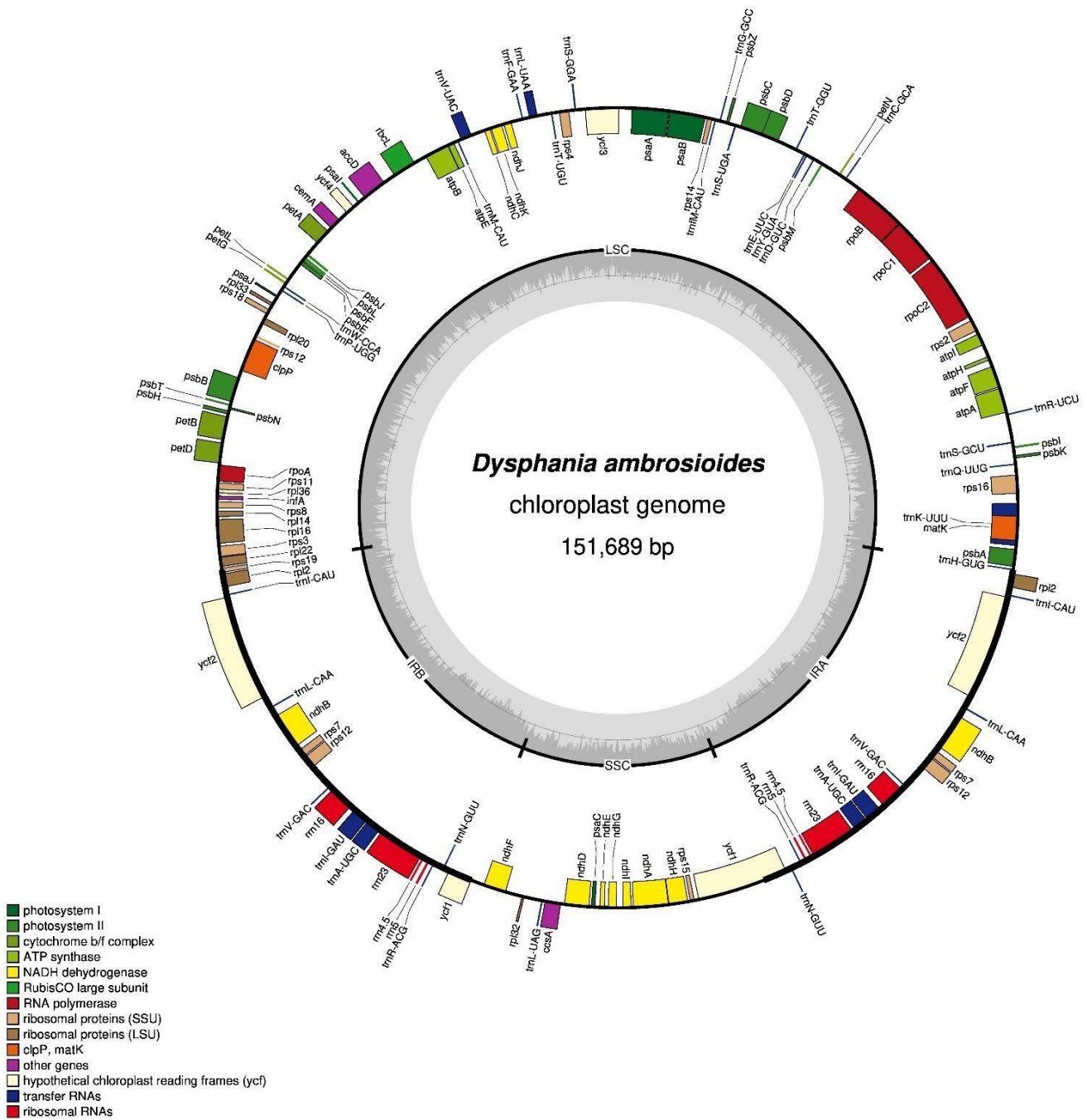


Figura 18. Estructura cuadripartita de *Dysphania ambrosioides* (Generada).



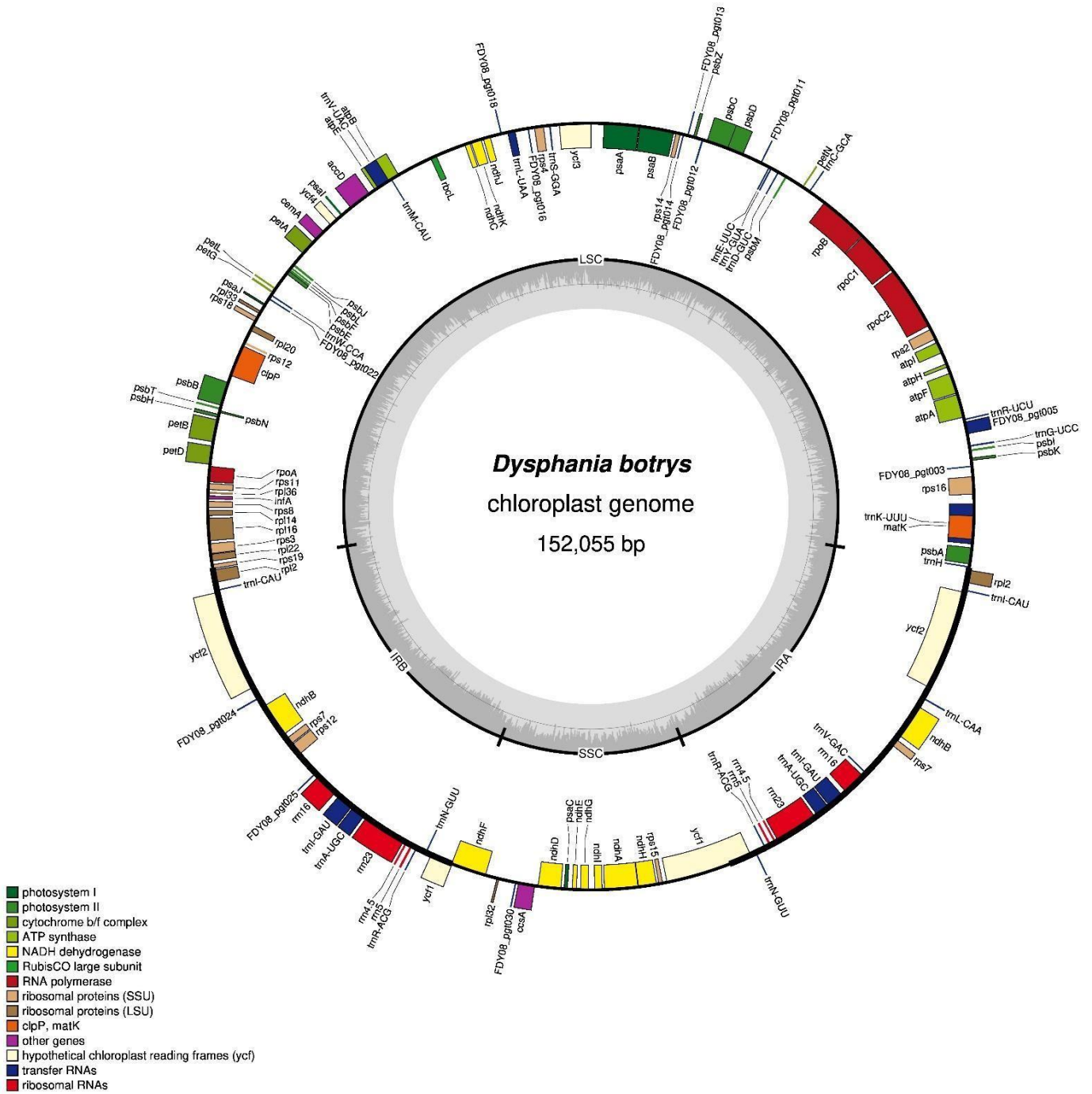


Figura 19. Estructura cuatripartita de *Dysphania botrys* (Generada).

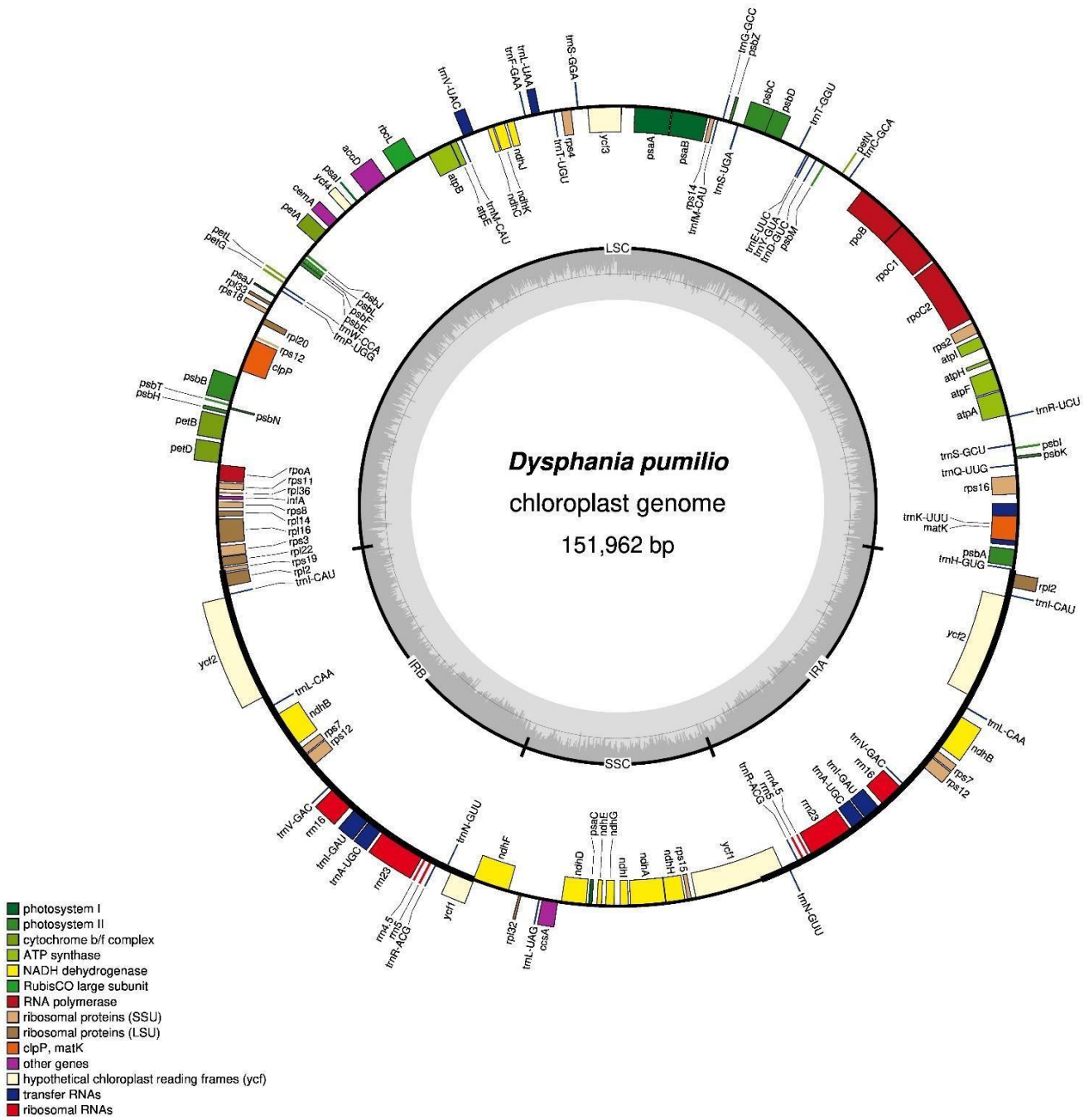


Figura 20. Estructura cuadripartita de *Dysphania pumilio* (Generada).

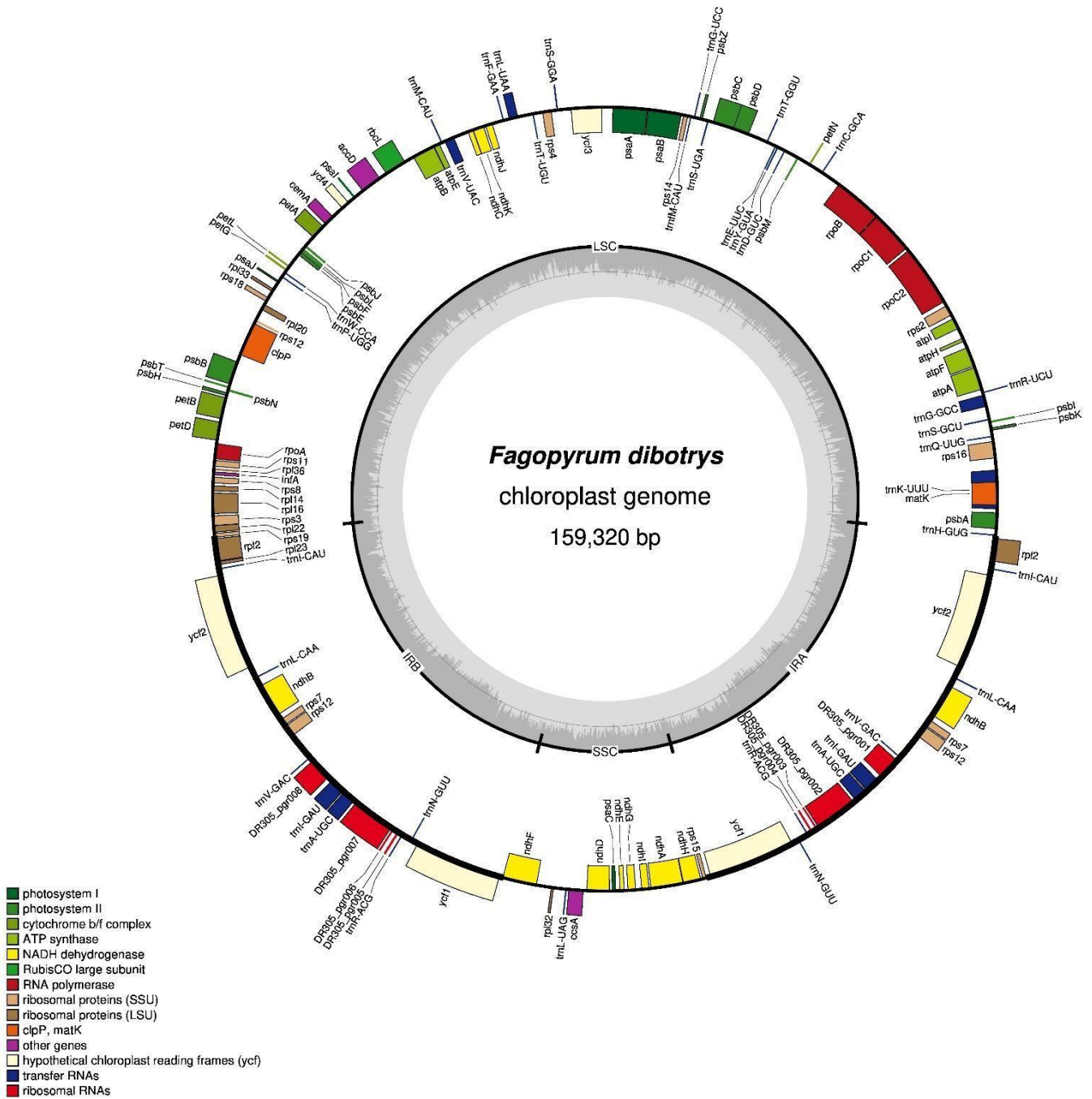


Figura 21. Estructura cuadripartita de *Fagopyrum dibotrys* (Generada).



Figura 22. Estructura cuatripartita de *Fagopyrum esculentum* (Generada).

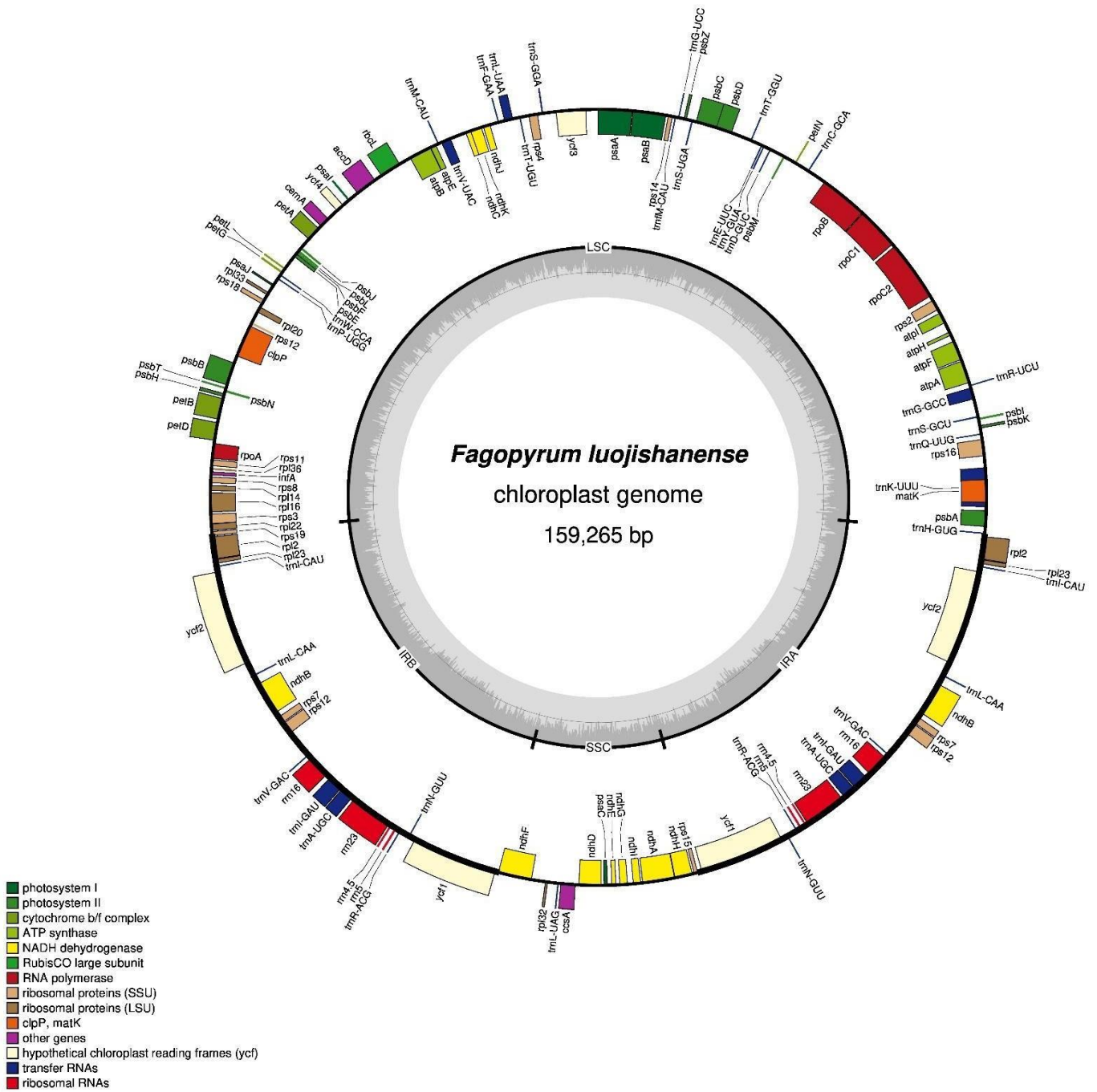


Figura 23. Estructura cuadripartita de *Fagopyrum luojishanense* (Generada).

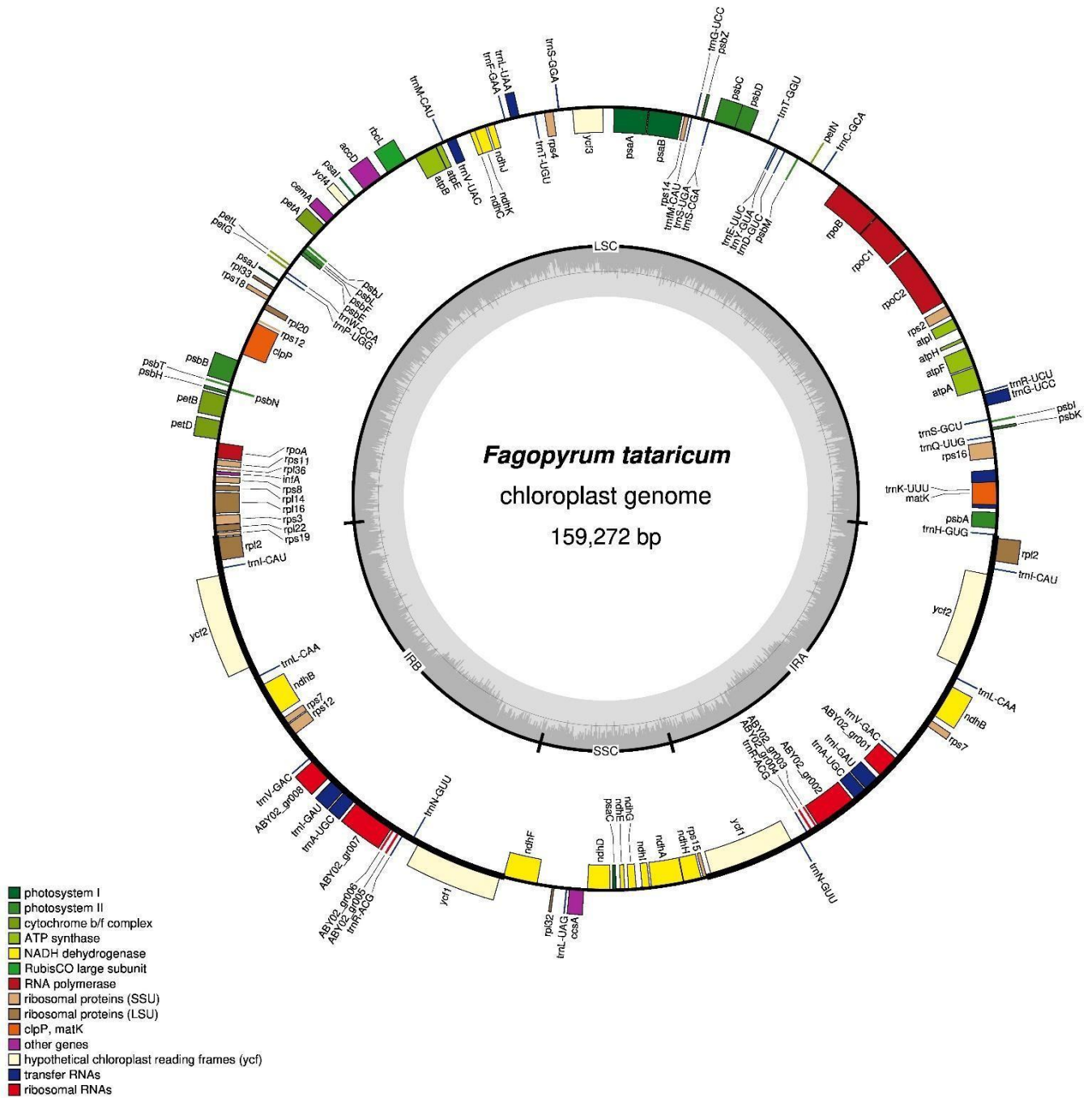
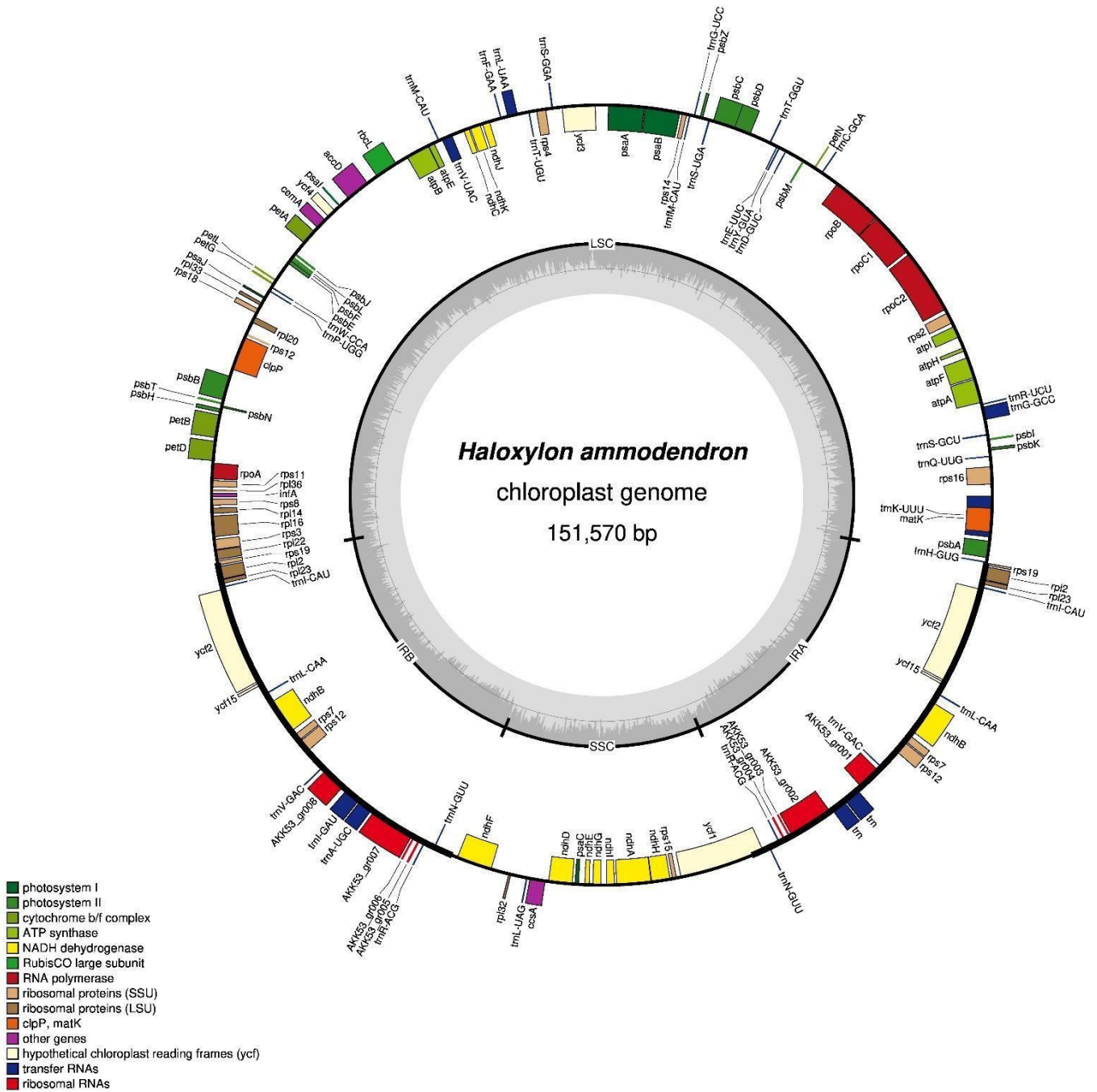


Figura 24. Estructura cuatripartita de *Fagopyrum tataricum* (Generada).







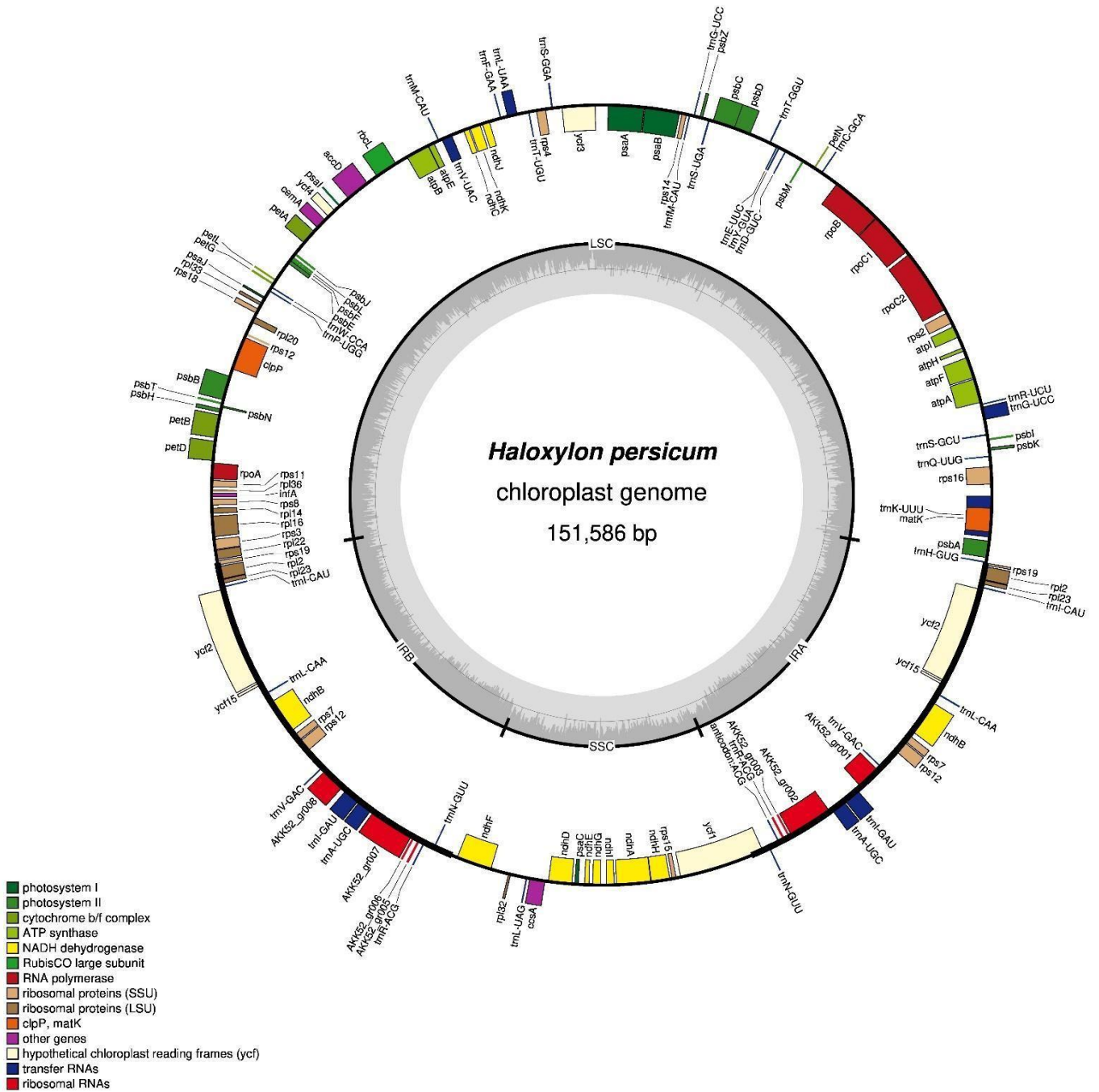


Figura 27. Estructura cuatripartita de *Haloxylon persicum* (Generada).



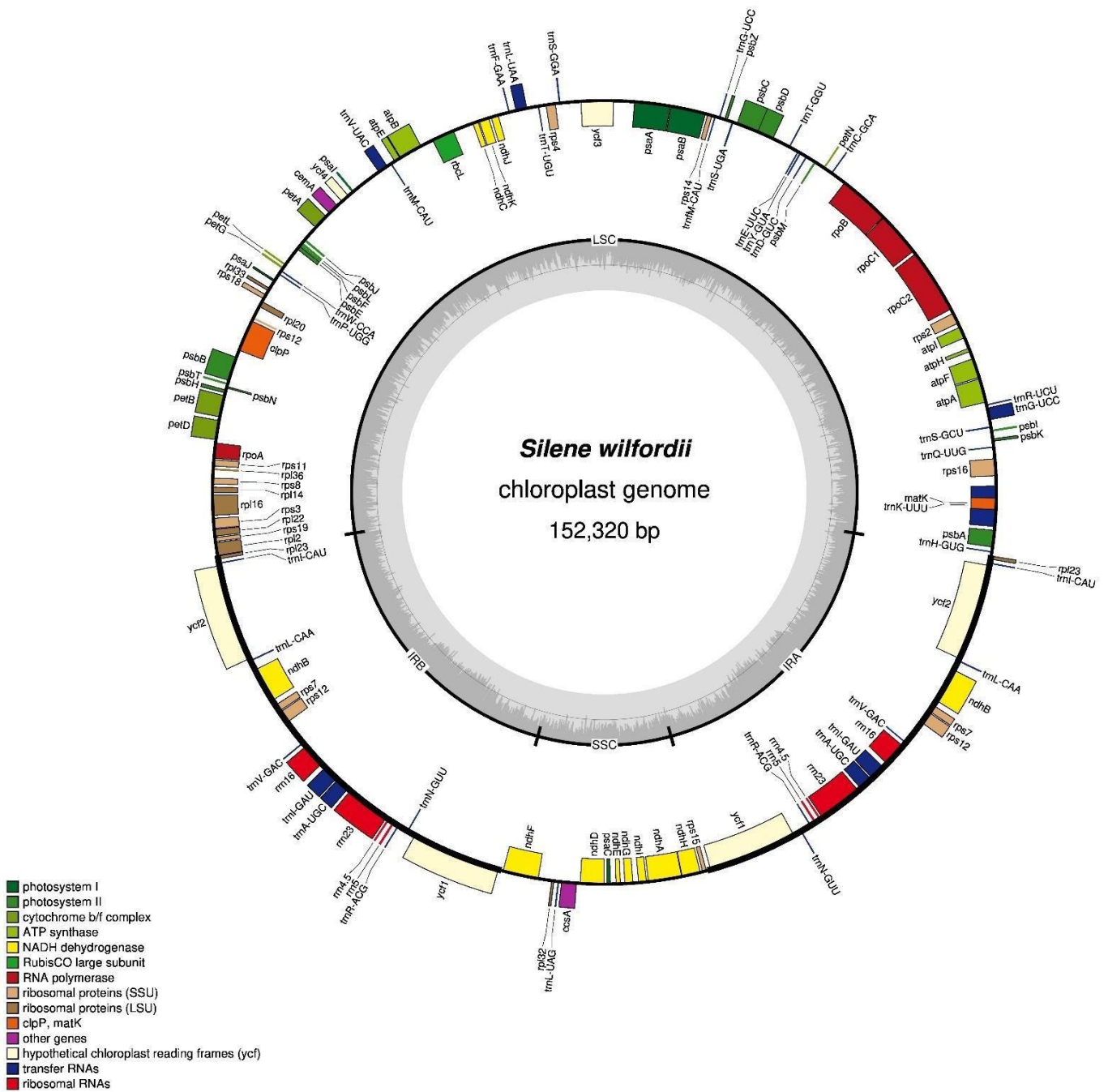


Figura 29. Estructura cuatripartita de *Lychnis wilfordii* (Generada).

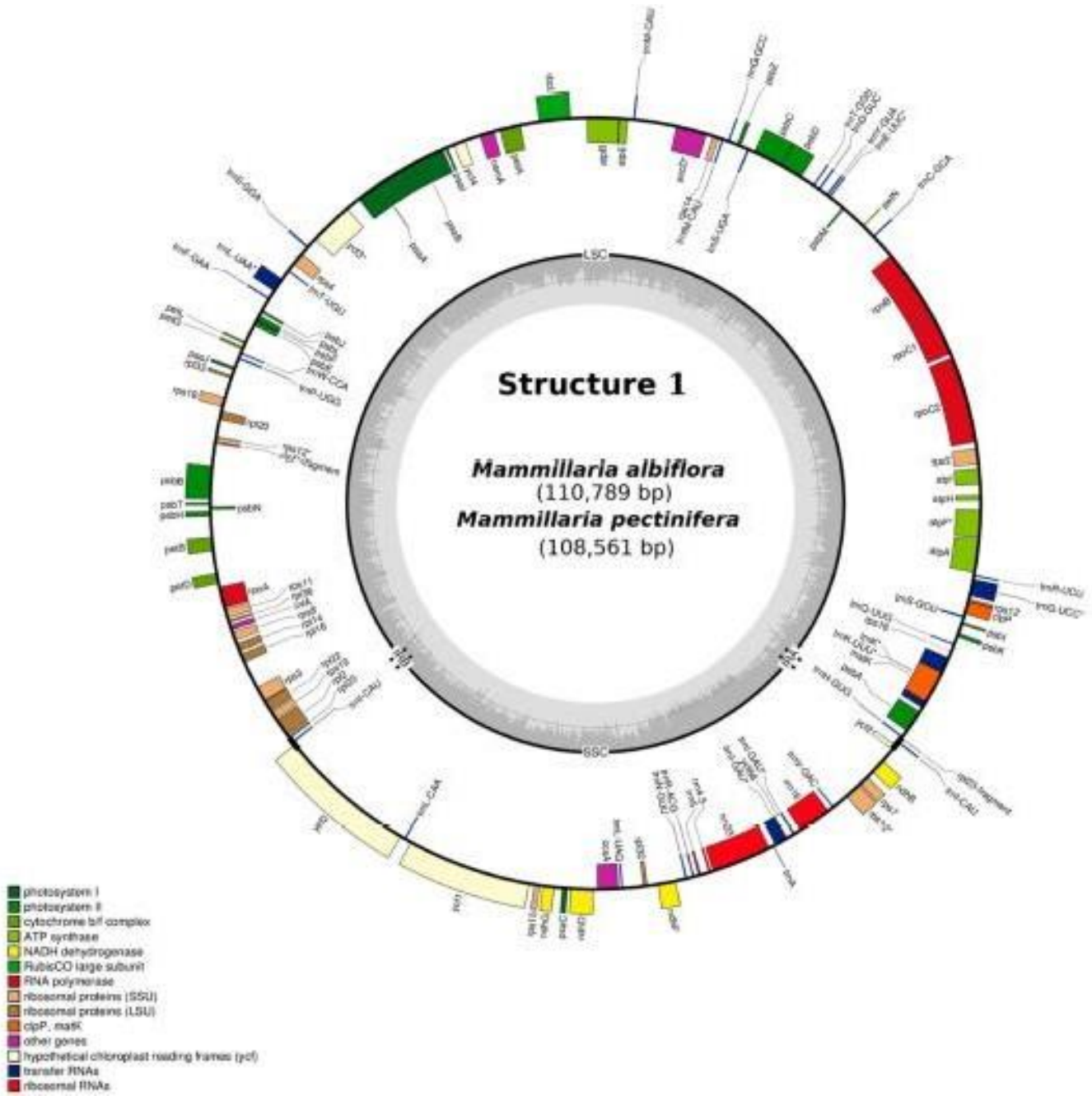


Figura 30. Estructura cuadripartita de *Mammillaria albiflora* y *Mammillaria pectinifera* tomada y modificada de Solórzano et al. 2019.

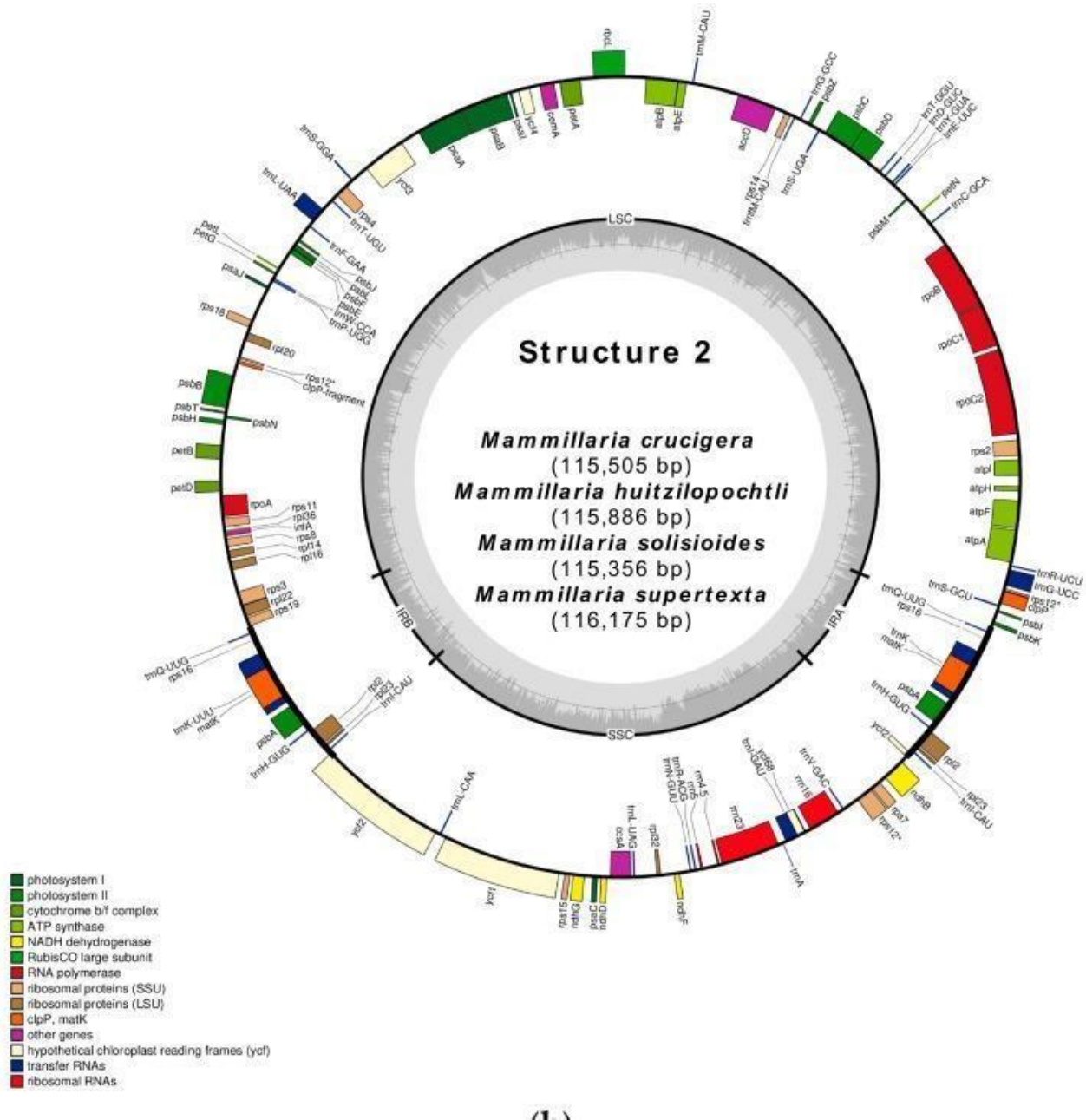
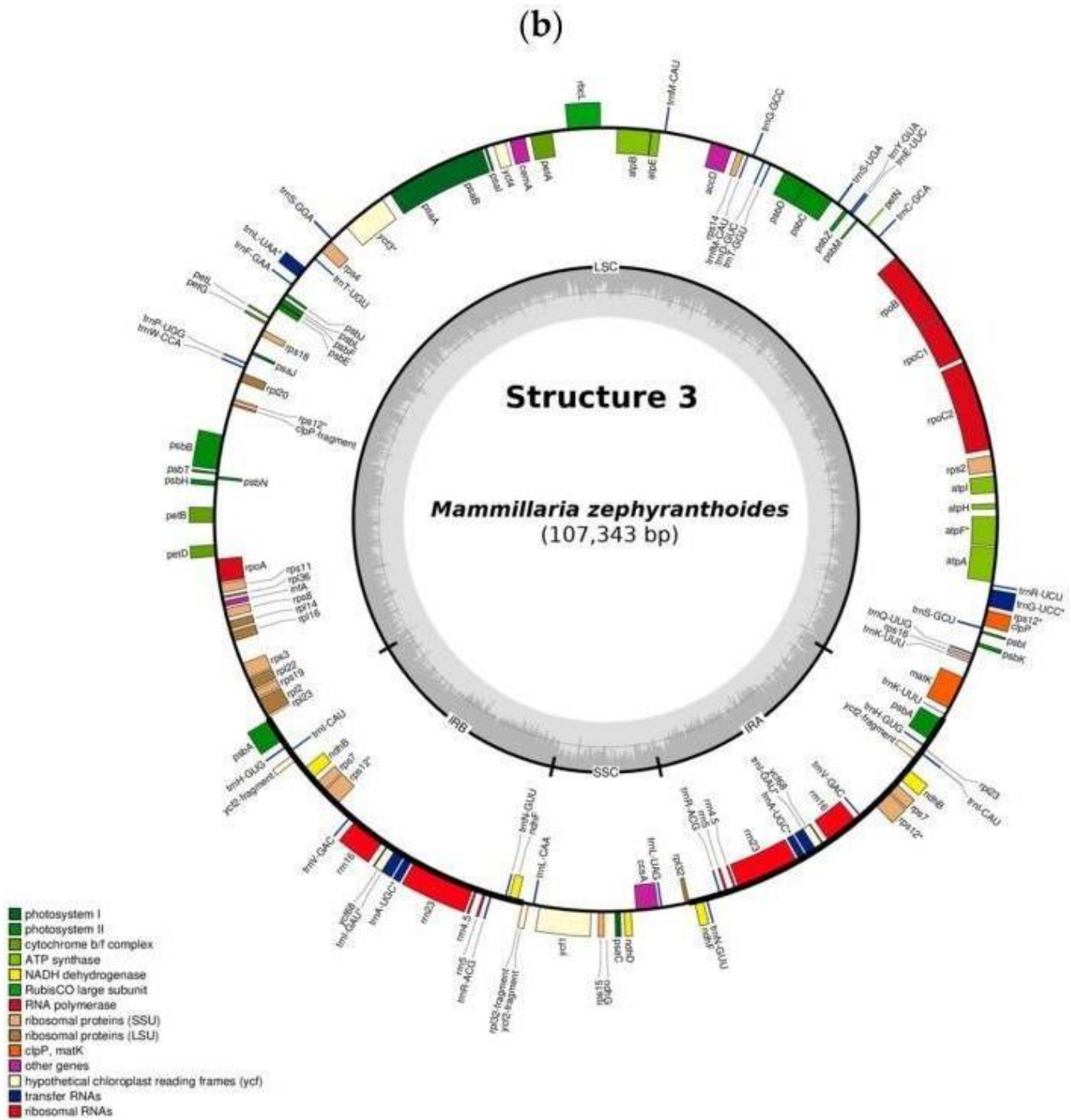
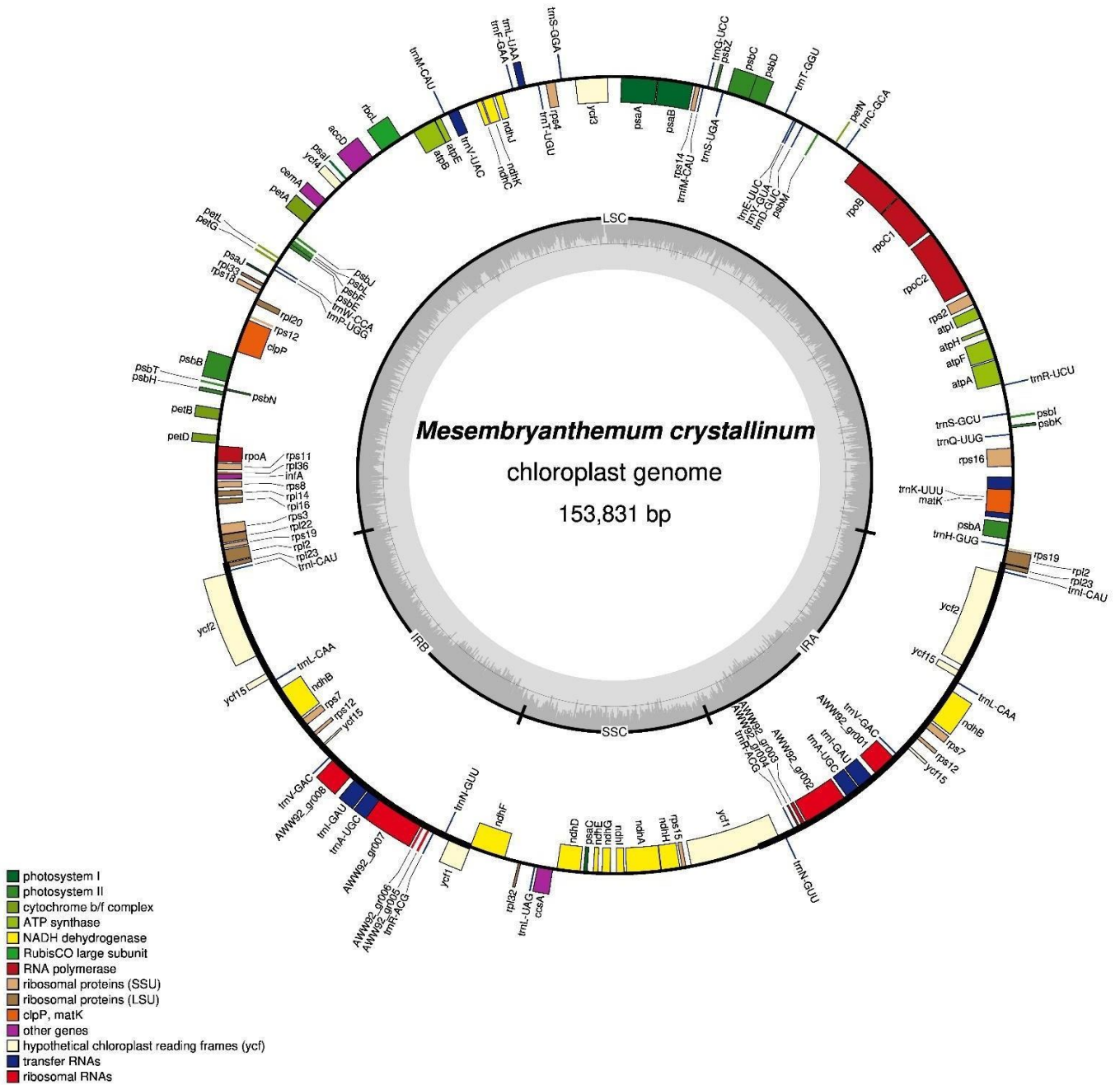


Figura 31. Estructura cuatripartita de *Mammillaria crucigera*, *Mammillaria huitzilopochtli*, *Mammillaria solisioides* y *Mammillaria supertexta* tomada y modificada de Solórzano et al. 2019.



(c)

Figura 32. Estructura cuatripartita de *Mammillaria zephyranthoides* tomada y modificada de Solórzano et al. 2019.



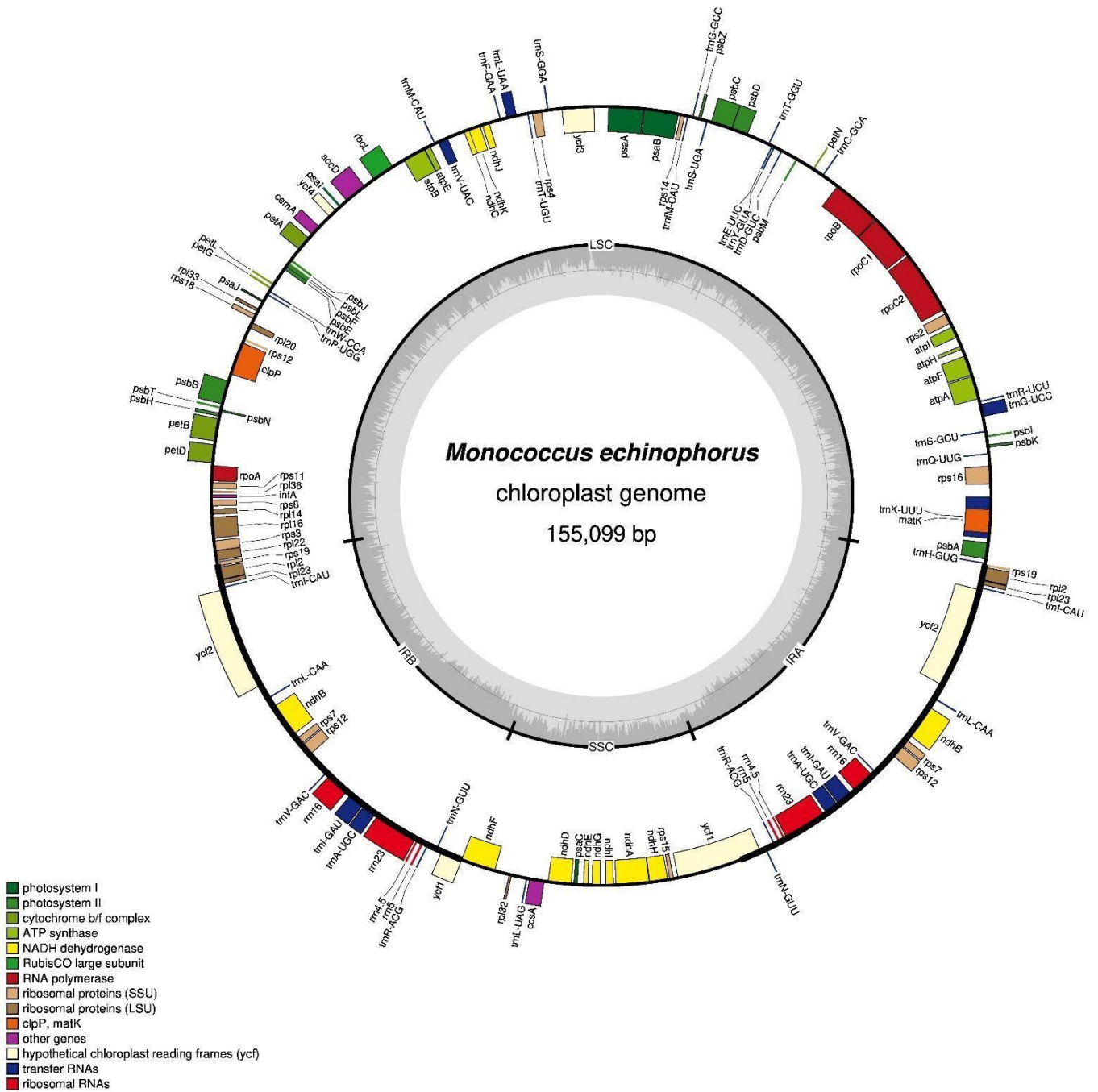


Figura 34. Estructura cuatripartita de *Monococcus echinophorus* (Generada).



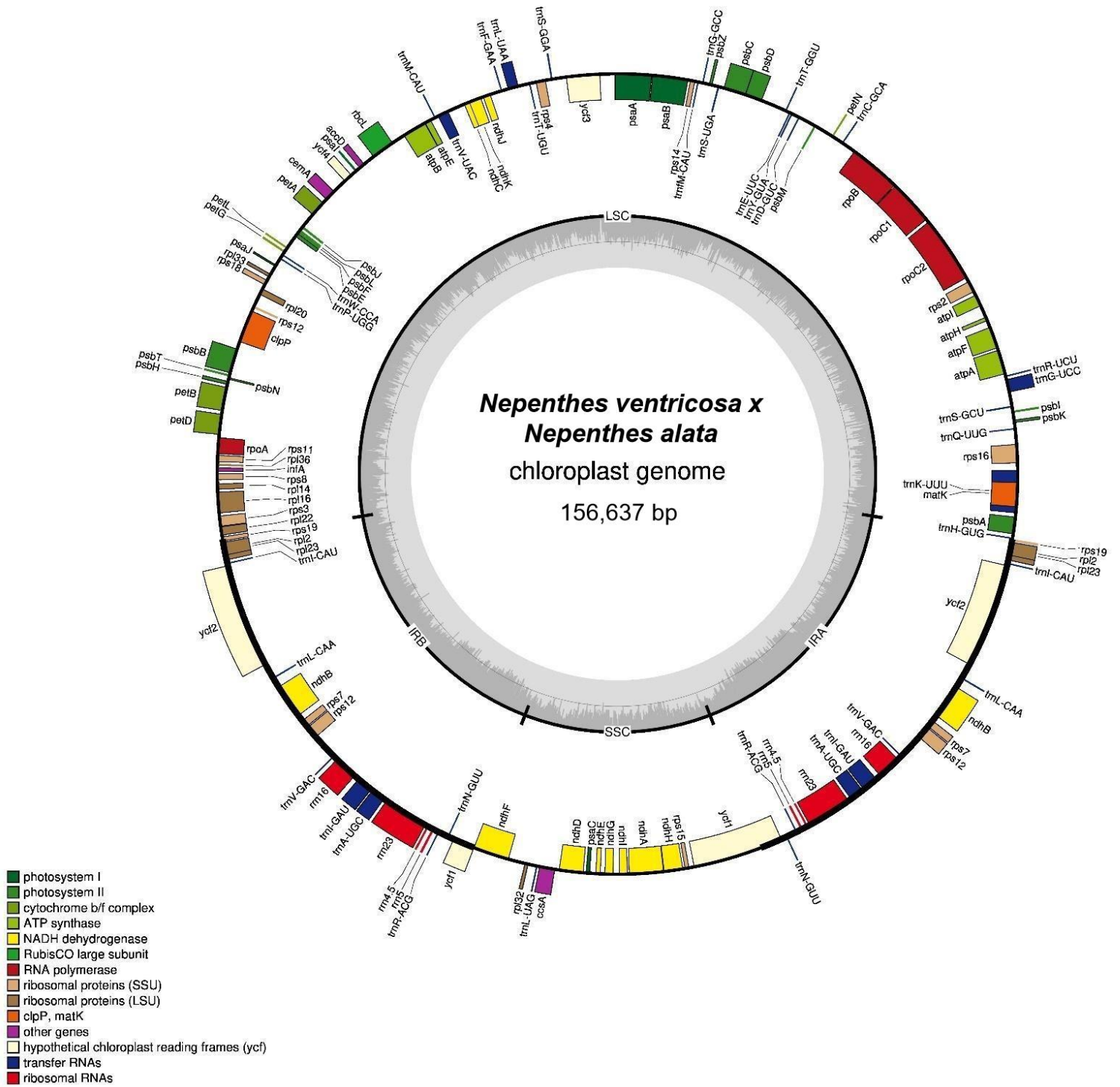


Figura 35. Estructura cuatripartita de *Nepenthes ventricosa* (Generada).

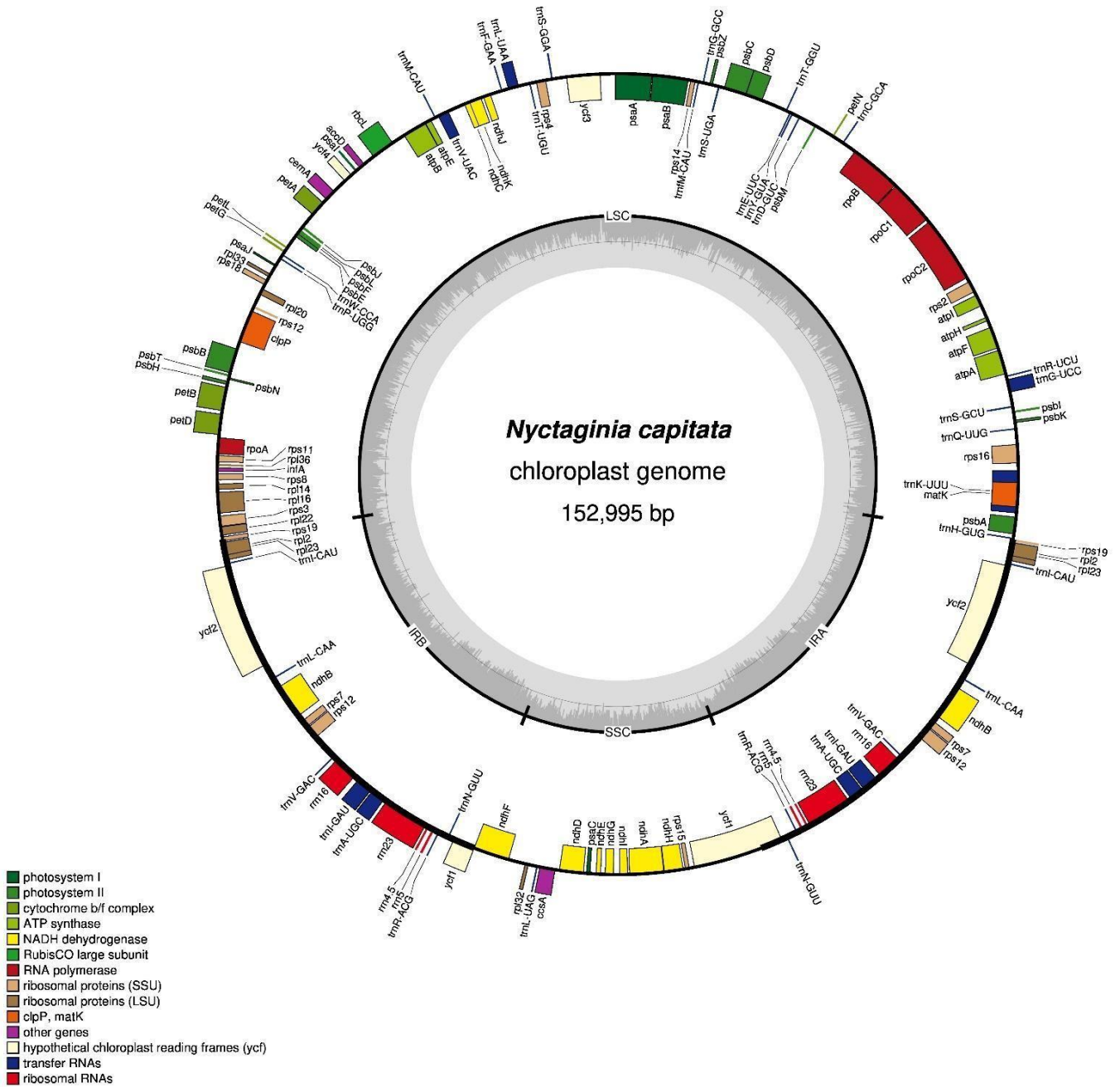


Figura 36. Estructura cuatripartita de *Nyctaginia capitata* (Generada).

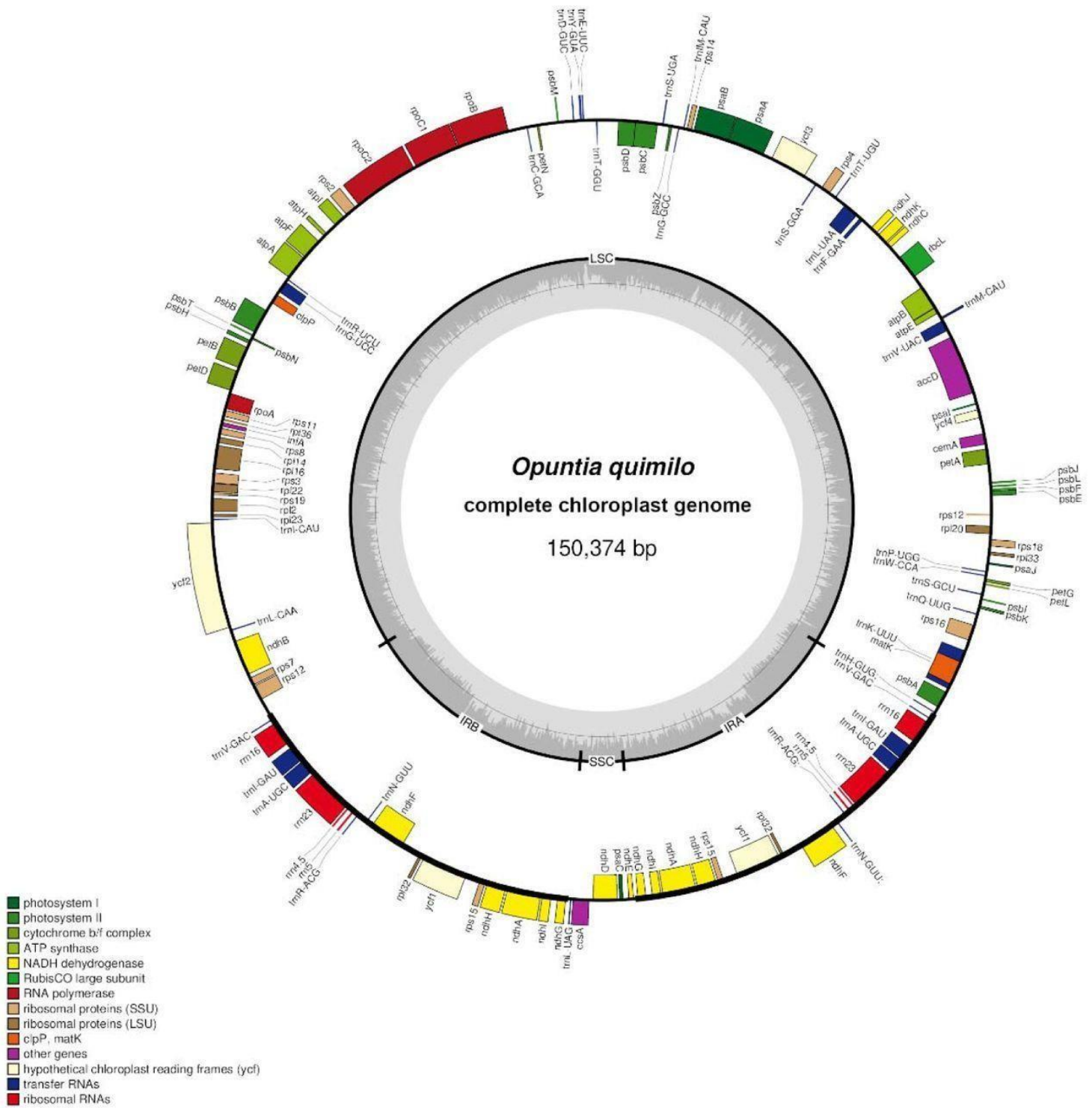


Figura 37. Estructura cuadripartita de *Opuntia quimilo* tomada y modificada de Kohler et al. 2020.



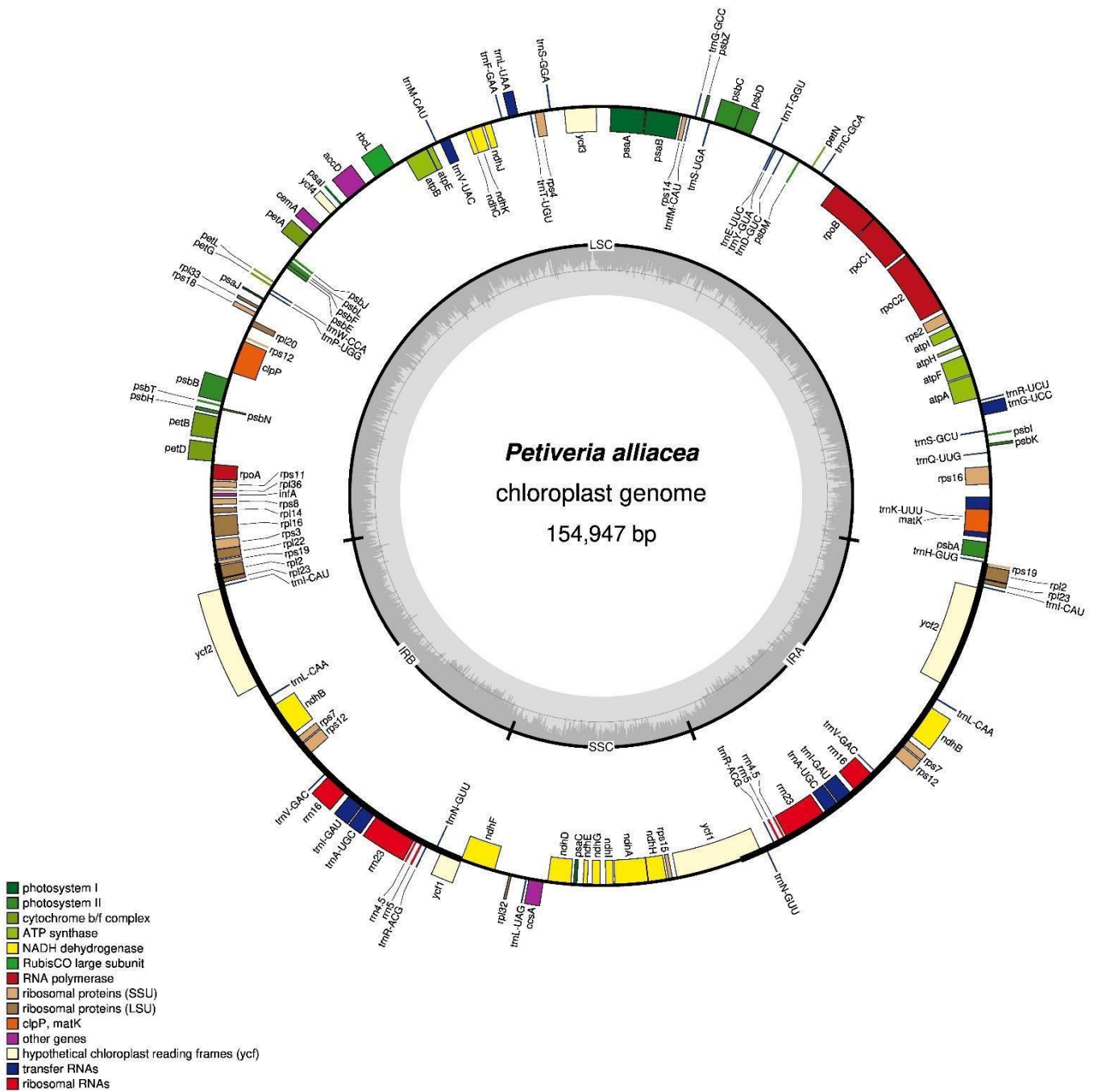


Figura 39. Estructura cuatripartita de *Petiveria alliaceae* (Generada).

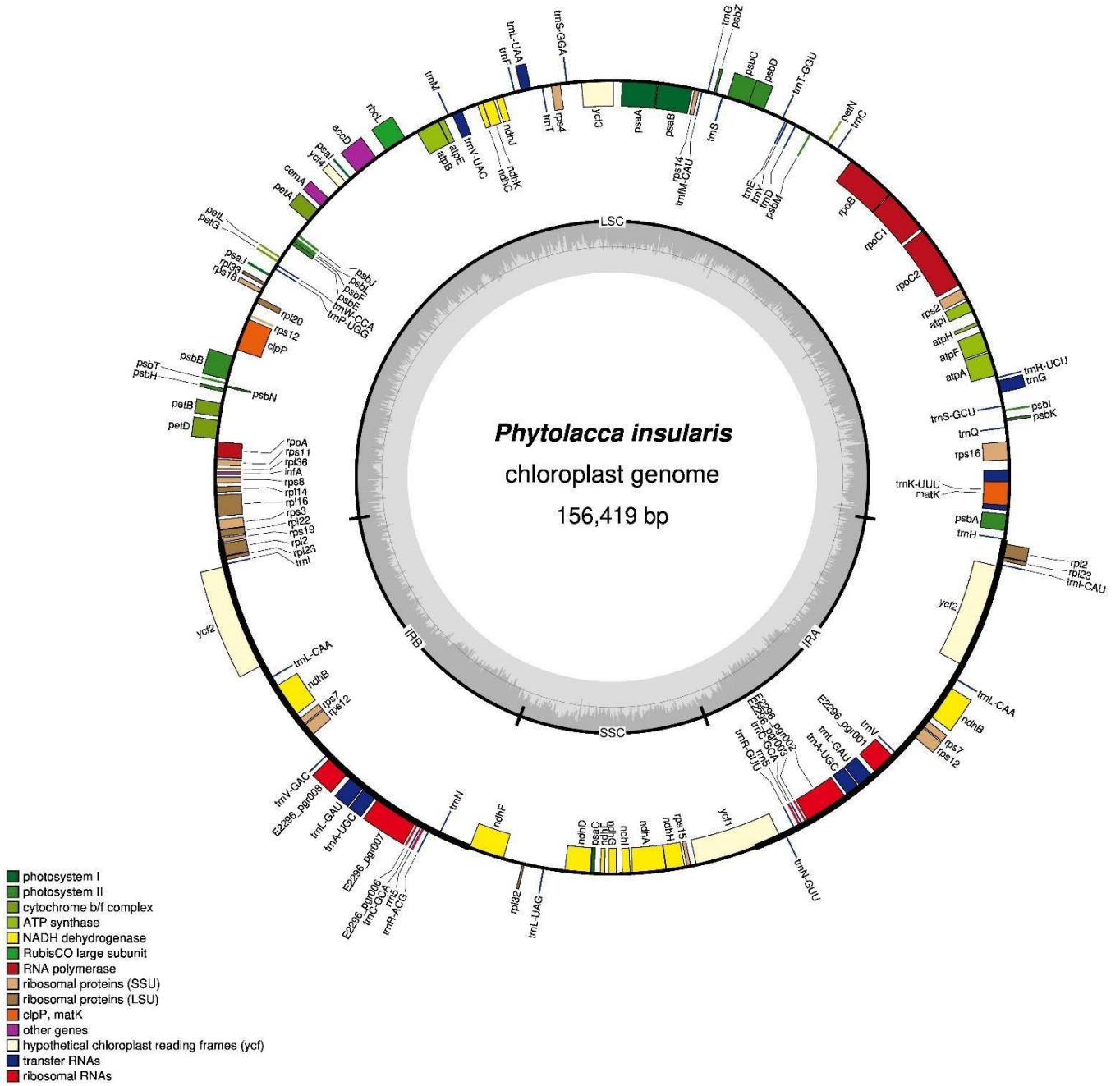


Figura 40. Estructura cuadripartita de *Phytolacca insularis* (Generada).



Figura 41. Estructura cuatripartita de *Plumbago auriculata* (Generada).

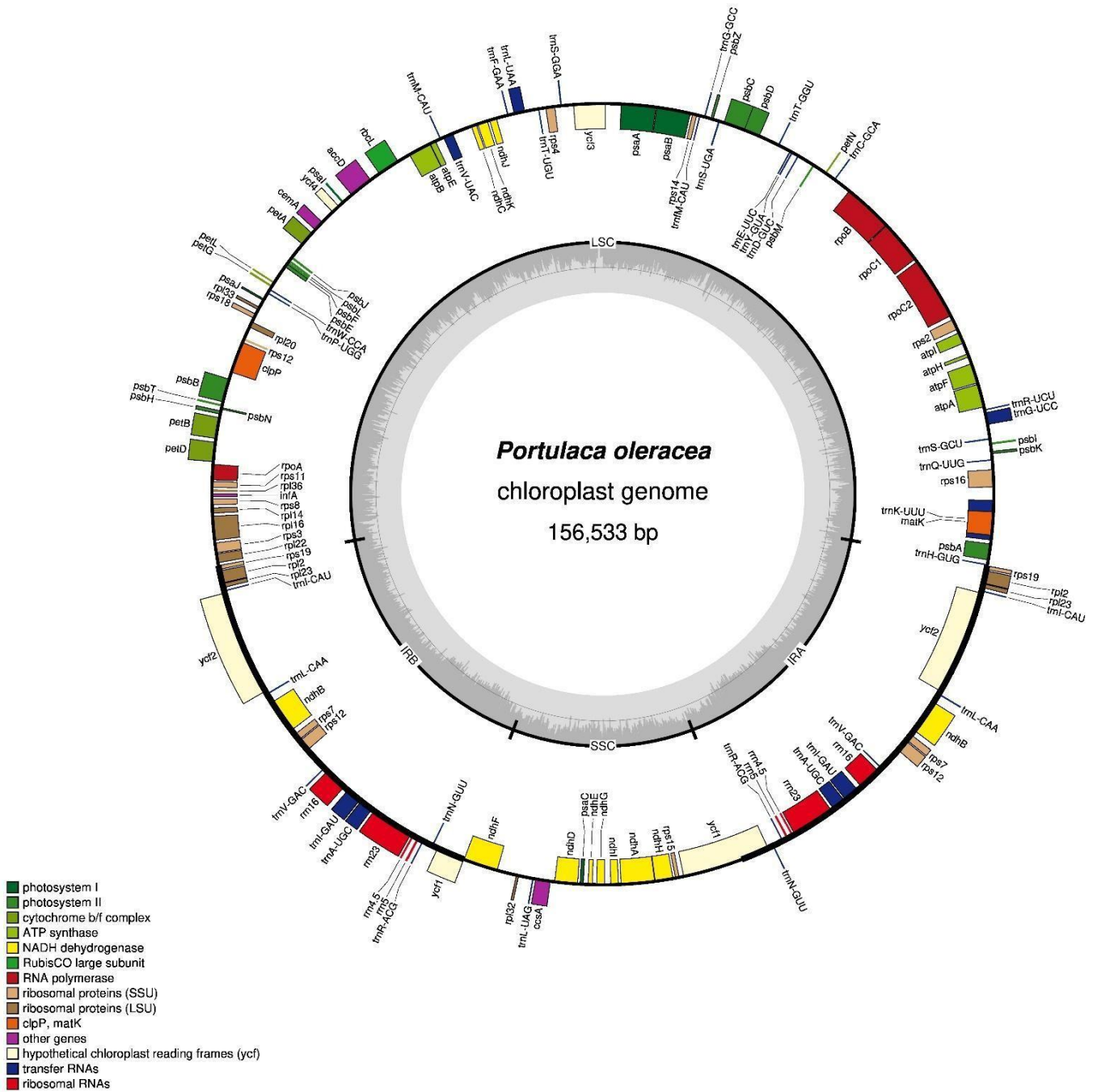


Figura 42. Estructura cuadripartita de *Portulaca oleraceae* (Generada).



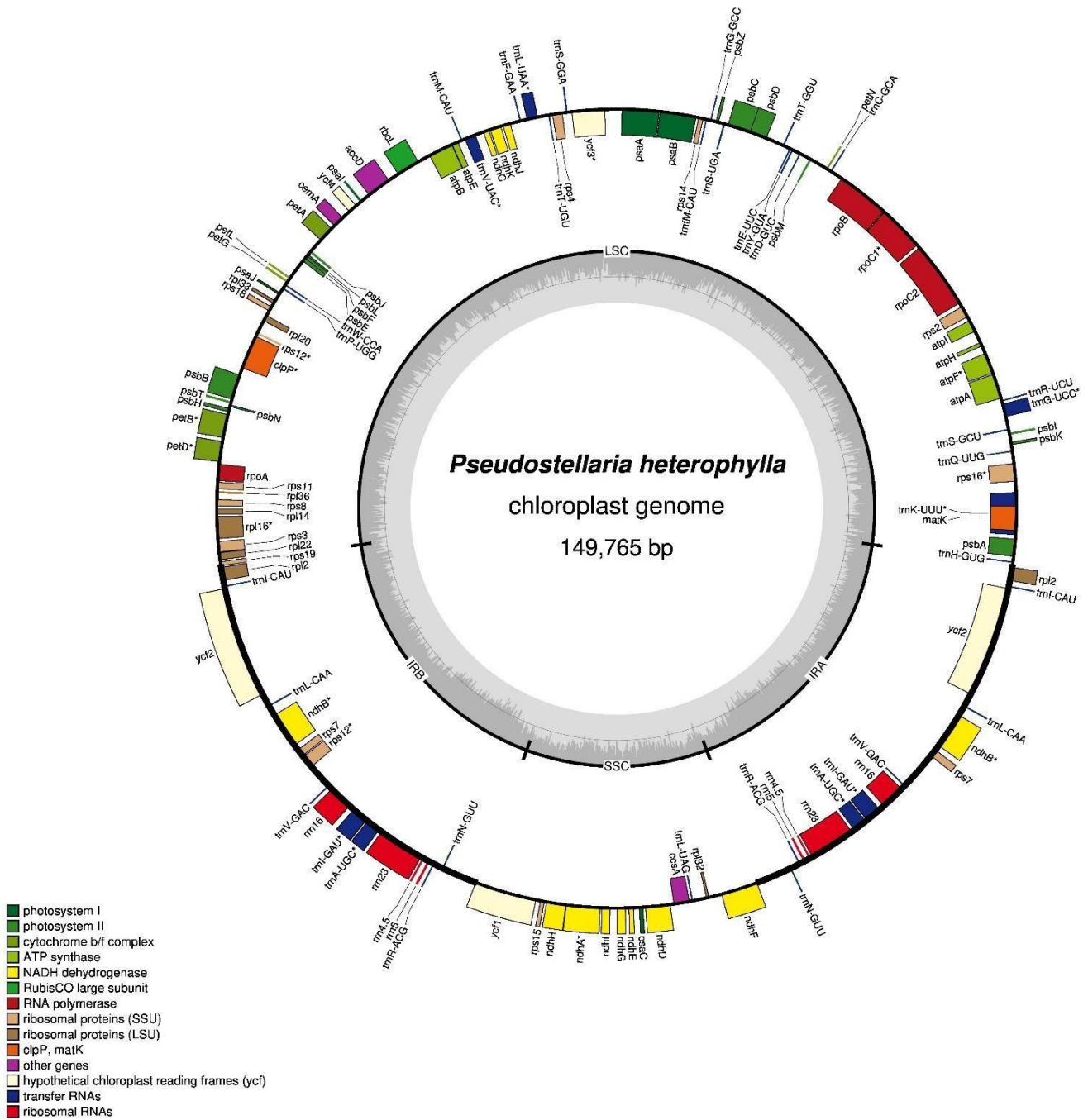


Figura 43. Estructura cuadripartita de *Pseudostellaria heterophylla* (Generada).

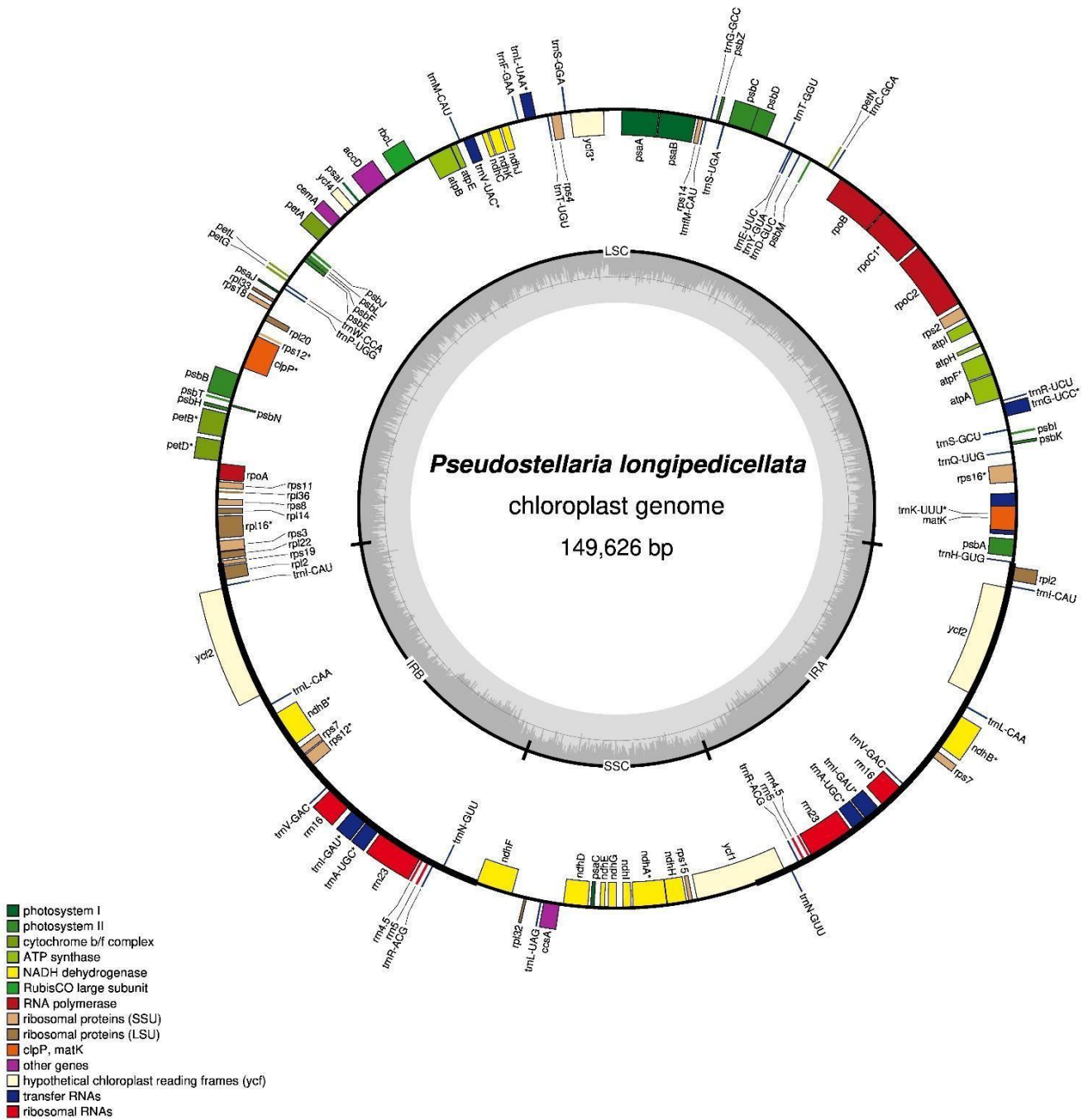


Figura 44. Estructura cuatripartita de *Pseudostellaria longipedicellata* (Generada).

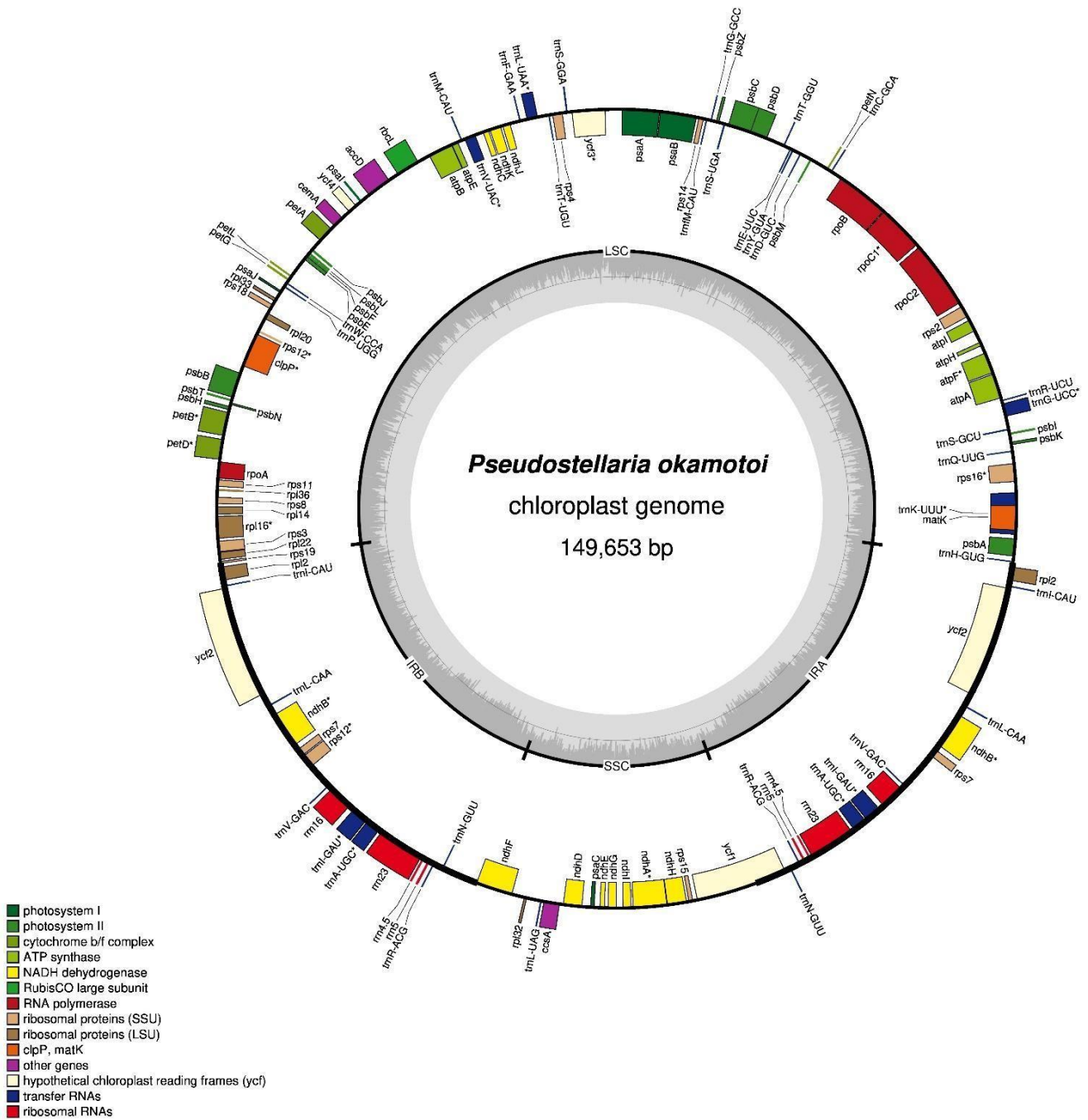


Figura 45. Estructura cuadripartita de *Pseudostellaria okamotoi* (Generada).

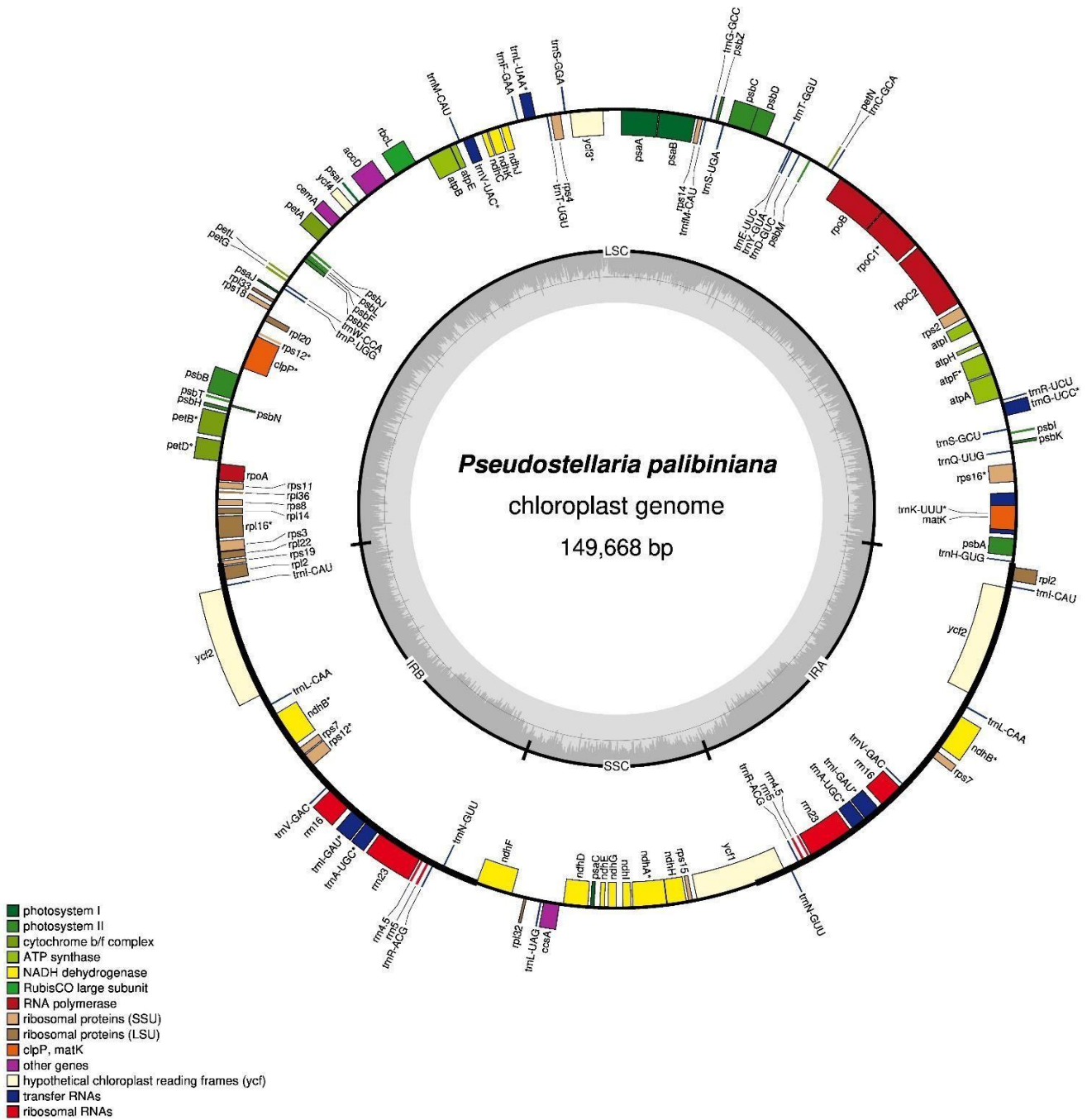


Figura 46. Estructura cuatripartita de *Pseudostellaria palibiniana* (Generada).

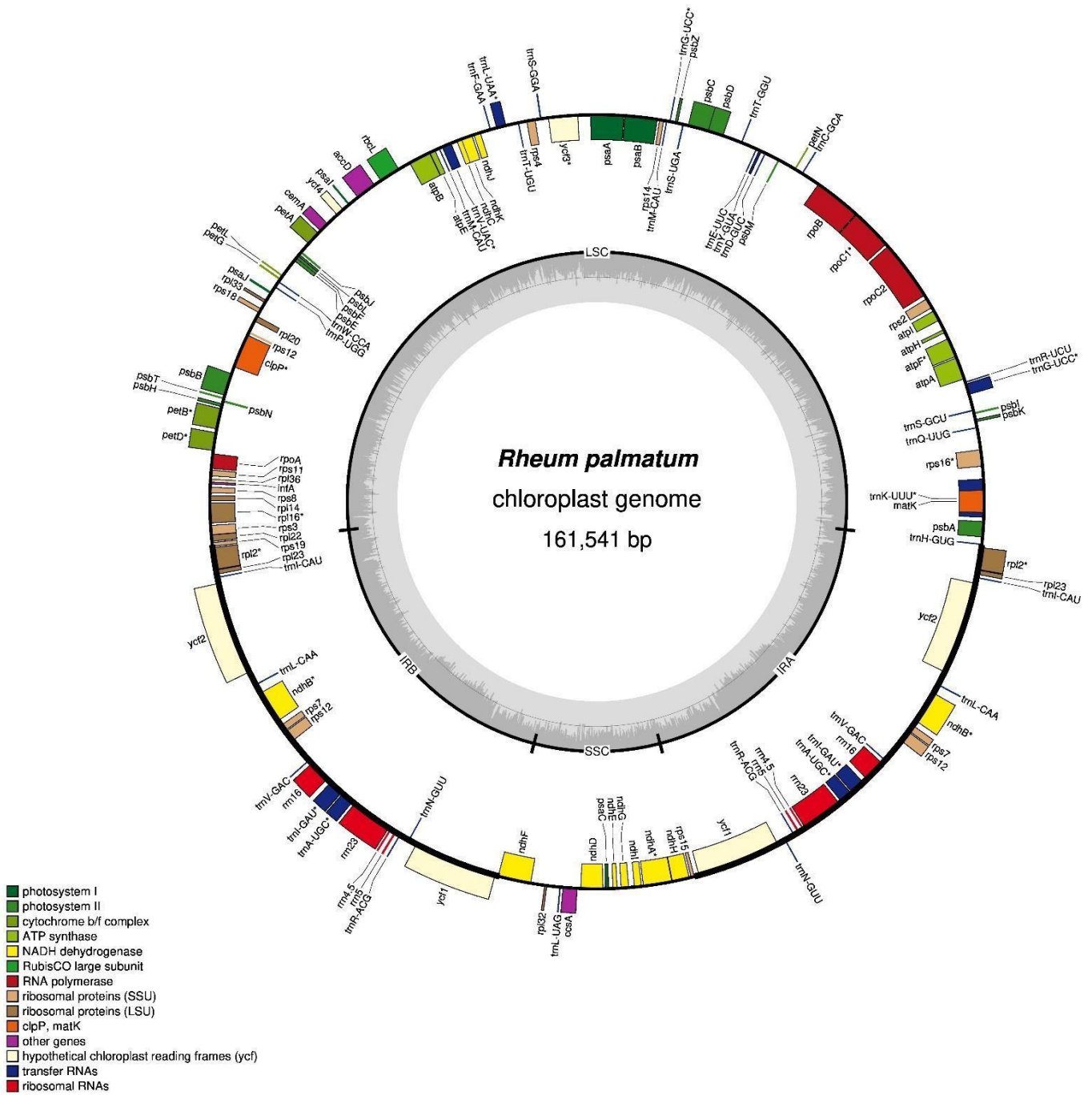


Figura 47. Estructura cuatripartita de *Rheum palmatum* (Generada).

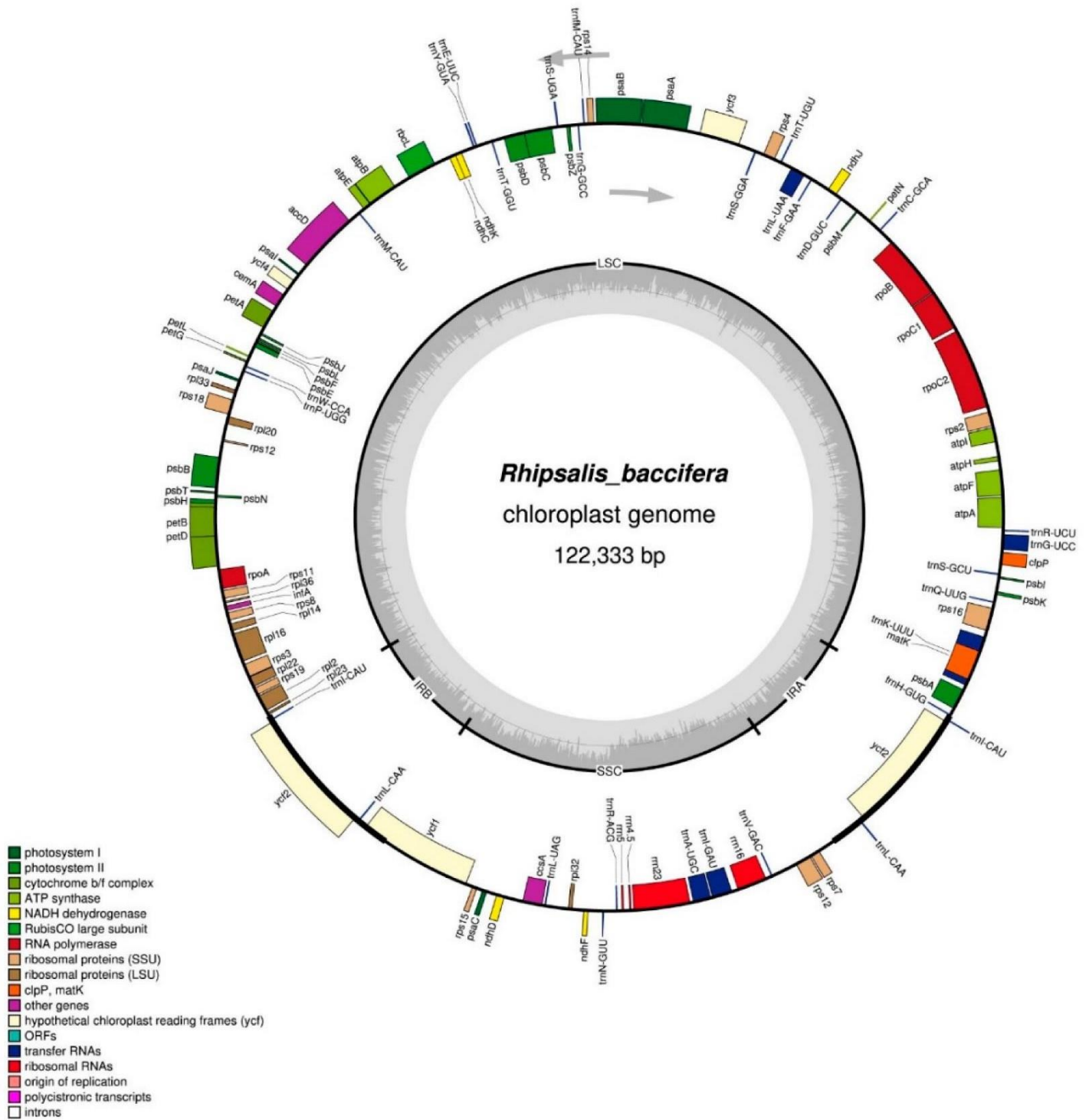


Figura 48. Estructura cuatripartita de *Rhipsalis baccifera* tomada y modificada de Oulo et al. 2020.

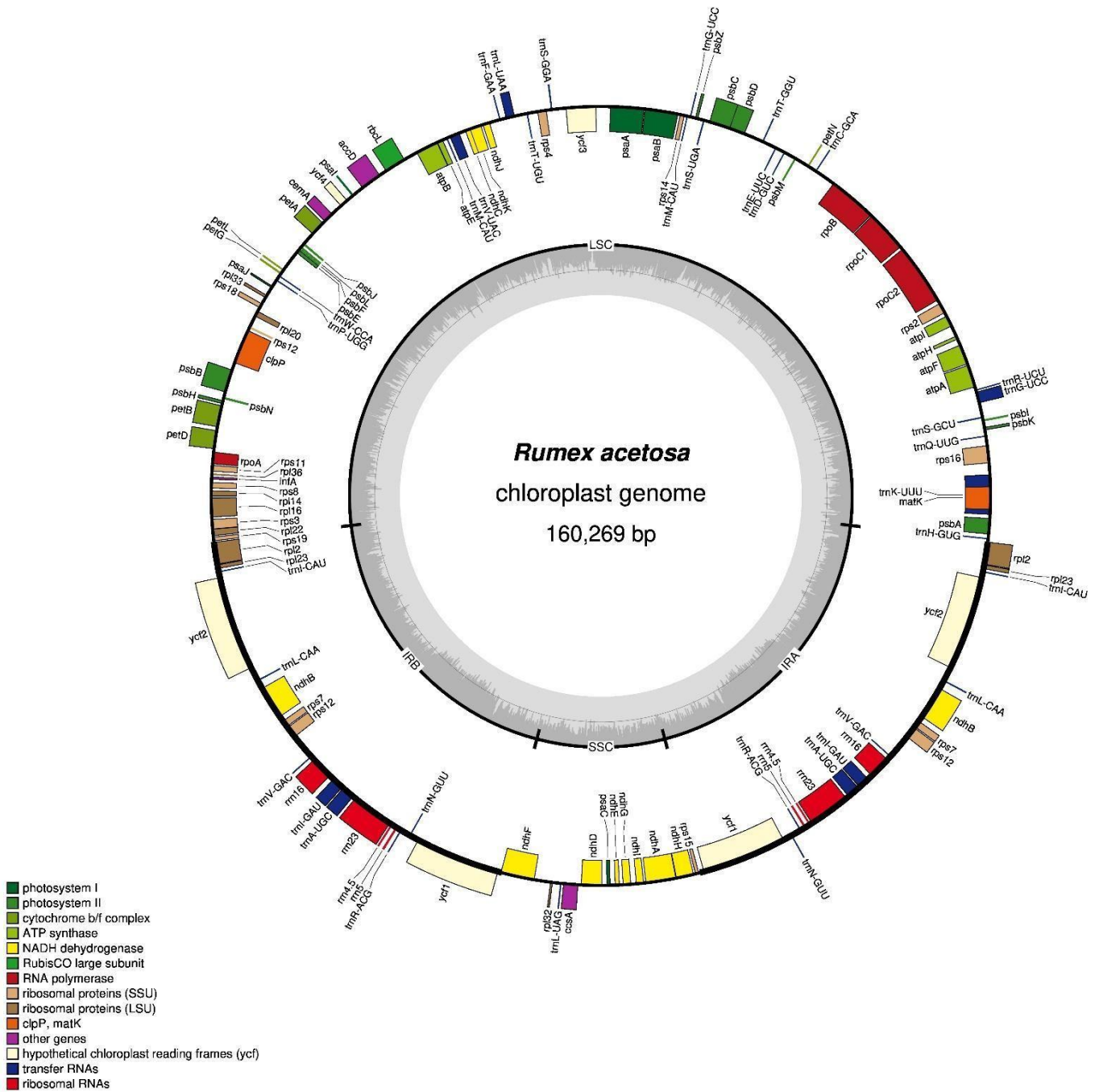


Figura 49. Estructura cuadripartita de *Rumex acetosa* (Generada).





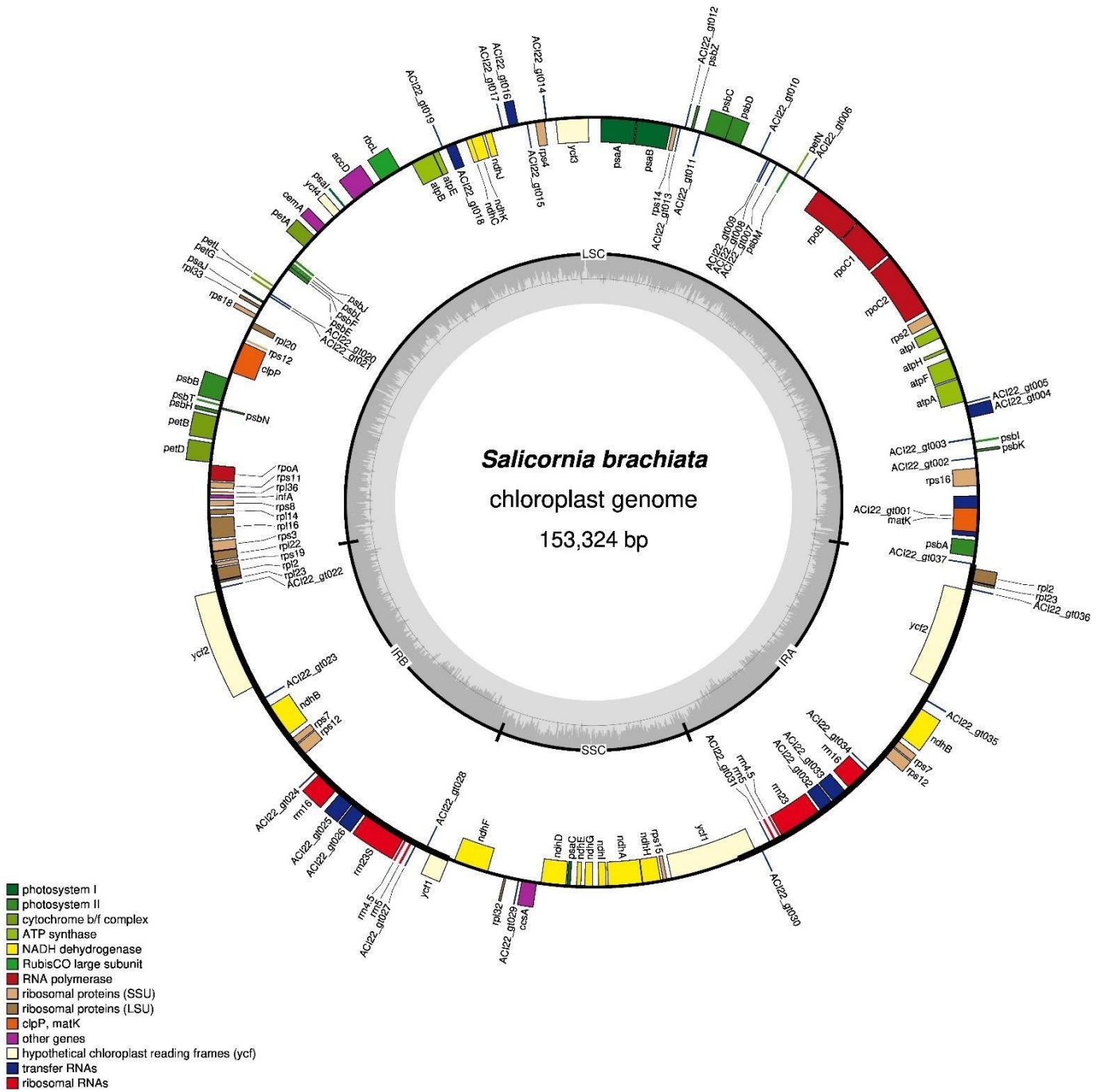


Figura 51. Estructura cuadripartita de *Salicornia brachiata* (Generada).

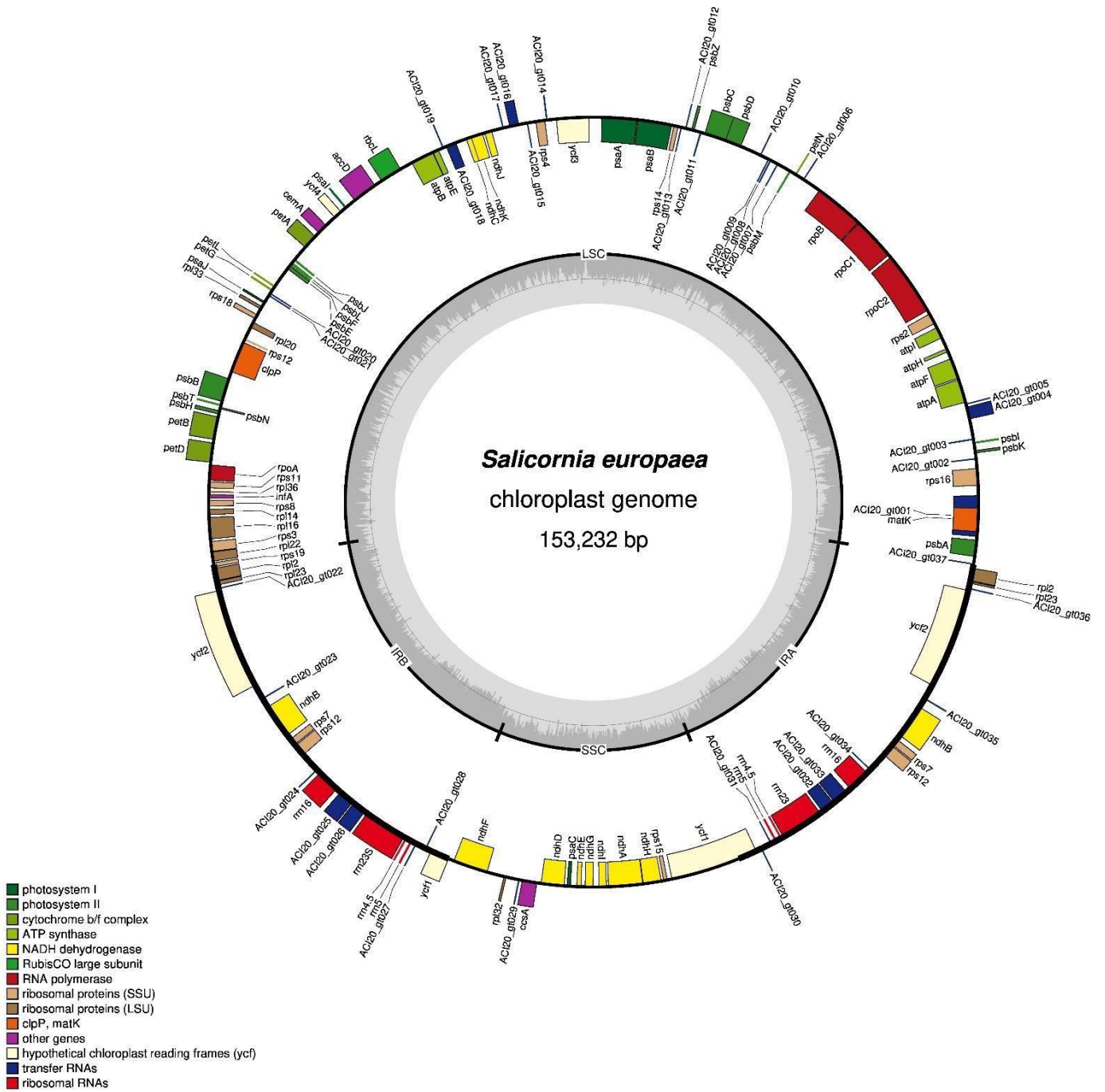


Figura 52. Estructura cuadripartita de *Salicornia europaea* (Generada).

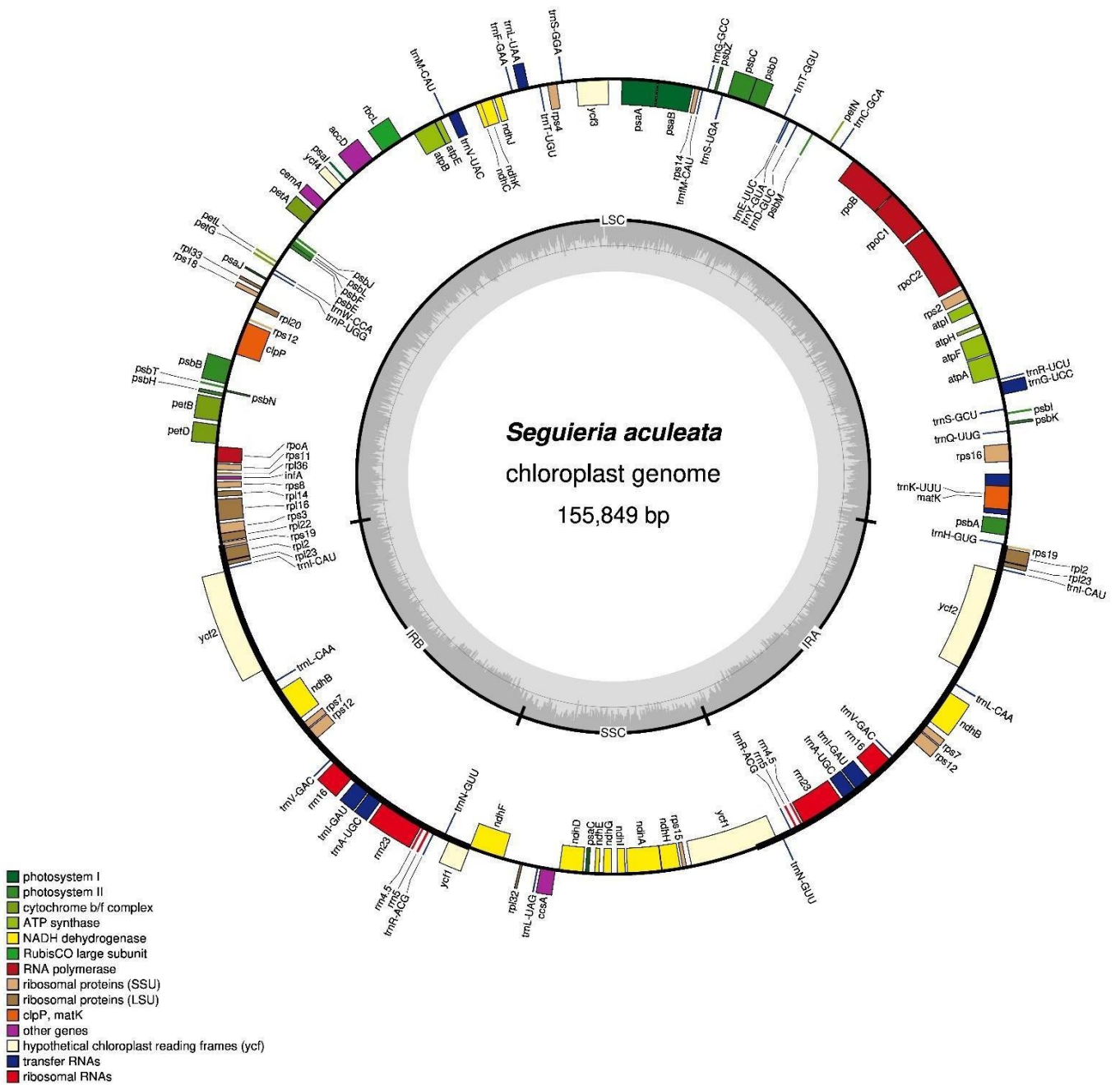


Figura 53. Estructura cuatripartita de *Segueria aculeata* (Generada).

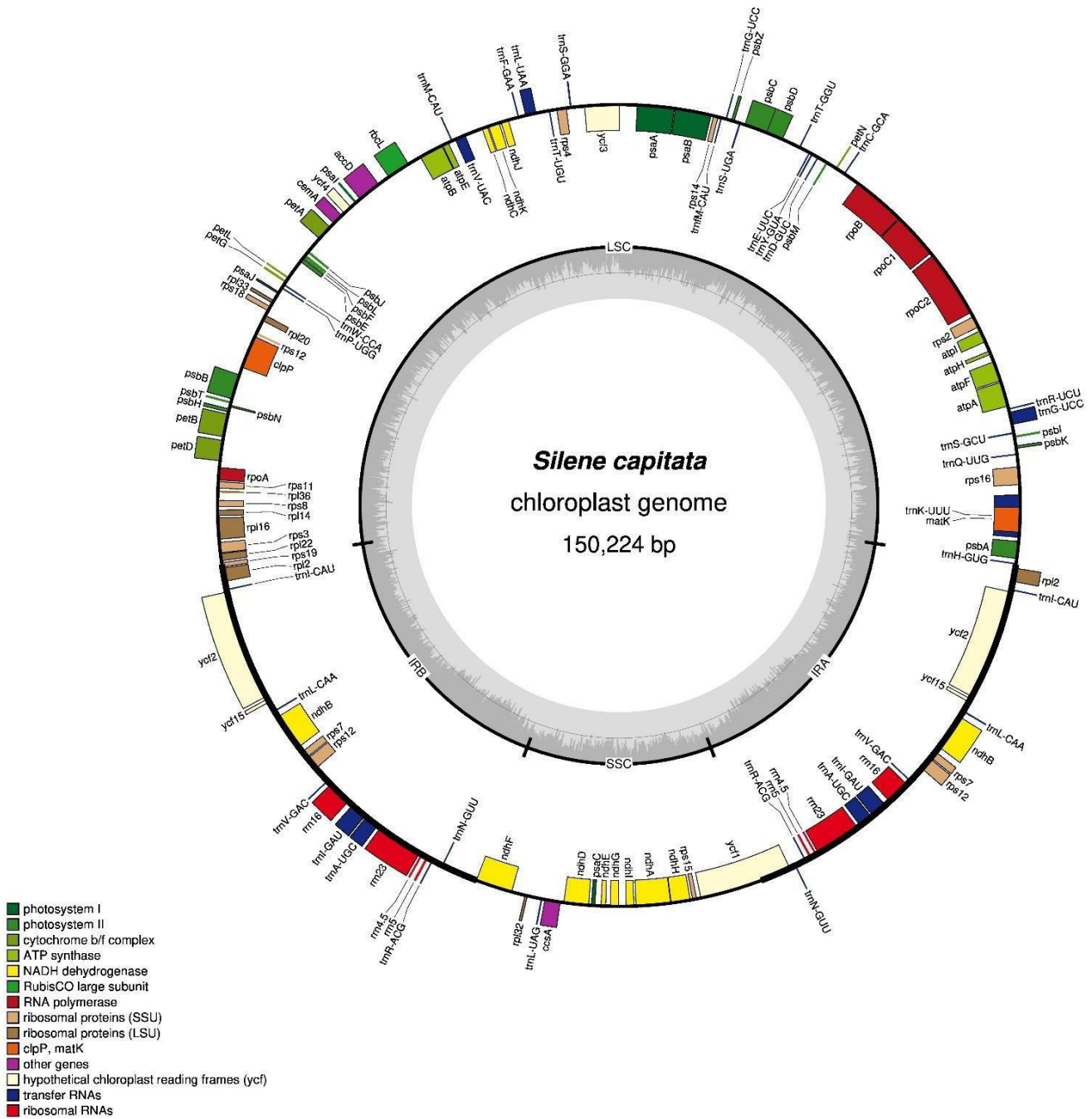


Figura 54. Estructura cuadripartita de *Silene capitata* (Generada).

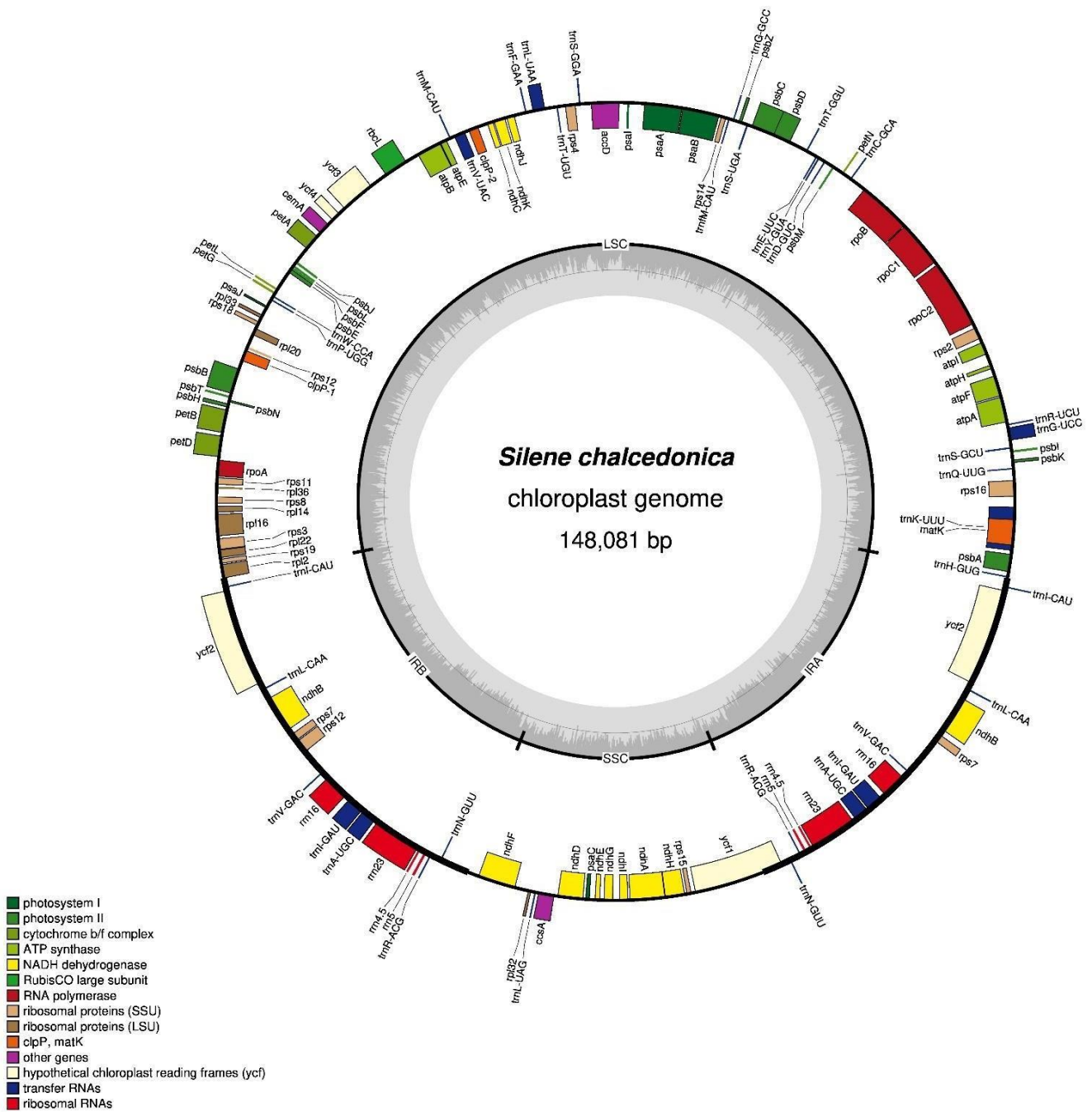


Figura 55. Estructura cuatripartita de *Silene chalcedonica* (Generada).

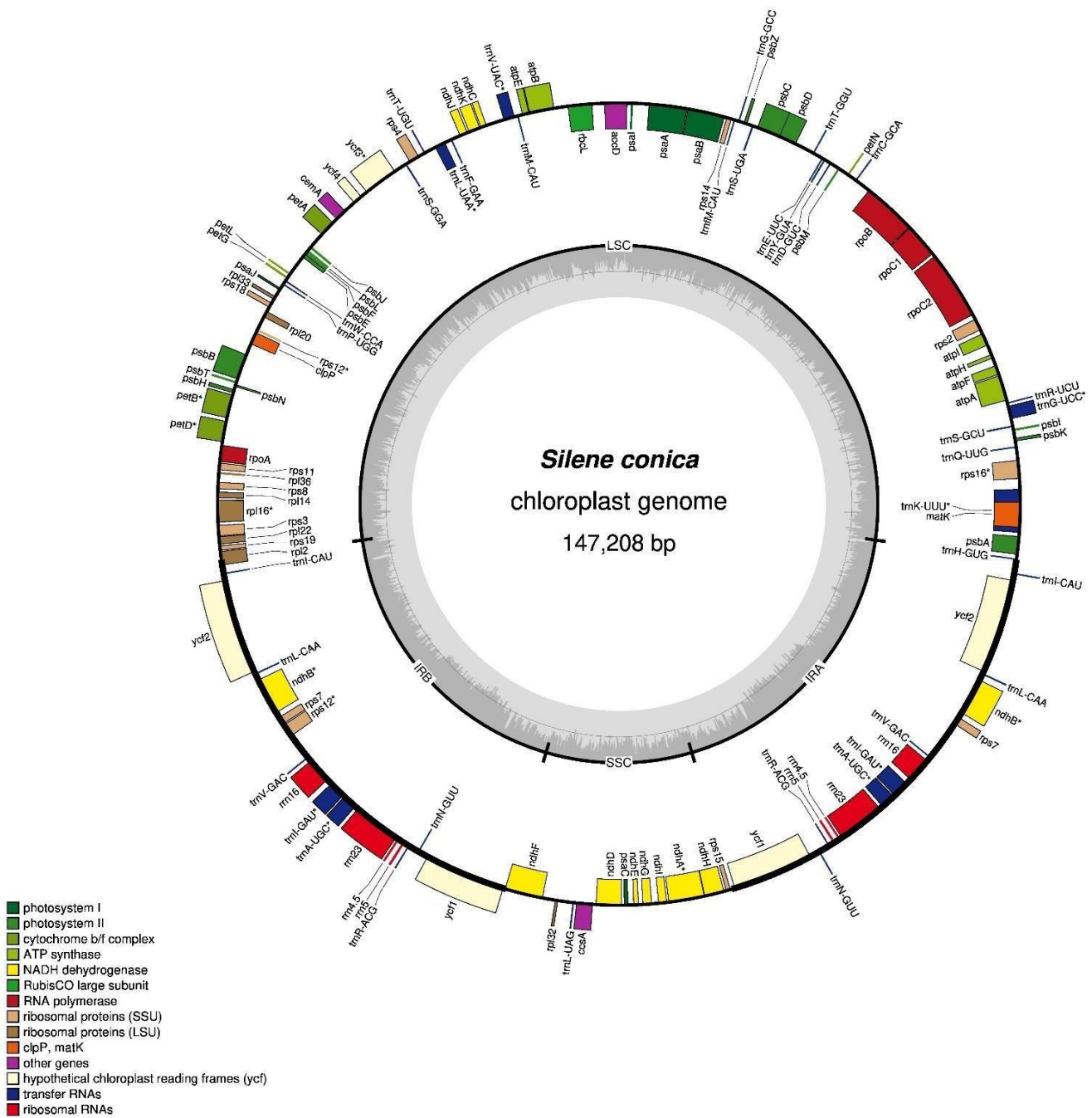


Figura 56. Estructura cuatripartita de *Silene conica* (Generada).

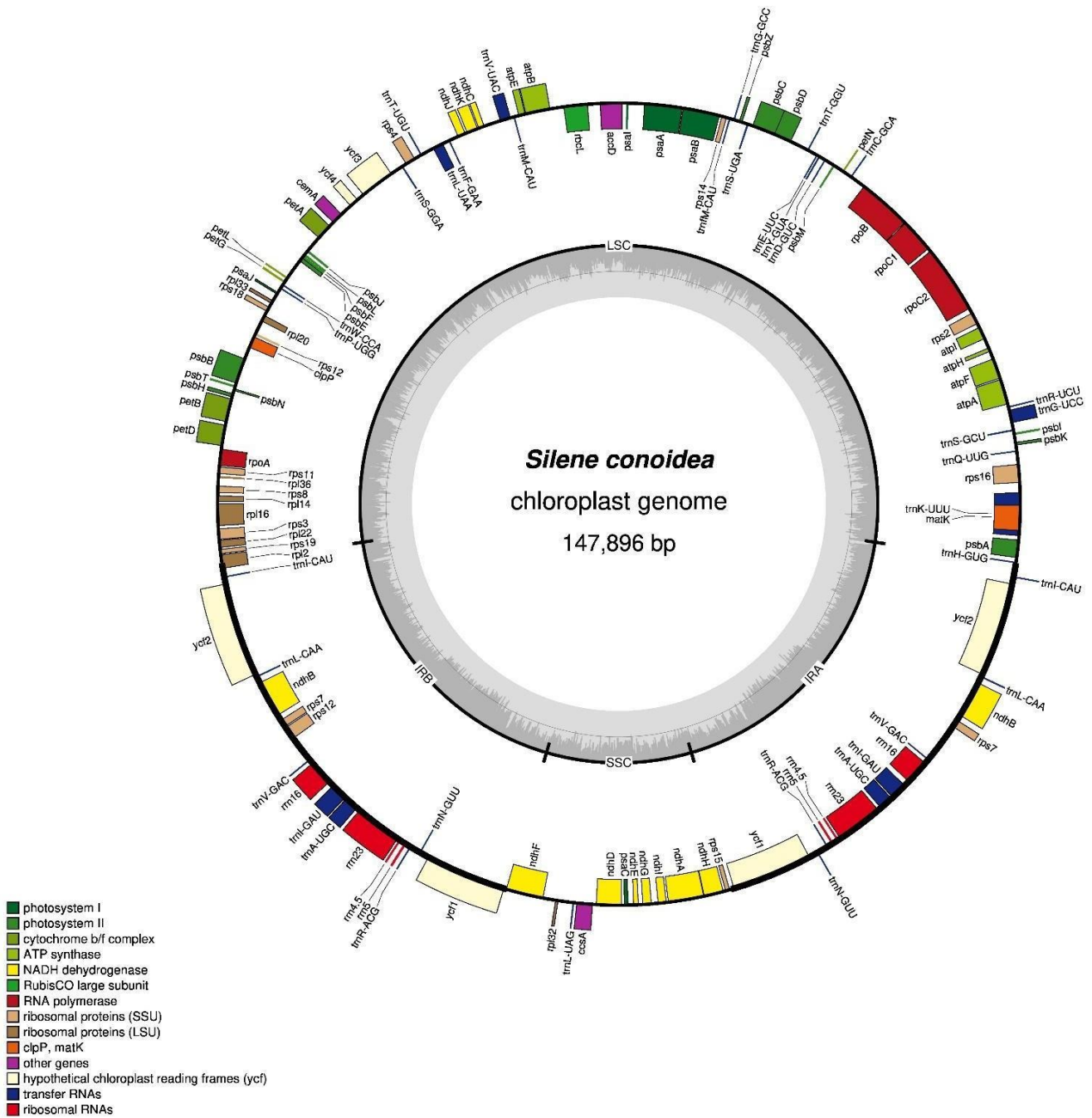


Figura 57. Estructura cuatripartita de *Silene conoidea* (Generada).

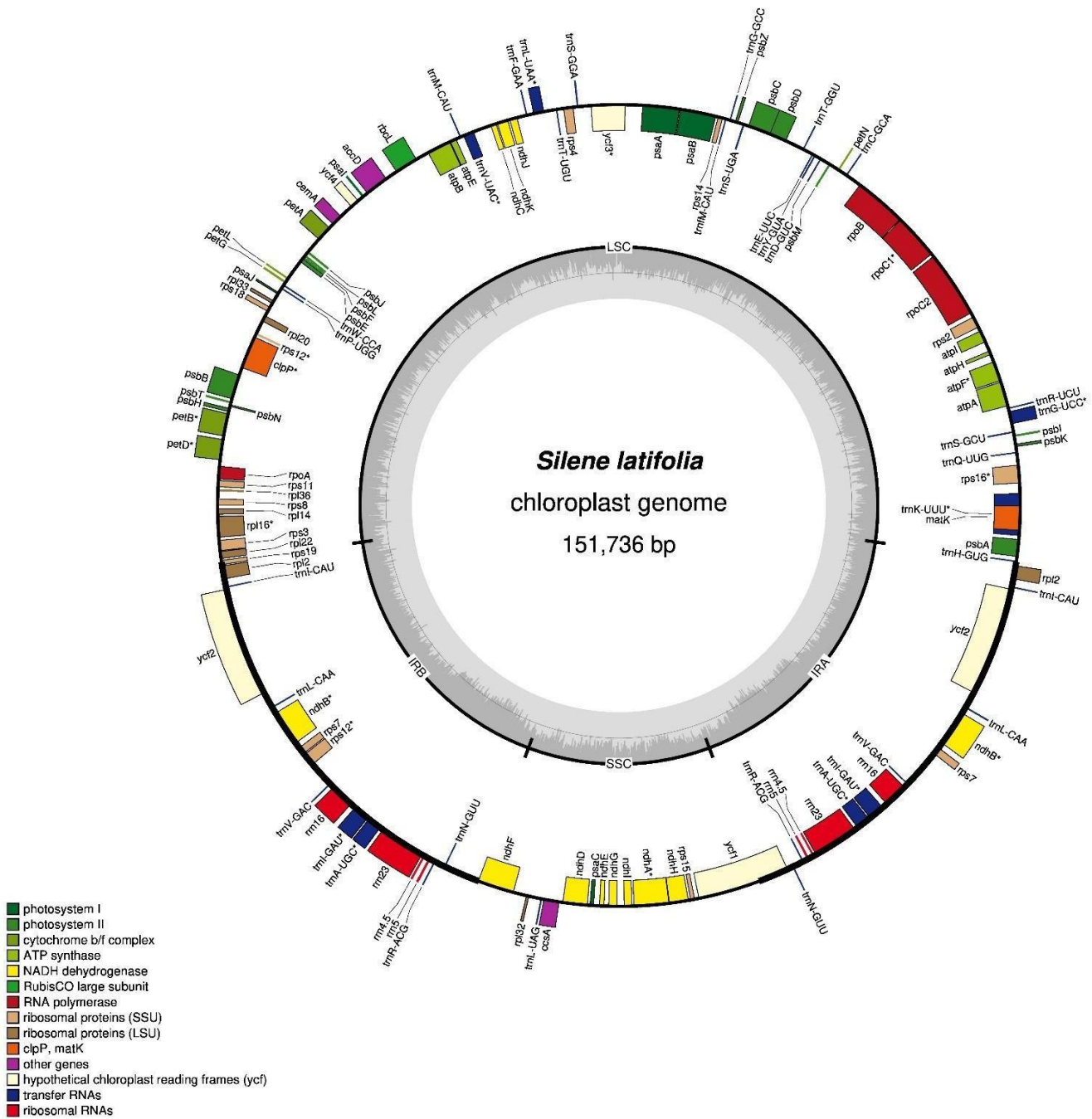


Figura 58. Estructura cuadripartita de *Silene latifolia* (Generada).



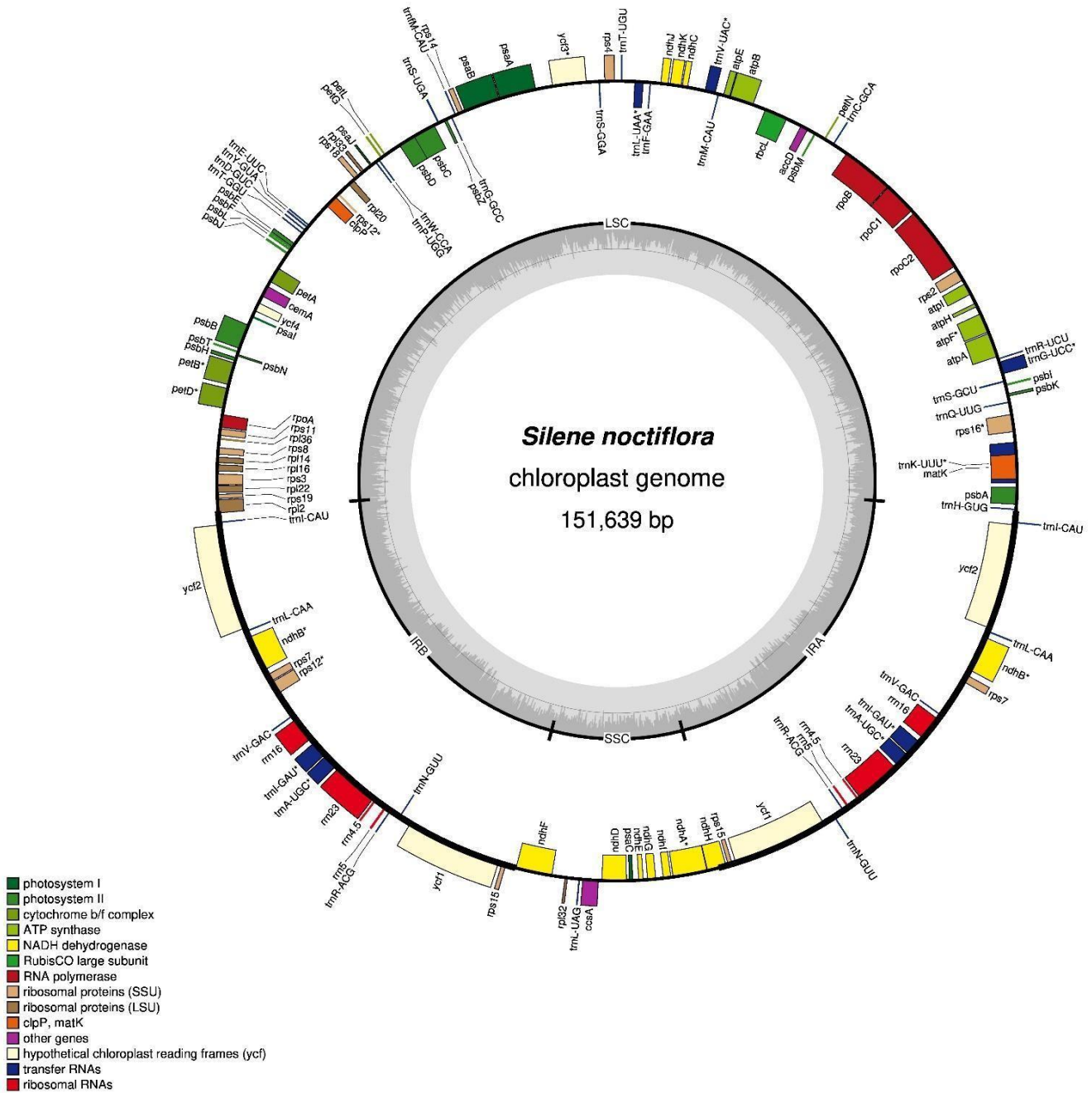


Figura 59. Estructura cuadripartita de *Silene noctiflora* (Generada).

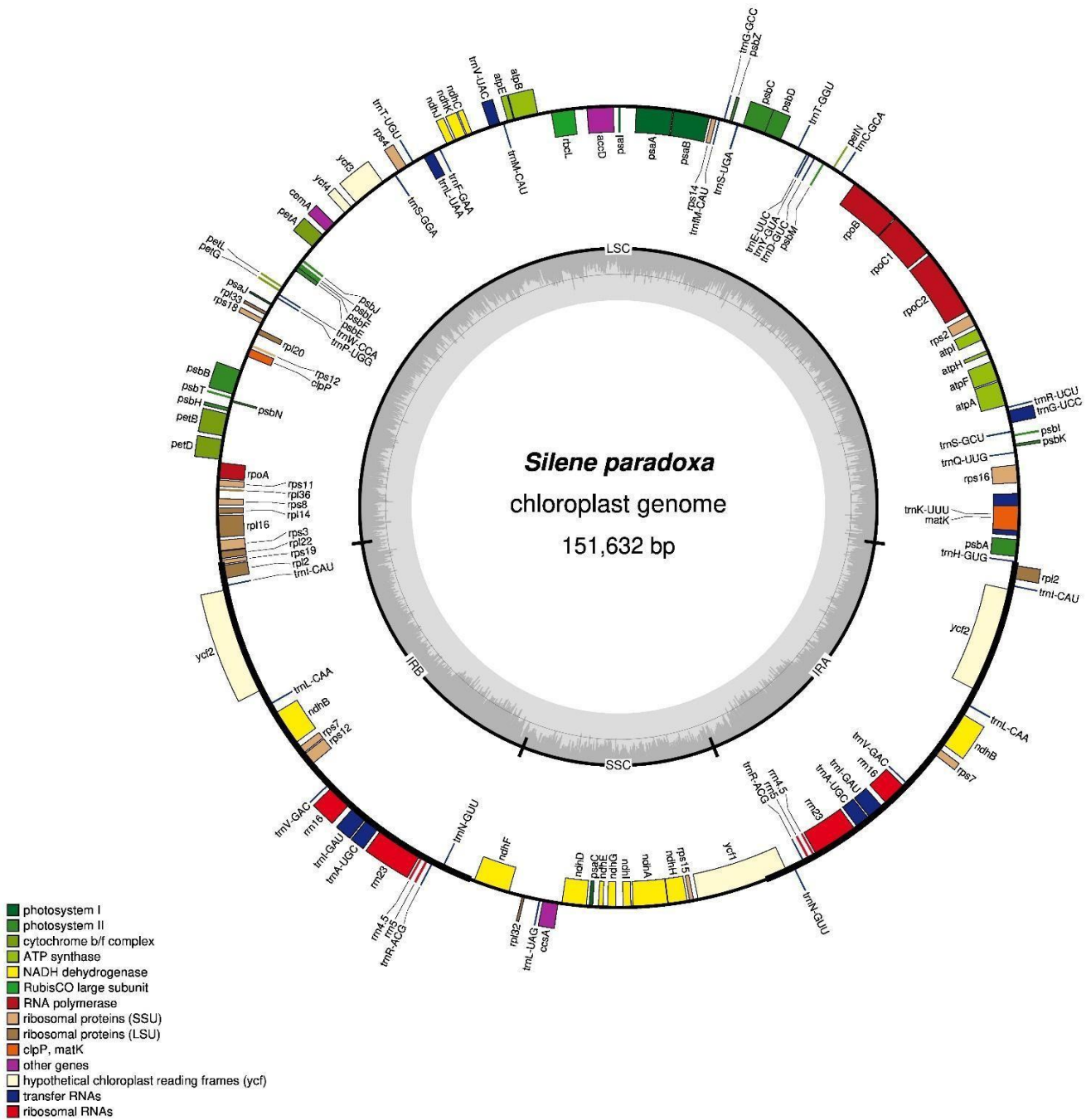


Figura 60. Estructura cuatripartita de *Silene paradoxa* (Generada).



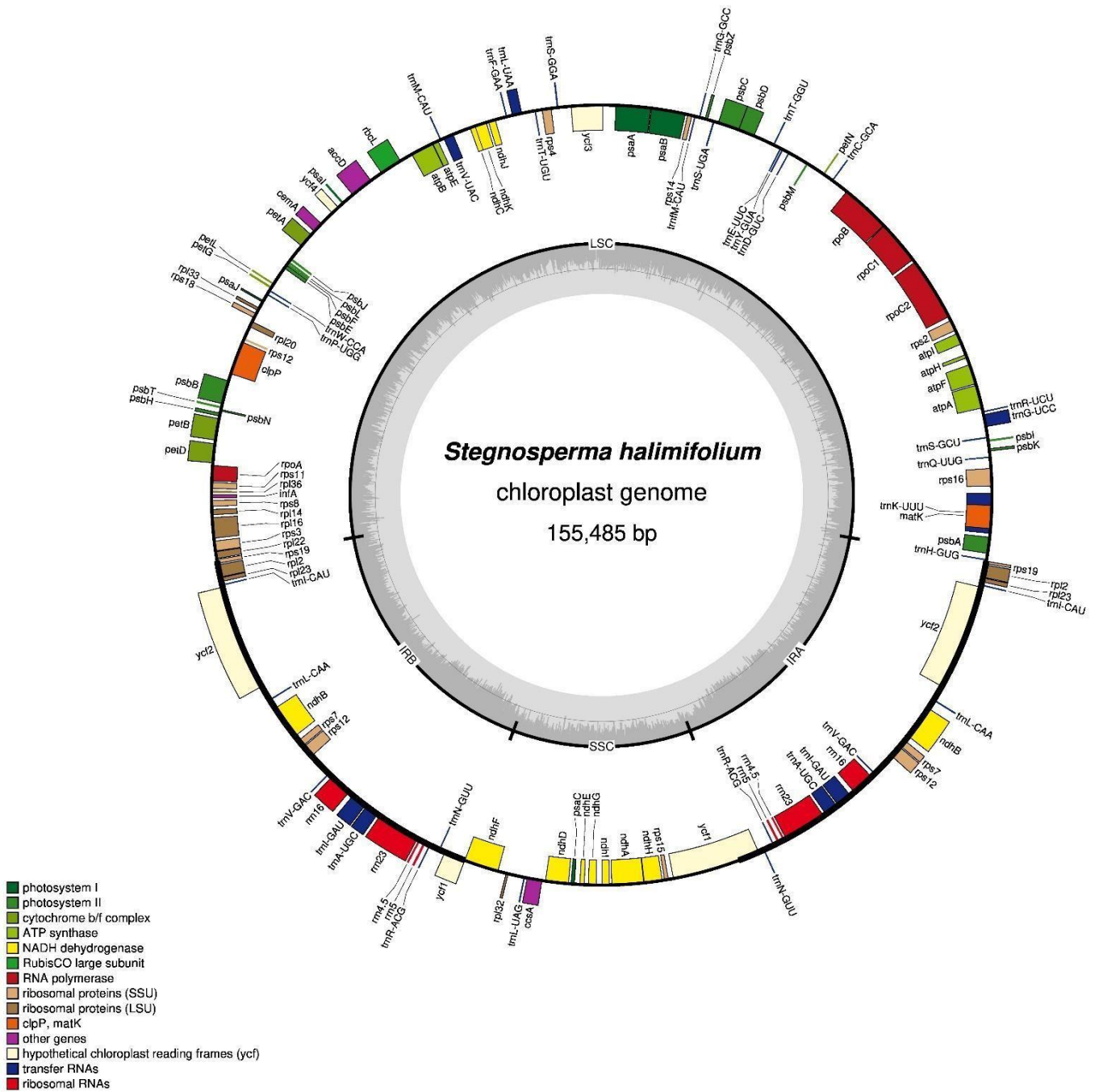


Figura 62. Estructura cuatripartita de *Stegnosperma halimifolium* (Generada)..

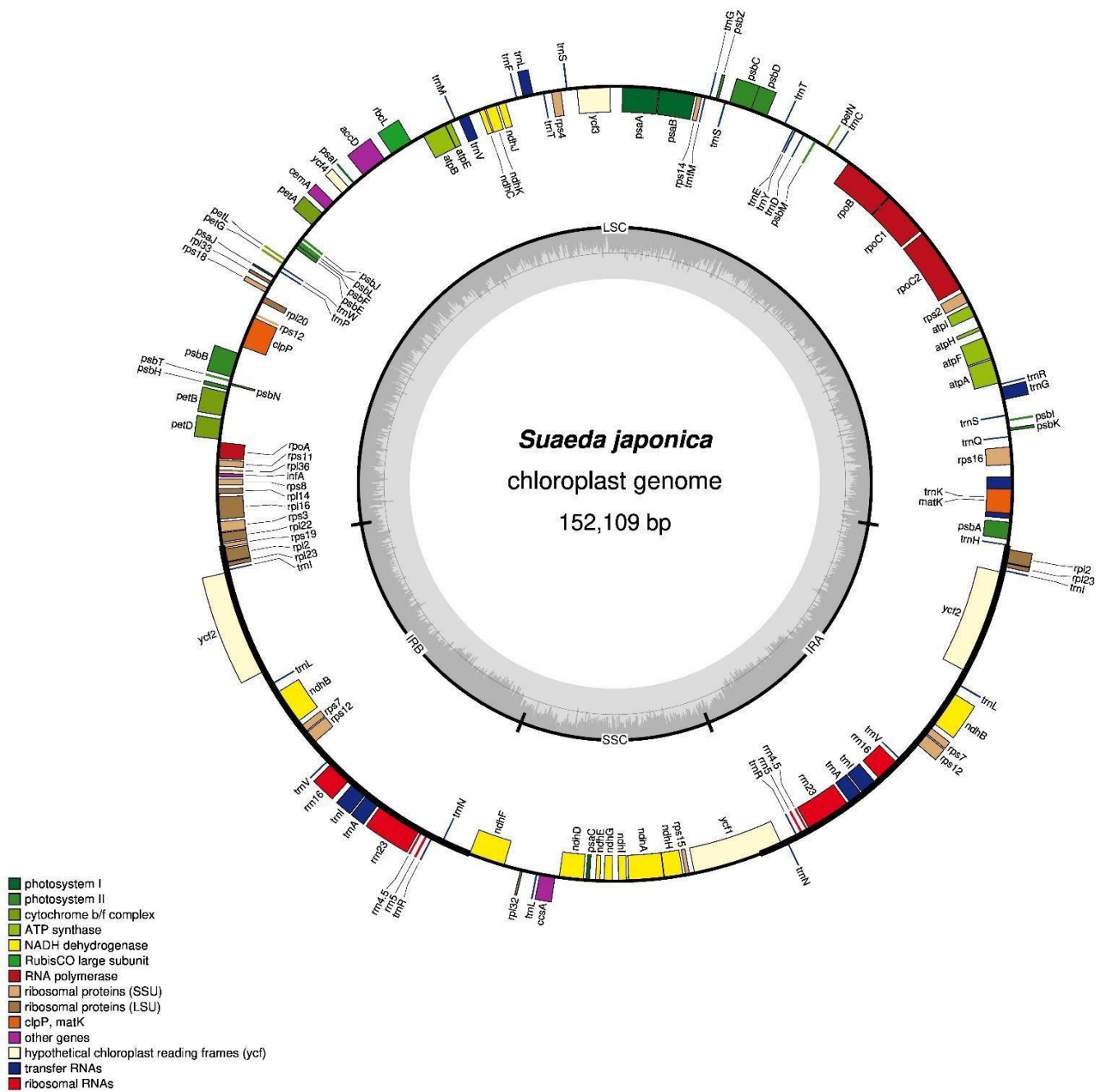


Figura 63. Estructura cuatripartita de *Suaeda japonica* (Generada).

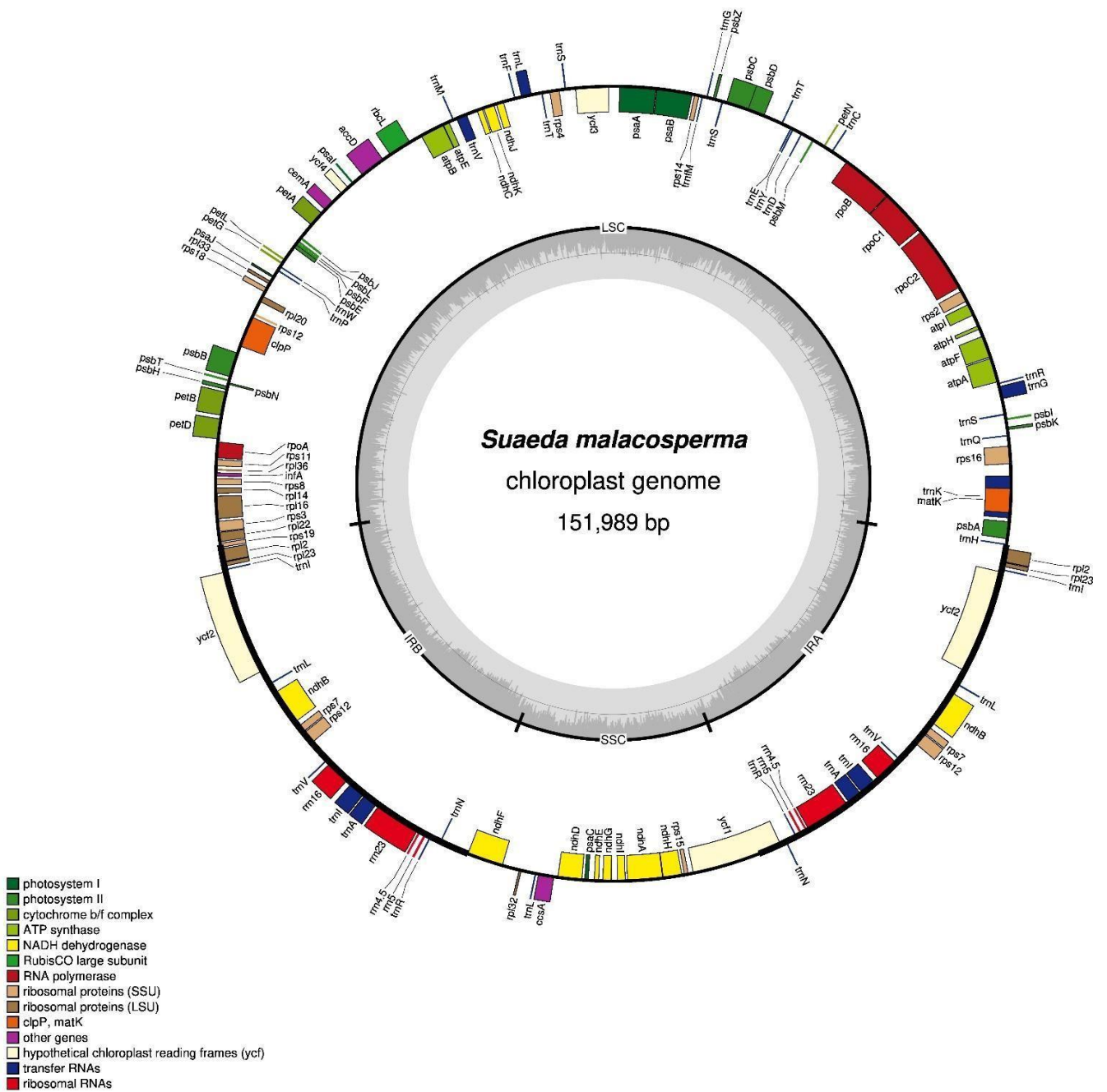


Figura 64. Estructura cuatripartita de *Suaeda malacosperma* (Generada).

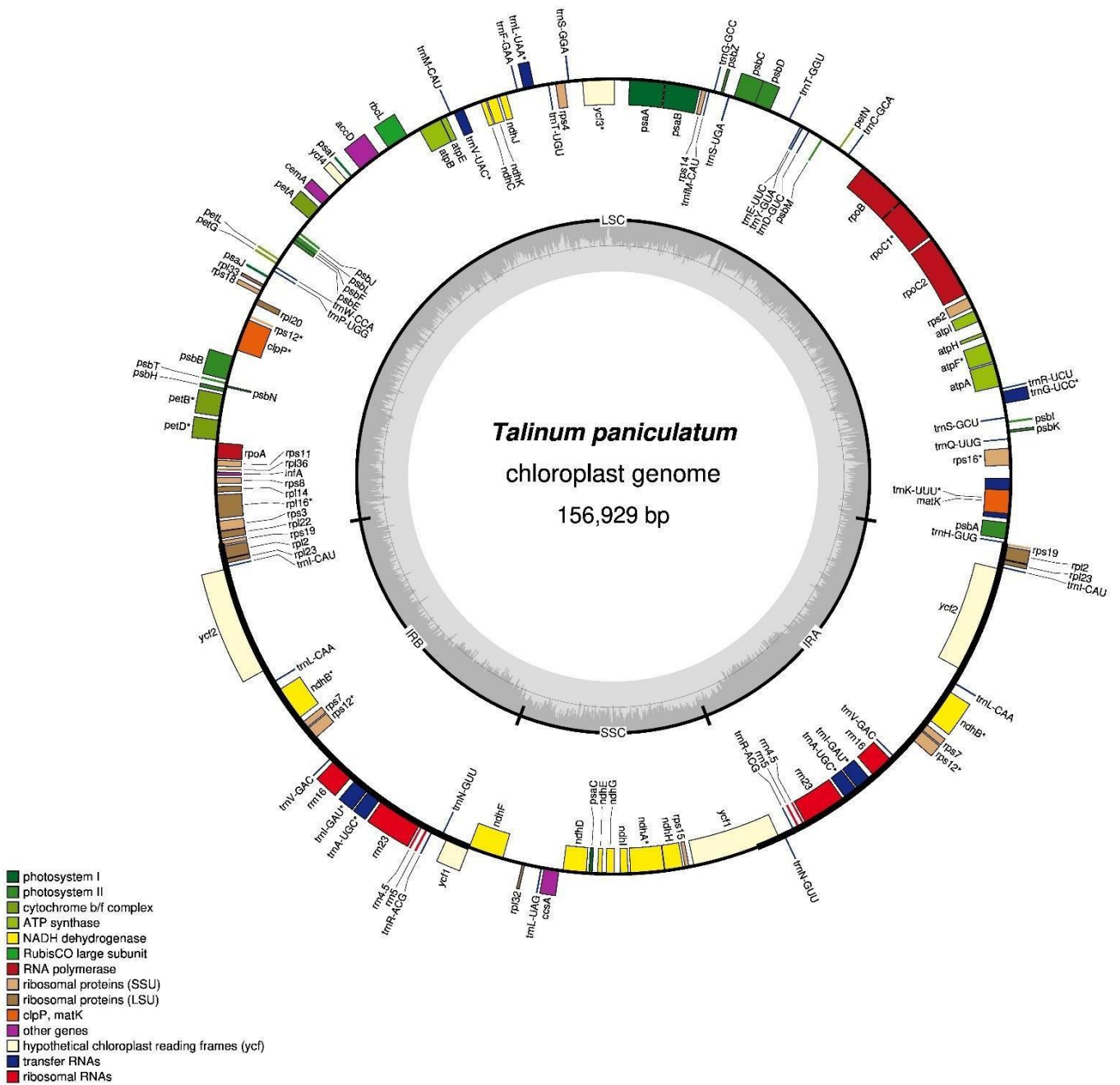


Figura 65. Estructura cuadripartita de *Talinum paniculatum* (Generada).

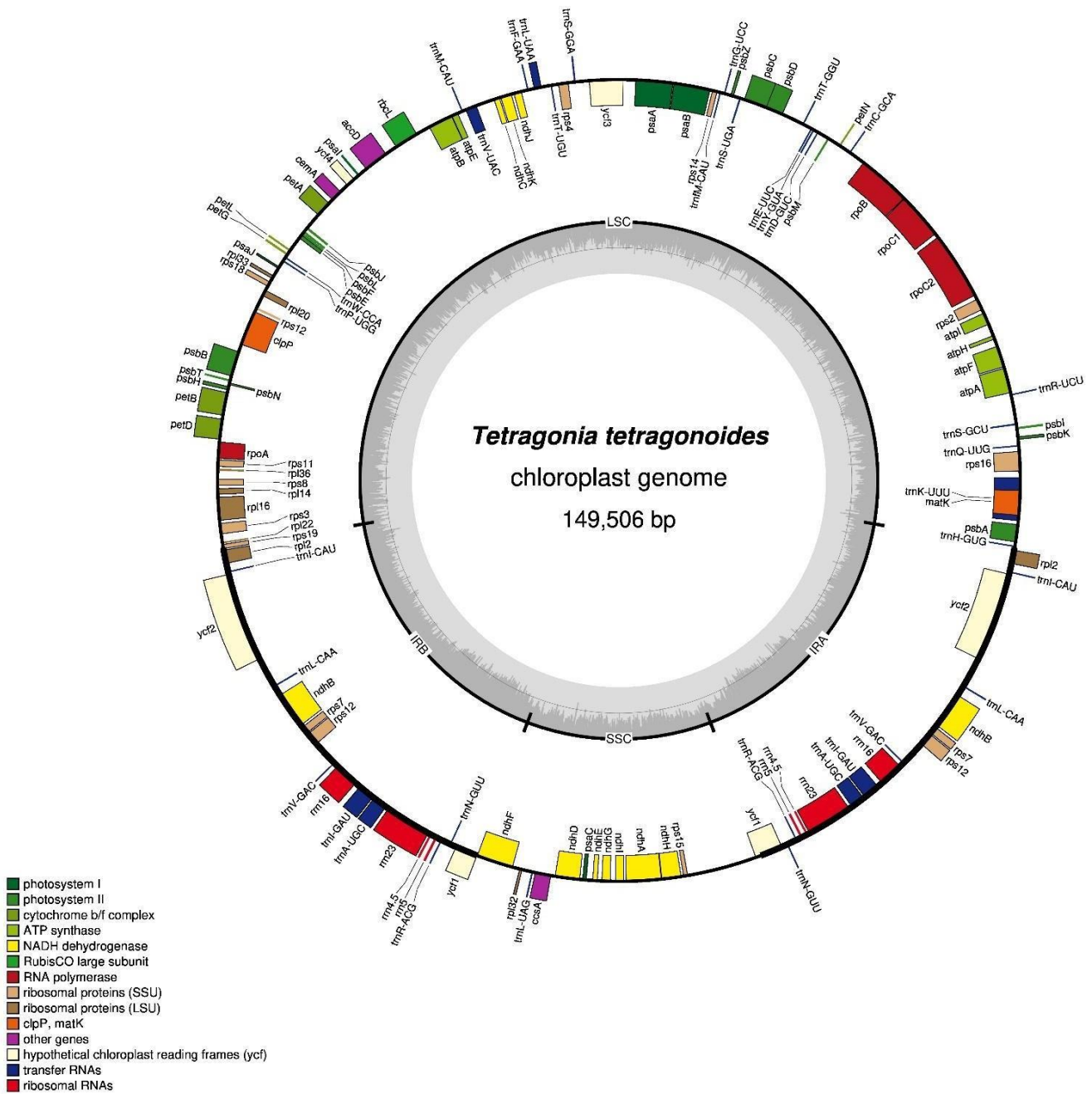


Figura 66. Estructura cuatripartita de *Tetragonia tetragonoides* (Generada).



## Anexo 2. Alineamientos estructurales generados a partir de MAUVE 2.4.0.

