

UNIVERSIDAD DE COSTA RICA
SISTEMA DE ESTUDIOS DE POSGRADO

EL GENOMA PLASTIDIAL DE LA PLANTA HOLOPARÁSITA
PROSOPANCHE COSTARICENSIS: REDUCCIÓN DEL PLASTOMA EN LA
FAMILIA HYDNORACEAE

Tesis sometida a la consideración de la Comisión del Programa de Estudios de
Posgrado en Biología para optar por el grado y título de Maestría Académica en
Biología con énfasis en Genética y Biología Molecular

MARIEL PAOLA OBANDO CORONADO

Ciudad Universitaria Rodrigo Facio, Costa Rica

2024

Dedicatoria

A mis papás, hermano y esposo
que estuvieron todo el proceso a
mi lado y nunca dudaron de mí.

A mi perrita Cocó que a su forma
me apoyó todo el tiempo.

A todos esos amigos que
estuvieron presentes y siempre
anuentes a brindarme una mano.

Agradecimientos

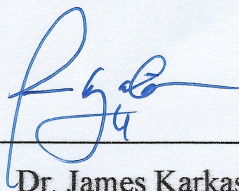
Le agradezco profundamente a toda mi familia, mis padres, hermano y esposo por estar siempre presentes y al tanto de los avances de mi proyecto. Por su anuencia a escucharme cada que necesité un oído y por siempre estar dispuestos a brindarme una mano. Les agradezco por valorar el esfuerzo que me tomó llegar hasta aquí y por brindarme todas las herramientas que lo permitieron.

Le agradezco a mi director de tesis, Federico Albertazzi, por el tiempo invertido, la paciencia y su dedicación en ayudar a formarme como científica y por nunca darse por vencido conmigo. Gracias por todo el apoyo que me brindó desde el primer día y al mismo tiempo por demostrarme que si se podía aunque fuera difícil. Le agradezco a mis lectores de tesis Diego Bogarín y Rafael Acuña por haber aceptado brindarme su apoyo en este proceso, por brindarme herramientas en el camino que me permitieron realizar un mejor trabajo en estos años de aprendizaje. Adicionalmente, le agradezco a Ivannia Atmella por toda la ayuda brindada en el proceso de secuenciación, su incansable ayuda me permitió finalizar la investigación satisfactoriamente.

También agradezco a las personas que contribuyeron de una u otra forma durante el desarrollo de esta investigación, especialmente a todos mis amigos que me ayudaron con información en el proceso de búsqueda de la planta de estudio y aquellos que leyeron cada parte de este documento.

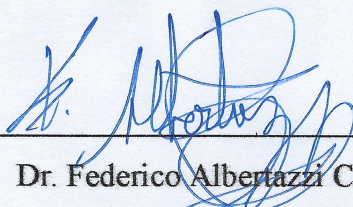
Finalmente, le agradezco al Sistema de Estudios de Posgrado por el apoyo económico brindado que me permitió llevar a cabo los experimentos necesarios para finalizar la investigación.

“Esta tesis fue aceptada por la Comisión del Programa de Estudios de Posgrado en Biología de la Universidad de Costa Rica, como requisito parcial para optar al grado y título de Maestría Académica en Biología con énfasis en Genética y Biología Molecular”



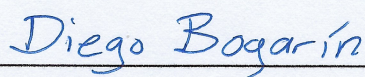
Dr. James Karkashian Córdoba

Representante del Decano Sistema de Estudios de Posgrado



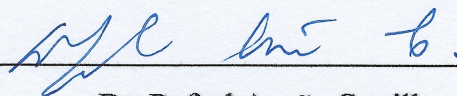
Dr. Federico Albertazzi Castro

Director de Tesis



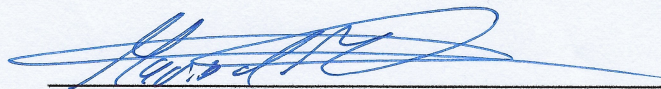
Dr. Diego Bogarín Chaves

Asesor



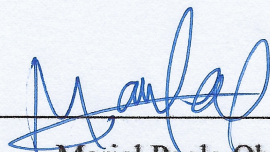
Dr. Rafael Acuña Castillo

Asesor



Dr. Mario Blanco Coto

Representante Programa de Posgrado en Biología



Mariel Paola Obando Coronado

Candidata

Tabla de Contenido

DEDICATORIA	II
AGRADECIMIENTOS	III
RESUMEN	VII
LISTA DE FIGURAS	VIII
Capítulo 1	VIII
LISTA DE CUADROS	IX
Capítulo 1	IX
Capítulo 1. El plastoma de <i>Prosopanche costaricensis</i> L.D Gómez.: El síndrome de la reducción parasitaria en la familia Hydnoraceae.....	1
INTRODUCCIÓN.....	1
OBJETIVOS	8
MATERIALES Y MÉTODOS.....	10
Material vegetal y extracción de ADN	10
Secuenciación masiva y ensamblado del plastoma	11
Diseño de imprimadores para los gaps del ensamblaje	13
Amplificación y Secuenciación de Gaps	16
Ensamblaje del plastoma de Novo por GetOrganelle.....	23
Anotación del Plastoma	24
Comparación de Plastomas y Análisis de Selección	25
Análisis filogenético	26
RESULTADOS	30
ADN para Secuenciaciones Masivas y Sanger.....	30
Calidad de las Lecturas de la Secuenciación Masiva	30
Genoma ensamblado de la secuenciación masiva	31
Amplificaciones y Secuenciaciones de los Gaps.....	31
Ensamblaje por GetOrganelle.....	38
Comparación de Plastomas y Análisis de Selección	46
Análisis filogenético	58
DISCUSIÓN	60

Secuenciación Masiva, Ensamblaje y Secuenciación de Gaps.....	60
Características de los plastomas de Hydnoraceae enfocado en <i>Prosopanche costaricensis</i>	64
Patrones de evolución	69
Selección sobre genes	73
Arboles filogenéticos	74
Nueva evidencia para considerar <i>Prosopanche americana</i> y <i>Prosopanche costaricensis</i> como dos especies distintas	75
CONCLUSIONES	78
RECOMENDACIONES	79
APÉNDICE 1.1 Primer ensamblaje a partir de la secuenciación masiva de <i>Prosopanche costaricensis</i> (30 192 pb).....	80
APÉNDICE 1.2 Secuencia intermedia de <i>Prosopanche costaricensis</i> con la inclusión de gap secuenciados (28489 pb).	89
APÉNDICE 1.3 Anotación final del plastoma de <i>Prosopanche costaricensis</i>	99
REFERENCIAS	114

RESUMEN

Desde el descubrimiento del síndrome de reducción parasitaria en los plastomas de plantas parásitas, los esfuerzos de investigación se han enfocado en describir cómo ocurre esta pérdida durante la transición hacia la vida heterótrofa. La familia Hydnoraceae, es uno de los 12 linajes de angiospermas holoparásitos que se han investigado para generar modelos de reducción. Los miembros de este linaje han sufrido una increíble reducción morfológica y es el segundo linaje parásito más antiguo, con un parasitismo que evolucionó hace aproximadamente ~ 55 Ma. Hydnoraceae contiene con dos géneros, *Hydnora* con ocho especies aceptadas en África y *Prosopanche* con seis especies aceptadas en América del Sur y una en América Central. De estas especies, se han ensamblado los plastomas de siete especies de *Hydnora* y tres de *Prosopanche*, y en esta investigación se ensambló el plastoma completo, de la única especie conocida de esta familia originario de América central, *Prosopanche costaricensis*. El ensamblaje del plastoma resultó ser un proceso difícil, lo que demuestra grandes desafíos asociados a la secuenciación y ensamblaje de plastomas inusuales, debido a sus características particulares de reducción de genoma. Así, el ensamblaje de estos plastomas requiere el uso de múltiples técnicas bioinformáticas y moleculares para ser exitoso. La obtención del plastoma de esta especie aportó evidencia crucial para apoyar el modelo de reducción plastidial desarrollado en el linaje Hydnoraceae. La estructura del plastoma de *P. costaricensis* es consistente con las de otras especies de la familia y coincidió con las características género. Este presentó la pérdida total de genes fotosintéticos y la retención de solo 27 genes esenciales, un bajo contenido GC (19.9%), una alta cantidad de sustituciones de nucleótidos en sus genes, y un tamaño que ronda los 29 mil pares de bases (28 705 pb). Además, presenta repeticiones directas (RD) contraídas (330 pb) y genes pseudogenizados. Las RD contraídas, las fuerzas de selección reducidas en los genes, la falta de genes fotosintéticos, la retención de genes esenciales, el tamaño disminuido y la presencia de genes pseudogenizados (*ycf2* y *trnI-AAU*), corroboran resultados similares en genomas tanto de la misma familia como de otras plantas holoparásitas. Tras perder todos los genes fotosintéticos, estas especies han empezado a perder genes esenciales mediante procesos de degeneración y conversión en pseudogenes. Esto respalda la hipótesis de que los plastomas de esta familia se encuentran probablemente en el final de la segunda etapa o el inicio de la tercera del modelo de reducción planteado. Además, apoya la existencia de repeticiones directas reducidas, presentando dos estructuras diferentes en el género, una con 3 estructuras codificantes (5' *ycf2*, *trnI*, y 5' *rpl2*) y otra con solo dos (*trnI*, y 5' *rpl2*). Adicionalmente, el genoma plastidial de *P. costaricensis* proporciona nueva evidencia para confirmar que esta y *P. americana* son especies diferentes. Finalmente, los análisis filogenéticos revelaron que *P. costaricensis* y *P. panguanensis* son especies hermanas lo que concuerda con el hecho de que ambas son las únicas dos especies del género que habitan en bosques húmedos, además refuerza el hecho de que *Prosopanche* y *Hydnora* son géneros hermanos y por último posiciona a la familia como hermana del clado de los Piperales.

Lista de figuras

Capítulo 1.

Figura 1. Fragmentos amplificaciones mediante la metodología de la prueba 1.....	32
Figura 2. Fragmentos amplificados y purificados con la metodología de la prueba 2.	33
Figura 3. Fragmentos con bandas múltiples obtenidos de la amplificación del plastoma de <i>P. costaricensis</i> con el protocolo de la prueba 3.	34
Figura 4. Fragmentos con bandas múltiples obtenidos de la amplificación del plastoma de <i>P. costaricensis</i> con el protocolo de la prueba 4.	35
Figura 5. Scaffolds obtenidos del ensamblaje de Novo del plastoma de <i>Prosopanche costaricensis</i> por GetOrganelle.	39
Figura 6. Mapeo por referencia de los scaffolds de <i>Prosopanche costaricensis</i> contra el plastoma <i>P. panguanensis</i> y la secuencia consenso circularizada.....	40
Figura 7. Mapa circularizado del genoma plastidial parcial de <i>Prosopanche costaricensis</i>	41
Figura 8. Alineamientos de los imprimadores diseñados para el primer ensamblaje que no se alinean correctamente en el ensamblaje final.....	42
Figura 9. ARNs de transferencia de triptófano e isoleucina2 de la repetición A de <i>Prosopanche costaricensis</i>	45
Figura 10. Comparación de los plastomas de <i>Prosopanche americana</i> , <i>P. panguanensis</i> , <i>P. bonacinae</i> y <i>P. costaricensis</i>	49
Figura 11. Comparación del plastoma de <i>Hydnora visseri</i> y <i>Prosopanche costaricensis</i>	50
Figura 12. Comparación de los plastomas de <i>Pilostyles aethiopica</i> , <i>Pilostyles hamiltonii</i> y <i>Prosopanche costaricensis</i>	51
Figura 13. Comparación de los plastomas de <i>Balanophora reflexa</i> , <i>Balanophora laxiflora</i> y <i>Prosopanche costaricensis</i>	52
Figura 14. Comparación de los plastomas de <i>Cytinus hypocistis</i> , <i>Mitrastemon kanehirai</i> y <i>Prosopanche costaricensis</i>	54
Figura 15. Comparación de los plastomas de <i>Cuscuta erosa</i> , <i>Conopholis americana</i> y <i>Prosopanche costaricensis</i>	55
Figura 16. Comparación de los plastomas lineares de <i>Aristolochia contorta</i> y <i>Prosopanche costaricensis</i>	56

Figura 17. Análisis de mutaciones sinónimas y no sinónimas de los genes codificantes de proteínas de todas las especies de <i>Prosopanche</i> con plastomas disponibles, utilizando a <i>Hydnora visseri</i> como grupo externo.	58
Figura 18. Reconstrucción filogenética por máxima verosimilitud con valores de bootstrap de máxima verosimilitud basada en una matriz concatenada de 26 genes (los 26 genes retenidos en <i>Prosopanche costaricensis</i>) de todas los taxones disponibles en GenBank de Hydnoraceae así como de los plastomas de las demás angiospermas.	59

Lista de cuadros

Capítulo 1.

Cuadro 1. Principales estudios que recopilan los plastomas disponibles para los 12 linajes parasíticos haustoriales.	4
Cuadro 2. Información de los imprimadores para secuenciar los gaps encontrados en el ensamblaje.	14
Cuadro 3. Protocolos de las primeras pruebas para la amplificación de los gaps.	17
Cuadro 4. Reacción de secuenciación de la primera prueba	18
Cuadro 5. Segunda prueba para la amplificación de los Gaps.	19
Cuadro 6. Reacción de secuenciación de la segunda prueba.....	21
Cuadro 7. Reacción de amplificación para PCR de gaps.	21
Cuadro 8. Taxones muestreados para análisis filogenéticos de este estudio con su respectivo orden y número de accesión, que se incluyen en los análisis filogenéticos.	27
Cuadro 9. Fragmentos obtenidos de las pruebas de amplificación y secuenciación para la determinación de los nucleótidos indeterminados del ensamblaje por referencia de <i>P. costaricensis</i>	36
Cuadro 10. Características de los plastomas altamente reducidos de holoparásitas estudiadas y de la especie fotosintética más cercana que fue analizada en detalle.	47
Cuadro 11. Genes plastidiales en <i>Prosopanche</i> con sus respectivos tamaños y orientación.	47

Capítulo 1. El plastoma de *Prosopanche costaricensis* L.D Gómez.: El síndrome de la reducción parasitaria en la familia Hydnoraceae.

INTRODUCCIÓN

A través de su historia evolutiva los genomas plastidiales o plastomas de plantas parásitas han experimentado reducciones funcionales y físicas con cambios en el contenido de genes, estructura (perdida o contracción de las repeticiones invertidas o directas) y tamaño (Krause, 2008; Yoshida et al., 2016). En contraste, los plastomas de plantas fotosintéticas se han conservado relativamente estables manteniendo una gran complejidad (Lockhart, 2013). Esta estabilidad y complejidad se debe a que la fotosíntesis genera una fuerte presión selectiva fuerte que favorece la conservación de las secuencias y funcionalidades de los genomas plastidiales. Por ende, al perder la capacidad autotrófica se reduce esta presión selectiva, lo que permite que los plastomas experimenten reducciones en sus secuencias, principalmente por la pérdida de la función de los genes, aunque esto generalmente no implica la pérdida completa del genoma plastidial (Krause, 2008; Yoshida et al., 2016), debido a que son organelas donde se producen las moléculas precursoras de hormonas, lípidos y otros metabolitos.

Esta pérdida de la capacidad fotosintética puede resultar sorprendente, debido a las múltiples ventajas evolutivas que ofrece la fotosíntesis (Sibbald & Archibald, 2020). No obstante, la vida heterótrofa también presenta ventajas sobre la vida autótrofa. Estas ventajas incluyen la capacidad de colonizar hábitats no disponibles para algunas plantas fotosintéticas (Těšitel, 2016) permitiéndole a las especies superar la competencia por recursos (Lyko & Wicke, 2021). Este estrecho vínculo entre parásito y hospedero facilita el acceso a nutrientes, sustancias complejas y minerales sin el coste energético que demanda sintetizar sustancias o sistemas para la absorción de minerales específicos (Muche et al., 2022). Por otro lado, la transferencia de genes también puede llegar a ser ventajosa. Debido a la naturaleza azarosa de estos eventos, el parásito tiene la posibilidad de adquirir genes de su hospedero que pueden incluso llegar a proporcionarle una mayor adaptabilidad física para mantener su asociación con el hospedero (Molina et al., 2014; Schneider, Chun, et al., 2018; Xi et al., 2012). No obstante, esta no es una ventaja *per se* de las

plantas heterótrofas, sino una característica del parasitismo que puede ocurrir con una alta probabilidad por lo íntimas que son las interacciones parásito-hospedero (Davis & Xi, 2015). A la vez, las especies parásitas presentan funciones ecológicas importantes para los ecosistemas que ocupan, algunas de ellas al infligir daños sobre sus hospederos varían las relaciones en las comunidades lo que puede aumentar la riqueza de especies no parásitas (Těšitel, 2016; Zhang et al., 2018). Adicionalmente, estas especies modifican la comunidad microbiana del suelo influyendo en el ciclo de los nutrientes y beneficiando a otros heterótrofos presentes en el ecosistema (Těšitel, 2016). Entonces, con las ventajas del parasitismo ya identificadas, las preguntas actuales se centran en determinar cómo se desarrollan los cambios en los genomas plastidiales de estas plantas parásitas (por qué se pierden unos genes y otros se conservan, en qué orden y a qué velocidad); así como entender por qué se conserva un genoma plastidial cuya función principal se ha perdido (Sibbald & Archibald, 2020), o qué partes del genoma plastidial se conservan ya que este aún puede estar funcionando como productor de moléculas precursoras de hormonas, lípidos y otros metabolitos.

Con respecto a las variaciones genéticas, los diversos estudios existentes se han enfocado en analizar y describir los cambios que ocurren en los plastomas de plantas parásitas en el proceso de transición a una vida heterótrofa. A este conjunto de variaciones se le ha denominado el “síndrome de reducción parasitaria” (Wicke & Naumann, 2018, Gruzdev, Kadnikov, Beletsky, Mardanov & Ravin, 2019). La investigación de este síndrome en la evolución plastidial inició con la secuenciación del plastoma de la holoparásita *Epifagus virginiana* (Barrett et al., 2014), especie perteneciente a la familia Orobanchaceae, una de las más estudiadas en este ámbito (Barbrook et al., 2006; Krause, 2008; Su et al., 2019). Sin embargo, para poder captar todo el espectro de variación, es necesario obtener mayor cantidad de plastomas de los diferentes linajes parásitos.

Los efectos del síndrome de reducción parasitaria están determinados por el nivel de capacidad heterótrofa (Wicke & Naumann, 2018). A menor capacidad fotosintética, mayor es la reducción observada en los plastomas. Es decir, el síndrome muestra variados niveles de afectación entre los linajes: las hemiparásitas, especies que fotosintetizan parcialmente, tienen efectos leves, mientras que las holoparásitas, que han perdido la capacidad de fotosintetizar,

presentan reducciones mayores (Krause, 2008, 2012; Li et al., 2013). La evolución independiente de cada linaje también modela el grado y la velocidad de la reducción del genoma plastidial. Factores adicionales como, el tamaño, función y la asociación física o transcripcional de cada gen con otros genes esenciales también determinan la pérdida o retención de cada gen (Wicke et al., 2016).

Hasta ahora, se han generado dos hipótesis que buscan explicar este fenómeno. La primera hipótesis se basa en la funcionalidad de los genes, por ejemplo, en los plastomas se encuentran codificados ARNs de transferencia (ARNt) esenciales para otros procesos metabólicos, aparte de la fotosíntesis, como la síntesis de tetrapirroles y proteínas, los cuales, al ser difícilmente reemplazables por ARNt citosólicos, obligan su retención en el plastoma. En la segunda hipótesis, la retención puede ocurrir por una ventana limitada de transferencia genética entre el genoma plastidial y el genoma nuclear, lo que evita la salida rápida de los genes del plastoma (Barbrook et al., 2006; Molina et al., 2014). Por ejemplo, investigaciones han evidenciado que en algunas familias parásitas (Balanophoraceae) los genomas nucleares han tendido hacia un tamaño mayor, pero no se ha probado que la causa de esto sea el estilo de vida heterótrofo (Lyko & Wicke, 2021).

No obstante, existen excepciones notables. Un caso destacado es el de la holoparásita *Rafflesia lagascae*, ya que presenta la única posible pérdida total del plastoma en una planta (Molina et al., 2014; Schneider et al., 2018). Se considera como pérdida del plastoma cuando los genes se han degradado, debido a que la definición de plastoma implica la existencia de genes codificantes y regiones intergénicas (Lee et al., 2021). Para *R. lagascae*, se han planteado dos posibles escenarios. El primero propone que los genes esenciales para funciones metabólicas se lograron trasladar al genoma nuclear y/o mitocondrial, causando la pérdida del plastoma. El otro caso coincide en que la degradación de los genes del plastidio a pseudogenes podría explicarse por la traslocación de estos al núcleo; pero adicionalmente, propone que *R. lagascae* pudo haber adquirido inicialmente los plastidios de su hospedero (Molina et al., 2014). Sin embargo, estos son solo escenarios hipotéticos y no se han verificado definitivamente, por lo que este tema de investigación aún requiere de mayor cantidad de evidencia para establecer las causas de la pérdida total o la retención de los plastomas reducidos en plantas parásitas.

Actualmente, están disponibles los plastomas de especies pertenecientes a los 12 linajes parasíticos haustoriales que han evolucionado independientemente (cuadro 1) (Sanchez-Puerta et al., 2023; Westwood et al., 2010; Wicke & Naumann, 2018). A partir de estos linajes, se han desarrollado diversos modelos de reducción para explicar la degradación del plastoma asociada al estilo de vida heterótrofo. Wicke & Naumann, (2018) destacan tres modelos principales: El primer modelo, inicialmente propuesto por Barrett & Davis, (2012) describe la pérdida de genes de los plastomas por su funcionalidad y cambios físicos (i.e. pérdida de hojas). Este modelo fue refinado por Barrett et al., (2014) y Naumann et al., (2016), lo cual generó dos versiones. En la versión de Barrett et al., (2014) en la pérdida de genes de los plastomas se contemplan cuatro etapas de reducción donde se pierden específicos grupos de genes en cada etapa (modelo 1), mientras que la versión de Naumann et al., (2016) propone cuatro etapas donde los genes se pierden de acuerdo con su funcionalidad, pero incluye una etapa final con la pérdida total del plastoma (modelo 2). En contraste, el modelo propuesto por Wicke et al., (2016), describe la evolución de plantas parásitas hacia el estilo de vida heterótrofo como una red de características, donde se involucra la funcionalidad de los genes, los cambios físicos y los cambios genéticos evolutivos y moleculares, como las tasas de evolución y la presión selectiva (modelo 3).

Cuadro 1. Principales estudios que recopilan los plastomas disponibles para los 12 linajes parasíticos haustoriales.

Linajes	Referencias
Apodanthaceae	(Bellot & Renner, 2016) (Arias-Agudelo et al., 2019)
Convolvulaceae (<i>Cuscuta</i>)	(Funk et al., 2007) (Banerjee & Stefanović, 2019)
Cytinaceae	(Roquet et al., 2016)
Cynomoriaceae	(Bellot et al., 2016)
Cytinaceae	(Roquet et al., 2016)
Hydnoraceae	(Naumann et al., 2016) (Jost et al., 2020)
Krameriaceae	(Gonçalves et al., 2019)
Lauraceae (<i>Cassytha</i>)	(Song et al., 2017) (Wu et al., 2017)
Lennoaceae	(Hasenstab-Lehman, 2017)
Orobanchaceae	(Fan et al., 2016) (Wicke et al., 2016)
Rafflesiaceae*	(Molina et al., 2014)
Santalales (familias como Balanophoraceae)	(Nickrent, 2020) (Westwood et al., 2010)(Su et al., 2019)

* Reporte de posible ausencia completa del plastoma

Entre los tres modelos principales, el propuesto por Naumann et al., (2016) y posteriormente ampliado por (Jost et al., 2022) con sus modelos de pérdida y conservación de repeticiones invertidas y directas, se basa en la familia Hydnoraceae con la especie *Hydnora visseri*. La importancia de este modelo, yace en que la familia Hydnoraceae es un linaje clave para alcanzar un consenso en la reducción plastidial en los linajes parasíticos, debido a tres razones principales: es una de las familias de plantas más peculiares del planeta por su alta modificación morfológica (Nickrent, 2020); está constituida exclusivamente holoparasitas, por lo que se espera una mayor reducción plastidial; y tercero, es uno de los linajes más antiguos de plantas parasitas (Musselman & Visser, 1989; Naumann et al., 2016; Thorogood, 2019), con una edad estimada de divergencia de su grupo hermano de aproximadamente 91 Ma, mientras que la edad estimada del grupo corona (todas las cuales son parasitas) es de aproximadamente 55 Ma (Naumann et al., 2013; Sanchez-Puerta et al., 2023).

A pesar de esto, los mismos investigadores mencionan que para validar el modelo era necesario obtener secuencias del plastoma del género *Prosopanche* (hermano de *Hydnora*) con el objetivo de comprender la evolución plastidial de las Hydnoraceae y comparar la reducción con la de otros linajes parasíticos. Actualmente están disponibles 11 plastomas de la familia Hydnoraceae, distribuidos en siete plastomas secuenciados para *Hydnora* y cuatro plastomas para *Prosopanche* (uno de los cuales fue generado en esta investigación). Los plastomas de *Hydnora* incluyen *Hydnora abyssinica*, *Hydnora africana*, *Hydnora arabica*, *Hydnora longicollis*, *Hydnora triceps*, y *Hydnora visseri*, que tiene repeticiones invertidas altamente reducidas, y el de *Hydnora esculenta* que carece de repeticiones invertidas (Jost et al., 2022; Naumann et al., 2016). En *Prosopanche* se encuentran tres plastomas disponibles públicamente y el secuenciado en esta investigación. En estos plastomas disponibles las cuatro especies (*Prosopanche americana*, *P. bonacinae*, *P. panguanensis* (Jost et al., 2020, 2022) y *P. costaricensis*) presentan repeticiones directas reducidas. Es importante mencionar que, de las especies no secuenciadas en esta investigación, la de *P. americana* (consultado mayo-2024) se encuentra aún sin verificar en el GenBank.

La familia Hydnoraceae y el género *Prosopanche* han presentado incertidumbre en el posicionamiento filogenético y aceptación de especies, respectivamente. En el caso de la familia, el posicionamiento filogenético dentro de las Piperales ha mostrado topologías contradictorias en diferentes estudios, dependiendo del origen y cantidad genes analizados (mitocondriales, nucleares y plastidiales). La mayoría de análisis han utilizado combinaciones de loci plastidiales, mitocondriales y nucleares (Jost et al., 2020; Massoni et al., 2014; Naumann et al., 2013, 2016; Nickrent et al., 2002), y pocos han utilizado exclusivamente genes plastidiales para estudiar su posición filogenética (Jost et al., 2020). Un estudio profundo es el realizado por Jost et al., (2021) quienes usaron 137 loci nucleares, plastidiales y mitocondriales de 23 taxones para reconstruir árboles filogenéticos según el origen de los loci. Es decir, obtuvieron arboles por separado para los genes plastidiales, los genes mitocondriales y para los genes nucleares, y mediante estos propusieron el posicionamiento de Hydnoraceae como una familia aparte de Aristolochiaceae en el orden Piperales. Otro estudio importante es el de Jost et al., (2020) que utilizaron 83 taxones con alineamientos de 82 genes plastidiales para posicionar a Hyndoraceae en el orden de Piperales. Aumentar el número taxones (Kapli et al., 2020) en un análisis solo plastidial podría generar información importante para ayudar a clarificar o reforzar el posicionamiento de esta familia. No obstante, es importante recordar que a pesar de que existe la noción de que los datos de plastidios arrojan estimaciones de filogenia más confiables, esto puede ser no del todo cierto. La confiabilidad sobre los datos de plastidios es probablemente generado por la poca evidencia sobre la transferencia horizontal de genes de plastidios o de plastidios completos entre plantas (Gitzendanner et al., 2018; Renner & Bellot, 2012), pero esto no quiere decir que los eventos no sucedan, actualmente hay disponibles reportes de este tipo de transferencia en plantas parásitas (Park et al., 2007). Entonces a la hora de obtener resultados provenientes de estos datos hay que tomar en cuenta estas interacciones e intercambio de datos genéticos (Hedenäs et al., 2021).

Otra problemática que existe en la familia es que, para el género *Prosopanche*, existe un debate sobre la aceptación de *P. costaricensis* como una especie diferente de *P. americana*. Según la revisión morfológica más completa del género, realizada por Cocucci & Cocucci (1996), solo se reconocen dos especies, *P. bonacinae* y *P. americana*; considerando a *P. costaricensis* como conspecífica de *P. americana*. Sin embargo, investigaciones posteriores por Ferreira-Machado &

Paganucci-De queiroz (2012), Funez, Ribeiro-Nardes, Kossmann, Peroni, & Drechsler-Santos, (2019) y Martel, Fernandez-Hilario, Tello, Arteaga, & Gerlach, (2018), han propuesto nuevas especies para el género (*P. caatingicola* y *P. demogorgoni* de Brasil y *P. panguanensis* del Perú, respectivamente), y consideran a *P. americana* y *P. costaricensis* como especies diferentes. La diferenciación fue propuesta con base en sus características morfológicas como: la longitud del tubo perigonal, tamaño de los tépalos y tamaño del sinandrio, características que en *P. costaricensis* son más pequeñas que en *P. americana*. De igual manera, Bolin, (2009) distingue entre ambas especies en sus análisis filogenéticos utilizando los genes *ITS* y *rpoB*.

En una reciente revisión del género (Hatt et al., 2023), se menciona que *Prosopanche* incluye siete especies, las dos especies originalmente tratadas por Cocucci & Cocucci (1996) *P. americana* y *P. bonacinae*, además de más otras cinco especies adicionales: *Prosopanche costaricensis* (Gómez-Pignataro & Gómez-Laurito, 1981), *P. caatingicola* (Ferreira-Machado & Paganucci-De queiroz, 2012), *P. panguanensis* (Martel et al., 2018), *P. demogorgoni* (Funez et al., 2019) y *P. cocuccii* (de Carvalho et al., 2021); la mayoría de estas descripciones han sido basadas en la morfología. La especificidad de hospedero no se ha sido útil como evidencia para apoyar la separación en diferentes especies debido a que la documentación del hospedero para este tipo de plantas no siempre es confiable, debido a la práctica usual de asumir que el hospedero es la planta más cercana a donde se encontró el parásito. Aunado a esto, varias de estas especies comparten hospederos: *P. demogorgoni* y *P. cocuccii* parasitan a especies de Asteraceae y Solanaceae (de Carvalho et al., 2021; Funez et al., 2019), mientras que *P. americana*, *P. caatingicola*, *P. panguanensis* y *P. costaricensis* se han encontrado parasitando especies de Fabaceae (Cocucci & Cocucci, 1996; Ferreira-Machado & Paganucci-De queiroz, 2012; Gómez-Pignataro & Gómez-Laurito, 1981; Martel et al., 2018). En el caso de *P. bonacinae*, se ha reportado como generalista pero no se ha observado asociada a Fabaceae (Cocucci & Cocucci, 1996). Por lo tanto, la información plastidial podría ser una alternativa valiosa para añadir evidencia que apoye la diferenciación entre especies de este género.

Esta investigación se enfocó en determinar el patrón de reducción de los plastomas en la familia Hydnoraceae, incluyendo en el análisis todos los plastomas disponibles de esta familia, y adicionando el plastoma de *Prosopanche costaricensis*, que fue secuenciado por primera vez en

este estudio. Luego de obtener el plastoma de *P. costaricensis* se comparó su constitución contra los plastomas de las otras especies de Hydnoraceae, así como con parientes fotosintéticos cercanos. Además, se generó evidencia plastídica para distinguir entre las especies *P. americana* y *P. costaricensis* (Ferreira-Machado & Paganucci-De queiroz, 2012). Se determinó la selección que se encuentra actuando sobre los genes retenidos en los plastomas de las especies de *Prosopanche* más el de *Hyndora visseri* como grupo externo, para validar el modelo de Naumann et al. (2016) y Jost et al., (2022). Finalmente, se comparó la reducción plastidial de a la familia Hydnoraceae contra otros linajes de angiospermas a través de su posicionamiento filogenético, confirmando su ubicación entre los Piperales, con una relación todavía no resuelta con respecto a cuáles familias de este orden son las más cercana a ella, de acuerdo a genes plastidiales.

OBJETIVOS

General

Determinar el patrón de reducción y cambio en los plastomas de Hydnoraceae, mediante la secuenciación de un nuevo plastoma (*Prosopanche costaricensis*) y su comparación con otros plastomas de la familia y parientes fotosintéticos cercanos.

Específicos

- 1 Generar la secuencia completa del plastoma de *Prosopanche costaricensis*.
- 2 Analizar las variaciones del plastoma de *Prosopanche costaricensis* en comparación con los plastomas disponibles de Hydnoraceae y los plastomas de holoparásitas altamente reducidos de otros linajes principales de angiospermas parásitas haustoriales.
- 3 Comparar los patrones de la selección que actúan sobre los genes retenidos en los tres plastomas de *Prosopanche* con relación a *Hydnora*.
- 4 Realizar un análisis detallado de los plastomas de *Prosopanche americana* y *Prosopanche costaricensis* para identificar diferencias significativas entre ambos.
- 5 Determinar la posición filogenética de Hydnoraceae entre las angiospermas, incluyendo el plastoma de *P. costaricensis*.

MATERIALES Y MÉTODOS

Material vegetal y extracción de ADN

Las plantas de la familia Hydnoraceae se desarrollan de manera subterránea y solo presentan rizomas, estructuras florales y frutos. Los rizomas y frutos crecen generalmente bajo tierra y las estructuras florales en algunas especies emergen a la superficie mientras que en otras no (Hatt et al., 2023; Nickrent et al., 2002). En particular la estructura floral de *Prosopanche costaricensis* es visible sobre la superficie (Gómez-Pignataro & Gómez-Laurito, 1981), por lo que en esta investigación se utilizó la flor de un espécimen como material para la extracción de ADN. El material fue previamente recolectado por Jorge Gómez Laurito en el Parque Nacional Braulio Carrillo en el sendero Botarrama en junio del 2003. Sitio donde esta especie también ha sido recolectada por Barry Hammel, el cual presenta dos entradas una en el Jardín Botánico de Missouri con el número de recolección 18037 y la otra en el Museo Nacional de Costa Rica con el número de recolección CR153815. Entonces, este material desde junio del 2003 se mantuvo congelado a -12 a -15 °C hasta su uso.

Este fue el material que combina ser el más reciente y mejor preservado disponible para este estudio ya que *P. costaricensis* es una planta muy raras veces recolectada y que ha sido observada en el campo en muy raras ocasiones. A nivel mundial solo se pudieron comprobar 14 recolectas (repartidas en ca. 45 especímenes), la más antigua en 1937 y la más reciente en 2008 (en las colecciones de BM, CR [incluyendo INB], F, L, MO, NY, U, US, USJ), con solo 3 recolectas desde el año 2000 a la actualidad. A pesar de los esfuerzos por recolectar material más reciente mediante giras de campo en el periodo del 2021 al 2022 al Parque Nacional Braulio Carrillo, Sendero Botarrama, y otros sitios donde se ha reportado su avistamiento en la zona de Limón: Siquirres, Finca la Lola, Callejón sección 18, y el Parque Nacional Hitoy Cerere, pero sin éxito.

El material recolectado en el 2003 fue procesado para la extracción del ADN total utilizando el protocolo de extracción CTAB con nitrógeno líquido (Doyle, JJ; Doyle, 1990) en conjunto con un tratamiento con RNasa (Benore-Parsons & Ayoub, 1997). La calidad y cantidad de ADN se evaluó por medio de espectrofotometría con el equipo NanoDrop 2000c (Thermo Scientific). Luego de cuantificarse la calidad y cantidad de ADN, se procedió a la precipitación

de este mediante el protocolo de acetato de sodio y etanol (Piaggio et al., 2014) y finalmente este se re-suspendió en agua libre de nucleasas.

Secuenciación masiva y ensamblado del plastoma

Con ADN total extraído, se realizó la secuenciación masiva del genoma de *Prosopanche costaricensis* con la empresa CD Genomics (Estados Unidos, Nueva York). Se solicitó la secuenciación del genoma completo por medio de la técnica de extremos emparejados (pair-end) para fragmentos de 300 pb con Illumina para una cobertura de 40X. Dicha plataforma fue seleccionada debido a su popularidad en los estudios de metagenómica por su mayor rendimiento en comparación a las otras técnicas de secuenciación de segunda generación (NGS por sus siglas en inglés). En esta plataforma las lecturas son cortas, no obstante al ser leídas de manera emparejada se superponen, lo que permite fusionarlas en una lectura más larga y así obtener un genoma mediante técnicas de ensamblaje (Werner et al., 2012).

CD Genomics (Estados Unidos, Nueva York) envió las lecturas pareadas, sentido (R1) y anti-sentido (R2), de la secuenciación masiva del genoma completo. Para comprobar la calidad de estos datos se utilizó el programa FastQC (versión 0.11.5) (Andrews, 2010) y así se determinó si los datos eran aptos para su uso en el ensamblaje y anotación del plastoma.

Una vez que la calidad de las lecturas fue determinada como apropiada, estas se emplearon para el ensamblaje inicial utilizando como referencia los plastomas de *Hydnora visseri* (NC_029358) y *Prosopanche americana* (MT075717.1), por medio del supercomputador del Centro Nacional de Alta Tecnología (CeNAT) Kabré, utilizando el programa Bowtie2 (Langmead & Salzberg, 2012). El ensamblaje se realizó utilizando a *P. americana* como referencia y no *P. panguanensis*, siendo la última más similar, debido a que el genoma de *P. panguanensis* fue depositado en mayo del 2022 al NCBI y el ensamblaje inicial de *P. costaricensis* se realizó antes de dicha fecha, por lo que no se logró incluir el genoma de esta especie en el momento y se utilizó el genoma sin verificar de *P. americana* y el de *H. visseri*.

Para realizar el ensamblaje se utilizó el código descrito a continuación. Primero se preparó el indexado de ambos genomas de referencias:

- *H. visseri*

- o bowtie2-build NC_029358.1.fasta Hydnora_index
- *P. americana*
 - o bowtie2-build MT075717.fasta Prosopanche_index

Luego se realizó el alineamiento de las lecturas pareadas de una manera altamente sensible contra los genomas de referencia. En el mismo código del alineamiento se empleó el programa Samtools 1.9 (Danecek et al., 2021) para transformar el alineamiento que se obtiene con Bowtie-2 de un formato SAM a un formato BAM que se puede visualizar en el programa IGV (Robinson et al., 2023).

- *H. visseri*
 - o bowtie2 --very-sensitive -x Hydnora_index -1 R1.fq -2 R2.fq | samtools view -b -o hydnora_proso_alingment2.bam
- *P. americana*
 - o bowtie2 --very-sensitive -x Prosopanche_index -1 R1.fq -2 R2.fq | samtools view -b -o proso_proso_alingment2.bam

Se indexaron los resultados de los alineamientos en formato BAM para poder visualizarlos en programas como IGV, con el siguiente código:

- *H. visseri*
 - o samtools index hydnora_proso_alingment2_sort.bam
- *P. americana*
 - o samtools index proso_proso_alingment2_sort.bam

Luego, se descargaron los alineamientos y se visualizaron utilizando el programa IGV para determinar la cobertura de las lecturas alineadas a los dos diferentes genomas de referencia. Con este paso se determinó que el alineamiento con *P. americana* fue mejor por lo que se procedió a obtener la secuencia consenso de este, con el uso de Samtools 1.9 y Bcftools 1.9 (Danecek et al., 2021).

- *P. americana*
 - o samtools mpileup -uf MT075717.fasta proso_proso_alingment2_sort.bam | bcftools call -c | vcfutils.pl vcf2fq > Proso_proso_cns.fastq

Como *Prosopanche americana* fue el mejor genoma de referencia, se realizaron dos protocolos de ensamblaje más con *P. americana* como genoma de referencia y con solo las lecturas sentido (R1) ya que estas lecturas tuvieron una leve mejor calidad que las anti sentido. El primero se hizo utilizando el ensamblador de Geneious Prime 2022.0.2

(<https://www.geneious.com>) utilizando la opción de mapeo con referencia; y el segundo se realizó con Geneious pero a través del plug-in de Bowtie2 (Langmead & Salzberg, 2012) como ensamblador. El código empleado para el uso de Bowtie2 incluyó un alineamiento altamente sensible, pero ahora también controlando que las lecturas utilizadas tuvieran tamaños entre los 10 pb (I) y los 600 pb (X).

```
- bowtie2-build-s BowtieCombinedReferenceSequences.fasta myIndex
- bowtie2-align-s -I 10 -X 600 -p 4 --very-sensitive-local -q -x myIndex -U forwardReads.fastq -S bowtie.sam
```

Una vez obtenidos los tres ensamblajes, estos fueron alineados mediante MAFFT en línea (Kato et al., 2019) y la secuencia consenso se conservó para el resto de análisis, dígame para el diseño de imprimadores para *gaps*, la inclusión de los gaps que se lograron secuenciar, la construcción del ensamblaje final con GetOrganelle y el proceso de anotación.

Diseño de imprimadores para los gaps del ensamblaje

Con esta secuencia consenso de los tres ensamblajes se identificaron sectores de la secuencia que presentaban nucleótidos no identificados correctamente (de aquí en adelante llamados “gaps”). En total, se agruparon estos gaps en 28 fragmentos con tamaños de 200 pb a 3000 pb, y se diseñaron imprimadores con Primer3Plus (Rozen & Skaletsky, 2000) para amplificarlos por medio de la técnica de amplificación de reacción en cadena de polimerasa (PCR) (cuadro 2).

Se mandaron a sintetizar 28 pares de imprimadores (56 oligos) a la empresa Biocientífica Internacional (San José, Costa Rica). De estos 56 oligos, tres fueron descartados ya que no se lograron obtener amplicones en ningún proceso de PCR. Se eliminaron los dos imprimadores del fragmento 10 (R y F) y el imprimador sentido del fragmento 11 (F) (cuadro 2). Para poder obtener las secuencias de los gaps de estos fragmentos, se emparejó el imprimador sentido del fragmento 2 (F) con el imprimador anti sentido del fragmento 11 (R) lo que generó una nueva pareja (10.1) y se eliminaron las parejas de imprimadores de los fragmentos 10 y 11. Entonces la nueva pareja

de imprimadores (10.1) cubrió los dos gaps faltantes y así se redujo el número a 27 fragmentos por secuenciar.

Adicionalmente, para la prueba 4 de la metodología de Amplificación y Secuenciación de Gaps, se diseñaron y sintetizaron de la misma manera, 5 pares de imprimadores más (últimas filas de la cuadro 2). Específicamente para intentar obtener las secuencias de genes en específico o gaps entre genes. Diferentes grupos de los primeros 27 pares de imprimadores fueron utilizados en todas las pruebas (1 a 4) y los 5 imprimadores diseñados para amplificar genes y/o gaps entre genes se emplearon solo en la prueba 4.

Cuadro 2. Información de los imprimadores para secuenciar los nucleótidos indeterminados encontrados en el ensamblaje por referencia de *Prosopanche costaricensis*.

ID del fragmento	Nombre del imprimador*	Tamaño de imprimador (pb)	Temperatura de fusión ° C	GC%	Secuencia del Imprimador (5'-3')
1	17882F	25	63,6	52	TACGCCCTTCTTCGTCTCTGAGTAC
	19433R	21	62,3	57,1	TACACACTGCCCGTCACAGTC
2	64F	20	57,3	50	ACCTCTAGACGATAAGGGCA
	361R	23	59,8	47,8	CGATGTTTCGATAAAAAGGTCGTCC
3	15662F	22	60,5	55	GAGGCCTACGCCCATATAAAC
	15873R	21	58,8	52,4	GATGCGGACTATTCACACCAG
4	4755F	22	57,1	40,9	ACAACATTAGGTTTTGGAGACC
	5651R	25	57,5	36	ACGTTTAGGAATGTGTATGAGATTG
5	16970F	25	61,4	44	TGACAATGAAGCTTATCCCCATTG
	17890R	22	55,7	41	TAACCTGGAGAACTGAAACATC
6	2415F	20	60	40	CGTTTGTGAAGAATGTGGTT
	4139R	20	59,9	45	CCTGTTTTTACAGCTTCAGG
7	8789F	23	56,3	34,8	TGTTCTTTCCATTGTGTATTGC
	9041R	23	57,2	43,5	GTAATGGTAGCCGCTTTAGTATG
8	16136F	24	65,8	38	CAGTCGTCTGGCCTGGTTATTGC
	16793R	23	58,8	34,8	TGCCGAAACTATGAGATTAACAA
9	24325F	20	56,1	45	TGAAGCTGACGGAATCATTG
	25160R	21	56,2	38,1	GAAGCATGCATGAAAAACACA
10	387F	25	59,7	32	TTCGAGGCAATTTTTGTACTACTTT
	886R	27	57,1	25,9	ACAGTAAATATCCCAGTTTTATTGTG
10.1	64F	20	57,3	50	ACCTCTAGACGATAAGGGCA

	1551R	27	55,6	25,9	TGTTTTACCACTTATATCATCTTTGAA
11	1185F	24	56,3	29,2	ACAATTTCAAACTACGCTTTCTT
	1551R	27	55,6	25,9	TGTTTTACCACTTATATCATCTTTGAA
12	1623F	26	57,2	30,8	GAGATTGGTTTTACATCATTAGAAGA
	2437R	23	58,6	39,1	GATAACCACATTCTTCACAAACG
13	2485F	22	58,1	36,4	GATACTTGGAAATCCATAAATG
	2825R	23	59,2	30,4	TTTGTGCATTGTCATTCTTTTTTC
14	5666F	20	53,3	40	AAGAGGAGCTGATGATTTTG
	5959R	22	53,1	36,4	CTCCTTTTTCTCATAATGGTTG
15	7173F	20	54	45	GCACCACTGTTATCTGTTAC
	7404R	20	55,7	50	GTCTGGTAAAGGTCCACTG
16	9018F	20	59,9	45,0	ACATACTAAAGCGGCTACCA
	9255R	22	57,1	31,8	AGTCCTGAAACGATTTTAACTT
17	12800F	22	53,1	36,4	TTAAGCTACAAATCCATCTAGG
	13286R	20	52,6	35	TTTCACCAAATTTGTATGCC
18	20633F	23	58,7	47,8	CAGACATAAATCCAGCCTAGAGC
	21224R	22	58,4	31,8	GATGCTTTTAAAGGAAATGCAA
19	25140F	21	56,2	38	TGTGTTTTTCATGCATGCTTC
	28415R	26	56	34,6	CCTCTATGTTTTACAGTTATTGTACC
20	6179F	23	52,1	30,4	TTCATCCATAATTGTGATTCTAG
	6511R	24	53,8	29,2	TTGGAGGGGAAATTTATTTTATG
21	9704F	23	52,7	34,8	CTACTATCAACTTTATTTCCACC
	9945R	22	53,1	36,4	CCTTGGCTTTTTCTATAAATC
22	21343F	20	51,6	35	TTACGTATTGATGCTTTTCC
	21667R	24	54	33,3	TTCAAAGACGTTCAAATATGGAG
23	13973F	26	51,8	26,9	GATATACTCAATCTCTTAAATATTGG
	14867R	25	55,2	36	CGATACTTTAAGAGGAATCTTATGG
24	21828F	20	57,3	50	GTCAATCGTATTCGGGCAAC
	23058R	23	50,8	30,4	GATACTTTACCAATTAATAACCC
25	29806F	23	52,3	30,4	CAAGCTATTGTTAATAAAGAAGC
	30140R	25	59,93	40	TTGACTTATGTGTCTACGTTGGTTC
26	10399F	22	51,2	27,3	TAAATTCGTCAGAATCAAAATG
	10629R	22	49,1	27,3	AGTTCTAATAATTTACATACCAG
27	10607F	22	49,1	27,3	TCTGGTATGAAATTATTAGAAC
	11843R	23	57,5	43,5	TAGATTAAGACCTTGGTACGACG
28	12403F	22	50,1	27,3	TTCTGAAAATCTAATATCATGC
	12715R	21	54	42,9	CAGAACCTCCGAAAGATAAAG
rps14/rps4	rps14.1F	23	63,7	47,8	ACCTCTAGGACGACCTTTTATCG

	rps4 R	25	56.6	28	ACAATTATACAGTAAATATTCCCAG
accD	Pc accD F	24	55.2	37.5	CTATAGATCCTCTAGAGATTGATG
	Pc accD R	25	64.5	40	CTGAATCCATTATACCAAGTGCAAC
ycf1	Pc ycf1 F	22	52.5	27.3	TCTGGTATGAAATTATTAGAAC
	Pc ycf1 R	23	71	52.2	GAGCAGGCAGAAATGTTGCTTGG
23s/16s	Pc 23S F	25	61.2	44	CGTCTCTGAGTACCAAGGTATTAC
	Pc 16S R	25	73.5	52	CTCGCACTCATGAAGTCGAAATCGC
ycf2/rpl2	Pc ycf2 F	25	48.3	24	CATAACTTTACATATAGATCTATAC
	Pc rpl2 R	23	66	39.1	CAAATTGTTGTATTCCGATTGGG

*F se refiere a sentido y R se refiere a anti sentido y los números en el nombre del oligo se refieren a la posición en la que se encuentran de acuerdo con el consenso del alineamiento de los 3 primeros ensamblajes. Adicionalmente, las secuencias de los imprimadores en negrita se refieren a las secuencias que no funcionaron en el ensamblaje final.

Amplificación y Secuenciación de Gaps

Utilizando el ADN extraído y los imprimadores diseñados se procedió a realizar la amplificación y secuenciación de los gaps. Para esto se tuvo que realizar 4 pruebas (protocolos distintos) donde se hicieron cambios en diferentes factores para lograr la secuenciación correcta de los gaps. Estas variaciones se resumen en: i) el uso de diferentes Taq polimerasas, ii) variaciones en los protocolos de amplificación, iii) cambios en las reacciones de amplificación, iv) uso de diferentes metodologías de secuenciación, y v) empleo de diferentes equipos de termociclado.

Prueba 1

Como primer paso se probaron cinco polimerasas diferentes para el fragmento más largo (fragmento 19) con un protocolo de amplificación con touch up/Hot Start (Burger et al., 2007) y con sus respectivas reacciones de amplificación (cuadro 3). Luego, se decidió utilizar solo la DreamTaq PCR Master Mix (2X) (Thermo Fisher Scientific) para los demás 26 fragmentos, con el mismo protocolo de amplificación. De los fragmentos amplificados, aquellos que amplificaron bandas múltiples se corrieron en un gel de electroforesis al 0.8% de agarosa y se seleccionaron las bandas de tamaño esperado. Dichas bandas se purificaron del gel de agarosa con el kit NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL GmbH & Co. KG). En el caso de las que rindieron solo una banda se purificaron con el kit Invitrogen™ Minikit de ADN genómico

PureLink™ (Fisher Scientific SL). Ambos kits se utilizaron de acuerdo con las instrucciones del fabricante, solo se modificó la elución final, realizada en agua libre de nucleasas en lugar de buffer TE.

Los productos purificados se secuenciaron, para su secuencia sentido y antisentido, en grupos de 2 a 6 productos a la vez con el kit Big Dye v3.1 Terminator (Applied Biosystems). Para esta primera prueba se utilizó como reacción de secuenciación las cantidades de la cuadro 4. Para todas las reacciones de secuenciación se utilizó el protocolo de termociclado reportado por el fabricante.

Luego de la secuenciación se procedió con la precipitación de los productos siguiendo el protocolo de etanol/acetato de sodio de ABI PRISM® dGTP BigDye™ Terminator v3.0 Ready Reaction Cycle Sequencing Kit para tubos de microcentrífuga con una serie de modificaciones. Los volúmenes de precipitación se utilizaron a la mitad ya que el producto de secuenciación generado fue de 20 µl y no 40 µl. Se precipitó por 30 min en vez de 15 min, y la primera centrifugación fue de 25 min en vez de 20 min. Todas las centrifugaciones se llevaron a cabo a -3° C.

Una vez precipitados, los productos fueron enviados al CIBCM (Centro de Investigación en Biología Celular y Molecular) de la Universidad de Costa Rica (San José, Costa Rica), para determinar las secuencias mediante el equipo Genetic Analyzer 3500 (Applied Biosystems ABI® 3500).

Cuadro 3. Protocolos de las primeras pruebas para la amplificación de los nucleótidos indeterminados del ensamblaje por referencia de *Prosopanche costaricensis*.

Polimerasa	Reacción de amplificación		Protocolo de amplificación
	Reactivos	1X (µl)	
DreamTaq PCR Master Mix (2X) (Thermo Fisher Scientific)	H ₂ O libre de nucleasas	8	Un ciclo de 5 min de desnaturalización a 95° C, luego 5 ciclos de touch up con 40 s a 94° C, 35 s a 45° C con +1° C cada ciclo y 2.5 min a 72° C, seguido de 30 ciclos con la desnaturalización a 94°
	MM2 Buffer	12,5	
	Imprimador Sentido 10 µM.	1	
	Imprimador antisentido 10 µM.	1	
	MgCl ₂	1	
OneTaq® High GC (NEB)	ADN	1	
	H ₂ O libre de nucleasas	11,3	
	Buffer 5X GC	5	
	MgSO ₄	2,5	
	2.5 mM dNTPs	2,5	

	Imprimador Sentido 10 μ M.	1	C por 35 s, 30 s a -5° C de la temperatura de hibridación del imprimador con la menor temperatura, y 1:30 min a 72° C de elongación; y finaliza todo el proceso con un ciclo de elongación a 72° C por 7 minutos.
	Imprimador antisentido 10 μ M.	1	
	ADN	1	
	Taq	0,2	
Long Range PCR (NEB)	H ₂ O libre de nucleasas	14,35	
	Buffer 5X	5	
	MgSO ₄	1	
	10 mM dNTPs	1	
	Imprimador Sentido 10 μ M.	1	
	Imprimador antisentido 10 μ M.	1	
	ADN	1,5	
	Taq	0,4	
Q5 [®] High-Fidelity DNA Polymerase (NEB)	H ₂ O libre de nucleasas	8,5	
	Taq Q5 buffer	12,5	
	Imprimador Sentido 10 μ M.	1	
	Imprimador antisentido 10 μ M.	1	
	ADN	1	
Platinum [™] Taq DNA Polymerase High Fidelity (Invitrogen [™])	H ₂ O libre de nucleasas	13,8	
	Buffer	5	
	2,5 mM dNTPs	1,5	
	Imprimador Sentido 10 μ M.	1	
	Imprimador antisentido 10 μ M.	1	
	ADN	1	
	MgCl ₂	1	
	Taq	0,2	

Cuadro 4. Reacción de secuenciación de la primera prueba para la amplificación de los nucleótidos indeterminados del ensamblaje por referencia de *Prosopanche costaricensis*.

Reactivos	1 x
Buffer	2 μ l
Imprimador Sentido (F) 10 μ M.	2,5 μ l
Imprimador Antisentido (R) 10 μ M.	2,5 μ l
ADN	4,5 μ l
Taq	1,3 μ l

Prueba 2

En la segunda prueba se probaron cuatro Taq polimerasas para encontrar la enzima que produjera la mejor calidad fragmentos. Se eliminaron dos de las usadas en la prueba 1, Long Range PCR (NEB) y Platinum[™] Taq DNA Polymerase High Fidelity (Invitrogen[™]), y se incluyó la polimerasa Phire II Hot Start (Thermo Fisher Scientific). Adicional a esto, se realizaron variaciones en los reactivos de la reacción de amplificación. Se añadió PEG8000 (polietileno glicol), MgSO₄ (sulfato de magnesio), DMSO 10% (dimetilsulfóxido), y OneTaq[®] High GC

Enhancer en diferentes casos y se realizaron cambios en la cantidad de ADN (de 1 μ l hasta 3 μ l) (ver reacciones de amplificación en cuadro 5).

Además, se probaron dos protocolos de amplificación diferentes (cuadro 5) para tratar de aumentar la probabilidad de amplificación de los fragmentos. El denominado Touch up/Hot start lleva a cabo: un ciclo de 5 min de desnaturalización a 95° C, luego 5 ciclos de touch up con 40 s a 94° C, 35 s a 45° C con +1° C cada ciclo y 2.5 min a 72° C, seguido de 30 ciclos con la desnaturalización a 94° C por 35 s, 30 s a -5° C de la temperatura de hibridación del imprimador con la menor temperatura, y 1:30 min a 72° C de elongación; y finaliza todo el proceso con un ciclo de elongación a 72° C por 7 minutos. El segundo protocolo se denomina Sin touch up/Hot start y este lleva a cabo un ciclo de desnaturalización inicial a 95° C por 5 min, 30 ciclos de desnaturalización a 94° C por 35 s, 30 s a -5° C de la temperatura de hibridación del imprimador con la menor temperatura, y 1:30 min a 72° C de elongación, y finaliza con un ciclo de elongación a 72° C por 7 minutos.

Cuadro 5. Protocolos de la segunda prueba para la amplificación de los nucleótidos indeterminados del ensamblaje por referencia de *Prosopanche costaricensis*.

Polimerasa	Gaps	Reacción de amplificación		Protocolo de amplificación
		Reactivos	1X (μ l)	
DreamTaq PCR Master Mix (2X) (Thermo Fisher Scientific)	11, 14	H ₂ O libre de nucleasas	8	Sin touch up
		MM2 Buffer	12,5	
		Imprimador Sentido 10 μ M.	1	
		Imprimador antisentido 10 μ M.	1	
		MgCl ₂	1	
		ADN	3	
		PEG8000 10%	1	
OneTaq® High GC (NEB)	10, 21	H ₂ O libre de nucleasas	10,8	Sin touch up
		Buffer 5X GC	5	
		MgSO ₄	3	
		2.5 mM dNTPs	2,5	
		Imprimador Sentido 10 μ M.	1	
		Imprimador antisentido 10 μ M.	1	
		ADN	1,5	
		Taq	0,2	

Q5® High-Fidelity DNA Polymerase (NEB)	1, 6, 7, 10, 5, 13, 14, 25, 9, 27, 11, 12,	Opción 1		Sin touch up
		H ₂ O libre de nucleasas	8	
		Taq Q5 buffer	12,5	
		Imprimador Sentido 10 µM.	1	
		Imprimador antisentido 10 µM.	1	
	ADN	1,5		
	Opción 2		Touch up	
	H ₂ O libre de nucleasas	8		
	Taq Q5 buffer	12,5		
	Imprimador Sentido 10 µM.	1		
Imprimador antisentido 10 µM.	1			
DMSO 10% ó OneTaq® High GC Enhancer	2,5			
ADN	1,5			
Phire II Hot Start (Thermo Fisher Scientific)	1, 9, 10, 10.1, 7, 25, 12, 13, 27, 2, 3, 4, 14,15, 16, 5, 24, 6, 12, 21, 22, 20, 26, 28, 18, 19, 17, 23	H ₂ O libre de nucleasas	13,1	Con y sin touch up
		Buffer	5	
		10 mM dNTPs	0,5	
		Imprimador Sentido 10 µM.	1	
		Imprimador antisentido 10 µM.	1	
		ADN	1,5	
		DMSO 10%	2,5	
		Taq	0,4	

De los fragmentos amplificados, aquellos que rindieron bandas múltiples se corrieron en un gel de electroforesis al 0.8% de agarosa y se seleccionaron las bandas de tamaño esperado. Dichas bandas se purificaron del gel de agarosa con el kit NucleoSpin Gel and PCR Clean-up kit (MACHEREY-NAGEL GmbH & Co. KG). En el caso de las que rindieron solo una banda se purificaron con el kit Invitrogen™ Minikit de ADN genómico PureLink™ (Fisher Scientific SL). Ambos kits se utilizaron de acuerdo con las instrucciones del fabricante, el único paso que se modificó fue que la elución se realizó en agua libre de nucleasas en lugar de buffer TE.

Posteriormente los productos purificados se secuenciaron, en ambos sentidos con el kit Big Dye v3.1 Terminator (Applied Biosystems). Para esta segunda prueba también se cambiaron las cantidades de la reacción secuenciación por las descritas en la cuadro 6. Además, para todas las reacciones de secuenciación se modificó el protocolo de termociclador reportado por el fabricante de la siguiente manera: Se agregaron 5 ciclos previos de elongación de productos a 45° C y luego se dejaron 25 ciclos de elongación a 50° C.

Cuadro 6. Reacción de secuenciación de la segunda prueba para la amplificación de los nucleótidos indeterminados del ensamblaje por referencia de *Prosopanche costaricensis*.

Reactivos	1 x
H ₂ O	0 µl o 1,7 µl
Buffer	2 µl
Imprimador Sentido (F)10 µM.	2,5 µl
Imprimador Antisentido (R) 10 µM.	2,5 µl
ADN	4,2 µl o 1,5 µl
DMSO10%	1 µl
Taq	1,3 µl

Luego de la secuenciación se procedió con la precipitación de los productos siguiendo el protocolo usado en la prueba 1 y una vez precipitados, los productos fueron enviados al CIBCM para la determinación de las secuencias mediante el Analizador 3500 Genetic Analyzer (Applied Biosystems ABI ® 3500).

Prueba 3

Para la prueba 3, partiendo de los resultados en las pruebas anteriores, se decidió utilizar solo la ADN polimerasa Phire II Hot Start (Thermo Fisher Scientific), con DMSO 10% y el reactivo OneTaq® High GC Enhancer (New England Biolabs). Con esta metodología se volvieron a amplificar los fragmentos que no se habían logrado secuenciar hasta este punto. La reacción de amplificación que se utilizó se puede observar en la cuadro 7 y protocolo de amplificación utilizado fue el programa Sin touch up descrito en la prueba 2.

Cuadro 7. Reacción de amplificación de la tercera prueba para la amplificación de los nucleótidos indeterminados del ensamblaje por referencia de *Prosopanche costaricensis*.

Reactivos	1 x
H ₂ O libre de nucleasas	12.62 µl

5X Phire Buffer de Reacción	5 µl
10 mM dNTPs	0.5 µl
Imprimador Sentido (F) 10 µM.	1 µl (0.4 µM concentración final)
Imprimador Antisentido (R) 10 µM.	1 µl (0.4 µM concentración final)
ADN	1 µl (24 ng concentración final)
DMSO 10%	1.8 µl
OneTaq® High GC Enhancer	1.8 µl
Taq	0.28 µl

Todos los fragmentos obtenidos mediante la metodología anterior se analizaron para determinar su tamaño y calidad por medio de geles de electroforesis del 0.8% al 1.2% con el uso del marcador molecular GeneRuler 1 kb Plus DNA Ladder (Thermo Fisher Scientific Inc.). Aquellos fragmentos de PCR que resultaron en una sola banda de amplificación con el tamaño esperado de acuerdo con el genoma ensamblado fueron secuenciados de manera diferente a la realizada en las pruebas 1 y 2. Los fragmentos fueron enviados directamente al CIBCM (San José, Costa Rica), junto con sus respectivos imprimadores sentido y antisentido a una concentración de 3,4 µM. La secuenciación se realizó igualmente con el kit Big Dye y la realizaron mediante el Analizador 3500 Genetic Analyzer (Applied Biosystems ABI ® 3500). Los productos se purificaron y secuenciaron rutinariamente según el procedimiento del CIBCM, debido a que en la pruebas 1 y 2, el proceso fraccionado de secuenciación y precipitación presentó una baja capacidad de secuenciación. Además, aquellos fragmentos que presentaron más de una banda de amplificación se descartaron para realizar una cuarta prueba con el objetivo de disminuir la cantidad de bandas múltiples previo a su purificación y secuenciación.

Prueba 4

Como se menciona en la prueba 3, los fragmentos de bandas múltiples son los que presentaron la mayor limitación para su secuenciación. Debido a lo anterior, se decidió hacer un cambio de termociclador para disminuir la cantidad de bandas obtenidas por fragmento lo más posible. Se pasó a utilizar un termociclador que no permitía el multiplex sino solo corridas a una temperatura.

En esta prueba se corrieron las muestras con la misma enzima, reacción de amplificación (cuadro 7), y protocolo de corrida Sin touch up. También se probaron 5 fragmentos más para los

cuales se diseñaron nuevos imprimadores a fin de amplificar regiones de genes específicos o gaps entre genes: *accD* (gen), *yfc1* (gen), *yfc2* y *rpl2* (gen *yfc2* y gap), *rps4* y *rps14* (gap), y 23s y 16s (gap), (cuadro 2). En total se amplificaron los 5 fragmentos nuevos y los 16 fragmentos de múltiples bandas: 10.1, 21, 24 y 27, 1, 5, 4, 22, 13, 14, 17, 18, 7, 6, 9 y 19.

Para los fragmentos con múltiples bandas se aplicó el mismo proceso de purificación empleado en la prueba 3. No obstante, al utilizar el kit, a diferencia de lo realizado en las otras pruebas, esta vez los fragmentos se eluyeron en buffer TE como indica el fabricante. Una vez purificados los fragmentos se enviaron a secuenciar al CIBCM (San José, Costa Rica), junto con sus respectivos imprimadores sentido y antisentido a una concentración de 3,4 μM (como en la prueba 3).

Las secuencias obtenidas se curaron manualmente y todas se incluyeron en el ensamblaje final del genoma. Para esto se realizaron alineamientos múltiples entre los fragmentos secuenciados de los gaps, la región con el *gap* del ensamblaje de *Prosopanche costaricensis* y la región correspondiente de los genomas de referencia (*Prosopanche panguanensis* y *Prosopanche americana*) por medio del servicio de alineamiento múltiple Clustal-Omega de EMBL-EBI (Madeira et al., 2022).

Ensamblaje del plastoma de Novo por GetOrganelle

Como el ensamblaje por referencia y proceso de secuenciación de gaps no logró generar un plastoma completo, se decidió utilizar el ensamblaje por medio del pipeline GetOrganelle (Jin et al., 2021) para complementar. Este ensamblaje permite un alineamiento de las secuencias contra un genoma o genomas de referencia, pero el alineamiento lo realiza por extensión. Con el o los genomas de referencia genera una “semilla” de datos que el proceso utiliza para recuperar un set de lecturas iniciales de un tamaño en específico que luego alinea contra los genomas de referencia. Luego del filtrado el programa genera una extensión de las primeras lecturas alineadas con la ayuda de la “semilla” usando las lecturas del genoma a ensamblar (Twyford & Ness, 2017).

De esta forma, para ensamblar a *P. costaricensis* este pipeline corrió en el supercomputador del Centro Nacional de Alta Tecnología (CeNAT) Kabré, con el siguiente código:

```
- python/home/mobando/software/GetOrganelle/get_organelle_fro
m_reads.py -s MT075717_NC_029358.seed.fasta --genes
MT075717_NC_029358.label.fasta -1 R1_val_1.fq -2 R2_val_2.fq
-o ProsoCost_pt4 -R 15 -k 21,45,65,85,105,125 -w 50 -F
embplant_pt
```

Se ejecutó el ensamblaje de *P. costaricensis* con los genomas de *Prosopanche americana* y *Hydnora visseri* como semilla, con un tamaño de palabra de 50, con 15 rondas y con los kmers (longitudes de fraccionamiento de las lecturas) con tamaños de 21,45,65,85,105,125. Los amplios tamaños de kmers se usaron para favorecer la circularización del plastoma ensamblado.

Con este ensamblaje se recuperaron 5 *scaffolds* en vez de un ensamblaje completo. Pero se tomaron estos 5 *scaffolds* y por medio de Geneious se mapearon por referencia al plastoma *P. panguanensis*. Se usó este plastoma y no el de *P. americana* ya que los análisis realizados con primer ensamblaje arrojaron que este tiene una mayor similitud al de *P. panguanensis*. Luego, a partir de este mapeo se extrajo una secuencia consenso circular con un solo gap, es decir otro ensamblaje parcial.

Ambos ensamblajes parciales (ensamblaje por referencia con los fragmentos secuenciados incluidos y el ensamblaje por GetOrganelle) obtenidos a este punto se alinearon gap por gap mediante Clustal-Omega (Sievers & Higgins, 2014) y así se cerraron todos los gaps.

Anotación del Plastoma

Una vez incluidos todos los gaps obtenidos en el ensamblaje por referencia hecho por Bowtie2(ver.2.4.5) y Geneious (ver. 2023.2), se procedió a anotar este primero. Para esto se usó el software GeSeq de CHLOROBOX con el anotador de apoyo Chloë v0.1.0 (Tillich et al., 2017). Herramienta específica para la anotación de secuencias organelares en plantas. En este caso se empleó como referencia el plastoma de *Hydnora visseri* (NC_029358) y no el de *Prosopanche americana*, porque el de *H. visseri* es el único plastoma de esta familia que contiene sus anotaciones curadas en la base datos RefSeq. Luego, para la anotación de ARN de transferencia se utilizó la opción de anotar por medio del software de terceros tRNAscan-SE v2.0.6 (Chan et al., 2019).

Las anotaciones obtenidas para este ensamblaje se refinaron y confirmaron de forma manual por medio de BLAST de nucleótidos y de proteínas (Altschul et al., 1990) y por medio de alineamientos con especies *Prosopanche*. Adicionalmente, la circularización del plastoma se llevó a cabo con GeSeq, utilizando la opción de secuencia circular; y para generar un mapa del plastoma el archivo resultante se procesó con el programa OrganellarGenomeDRAW (OGDRAW), también del portal de CHLOROBOX, (Greiner et al., 2019).

Una vez con este primer ensamblaje parcial anotado, se procedió a alinearlo con el segundo ensamblaje generado por GetOrganelle para rellenar los gaps del primer ensamblaje. Manualmente se completaron los gaps y las anotaciones para generar un ensamblaje consenso completo sin discontinuidades. Posteriormente, se depositará la secuencia de este plastoma de *Prosopanche costaricensis* en GenBank.

Comparación de Plastomas y Análisis de Selección

Con el plastoma de *P. costaricensis* ensamblado y anotado se llevó a cabo la comparación de este contra los genomas plastidiales (código de GenBank y tamaño entre paréntesis) de *Hydnora visseri* ([NC_029358](#), 27 233 pb); *Prosopanche americana* ([MT075717.1](#), 28 191 pb); *Prosopanche panguanensis* ([NC_065054](#), 28 262 pb); *Prosopanche bonacinae* ([NC_065053](#), 28 246 pb); los plastomas altamente reducidos de holoparásitas *Balanophora laxiflora* ([KX784265](#), 15 506 pb) y *Balanophora reflexa* ([KX784266](#), 15 507 pb) (Balanophoraceae), *Cytinus hypocistis* (Cytinaceae) ([NC_031150](#), 19 400 pb), *Conopholis americana* (Orobanchaceae) ([NC_023131](#), 45 673 pb), *Cuscuta erosa* (Convolvulaceae) ([NC_042948](#), 60 959 pb), *Pilostyles aethiopica* ([NC_029235](#), 11 348 pb) y *Pilostyles hamiltonii* ([NC_029236](#), 15 167 pb) (Apodanthaceae), *Mitrastemon kanehirai* (Mitrastemonaceae) ([MF372930.1](#), 25 740 pb) y con un pariente fotosintético cercano de la familia cercanamente emparentada Aristolochiaceae: *Aristolochia contorta* ([NC_036152](#), 160576 pb). En la comparación se determinaron las diferencias en el tamaño de secuencia, el contenido de genes y de GC, y la estructura de los genomas plastidiales como lo indica Jost et al., (2020) y Banerjee & Stefanović (2020); esto mediante el software Geneious.

Por otro lado, para el análisis de selección se utilizaron los genes que codifican proteínas que se comparten entre los plastomas de *Hydnora visseri*, *Prosopanche bonacinae*, *Prosopanche panguanensis*, y *Prosopanche costaricensis*. Este análisis se llevó a cabo de acuerdo con lo realizado por Banerjee & Stefanović (2020) mediante el uso de alineamientos pareados por medio de traducción *in silico* de regiones codificantes de proteínas y *Hydnora visseri* como grupo externo. La única diferencia con el protocolo Banerjee & Stefanović (2020), es que se utilizó el programa MatLab (The MathWorks Inc, 2023) para realizar los cálculos de mutaciones no sinónimas y mutaciones sinónimas (dN/dS).

Análisis filogenético

Para los análisis filogenéticos se incluyeron la mayoría de los genomas de los taxones utilizados por Jansen et al., (2007) y Jost et al., (2020) en sus investigaciones. Estos taxones incluyen genomas plastidiales que representan los principales linajes de angiospermas, los plastomas de mayoría de las Piperales disponibles en GenBank, y las secuencias plastidiales de tres gimnospermas de los órdenes Cycadales, Ginkgoales y Pinales como grupos externos. Adicionalmente, para esta investigación se incluyeron los plastomas de todas las especies disponibles de los géneros *Hydnora* y *Prosopanche*, un taxón de Chloranthales (*Chloranthus spicatus*), otro taxón de Nymphaeales (*Cabomba caroliniana*), esto para un total de 85 especies (cuadro 8 más *Prosopanche costaricensis*).

Para dicho análisis se extrajeron 26 genes (los 26 genes retenidos en *Prosopanche costaricensis*) de cada uno de los taxones a partir de sus registros en GenBank. Se concatenaron y se realizaron alineamientos con MAFFT (Kato et al., 2002; Kato & Standley, 2013) mediante Geneious y se obtuvieron los alineamientos en formato FASTA .

Seguidamente, a partir de este alineamiento se construyó el árbol por el método de máxima verosimilitud. Primero, se determinó el modelo evolutivo que mejor se ajustara a los datos con JModelTest2 (Guindon & Gascuel, 2003; Darriba et al., 2012) por medio del portal CIPRES Science Gateway V. 3.3 (Miller et al., 2010). Seguidamente se generó el árbol utilizando la herramienta RAxML-NG con un bootstrap rápido de 1000 (Kozlov et al., 2019) por medio del portal CIPRES Science Gateway V. 3.3 (Miller et al., 2010).

Cuadro 8. Taxones muestreados para análisis filogenéticos de este estudio con su respectivo orden y número de accesoión, que se incluyen en los análisis filogenéticos.

Número	Taxón	Orden	Número de accesoión GenBank
Gimnospermas - Grupo externo			
1	<i>Cycas micronesica</i>	Cycadales	NC_064311
2	<i>Pinus thunbergii</i>	Pinales	NC_001631
3	<i>Ginkgo biloba</i>	Ginkgoales	NC_016986
“Grado ANA”			
4	<i>Amborella trichopoda</i>	Amborellales	NC_005086
5	<i>Nuphar advena</i>	Nymphaeales	NC_008788
6	<i>Nymphaea alba</i>	Nymphaeales	NC_006050
7	<i>Cabomba caroliniana</i>	Nymphaeales	NC_031505
8	<i>Illicium oligandrum</i>	Austrobaileyales	NC_009600
Magnólidas			
9	<i>Chloranthus spicatus</i>	Chloranthales	NC_009598
10	<i>Calycanthus floridus</i>	Laurales	NC_004993
11	<i>Drimys granadensis</i>	Canellales	NC_008456
12	<i>Piper coenocdatum</i>	Piperales	NC_008457
13	<i>Piper nigrum</i>	Piperales	NC_034692
14	<i>Piper kadsura</i>	Piperales	NC_027941
15	<i>Piper longum</i>	Piperales	NC_047247
16	<i>Piper auritum</i>	Piperales	NC_034697
17	<i>Piper laetispicum</i>	Piperales	NC_042254
18	<i>Saururus chinensis</i>	Piperales	NC_050853
19	<i>Aristolochia kwangsiensis</i>	Piperales	MN965793
20	<i>Aristolochia manshuriensis</i>	Piperales	NC_046766
21	<i>Aristolochia contorta</i>	Piperales	NC_036152
22	<i>Aristolochia debilis</i>	Piperales	NC_036153
23	<i>Aristolochia tagala</i>	Piperales	NC_041455
24	<i>Aristolochia tubiflora</i>	Piperales	NC_041456

25	<i>Aristolochia kunmingensis</i>	Piperales	NC_041451
26	<i>Aristolochia moupinensis</i>	Piperales	NC_041454
27	<i>Aristolochia macrophylla</i>	Piperales	NC_041453
28	<i>Aristolochia kaempferi</i>	Piperales	NC_041452
29	<i>Aristolochia mollissima</i>	Piperales	NC_041457
30	<i>Saruma henryi</i>	Piperales	NC_039933
31	<i>Houttuynia cordata</i>	Piperales	NC_047437
32	<i>Asarum sieboldii</i>	Piperales	NC_037190
33	<i>Asarum heterotropoides</i>	Piperales	MK577409
34	<i>Prosopanche americana</i>	Piperales	MT075717
35	<i>Prosopanche bonacinae</i>	Piperales	NC_065053
36	<i>Prosopanche panguanensis</i>	Piperales	NC_065054
37	<i>Hydnora visseri</i>	Piperales	NC_029358
38	<i>Hydnora abyssinica</i>	Piperales	NC_065049
39	<i>Hydnora africana</i>	Piperales	NC_065136
40	<i>Hydnora esculenta</i>	Piperales	NC_065050
41	<i>Hydnora longicollis</i>	Piperales	NC_065051
42	<i>Hydnora triceps</i>	Piperales	NC_065052
43	<i>Hydnora arabica</i>	Piperales	NC_065211

Monocotiledóneas

44	<i>Acorus calamus</i>	Acorales	NC_007407
45	<i>Dioscorea elephantipes</i>	Dioscoreales	NC_009601
46	<i>Elaeis guineensis</i>	Arecales	NC_017602
47	<i>Hordeum vulgare</i>	Poales	NC_008590
48	<i>Musa acuminata</i>	Zingiberales	HF677508
49	<i>Oryza sativa</i>	Poales	NC_001320
50	<i>Phalaenopsis aphrodite</i>	Asparagales	NC_007499
51	<i>Saccharum officinarum</i>	Poales	NC_035224
52	<i>Sorghum bicolor</i>	Poales	NC_008602
53	<i>Triticum aestivum</i>	Poales	NC_002762
54	<i>Typha latifolia</i>	Poales	NC_013823
55	<i>Yucca schidigera</i>	Asparagales	NC_032714
56	<i>Zea mays</i>	Poales	NC_001666

Eudicotiledóneas

57	<i>Anethum graveolens</i>	Apiales	NC_029470
58	<i>Arabidopsis thaliana</i>	Brassicales	NC_000932

59	<i>Atropa belladonna</i>	Solanales	NC_004561
60	<i>Brassica rapa</i>	Brassicales	NC_040849
61	<i>Buxus microphylla</i>	Buxales	NC_009599
62	<i>Coffea arabica</i>	Gentianales	NC_008535
63	<i>Eucalyptus globulus</i>	Myrtales	NC_008115
64	<i>Gossypium hirsutum</i>	Malvales	NC_007944
65	<i>Helianthus annuus</i>	Asterales	NC_007977
66	<i>Ipomoea purpurea</i>	Solanales	NC_009808
67	<i>Jasminum nudiflorum</i>	Lamiales	NC_008407
68	<i>Conopholis americana</i>	Lamiales	NC_023131
69	<i>Lathraea squamaria</i>	Lamiales	NC_027838
70	<i>Lactuca sativa</i>	Asterales	NC_007578
71	<i>Nandina domestica</i>	Ranunculales	NC_008336
72	<i>Nicotiana sylvestris</i>	Solanales	NC_007500
73	<i>Nicotiana tabacum</i>	Solanales	NC_001879
74	<i>Nicotiana tomentosiformis</i>	Solanales	NC_007602
75	<i>Cuscuta erosa</i>	Solanales	NC_042948
76	<i>Oenothera elata</i>	Myrtales	NC_002693
77	<i>Panax ginseng</i>	Apiales	NC_006290
78	<i>Pelargonium x hortorum</i>	Geraniales	NC_008454
79	<i>Platanus occidentalis</i>	Proteales	NC_008335
80	<i>Ranunculus macranthus</i>	Ranunculales	NC_008796
81	<i>Solanum lycopersicum</i>	Solanales	NC_007898
82	<i>Solanum bulbocastanum</i>	Solanales	NC_007943
83	<i>Solanum tuberosum</i>	Solanales	NC_008096
84	<i>Spinacia oleracea</i>	Caryophyllales	NC_002202
85	<i>Trachelium caeruleum</i>	Asterales	NC_010442

RESULTADOS

ADN para Secuenciaciones Masivas y Sanger

CD Genomics (Estados Unidos, Nueva York), analizó la calidad del ADN enviado, y encontró que esta cumplía con los requerimientos para la construcción de bibliotecas para la secuenciación, y que la muestra permitía la construcción de hasta 3 bibliotecas o más. La concentración del ADN enviado fue de 54.214 ng/ μ l, con una mayor concentración de fragmentos de 11 000 pb.

Por otro lado, el ADN extraído para la secuenciación de los fragmentos por la metodología Sanger para completar los gaps presentó una concentración de 654.6 ng/ μ l con una absorción de 260/280 de 1.88. Por lo que cumplió con los parámetros necesarios para realizar los PCR.

Calidad de las Lecturas de la Secuenciación Masiva

La calidad de las lecturas sentido (R1) y antisentido (R2) obtenidas de la secuenciación masiva fue aceptable. De manera general, FastQC determinó que las puntuaciones de calidad en cada posición de la lectura para las secuencias R1 y R2 fue adecuada, pero las puntuaciones de las lecturas R1 en las posiciones de 90 a 124 pb fueron ligeramente mejores que las de las lecturas R2. Para R2, en las posiciones 90 a 124 pb los valores de calidad de los percentiles 90 tuvieron una calidad menor, de 28 hasta llegar a 24 (menor del umbral para buena calidad). Mientras que en R1 los percentiles 90 para estas mismas posiciones presentaron valores de calidad mayores a 28. No obstante, los valores de calidad de ambas (R1 y R2) mantuvieron una tendencia de calidad alta con valores arriba de 34 a lo largo de todas las posiciones.

En el caso del resto de factores evaluados para determinar la calidad de las lecturas, R1 y R2 presentaron los mismos valores de buena calidad. El contenido GC de ambas fue de 30% lo cual es esperado para secuenciaciones de genoma completo de plantas. La distribución del tamaño de las lecturas tuvo un rango de 149 a 151 pb con un pico de 150 pb. También presentaron una baja duplicación de secuencias y no contaron con secuencias sobre representadas. Por lo que todos los factores indicaron una buena calidad para proceder con el ensamble del plastoma.

Genoma ensamblado de la secuenciación masiva

En el apéndice 1.1, se encuentra el ensamblaje provisional de *Prosopanche costaricensis*, es decir el producto del alineamiento de los tres ensamblajes realizados con las lecturas pareadas entregadas por CD Genomics (Estados Unidos, Nueva York) con el plastoma de *P. americana* como referencia. Este primer ensamblaje tiene una longitud de 30 192 pb.

Este es un ensamblaje parcial, con 7421 nucleótidos indeterminados “N” (gaps) que no lograron identificarse por medio de el ensamblaje por referencia realizado con Bowtie2 y algoritmo de Geneious. La primera medida que se tomó para identificar los nucleótidos desconocidos fue el diseño de los imprimadores presentes en la cuadro 2 y luego llevar a cabo los protocolos de amplificación y secuenciación descritas en las pruebas 1 a 4. La segunda medida, fue realizar un ensamblaje *de Novo* con GetOrganelle usando a *Prosopanche americana* y *Hydnora visseri* como semilla y luego mapear los *scaffolds* por referencia a el plastoma de *P. panguanensis*. A continuación, se describen los resultados de la secuenciación de los gaps y luego se examinan los resultados del ensamblaje final y su anotación.

Amplificaciones y Secuenciaciones de los Gaps

Prueba 1

Para la amplificación del fragmento más largo, el fragmento 19, las enzimas DreamTaq PCR Master Mix (2X) (Thermo Fisher Scientific) y OneTaq® High GC (NEB) dieron los mejores resultados, siendo la primera mejor que la segunda, no obstante el fragmento 19 no se logró secuenciar. A partir de los resultados anteriores se decidió utilizar la DreamTaq PCR Master Mix (2X) (Thermo Fisher Scientific) para amplificar el resto de los 26 fragmentos y con esta se lograron amplificar 16 de los 26 fragmentos. De estos 16 fragmentos 2 y 3 dieron bandas múltiples y 4, 8, 9, 15, 16, 17, 20, 22, 23, 24, 26, y 28 rindieron bandas únicas, pero solo se logró secuenciar correctamente los fragmentos: sentido (F) de 2 y 3 (cuadro 9).

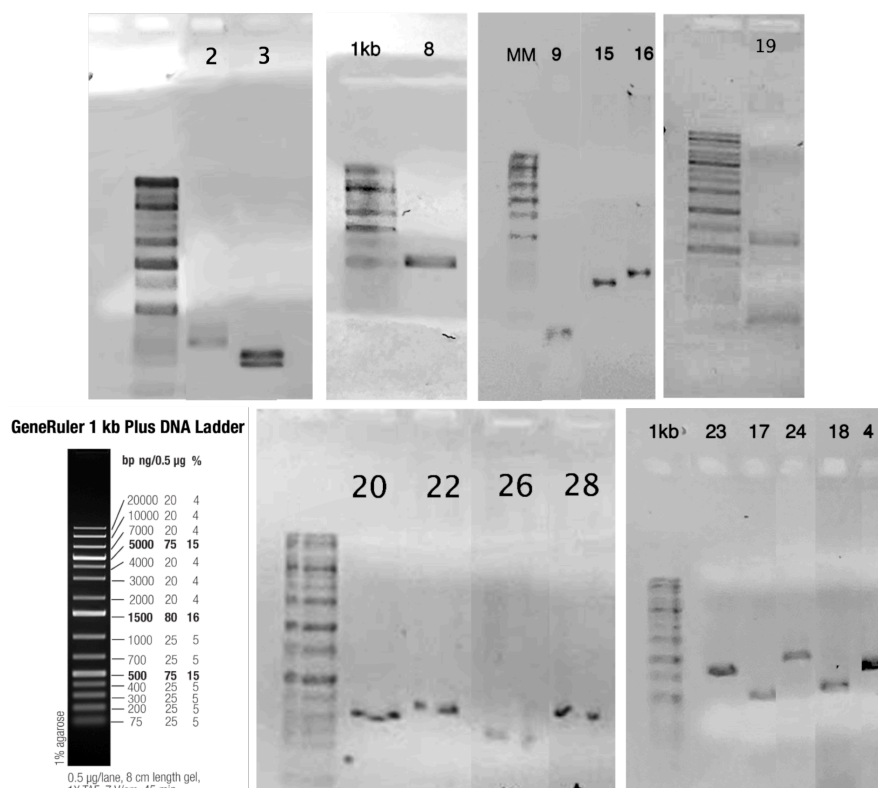


Figura 1. Fragmentos ampliaciones mediante la metodología de la prueba 1.

Nota: El fragmento 19 se amplificó con la taq Onetaq High GC NEB y los fragmentos 2, 3, 4, 8, 9, 15, 16, 17, 18, 20, 22, 23, 24, 26, 28 se amplificaron con la DreamTaq PCR Master Mix (2X). El marcador molecular utilizado fue el GeneRuler 1kb Plus DNA Ladder (Thermo Fisher Scientific Inc.).

Prueba 2

La figura 2 A1 y A2 muestra los fragmentos obtenidos sin touch up con las enzimas High Fidelity Q5 (NEB) y Phire II Hot Start (Thermo Fisher Scientific), estos dieron una baja intensidad, por ende, baja concentración. Luego, en la figura 2B, se puede observar los mejores fragmentos ya purificados que se obtuvieron con la enzima Phire II Hot Start (Thermo Fisher Scientific), con touch up y con DMSO al 10%. Su mejor calidad se evidencia por una mayor intensidad de bandas que indica mayor concentración de producto.

De todos estos fragmentos solo se logró secuenciar correctamente aquellos con mejor concentración (intensidad de banda), estos fueron: los fragmentos F de 8 (figura 2 A.1) y 28 (figura 2 B), y los fragmentos F y R de 23 (figura 2 A.2), 15 y 25 (figura 2 B); es decir la mayoría

se obtuvo con el protocolo con la enzima Phire II Hot Start (Thermo Fisher Scientific) y con touch up. Las secuencias de estos fragmentos pueden observarse en la cuadro 9, y para el resto de los fragmentos obtenidos en esta prueba no se obtuvo secuencias aceptables de analizar.

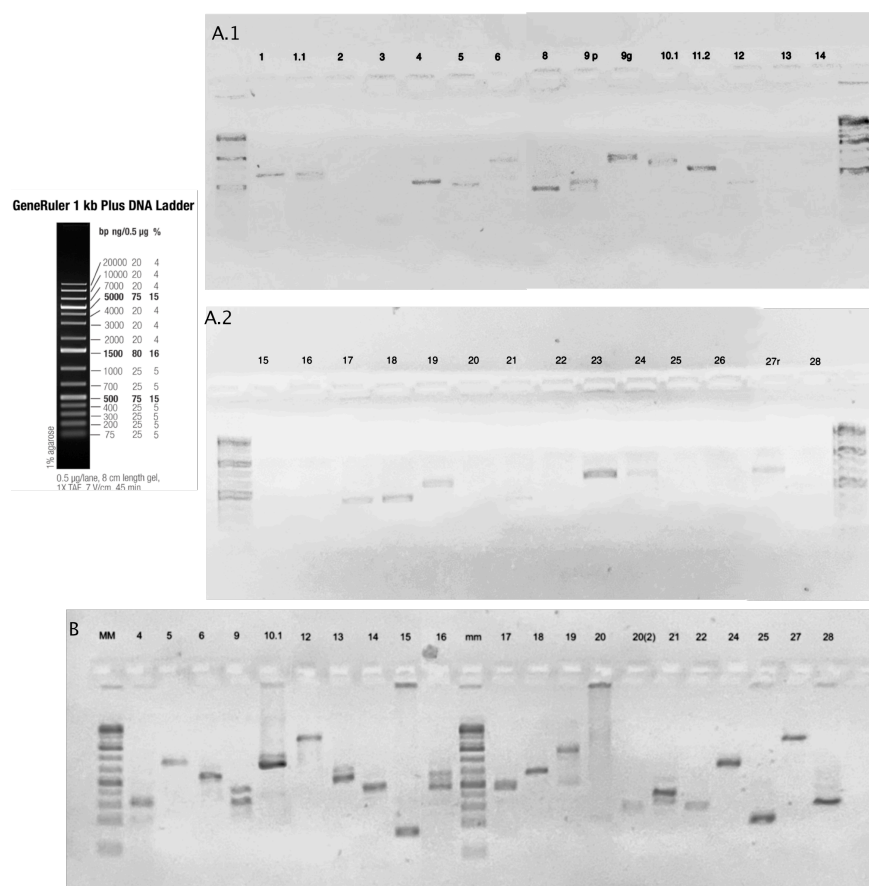


Figura 2. Fragmentos amplificados y purificados con la metodología de la prueba 2.

Nota: Amplificaciones de recuadro B se realizaron con la Phire II Hot Start (Thermo Fisher Scientific) con touch up y con DMSO al 10%. Las amplificaciones de A.1 y A.2 se realizaron con High Fidelity Q5 NEB y Phire II Hot Start (Thermo Fisher Scientific) sin touch up.

Prueba 3

Como se observa en la figura 3 el problema en esta prueba ya no fue la intensidad de las bandas sino la numerosidad de estas. Para este punto aún quedaban 20 fragmentos por secuenciar y mediante la metodología de esta prueba se lograron secuenciar tres fragmentos que presentaron

una sola banda, F y R de 20 y 26, y el fragmento R de 12. El resto de los fragmentos amplificados en esta prueba resultaron en múltiples bandas fueron desechados (10.1, 21, 24 y 27, 1, 5, 4, 22, 13, 14, 17, 16, 18, 7, 6, 9 y 19) para intentar obtener, en la siguiente prueba, productos con menos bandas previo a su purificación. Además, se logró amplificar las combinaciones 1, 6 y 7 ya que en pruebas anteriores su amplificación siempre fue difícil, sin embargo al secuenciarlos ninguno mostró resultados adecuados.

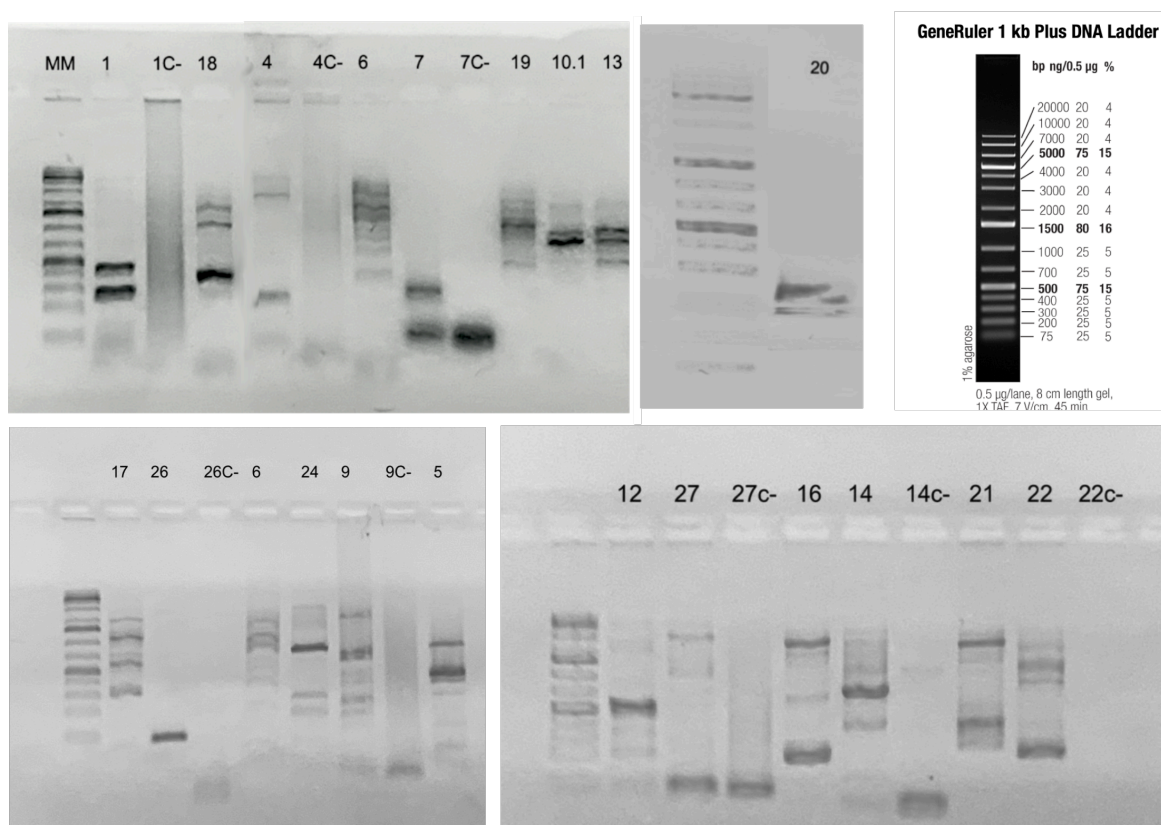


Figura 3. Fragmentos con bandas múltiples obtenidos de la amplificación del plasmoma de *P. costaricensis* con el protocolo de la prueba 3.

Nota: 12, 20 y 26 fueron los únicos fragmentos que rindieron una única banda al amplificarlos con la metodología de la prueba 3.

Prueba 4

Para los 21 fragmentos probados, se lograron obtener bandas para 18, específicamente 10.1, 21, 24, 27, 5, 22, 13, 14, 17, 9, 19, 4, 18, accD, ycf1, yfc2/rpl2, rps4/rps14, y 23s/16s. Los

fragmentos 1, 6 y 7 no se lograron amplificar de nuevo. De los amplificables, 4 fragmentos presentaron una única banda (21, rps4/rps14, 13, y 4) y el resto presentaron bandas múltiples, pero menos numerosas y la mayoría presentó una o dos bandas de tamaño esperado con una alta intensidad (figura 4). Para los fragmentos con bandas múltiples, debido a su mayor intensidad de bandas, se les eluyó y purificó. Además en esta prueba el uso de buffer TE mejoró la cantidad de ADN obtenido luego de la purificación de los fragmentos con bandas múltiples. Sin embargo, de nuevo no todos lograron secuenciarse. En total de estos 18 fragmentos se lograron secuenciar con éxito a F y R de 5 y 4 (Figura 4 C y D); los fragmentos F de 24 y 17 (Figura 4 B y D), y los fragmentos R de 9, 14, y 22 (Figura 4 E y D);. Ninguno de los fragmentos diseñados para secuenciar genes en específico se logró secuenciar exitosamente a pesar de que si se lograron amplificar con la metodología de la prueba 4.

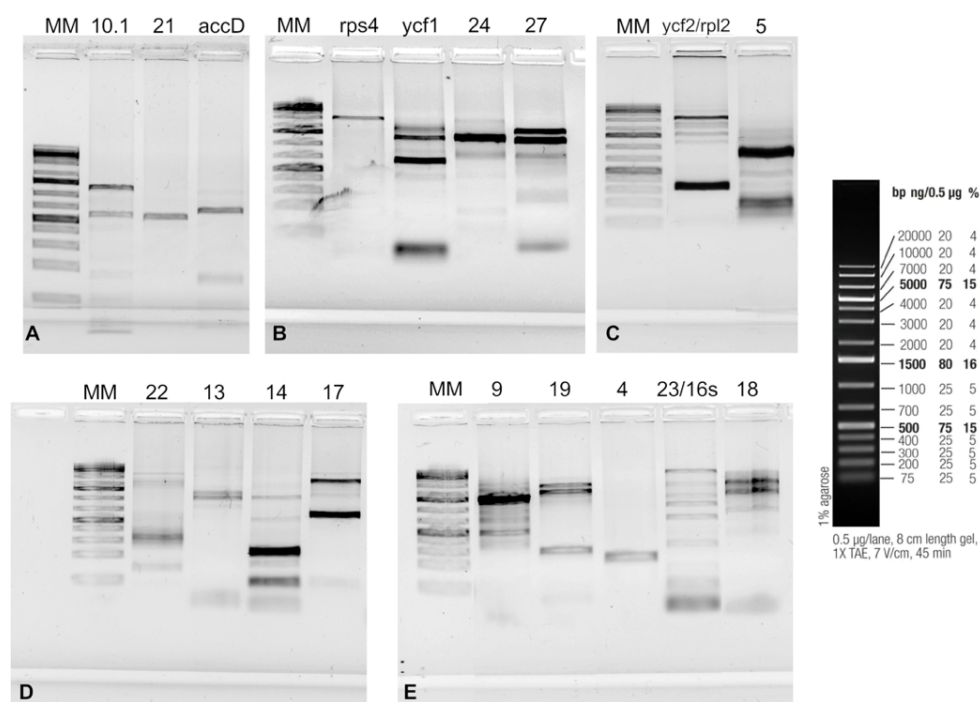


Figura 4. Fragmentos con bandas múltiples obtenidos de la amplificación del plastoma de *P. costaricensis* con el protocolo de la prueba 4.

Nota: Se realizaron 5 corridas de amplificación diferentes (A,B,C,D y E) debido al cambio de termociclador que no permitía realizar multiplex-

En total se lograron secuenciar 18 fragmentos (cuadro 9). Dichos fragmentos se incluyeron en ensamblaje inicial del plastoma, pero aun así no se logró generar un ensamblaje total. Por lo que a continuación se presentan los resultados obtenidos con el ensamblaje *de Novo* por extensión de lecturas con GetOrganelle.

Cuadro 9. Fragmentos obtenidos de las pruebas de amplificación y secuenciación para la determinación de los nucleótidos indeterminados del ensamblaje por referencia de *P. costaricensis*.

Fragmentos (5'-3')	
2	ACCTCTAGACGATAAGGGCAATTAATAATATAAACTTAATAATGTATTGTGCAATGTTGCCTAGTGCGAAGGTGGG AATTGAACCCACGACCTCAAGGTCATGAGCCTTGCAGCTACCAAACTGCTCCACTTCGCGCTATATTTTTATTTAATT ACCAACTGTTATTTAATTACCAACTGATATTTTGTCTCGTAAATTTACATGAATTAATTTGTCATAAAATTAATTT CTAGATTGTCCAAAATAACGATAATAACCTCTAGGACGACCTTTTATCGAACATCG
3	GAGGCCTACGCCCATATAAACTGTCTACCAAGACTGTCTTGTCTAAAAACAAGTTAGAAGTTAACTTAATTAGGG TGGTATTCCTGATGACTTAATTTGCCTCCACCTATATTTACAAAAAAGACCAACATCAATCACAGGGAACAGT AAAGCTTCATAGGGTCTTTCTGTCTGGTGTGAATAGTCCGCATC
4	ACAACATTAGGTTTTGGAGACCTAAATCTTACTAATTGAATTAAGAGCGCGTATAAAAAATATTAATAAAATGAAAAATC AAATAAGAAATAAAAAACTATAAAAAATTTAAATTTTTAAAAATTTAAATATAAATAAAATAAATATAAAGAAAT AAAATTAATTTACAAATTTTTAAGTATAAAAGGAAAAATATTATCAAGAAAAGTAACTAAATTTAAAAAACA AAAAGCATTACTATAGCTATAAAGAGAGCTCGTATATTAATCTTTTACCTTTAATAAATCAATTTACTTAAATAAAA ATAAACGATGCAATAAAAAATTAATAATAAATAAATTTAATTTAATTTAAAAAATTATAAAAAACTAATAATCTTTTAT CCAATATAATAATAAATAAATAAATTAACCTTAAATAACCTTCTTTTTTTTATTAGTTTTAAGTTTATAGTATAAACCAATA CTATATTTTTTATATCGTTTATATGTTAATTAATTTAAACAGTTAGTCTTTTTTGTATAAAAAATTTTGTATAAAA TACCGGAGTTTAAATTTAGTATTAACCTGTATAAATAAATTTTACTCTATAAAAAATAAATTAAGTAACGAT ATTATAAAAAATTAGTTTACTGAACCTTAAATCTGTAAATTTATATAATATAAATATACATTTATAAATTTCCGCTTAT TTAATAAAAAATAAATTAAAAAATGGAATTTAATTTTTAAAACAATTAATAAATAAATTAATAAATAAACATACGAA TCACACATACAATTCATACACATTCCTAAACGT
5	TGCAATGAAGCTTATCCCCATTGTCTCACTAGTTAATTTATAATTGTACAGTATTATTAGTTGTATAAAAAATTTAAAT TTAAACAGTGCTTACCCTGTATCATAATTAACCTGTGTCTCGACACATTTCCGGGGGAACCAGCTAGCTCGGGT TCGAGTGGCATTTCACCGCTATCCACAATTCATCCGCTGATTCTGCATTATCAGTCGGTTCGGACCTCCACTTAGTTTTAA CTAAATTTTCATCCTGACCATAGATAGATCATCCAGTTTCGGGTACATAAGTAATGACTTATATATATATATATATA TTTATTGCCGTTATAGACTTGCTTTCCGCTATGGCTACGACTTATATTCTTAACCAAGCCATTAATTTAGTTCGCTG GCTCCTTCTCAACAAGCACGAGTCAGAGATAATAAAAAATTCCTCCTACAATTTGACAGCTTTTGGTTTCATTTAAAT ATATCTTTTAAAAGTTATATTACCTTTCCCTCACGGTACTAGTTCGCTATCGGTTACTCAGGAGTATTTAGTCTTACAA GGTGGTCTTGTAGATTCACACGGAATTTAACTTGTCCAGGTTACTTAGATAAGTTAATTTAAGTATTAATATAA TTAAGTATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA TTTTATCTGTTTTATCTAATAATATTATTTAAATTTACATAAAAAATTTAGTTTTATATAAATAAATAAATAAATAA TAATGATAATGATTTAGACTATTTCCATTTCCGCTCACCCTACTAAGGAAATTAATTTTGTTTTTATTTCTCTGGATACTA AGATGTTTCAGTTCTCCAGGTTA
8	CAGTCGTCTGGGCTGGTTATTGCAATCCTTTTTAAGGACACCCTTTTCCCGAAGTTACAGGGCTATTTGCGGAGTTCC TTAGAGAGAGTTATCTCGATCCCCTTAGTATTCTACTCATCTACCTGTGTAGGTTTACGGTACAGGTATTTATTTATTT TTTTTCAATTTAGTTTTTCTGGAAGTATGACCTCAGTTATATTAATACAATAACAATACATATAAACGATTTAATAT ACTAGTATATTTTAAAGACTTTACAAAATGTTTAAATTTTTGCTTATATCTTTGACTGATAACCATTTATCAGCTAAC TAGCCTTCTCGTCCCTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA GCCTTATCTTAGGTCCTGACTAACTTTCCGATATAAATAATTTGAAAAGAAATCCTTAGATTGTCGGGCATAGGATTC ACCTATGTTTTGTTACTCAAGCCAACATTTCTCACTTCTGACATAAAAAATAAAGCAGAACGCCCTACCGATTGTTAA TCTCATAGTTTCGGCA

TGAAGCTGACGGAATCATTGTTAATGAAAAAATAAGGTTTTGTCTTATTTTATATCTGTTTTTTTTAAAAATAAAT
 ATTCTGTTATTTTTTATAATTTGTAATAAAAAACGATTTTTCTTTTATTTAATATATTTTGTCTTTGTATTAAT
 CTTAACATATCTAAAAAATATATTAATATAAATAACCATAATCCACAGAGTTTTTCATACTTTTTATAGATATATTTTTA
 AAAACATATTTAATAAGTTTTGTTCATTTTTTATTTAAAAAATATTTATAAATTGAATTAATAAATTTCTTTCTTCA
 TCAAAAAATTAATCATTAGTATTAATCCAAGATTCTAAATCAAATTTGGTTTCATTTTTTTTTAATATAGGTTTAGTATAT
 9 GTAGTTTCTTTTATAGATGTATATTCTTGATCTTTATTTGTTAAGAATACATTGATAAAAAATCTTCAGTTATATCTTTGTGAA
 AAACATTTAAAAAATTTACTTTATGGAGTTTCTTAAATAATAAAAAATCATAAATAAAAAATTTTATACCAATTTTTTAA
 ATTTGATTCAGTATAAAAAATATAAATATTTCTTTATTTAATTTTTTATTTTATATATATATTTTTTAAATTTACTTTTATGA
 TTATAAGTAGTTTTACTTTGTCTATTTCTAAAAATTTAAATAAAATGTCAACAATAAAAAATCGAATAATATATTTATT
 TATTTACAGTTAGTAAATTAATAATTTCAATTTATAAATAAATAAAAAAATATTGTTATAAAGATAAATTTGTGTTTTTCATGC
 ATGCTT

GAGATTGGTTTTACATCATTAGAAGATTTAGTATATAATTTTTTTGTTTCAGAATTTGAATCAAAAATTTAGTGATTTAAT
 ATAAATTTTTTCATATGTAAATAATAGTAAAGAGGAAGAAATTTTTGAAGAAAAAATAAATAAATAAATAAATAAATAA
 12 GAAAAATATTATGATTATATATAGATGAATGTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
 AATATATAAAAAATTTTTTAAATGAAAAATTTGTTAATAATAAATTTTATAAATTTTAAATCATTAAATCATATGAAAAA
 ATTTTAAAAAGAGCTTATTATTTTCAGAATGTACTTATCTAACAATTGATGATTTTATATTTGACTTAGACAAAGAAAAAT
 TTATAAAGATTTTTAAATATAAAAAATATTATAACTTCAGATGAATATTTTATTATCAAAAAAAGAAATTTAAGTGAA
 GTATTAGATAAAAAAGAAAAATTTGATTTTACAAAATTAAGGGTTTTTTGTGAAAACGTGAATTTCTTAAATTTAAAAA
 AGATTTTAAAGTCTAAATTTATACGTTTGTGAAGAATGTGGTTATC

AAGAGGAGCTGATGATTTTGTAAATTATCTTATTTAATTTCTGTTTTTTTTCTAAAAATTTGTTAATGTTTGTGCATAAAT
 14 TACTTTTATTTTTTATATTTATTTTATATTATATATTATTATAAATTATATATAAATTTATATATAAATTTAATAATATATTAAT
 AATATATTAATTTATATATATTAATTATATATTATTTTTTATTTACTTATTTACATATTATCTTACTTTTTTAAAGTAG
 ATTTACAACCATTATGAGAAAAAGGAG

GCACCACTGTTATCTGTTACATTAAGATAAGTTTTGTTGTGAATCATATTTTTTAACCATAAAAAAATTAGGTTTTAATAA
 15 AAAATTTTGTTTTTACAGGTAATTTAGATGAAACTATTTTAAAAGTATTTTTTGCAAGTATTTCCGATACGTTATCTACTT
 CATAAAGTATTCGTCAGGTTAATAACAGCAACCAATTTTCAGTGGAACCTTTACCAGAC

ACATACTAAAGCGGCTACCATTACTTGTATTTAAGCAATAACGATTTATGCAAGCTAAGTCTTCTAAAGAATAAATTAGG
 16 CCAGAGTAATTTTATGAAAAATTTTATTATCTATTTTATTTAATTTATGATCTTGATACTTAATTAATTAATTTATT
 AAATTTATATTTATTTTGTAAATTTTTTAAAAGTTTTAAAATTTCTAATTGTCTACGAAATTTATAAGTTAAAATCGT
 TTCAGGACT

TTAAGCTACAAATCCATCTAGGAATATTTTTTCTTCGATAATAAAAAATTTGAGTTTCTTCATATTTTTTTTTATTTTTCAA
 AAATTTTAAAAATAGGTGGTAGATTTTCATATATATATTTTTTTATTGCTTTTCATTTTCATTAATTTATTATGTCTGAT
 17 TTATTTATATTTATTTTATTTAATTTATCAATTTCAATAAATTTTATGATGATTTATTTATATTTATTTTTCTTTATT
 TTCATTAGTTTTTCTTTTAAAAAATACCCTTTAAAATATCATCTAATTTTATTAAGCTTTTTTTCATTATCTTTATCA
 AAATAAATAAAAAATCTATTTTTTATTTCAATAAATTTAAATTTTTTGATTTGATCAAAAAGATTCTATACGTTTTTAAAT
 TCATTTCAAAAATTTAAGTTTTTTTTTAAATTAATCTGTCATAAGGATATATTTTTTTTTAAAAACCTTGTGTTTTTT
 TTATTAATTTAAAAAGGCATACAAATTTGGTGAAA

TTCATCCATAATTTGTGATTCTAGTATTATTAATACTAGCTAAAATAATAAATTCCTTTTTCAAGGAAATTTATTTTTGG
 20 TTGGCGCTTATGACGATAAAGAACTTGTATAGGTTTTATTTAATTTCTTTTTCTCATATATTTTTTATTTTTGTTTTGTTT
 ATGCTTTTATTATTACAAATAATTAATAACGCCTATATCTTTAACTAATTTACATTTCTGGCAAATTTTTTCGAATAGA
 TGTTCTTTTTTTCATTTTTATACAACATAAAAAATAAATTTCCCTCCAA

TTACGTATTGATGCTTTTCTTTTTACGTTTTTTTTTAATATTTGTGTTTTACTATTTTTTTTTAGTTATTTTATTTAGT
 22 GCTTTTTGTACAACATAAATTTGGATTTAATTTTATTTTTTTTTTATTCTAGTAAAGATTTATAAATTTATTTATATGATA
 TTGATTTTTTCTTTTCTTAAATGTAATTAGATAGATTATAAATTTGATTTTTTAAAAAATTTGGATCATATTTTACTTT
 ATAAGTCATTTTTATTTTTAATTTTTAATTTTTGTTTTTTAGCTCCATATTTGAACGCTTTTGAA

GATATACTCAATCTCTTAAATATTGGGTTTGGTAAATTAATAAATTTGAAATATAAAAAAGTTTTATAATTATATC
 TAAAAATAAATTAATTTAAATTTAATATCTTTATAAATTTTAAATTAATATATTATAAATTTTCGATATATTTTTTAA
 23 TTTTATTTTCATTTCAATATAGTATGATTTAAATACATATAAAGTTTTATAGTTTAAATTTTATAGTATTTTTTTCATAA
 ATATATTTACCTTTAATAAAGAAAAATAAAGTTTTATTTTCGGTCTCATTTTAAATAATTTTTAATATGTGTTTTTTTT
 GTTTTTTAAATATAAATTTGGTTCTAAATATCAAAAATTTAATCTAAAACCGTATAAAAAGTTAATTAATAAATATATC
 CAGTATATAACCTTGTGTAATATACCTTTATAAAAAAGTATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA
 AAAAAAGAAACAAAGCGTATAATATATAGATAAACTCATTTAATTTGACCTAAAACGTAGAGCTTATTATCACAAA
 TATTTTTCTGTTTTTTTTCAAAATCTTTTTAAATAAGTGCTTATAAAAAAATAAAGAAAGGTATCACACAGAATG
 AACTAAAAAACCAAAATAAAAACTATAAAAAGATAAAAAAAGTTTATTATAAATAAATTTCTTAATATTAATA
 AAAAAATGATTATTACTTTACCATAATTTTTTATTTGAAATGTTATTTAAGTTGTATTTTTTAAATTAATAAAGG
 TCCCACCACTAGTTTTTTTTATTTCGGCATAAATTTTTTTCCATAAGATTTCTCTTAAAGTATCG

24 GTCATCGTATTCGGGCAACTTTTCTTGAGCCAGAATTAGGTTTTTTGGATGAATAATGTAGTTAACAGATAAAAAACT
 GTCACCTTACAGAACCTATACATGAAAGTTTCCTTTCATATAGCTCCTCATTAAATATTTATTAACCTTAAATTTGTT
 TTTATTAATAAATTATTAATAAAGCATAAATTATATATGTAAAAATTTTATATTTATTTATTTATTTATTTTATTTTATTT
 ATAATTTATATAATTTATTTATAATTTATAATTTATTTTATATATTTTATATTTAATTTAATTTAATTTAATTAAGTTTA
 ATAATTTAAACGGGTTAATGTAAGTAATTTAAATTTATATCTTAATATTTTTTATTATTTAATTTATTTTTAAAAATTA
 ATATTTTATATTTTATTTGCTTTAATACTTAATTTTATATATAATTTATTTTATAGTTTATATTATTATTTAATTTTAAATTA
 AAATATTAAGATTAATAATGAATATAAATATAAAAAATATATAACAATATAACAAATAATATTATAATTATAAAATTA
 AAAATTTAAATTTAATTTAAATTTAAATAAGGTCTTGAAAAATTTTGGCCATAAAAAATCTTCTTAATTTAATTTTAA
 AAGTTTATTAATAAGTTTACTATTATTATATAATTTTTGTAAATAATCAATTGCTTCATAATGAATTATAATTGAAAT
 ATTATGTTTATTCTTATATTATAATTAGATTTGTAAGAATCAGATTTATTATTAACATAAAAAATTTTATATAGTAGAGG
 TTTTAAATGTTATATACAAAAATCTTAGTTGGTCAGAGAATCTATAATCTTCATACATAGCATAACATAAATTTTTCAT
 AAGGAATATAAATATCTTTTTTTTTTTTTTTTATTTGATGGGTTATTAATTGGTAAAGTAC

25 CAAGCTATTGTTAATAAAGAAGCTATTTTTGTAATATCTATAATATAATTTTTAAATTAATATTAATAAATAATATAT
 CATTTTAAATTTAATATTTTTGTATTTAAAAAATGAATTTTTCTGTTAACATATCATTTAATTTTATTTTCCAATCTATT
 TTTTTTTCATTTAACATTTTAAATTTAACATTTTCATTTAACATTTTAAATACAAATCCATTTATACTATCTTTTATA
 TAAATGGATTGTGGATCGAGTTGGATTTGAACCAACGTAGACACATAAGTCAA

26 TAAATTCGTCAGAATCAAAATGTTTTTTTTCTTTTTTAAAAAGTATATTTGAATAGATAAAATTTCTTTATTGTTTTTATT
 TTTAATAGTTAACTTAATCGGTAATATTTTTTCATTTAACTATATTTTTTAAATAAATTTATAGATTGATAATTTTTTTTT
 ATATTTTTTTTTATTTAAAAAATTTCTGGTATGAAATTATTAGAAT

28 TTCTGAAAATCTAATATCATGCAATTCCTTTTTTTTTTTCGTCTTCTCAATCTTTTTTCAAATTTATAACCAAAAAAG
 TTATTTTATCCACTTTTTATATTCTTTATCTTCAATTTTTTATTACATTTAAAGAGTCAGATATACAAGCTAATTTTA
 ATTTACTTAAAGGTACTTTATATTCTATTTTAAATTTTTTATCTGTTATTTCTTTATCTTTATCTGTTATTTCTTTATCTTCT
 TATCTTTATCTGTTATTTCTTTATCTTTATCTGTTATTTCTTTATCTTTATCTGTTATTTCTTTATCTTTATCTTTATCTT

Nota: Fragmentos del primer ensamblaje que se lograron secuenciar correctamente.

Ensamblaje por GetOrganelle

Por medio de la secuenciación de los 18 fragmentos que se lograron secuenciar se disminuyó el tamaño total del ensamblaje provisional de 30 192 pb a 28 489 pb generándose un ensamblaje intermedio (Apéndice 1.2). Algunos de los gaps del ensamblaje inicial eran en realidad un artefacto del ensamblaje y en su lugar no había nucleótidos detectables. En este ensamblaje intermedio, a pesar de los múltiples esfuerzos realizados, aun se contaba con 5110 pb desconocidos. Por lo que este solo contenía el 82% del genoma plastidial de *Prosopanche costaricensis*.

Para identificar estos gaps en el ensamblaje intermedio, se utilizó el ensamblador de Novo de GetOrganelle con los genomas de *Hydnora visseri* y *Prosopanche americana* como referencias para la semilla de la extensión de lecturas. De esta manera se lograron ensamblar 5 scaffolds con los siguientes tamaños: 3 702 pb, 4 774 pb, 5 736 pb, y 14 730 pb (Figura 5).

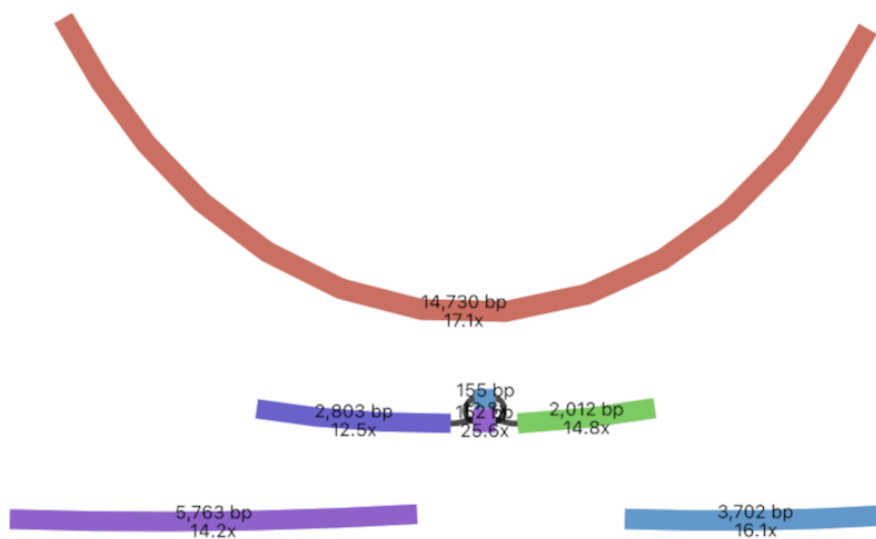
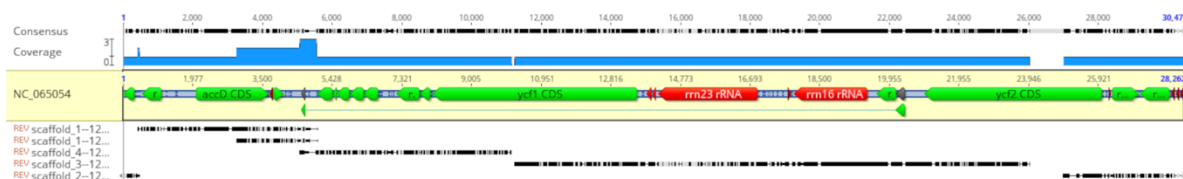


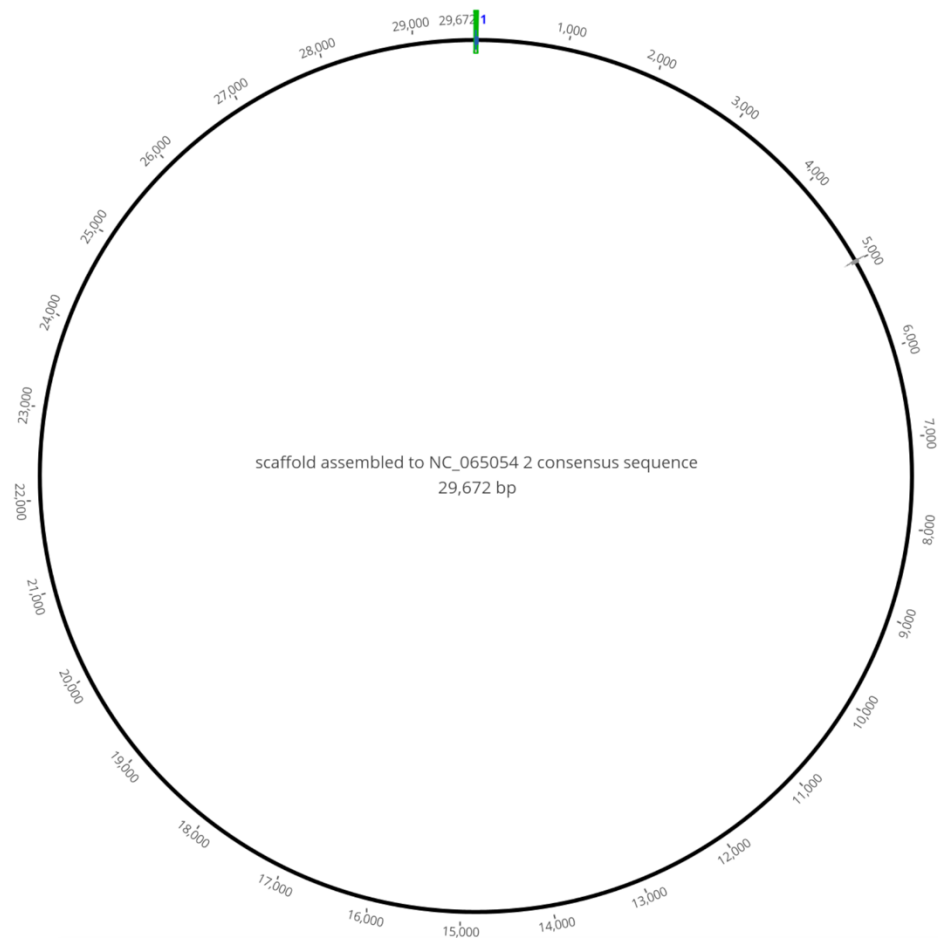
Figura 5. Scaffolds obtenidos del ensamblaje de Novo del plastoma de *Prosopanche costaricensis* por GetOrganelle.

Nota: Cada scaffold presenta su longitud y cobertura. La estructura del scaffold de 2803 pb - 152 pb / 152 pb - 2012 pb indica una posible existencia de repetición de secuencias (152 pb y 152 pb).

Al mapear estos scaffolds contra el plastoma de *P. panguanensis* (Figura 6A) se logró obtener un ensamblaje circularizado con solo dos gaps (Figura 6B) de 29 672 pb.



A



B

Figura 6. Mapeo por referencia de los scaffolds de *Prosopanche costaricensis* contra el plástoma *P. panguanensis* y la secuencia consenso circularizada.

Nota: En la sección A: los scaffolds son las secuencias representadas en negro. El genoma de referencia NC_065054 es el de *P. panguanensis* con sus respectivas anotaciones. La cobertura del mapeo se ve representada por la zona celeste y arriba se ilustra la secuencia consenso. En la sección B se muestra la secuencia consenso circularizada, y su tamaño.

Los dos gaps se lograron corregir mediante alineamientos gap por gap entre: i) el primer ensamblaje que incluye los fragmentos secuenciados, ii) el genoma de *P. panguanensis*, iii) este ensamblaje. Además, al identificar las repeticiones directas se terminaron de cerrar los gaps. De esta manera se obtuvo el ensamblaje final anotado de *P. costaricensis* con un tamaño total de 28 705 pb (Figura 7).

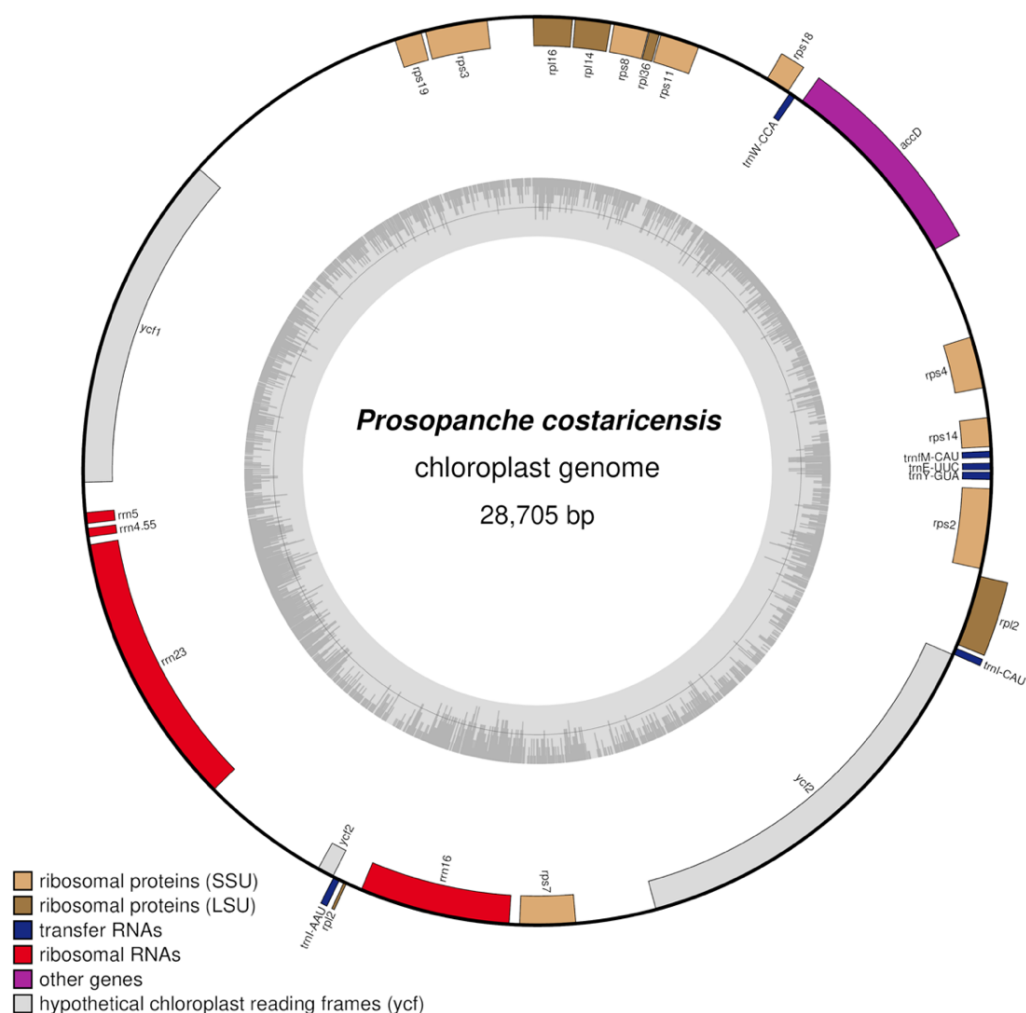


Figura 7. Mapa circularizado del genoma plastidial parcial de *Prosopanche costaricensis*.

Nota: Los genes en el círculo externo son aquellos que se transcriben en la dirección de las manecillas del reloj, y los del círculo interno se transcriben en contra las manecillas del reloj. El segundo círculo interno, color gris, contiene la representación visual del contenido GC a lo largo de todo el plastoma. El plastoma contiene 27 genes únicos: 14 genes de proteínas ribosomales, 4 genes de ARN ribosomales, 6 genes de ARN de transferencia (*trnI-AAU*

pseudogén), 3 genes de proteínas con otra función o con funciones desconocidas (*ycf2* pseudogén) y 1 gene trans-empalmado (*rps12*) en dos exones.

En la figura 7, se puede observar el mapa circularizado del ensamblaje final. Aquí se evidencia como el plastoma detectado carece de la estructura tradicional cuatripartita (repeticiones invertidas, única copia pequeña y única copia larga) de los genomas plastidiales en plantas autótrofas; pero si contiene repeticiones directas (*ycf2-trnI-rpl2*).

Además, en este plastoma final se logró identificar que de todos los imprimadores diseñados para el primer ensamblaje, solo tres no funcionaban para ensamblaje final (6R, 13R y 21F) (ver figura 8).



Figura 8. Alineamientos de los imprimadores diseñados para el primer ensamblaje que no se alinean correctamente en el ensamblaje final.

Nota: El imprimador 6R tiene una diferencia de dos nucleótidos entre el ensamblaje inicial y el final. El imprimador 13R tiene una mayor diferencia de nucleótidos (10) entre el ensamblaje inicial y el final. El imprimador 21F muestra una diferencia de 3 nucleótidos entre el ensamblaje inicial y el fina.

Por otro lado, los gaps del ensamblaje inicial se lograron corregir en el proceso de ensamblaje final. En el primer ensamblaje se tenían cuatro gaps grandes (> 300 pb) y 12 gaps pequeños (< 300 pb). Tres de los grandes gaps se encontraban en las secuencias de los genes con otra función (*accD*) o con funciones desconocidas o de transporte de proteínas (*yfc1* y *yfc2*), y uno entre los dos genes ribosomales; mientras que todos los gaps ligeros se encontraban en zonas intergénicas.

Específicamente, los gaps se encontraban distribuidos de la siguiente forma: El gap grande 1 de 802 pb estaba en medio del gen *accD*, el gap grande 2 de 358 pb estaba ubicado en el gen *yfc1*, y el gap grande 3 de 2746 pb estaba el final de la zona codificante de *yfc2* y la zona intergénica previa al gen *rpl2*. El gap ligero 1 de 65 pb estaba entre los genes *rps4* y *rps14*. El gap ligero 2 se encontraba entre los genes *rps14* y *accD* y era de 81 pb. Los gaps ligeros 3 y 4 son eran de 3 pb cada uno y se encontraba entre *accD* y *trnW-CCA* y entre *trnW-CCA* y *rps18*. Luego, en las zonas intergénicas de las parejas de genes *rrn4.5* y *rrn5*, *rrn4.5* y *rrn23*, *rrn16* y *yfc2*, y *trnY-GUA* y *trnE-UUC* se encontraban los gaps ligeros 8, 9, 10, y 11 que todos eran de 1 pb, respectivamente. El gap ligero 5 era de 11 pb y estaba entre los genes *rpl36* y *rps3*. El gap ligero 6 era de 2 pb y se observaba entre los genes *rps3* y *rps19*. El gap ligero 7 estaba entre *rps19* y *yfc1* y contaba con 4 pb desconocidos. Por último, el gap ligero 12 se encontraba entre los genes *rpl2* y *rps2* y era de 41 pb.

La corrección de estos gaps por medio de alineamientos permitió la determinación de las repeticiones directas de *P. costaricensis*. Estas repeticiones se confirmaron en *P. costaricensis* por alineamientos entre las zonas donde se esperaba que estuvieran de acuerdo con los genomas de las otras especies del mismo género y así fueron asignadas correctamente. La repetición directa A de *P. costaricensis* presenta 330 pb, donde el fragmento 5' de *yfc2* cubre 176 pb, el *trnI* cubre 73 pb y el fragmento 5' de *rpl2* abarca 36 pb, y la repetición directa B, cuenta con 318 pb, donde de igual manera el fragmento 5' de *yfc2* cubre 176 pb, el *trnI* cubre 73 pb y el fragmento 5' de *rpl2* comprende 36 pb. No obstante, *P. costaricensis* presenta dos secuencias diferentes para el *trnI* (*trnI-AAU* repetición A y *trnI-CAU* en la repetición B), ocasionado por tres substituciones encontradas en la secuencia de *trnI* de la repetición directa A con respecto a la secuencia de la repetición B.

Adicionalmente, el poder corregir los gaps permitió la anotación del plastoma final de *P. costaricensis* donde se identificaron un total de 27 genes únicos. Se lograron anotar 6 ARNs de transferencia, 4 ARNs ribosomales, 3 genes de proteínas con otra función o con funciones desconocidas y 14 genes de proteínas ribosomales: *trnE-UUC*, *trnM-CAU*, *trnW-CCA*, *trnY-GUA*, *trnI-AAU*, *trnI-CAU*, *rrn4.5*, *rrn5*, *rrn23*, *rrn16*, *accD*, *ycf1*, *ycf2*, *rps14*, *rps4*, *rps18*, *rps11*, *rps12* (trans-empalmado), *rpl36*, *rps8*, *rpl14*, *rpl16*, *rps3*, *rps19*, *rpl2*, *rps2*. Lo cual permite afirmar que el plastoma de *P. costaricensis* carece de genes implicados en la fotosíntesis, como los genes de los fotosistemas I y II, subunidad grande de RuBisCO, ATP sintasa, genes de NADH deshidrogenasa, así como subunidades de ARN polimerasa, *clpP* y *matK*.

En el caso de genes con diferencias en sus marcos abiertos de lecturas, se encontró que el *ycf1* presentó un marco abierto de lectura más corto en comparación al de los genes de las especies del mismo género, presentando un codón de terminación más temprano. Mientras que el gen *ycf2* no contaba con un codón de inicio en su marco abierto de lectura, por lo que se le clasificó como un pseudogén.

Con respecto a otros pseudogenes, en *Prosopanche costaricensis* existía la posibilidad de que el ARN de transferencia *trnW* fuera un pseudogén, debido a que en *P. americana* el ARN de transferencia para el triptófano (*trnW*) se clasifica como pseudogén porque su estructura secundaria presenta como anticodón la triada ACA en vez de la triada CCA. La sospecha se eliminó en *P. costaricensis* ya que el *trnW* si presentó el anticodón tradicional CCA (figura 9). Por otro lado, está el caso del ARN de transferencia para isoleucina ubicado en la repetición A de *P. costaricensis*. De acuerdo con la predicción de su estructura secundaria y anticodón este es el ARN de transferencia *Ile-AAU* (Ile2) (Figura 9), por lo que es diferente al anticodón y estructura del *trnI-CAU* de la repetición B.

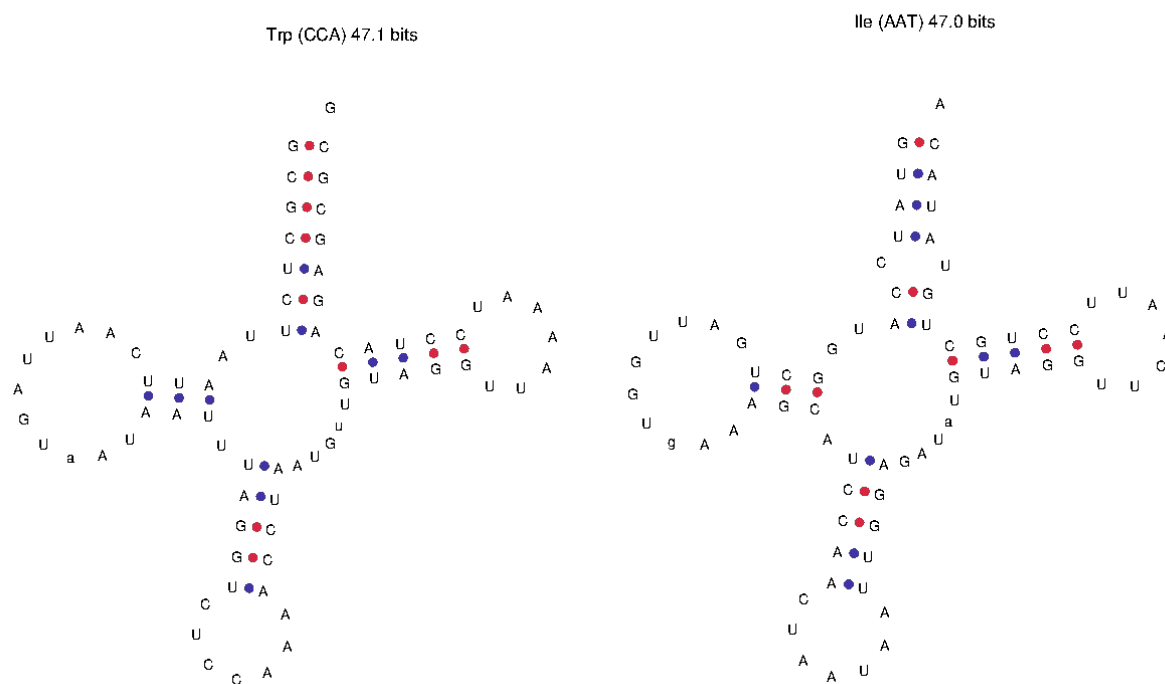


Figura 9. ARNs de transferencia de triptófano e isoleucina2 de la repetición A de *Prosopanche costaricensis*.

Nota: La estructura secundaria del *trnW-CAA* está al a derecha y a la izquierda se presenta la estructura secundaria del *trnI-AAU*^(Ile2).

Ahora bien, el contenido GC del plastoma final es de 19.9%. Las regiones codificantes, que son el 73.1% del genoma total, es decir 20 974 pb contienen un 22.7% de GC; los ARN ribosomales que comprenden 4 574 pb es decir el 15.9% del genoma total, contienen un 37.1% GC; y las regiones no codificantes que representan 26.9% del genoma total con 7 731 pb tienen un 12.14% de GC. Por lo que zonas codificantes de ARN ribosomales son las que cuentan con el porcentaje más elevado de GC de todo el plastidio.

Comparación de Plastomas y Análisis de Selección

Comparación de Plastomas

El plastoma de *Prosopanche costaricensis* (28 705 pb) cuenta con una mayor longitud al compararlo con sus especies hermanas, este tiene 514 pb más que el de *P. americana* (28 191 pb), 459 pb más que *P. bonacinae* (28 246 pb) y 443 pb más que el de *P. panguanensis* (28 262 pb) (cuadro 10).

La estructura del plastoma de *Prosopanche costaricensis* coincide con la de sus especies hermanas: *P. americana*, *P. panguanensis* y *P. bonacinae*. Tanto la ubicación y dirección de transcripción de sus genes así como el número de genes y estructura general entre las 3 especies es idéntica (cuadro y figura 10). Dentro de las diferencias entre los genes de estas especies está el gen *trnI-AAU* que solo se encuentra en *P. costaricensis* en la repetición directa A, el cual se considera un pseudogén por su codón diferente para *ile2*, y luego está el tamaño de los genes *ycf1* y *ycf2* que en *P. costaricensis* presentan un menor tamaño (*ycf1* 2000 pb menos y *ycf2* 1000 pb menos) en comparación a estos genes en las especies del mismo género. El tamaño del gen *ycf1* en *P. costaricensis* es menor que en las otras especies, ya que este gen tiene un marco abierto de lectura más corto. En el caso de *ycf2*, en *P. costaricensis* no tiene un codón de inicio en su marco abierto de lectura mientras que en las otras especies sí (figura 10). El resto de los genes de las 4 especies del género cuentan con tamaños muy similares, con diferencias de +/- 10 a 200 pb (cuadro 11).

Adicionalmente, existen diferencias con respecto a las repeticiones directas en estas especies. En *Prosopanche americana* y *P. panguanensis* las repeticiones directas son iguales, estas contienen el *trnI* y el fragmento 5' de *rpl2*, mientras que las repeticiones directas de *P. bonacinae* y *P. costaricensis* contienen el *trnI*, el fragmento 5' de *rpl2* y el fragmento 5' *ycf2*.

Cuadro 10. Características de los plastomas altamente reducidos de holoparásitas estudiadas y de la especie fotosintética más cercana que fue analizada en detalle.

Especie	Accesión	Longitud (pb)	% GC	Genes totales	Genes que codifican proteínas	Genes de ARNr	Genes de ARNt	Presencia de genes fotosintéticos
<i>Aristolochia contorta</i>	NC_036152.1	160 576	38.30%	129	84	8	37	si
<i>Cuscuta erosa</i>	NC_042948.1	60 959	36.90%	71	36	8	27	si
<i>Conopholis americana</i>	NC_023131.1	45 673	33.90%	52	29	4	19	si
<i>Prosopanche costaricensis</i>	-	28 705	19.9%	27	17	4	6	no
<i>Prosopanche panguanensis</i>	NC_065054.1	28 262	20.80%	27	17	4	6	no
<i>Prosopanche americana</i>	MT075717.1	28 191	20.40%	27	17	4	6	no
<i>Prosopanche bonacinae</i>	NC_065053	28 256	23.9%	27	17	4	6	no
<i>Hydnora visseri</i>	NC_029358.1	27 233	23.40%	25	17	4	4	no
<i>Mitrastemon kanehirai</i>	MF372930.1	25 740	22.40%	24	18	2	4	no
<i>Cytinus hypocistis</i>	NC_031150.1	19 400	29.90%	26	16	4	6	no
<i>Balanophora reflexa</i>	KX784266.2	15 507	11.60%	19	15	3	1	no
<i>Balanophora laxiflora</i>	KX784265.1	15 505	12.20%	19	15	3	1	no
<i>Pilostyles hamiltonii</i>	NC_029236.1	15 167	22.70%	7	4	2	0	no
<i>Pilostyles aethiopica</i>	NC_029235.1	11 348	24.20%	8	6	2	0	no

Nota: en el conteo de genes se incluyen todos los genes no solo los genes únicos. Es decir se contaron las múltiples copias de genes los presentes en repeticiones directas o inversas en los plastomas.

Cuadro 11. Genes plastidiales en *Prosopanche* con sus respectivos tamaños y orientación.

Nombre completo del gen	Gen	Dirección	Especie			
			<i>Prosopanche costaricensis</i>	<i>Prosopanche americana</i>	<i>Prosopanche panguanensis</i>	<i>Prosopanche bonacinae</i>
			Tamaño pb	Tamaño pb	Tamaño pb	Tamaño pb
Proteína hipotética RF2	<i>ycf2</i>	antisentido	4074	5181	5046	5049
Proteína hipotética RF1	<i>ycf1</i>	antisentido	3678	5322	5379	5337
ARN de transferencia Tirosina	<i>trnY-GUA</i>	antisentido	85	85	85	85
ARN de transferencia Triptófano	<i>trnW-CCA</i>	antisentido	74	73	74	74
ARN de transferencia Isoleucina *	<i>trnI-CAU</i>	sentido	73	73 (2x)	73 (2x)	73 (2x)

ARN de transferencia Isoleucina**	<i>trnI-AAU</i>	sentido	73	-	-	-
ARN de transferencia N – Formilmetionina	<i>trnFM-CAU</i>	antisentido	73	72	73	73
ARN de transferencia Ácido Glutámico	<i>trnE-UUC</i>	antisentido	73	73	73	73
ARN ribosomal 23S	<i>rrn23</i>	antisentido	2815	2852	2746	2951
ARN ribosomal 16S	<i>rrn16</i>	antisentido	1543	1563	1536	1543
ARN ribosomal 5S	<i>rrn5</i>	antisentido	119	121	119	121
ARN ribosomal 4.5S	<i>rrn4.55</i>	antisentido	97	97	97	97
Proteína ribosómica S19	<i>rps19</i>	antisentido	282	279	282	369
Proteína ribosómica S18	<i>rps18</i>	sentido	252	279	279	291
Proteína ribosómica S14	<i>rps14</i>	antisentido	303	312	303	303
Proteína ribosómica S12 (transempalmado)	<i>rps12</i>	antisentido	390	384	390	390
Proteína ribosómica S11	<i>rps11</i>	antisentido	408	408	408	408
Proteína ribosómica S8	<i>rps8</i>	antisentido	360	360	360	360
Proteína ribosómica S7	<i>rps7</i>	antisentido	576	552	522	570
Proteína ribosómica S4	<i>rps4</i>	antisentido	534	534	534	525
Proteína ribosómica S3	<i>rps3</i>	antisentido	642	636	642	636
Proteína ribosómica S2	<i>rps2</i>	antisentido	822	846	807	825
Proteína ribosómica L36	<i>rpl36</i>	antisentido	114	114	114	114
Proteína ribosómica L16	<i>rpl16</i>	antisentido	405	414	405	402
Proteína ribosómica L14	<i>rpl14</i>	antisentido	369	369	369	369
Proteína ribosómica L2	<i>rpl2</i>	sentido	731	738	741	732
subunidad beta de la acetil-CoA carboxilasa	<i>accD</i>	sentido	2022	1851	1584	1551

Nota: El ARN de transferencia Isoleucina con un asterisco (*) significa que es el ARNt que se encuentra duplicado en las repeticiones directas de todas las *Prosopanche* menos *P. costaricensis*. El ARN de transferencia Isoleucina con dos asteriscos (**) significa que es el único ARNt con este anticodón en la repetición invertida de *P. costaricensis*.

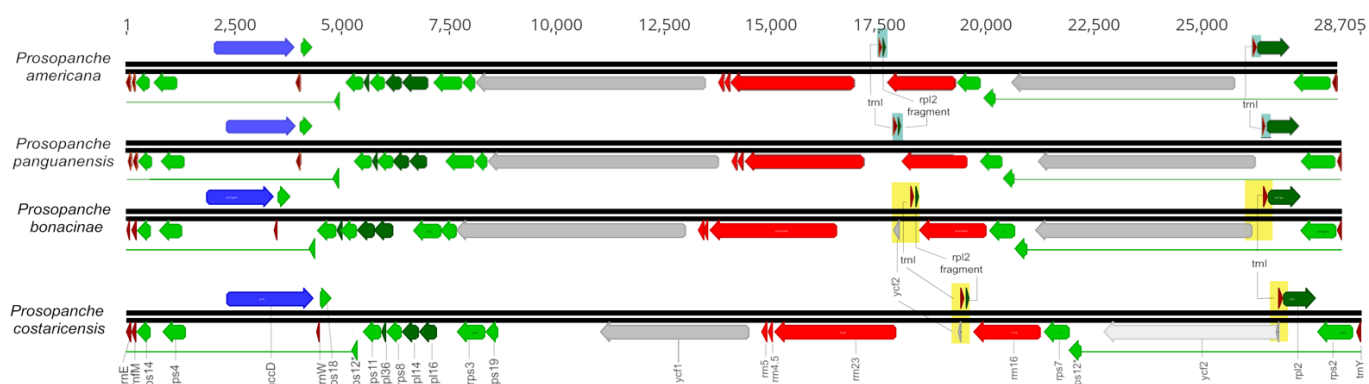


Figura 10. Comparación de los plastomas de *Prosopanche americana*, *P. panguanensis*, *P. bonacinae* y *P. costaricensis*.

Nota: Los genes se representan por bloques con flechas que indican la orientación de su transcripción. Los colores de los bloques corresponden a diferentes grupos de genes: *accD* es azul, el grupo de *ARNt* es rojo oscuro, los *ARNr* son rojo claro, los genes *rpl* son verde oscuro, el grupo de *rps* es verde claro (en *P. americana* el gen *rps19* es un pseudogén), los *ycf* son gris y las repeticiones directas son celestes para *P. americana* y *P. panguanensis*, mientras que para *P. bonacinae* y *P. costaricensis* son amarillas. El asterisco del gen *rps12* indica los exones del gen transempalmado. El gen *ycf2* de *P. costaricensis* es un pseudogén, por lo que se ilustra en la imagen con un gris más claro. El *trnI* de la repetición directa A de *P. costaricensis* es un pseudogén.

Al comparar el plastoma de *Prosopanche costaricensis* con el de *Hydnora visseri*, se observa que el de *P. costaricensis* es más grande por 1 472 pb. Con respecto a genes, *H. visseri* presenta dos genes menos que *P. costaricensis*, este carece del *trnW* y el *trnY*. Además, la mayoría de los genes de *P. costaricensis* son putativamente funcionales a excepción de *ycf2* que no tiene codón de inicio en su marco abierto de lectura y *trnI-AAU* con su anticodón no tradicional para Ile2. En contraste, en *Hydnora visseri* el único gen que no es putativamente funcional es *rps7*.

Con respecto a estructura, el plastoma de *Hydnora visseri* presenta el típico acomodo cuadripartito, mientras que *Prosopanche costaricensis* tiene repeticiones directas en vez de invertidas. Las repeticiones invertidas de *H. visseri* están compuestas por un fragmento 5' del gen *rpl2* de 111 pb, un fragmento 5' del gen *ycf2* de 951 pb y el ARNt *trnI*. Dicha composición es similar a la de las repeticiones directas del género *Prosopanche* que contienen el *trnI*, el fragmento 5' de *rpl2*, y el fragmento 5' de *ycf2*. Luego, la copia larga de *H. visseri* tiene los genes

rps12 y *rps7* en dirección opuesta a como se encuentran en *P. costaricensis* y en ésta misma región a *H. visseri* le faltan los genes *trnW* y *trnY* que si se observan en *P. costaricensis*. En la copia corta *H. visseri* presenta otra inversión con respecto a *P. costaricensis* con los genes *rps2* y *rpl2* (Figura 11).

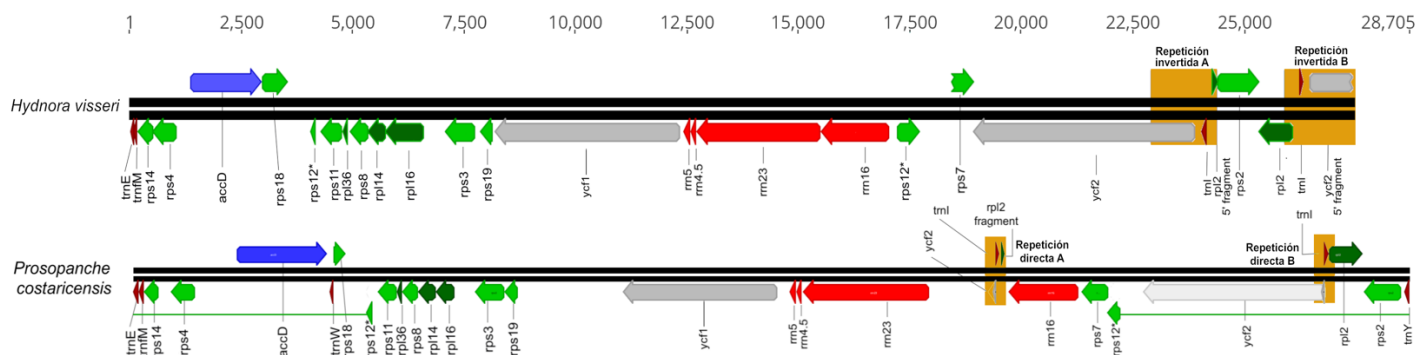


Figura 11. Comparación del plasmoma de *Hydnora visseri* y *Prosopanche costaricensis*.

Nota: Los genes son bloques con flechas que indican la orientación de su transcripción. Los colores de los bloques corresponden a diferentes grupos de genes: *accD* es azul, el grupo de *ARNt* es rojo oscuro, los *ARNr* son rojo claro, los genes *rpl* son verde oscuro y los *rps* son verde claro, los *yef* son gris y las repeticiones invertidas son anaranjadas. El asterisco del gen *rps12* indica los exones del gen trans-empalmado. El gen *ycf2* de *P. costaricensis* es un pseudogén, por lo que se ilustra en la imagen con un gris más claro. El *trnI* de la repetición directa A de *P. costaricensis* es un pseudogén.

En el caso de *Prosopanche costaricensis*, comparado contra los plasmomas más pequeños de *Pilostyles aethiopica* y *Pilostyles hamiltonii* (Cucurbitales), el plasmoma de *Prosopanche costaricensis* es casi el doble de grande que ambos, superándolos en tamaño por 17 357 pb y 13 538 pb respectivamente. En el caso de estructura, los plasmomas de *Pilostyles aethiopica* y *Pilostyles hamiltonii* no tienen repeticiones invertidas ni directas y no retienen dos de los genes que son usuales en las repeticiones invertidas y directas en las especies de la familia Hydnoraceae y uno es un pseudo gen (*ycf2* y *trnI* ausentes y el gen *rpl2* es un pseudo gen). Las diferencias en la cantidad de genes entre las especies del género *Pilostyles* y *Prosopanche costaricensis* también es amplia, *Pilostyles aethiopica* y *Pilostyles hamiltonii* tiene 7 y 8 genes totales (contando pseudogenes) respectivamente, mientras que *P. costaricensis* conserva 27 (cuadro 10). Los genes

que comparten las tres especies son 6: *rpl2*, *rps3*, *rrn23s*, *accD*, *rps4*, y *rrn16s*. En el caso de *rps12*, *P. costaricensis* y *Pilostyles aethiopica* lo conservan completo, pero este es un pseudogén *Pilostyles hamiltonii*. Por otro lado, *Pilostyles aethiopica* cuenta con el gen *rpoC2* que no está presente en ninguna de las otras dos especies. En este caso, tanto la ubicación como la dirección de los genes compartidos entre las de las especies del género *Pilostyles* y, *Prosopanche costaricensis* son muy diferentes, como se muestra en la (figura 12).

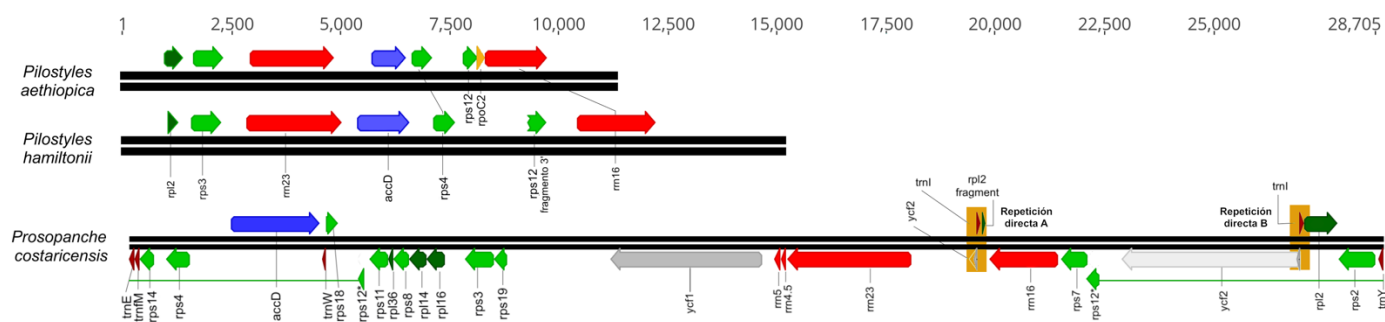


Figura 12. Comparación de los plasmidos de *Pilostyles aethiopica*, *Pilostyles hamiltonii* y *Prosopanche costaricensis*.

Nota: Los genes son bloques con flechas que indican la orientación de su transcripción. Los colores de los bloques corresponden a diferentes grupos de genes: *accD* es azul, el grupo de *ARNt* es rojo oscuro, los *ARNr* son rojo claro, los genes *rpl* son verde oscuro y los *rps* son verde claro, los *ycf* son gris y el gen *rpoC2* es naranja. El asterisco del gen *rps12* indica los exones del gen trans-empalmado. El gen *ycf2* de *P. costaricensis* es un pseudogén, por lo que se ilustra en la imagen con un gris más claro. Las repeticiones directas son naranjas. El *trnI* de la repetición directa A de *P. costaricensis* es un pseudogén.

Al comparar el plasmido de *Prosopanche costaricensis* contra los de *Balanophora laxiflora* y *Balanophora reflexa* (Balanophoraceae) las diferencias encontradas son similares a las vistas contra las especies del género *Pilostyles*. El plasmido de *P. costaricensis* cuenta con 13 200 pb y 13 198 pb más que los plasmidos de *B. laxiflora* y *B. reflexa*, respectivamente (cuadro 10). Las especies del género *Balanophora* no presentan la estructura tradicional cuatripartita de los plasmidos, y tampoco cuentan con repeticiones directas como la mayoría de especies de *Prosopanche* (Jost et al., 2022). Sin embargo, estas especies de *Balanophora* si conservan una

copia de dos de los genes que usualmente tienen fragmentos en las repeticiones invertidas y directas en las especies de Hydnoraceae: *ycf2* y *rpl2*; el único gen no presente es el ARNt *trnI*.

La cantidad de genes de ambas especies de *Balanophora* es de 19 genes en total, 8 menos que los que retiene *P. costaricensis* (cuadro 10). Las tres especies comparten los siguientes 18 genes: *rrn4.5s*, *rrn16s*, *rrn23s*, *accD*, *ycf1*, *ycf2*, *rpl2*, *rpl14*, *rps2*, *rps3*, *rps4*, *rps7*, *rps11*, *rps12* (*trans-empalmado en todas*), *rps14*, *rps18*, *rps19*, y *trnE*. Ambas especies de *Balanophora* cuentan con el gen *clpP*, mientras que *P. costaricensis*, y ninguna de las Hydnoraceae analizadas aquí, presentan este gen. Igual que con las especies de *Pilostyles* la ubicación y dirección de los genes compartidos entre estas especies es muy diferente con respecto a *P. costaricensis*. Pero a diferencia de lo que ocurre con *H. visseri*, estas especies de *Balanophora* si tiene los genes *rps12* y *rps7* en la misma dirección que *P. costaricensis* y el gen *rps12* también se encuentra trans-empalmado (figura 13).

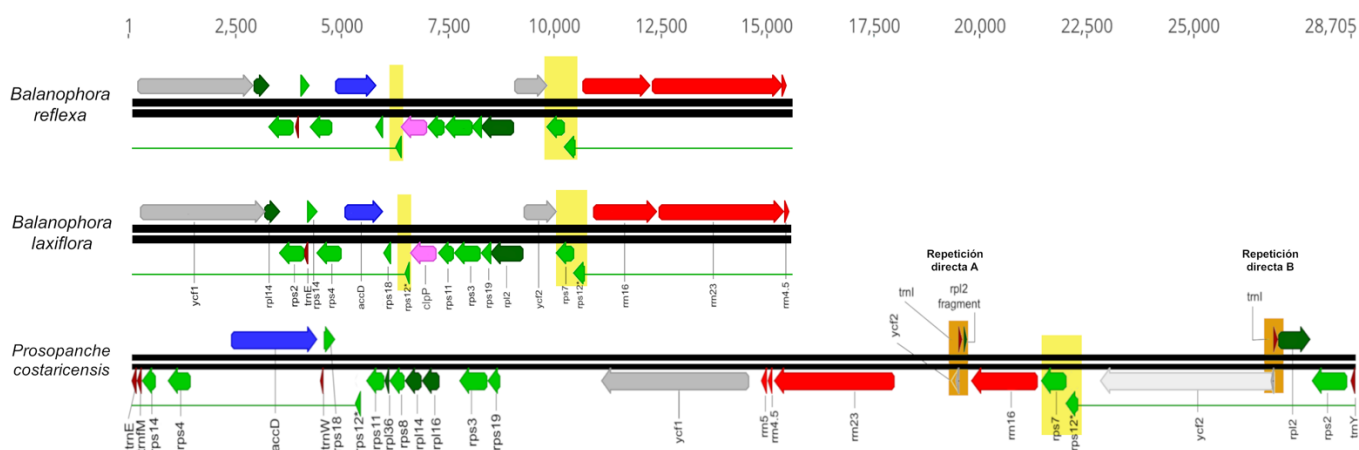


Figura 13. Comparación de los plastomas de *Balanophora reflexa*, *Balanophora laxiflora* y *Prosopanche costaricensis*.

Nota: Los genes son bloques con flechas que indican la orientación de su transcripción. Los colores de los bloques corresponden a diferentes grupos de genes: *accD* es azul, el grupo de ARNt es rojo oscuro, los ARNr son rojo claro, los genes *rpl* son verde oscuro y los *rps* son verde claro, los *ycf* son gris y el gen *clpP* es rosado. El asterisco del gen *rps12* indica los exones del gen trans-empalmado. Las zonas amarillas son aquellas que tienen el mismo acomodo en los tres genomas. El gen *ycf2* de *P. costaricensis* es un pseudogén, por lo que se ilustra en la imagen con un gris más claro. El *trnI* de la repetición directa A de *P. costaricensis* es un pseudogén. Las repeticiones directas son naranjas.

Por otro lado, el plastoma de *Prosopanche costaricensis* también es más largo que los plastomas de *Mitrastemon kanehirai* y *Cytinus hypocistis* por 2 965 pb y 9 305 pb respectivamente. Respecto a la estructura, *M. kanehirai* y *C. hypocistis*, tampoco presentan la estructura tradicional cuatripartita de los plastomas con repeticiones invertidas y no tienen repeticiones directas como en las especies de *Prosopanche*. Pero en *M. kanehirai* y *C. hypocistis* si conservan genes que se incluyen normalmente en las repeticiones invertidas/directas de Hydnoraceae; *M. kanehirai* conserva una copia de *trnI* y *ycf2*, y ambas especies tienen una copia de *rpl2*.

Si se comparan los números totales de genes de las 3 especies, *P. costaricensis* tiene 3 gen más que *M. kanehirai* y 2 genes más que *C. hypocistis*. Específicamente *M. kanehirai* cuenta con los genes *trnC*, *trnI*, *clpP*, y *infA*, que no están en el plastoma de *P. costaricensis*; y *P. costaricensis* tiene los genes *rpl14*, *rrn4.5S*, *rrn5S*, *trnI* (segunda copia), *trnW*, y *trnY* que no están en el plastidio de *M. kanehirai*. Ahora con respecto a *C. hypocistis*, ésta especie tiene los genes *clpP*, *rpl20*, *rpl22*, *trnC*, y *trnQ* que no están en *P. costaricensis*, mientras que *P. costaricensis* tiene los genes *rps4*, *trnW*, *ycf1*, *ycf2*, y *accD* que no están en *C. hypocistis*. Adicionalmente, *rps12* en *M. kanehirai* y *C. hypocistis* está trans-empalmado en ambas como en las especies de *Prosopanche*. Continuando la tendencia de la ubicación y dirección de los genes encontrados con las 4 especies mencionadas anteriormente, la estructura de *C. hypocistis* y la de *M. kanehirai* es muy diferente a la de *P. costaricensis* (figura 14).

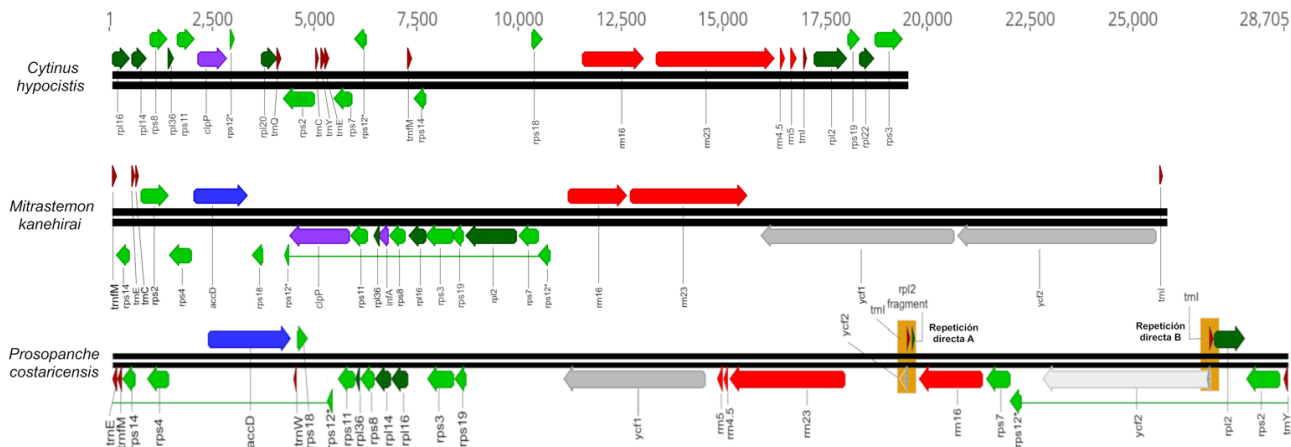


Figura 14. Comparación de los plastomas de *Cytinus hypocistis*, *Mitrastemon kanehirai* y *Prosopanche costaricensis*.

Nota: Los genes son bloques con flechas que indican la orientación de su transcripción. Los colores de los bloques corresponden a diferentes grupos de genes: *accD* es azul, el grupo de *ARNt* es rojo oscuro, los *ARNr* son rojo claro, los genes *rpl* son verde oscuro y los *rps* son verde claro, los *ycf* son gris y el gen *clpP* y el gen *infA* son morados. El asterisco del gen *rps12* indica los exones del gen trans-empalmado. El gen *ycf2* de *P. costaricensis* es un pseudogén, por lo que se ilustra en la imagen con un gris más claro. El *trnI* de la repetición directa A de *P. costaricensis* es un pseudogén. Las repeticiones directas son naranjas.

Al comparar los plastomas más largos de las especies holoparásitas contempladas en este análisis, se observa como los de *Conopholis americana* y *Cuscuta erosa* son 16 968 pb y 32 254 pb más largos que el de *P. costaricensis*, respectivamente. Con respecto a la estructura, *Conopholis americana* no presenta la estructura cuadripartita ni repeticiones directas (figura 15). Pero si conserva los genes incluidos normalmente en las repeticiones invertidas/directas de Hydnoraceae, *ycf2*, *rpl2*, y *trnI*. Además *Conopholis americana* tiene el gen *rps12* trans-empalmado.

En el caso de *Cuscuta erosa*, este plastoma sí muestra el arreglo tradicional cuadripartito de los plastomas de plantas autótrofas. Sus repeticiones invertidas contienen los genes *trnN*, *rrn5s*, *rrn4.5s*, *rrn23s*, un remanente de *trnA*, *rrn16s*, *trnV*, exón 2 de *rps12*, *rps7*, *trnL*, y *ycf2* (figura 15). La copia corta tiene los genes *trnL*, *rps15* y *ycf1*; y la copia larga de 29 596 pb contiene el resto de los 71 genes totales contenidos en su plastoma. Igual que *Conopholis americana*, *Cuscuta erosa* también contiene el gen *rps12* trans-empalmado. Pero las repeticiones

invertidas de *Cuscuta erosa* solo comparten, con las repeticiones directas de *Prosopanche*, el gen *ycf2*; los genes *rpl2* y *trnI* no se encuentran incluidos en estas regiones del plastoma de *Cuscuta erosa*.

Con respecto al número de genes, ambas especies presentan muchos más genes que *Prosopanche costaricensis*. *Conopholis americana* tiene 52 genes, es decir 25 genes más que *P. costaricensis*; además contiene genes fotosintéticos y una mayor cantidad de ARNs de transferencia en comparación a *P. costaricensis*. En el caso de *Cuscuta erosa*, este cuenta con 71 genes, es decir 44 genes más que *P. costaricensis* y también retiene genes fotosintéticos. Al igual que con *Conopholis americana*, *Cuscuta erosa* tiene una gran cantidad de ARNs de transferencia si se compara con los disponibles en el plastoma de *P. costaricensis*. Debido a esto, la similitud entre los plastomas de *Conopholis americana*, *Cuscuta erosa* y *P. costaricensis* es baja.

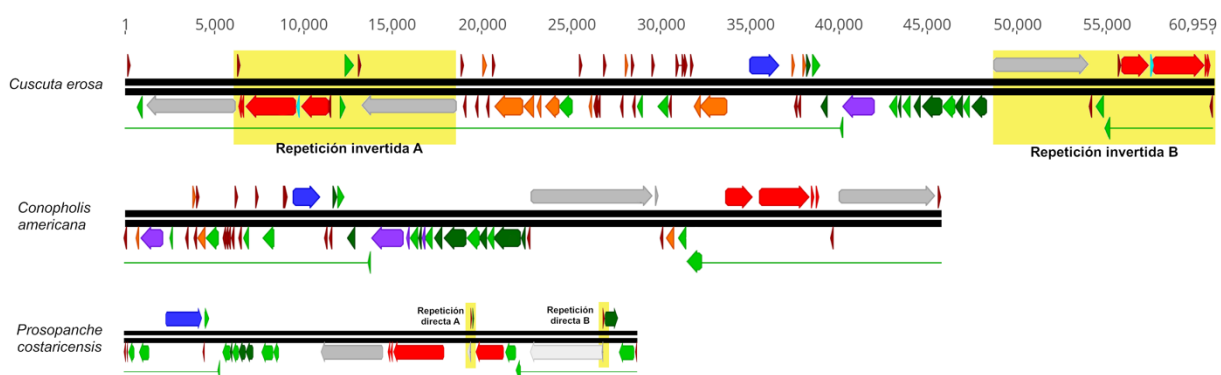


Figura 15. Comparación de los plastomas de *Cuscuta erosa*, *Conopholis americana* y *Prosopanche costaricensis*.

Nota: Los genes son bloques con flechas que indican la orientación de su transcripción. Los colores de los bloques corresponden a diferentes grupos de genes: *accD* es azul, el grupo de *ARNt* es rojo oscuro, el turquesa indica el remanente de *trnA*, los *ARNr* son rojo claro, los genes *rpl* son verde oscuro y los *rps* son verde claro, los *ycf* son gris, los genes relacionados con la fotosíntesis son los anaranjados y el gen *clpP*, el gen *matK*, y el gen *infA* son morados. El gen *rps12* indica ser trasn-empalmado mediante la línea verde que conecta los 2 exones. El gen *ycf2* de *P. costaricensis* es un pseudogén, por lo que se ilustra en la imagen con un gris más claro. El *trnI* de la repetición directa A de *P. costaricensis* es un pseudogén. Finalmente los recuadros amarillos indican las repeticiones invertidas o directas.

Finalmente, en la comparación entre el plastoma de *P. costaricensis* y el taxon fotosintético más cercano *Aristolochia contorta*, se observa como esta especie presenta un total de 160 576 pb y 131 genes, es decir, tiene 131 871 pb y 104 genes más que *P. costaricensis*. Además, la estructura del plastoma de *A. contorta* es la cuadripartita tradicional, cada una de sus repeticiones invertidas contienen 25 459 pb, la copia corta es de 19 882 pb y la copia larga tiene 89 756 pb. Una similitud entre estas dos especies es que dentro de las repeticiones invertidas de *A. contorta* también se incluyen los genes que normalmente están en las repeticiones invertidas/directas de Hydnoraceae, *ycf2*, *rpl2*, y *trnI*. Con respecto a la diferencia del número de genes entre estas especies, *A. contorta* supera a *P. costaricensis* en cantidad de genes de ARNs de transferencia y en genes que codifican proteínas. En el caso de los genes que codifican proteínas, el número de *A. contorta* es mucho más elevado debido a los genes fotosintéticos que si se conservan en su plastoma (cuadro 10). Por lo tanto, la similitud entre los plastomas de *P. costaricensis* y *A. contorta* es la más baja.

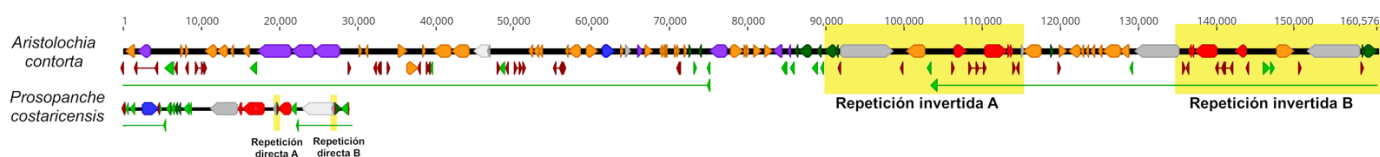


Figura 16. Comparación de los plastomas lineales de *Aristolochia contorta* y *Prosopanche costaricensis*.

Nota: Los genes son bloques con flechas que indican la orientación de su transcripción. Los colores de los bloques corresponden a diferentes grupos de genes: *accD* es azul, el grupo de *ARNt* es rojo oscuro, el turquesa indica el remanente de *trnA*, los *ARNr* son rojo claro, los genes *rpl* son verde oscuro y los *rps* son verde claro, los *ycf* son gris, los genes relacionados con la fotosíntesis son los anaranjados y el gen *clpP*, el *matK*, el *infA*, el *cemA* y los *rpo* son morados. El gen *rps12* indica ser trans-empalmado mediante la línea verde que conecta los 2 exones. El gen *ycf2* de *P. costaricensis* es un pseudogén, por lo que se ilustra en la imagen con un gris más claro. El *trnI* de la repetición directa A de *P. costaricensis* es un pseudogén. Finalmente, los recuadros amarillos indican las repeticiones invertidas o directas.

Análisis de Selección

El análisis de selección realizado con todas las especies con plastoma disponible del género *Prosopanche* se observa en la figura 17. Los 16 genes codificantes de proteínas compartidos entre los cuatro plastomas de *Prosopanche* se alinearon, mediante alineamiento pareados, contra los genes de *Hydnora visseri* ya que éste fue la especie utilizada como grupo externo para el análisis. Solo tres genes mostraron una selección neutral o positiva (>1), estos fueron el gen *rps7* para *Prosopanche americana* y *Prosopanche costaricensis* con valores de relación de sustitución (dN/dS) de 1.11 y 1.22 respectivamente; el gen *rps3* para *Prosopanche panguanensis* con una relación de 1.01 y el gen *ycf1* de *Prosopanche costaricensis* con un valor de 1.19. Luego en el caso de *ycf1* de *Prosopanche costaricensis*, este presentó un marco de lectura mucho más corto lo que pudo generar este resultado.

En el caso del análisis del gen *rps2*, este mostró valores de la relación de sustitución también altos, pero sin sobrepasar el umbral de >1 para todas las especies, en un rango de 0.83 a 0.96. Luego, entre los genes que mostraron mayor selección purificadora en *Prosopanche*, se tiene al gen *rpl36* con valores de 0.17 a 0.27, y los genes *rpl14* y *rps19* con valores de 0.29 a 0.35. El resto de genes codificantes de proteínas presentaron valores de relación de sustitución entre 0.4 a 0.6.

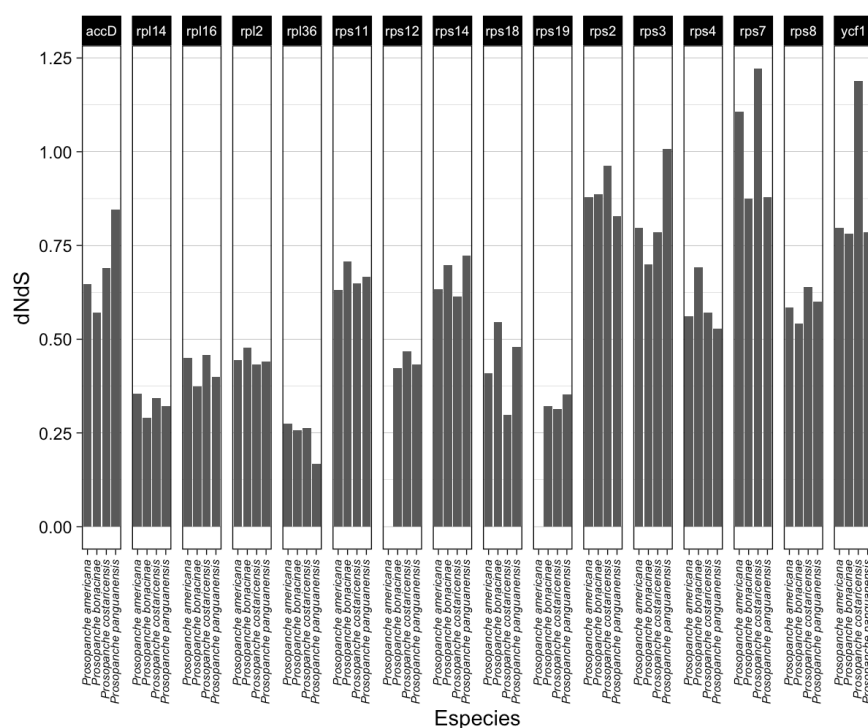


Figura 17. Análisis de mutaciones sinónimas y no sinónimas de los genes codificantes de proteínas de todas las especies de *Prosopanche* con plastomas disponibles, utilizando a *Hydnora visseri* como grupo externo.

Nota: Los valores de la relación de sustitución dN/dS para los genes *rps12* y *rps19* de *Prosopanche americana* no se reportan ya que su secuencia anotada en el GenBank no los presenta. Además, en la publicación de Jost et al., (2022) estos genes si se reportan presentes en el plastoma de esta especie pero los autores no indican la secuencia anotada.

Análisis filogenético

El árbol filogenético construido con alineamientos de 26 genes concatenados de 86 taxones estimados por máxima verosimilitud con el modelo de sustitución GTR+I+G4 con un Bootstrap rápido de 1000 replicas, se recuperó a los clados más grandes como monofiléticos (eudicotiledóneas y monocotiledóneas) con un soporte bajo. Luego los géneros de *Prosopanche* y *Hydnora*, mostraron el soporte máximo en su posicionamiento como géneros hermanos y la familia como tal obtuvo un soporte de 51 replicas para su inclusión en el orden de los Piperales. Además la familia Hydnoraceae presentó mostro ramas largas, por su alta diferencia con respecto a los taxones fotosintéticos, debido a la alta modificación de las secuencias de sus genes.

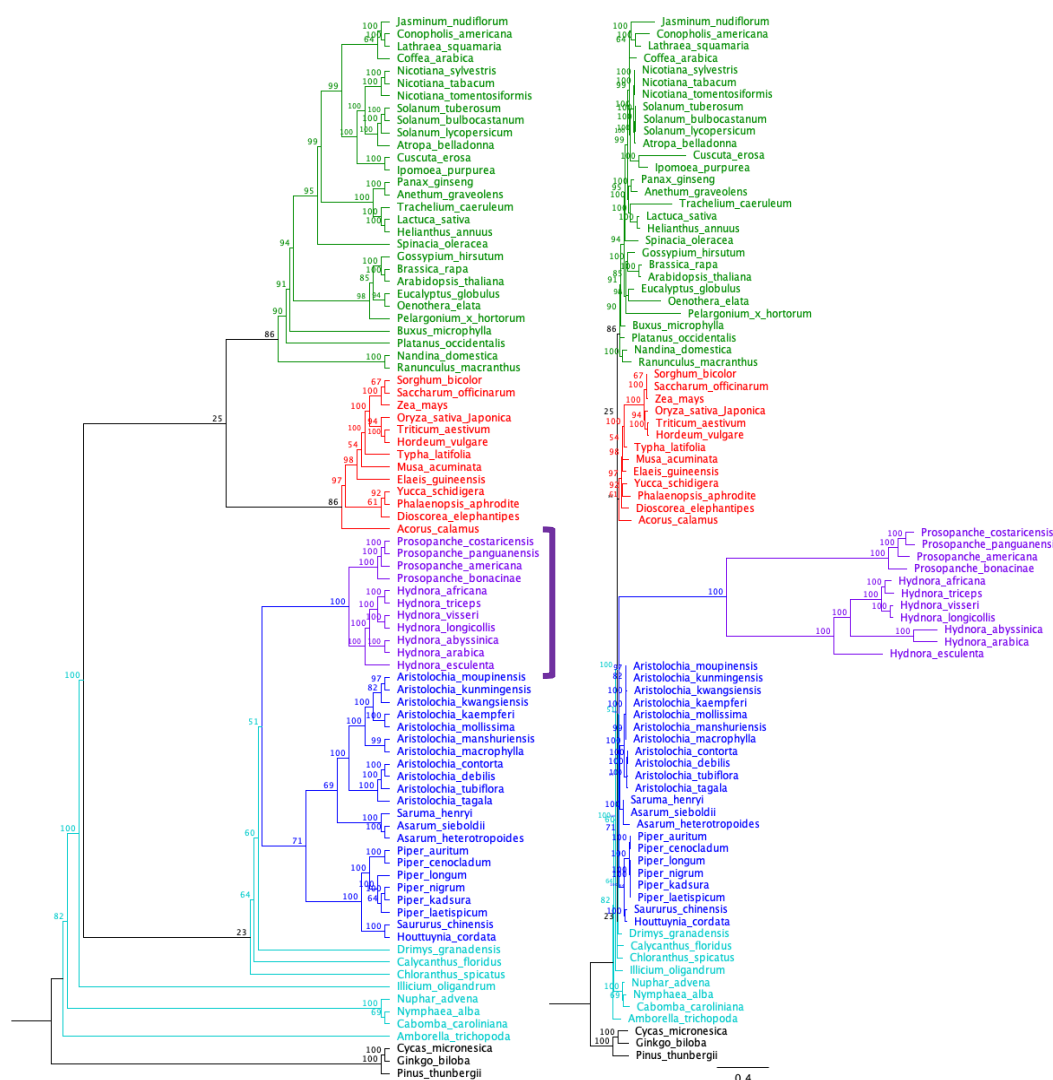


Figura 18. Reconstrucción filogenética por máxima verosimilitud con valores de bootstrap de máxima verosimilitud basada en una matriz concatenada de 26 genes (los 26 genes retenidos en *Prosopanche costaricensis*) de todas los taxones disponibles en GenBank de Hydnoraceae así como de los plastomas de las demás angiospermas.

Nota: Reconstrucción por máxima verosimilitud bajo el modelo de sustitución GTR + I + G4. El de la izquierda muestra el cladograma y el de la derecha el filograma. Aquí se muestra a *Hydnora* y *Prosopanche* como géneros hermanos con un soporte alto y a Piperales (incluyendo a Hydnoraceae) con un soporte de BS 51. El soporte con valores de BS se dispone sobre los nodos. Los grupos son codificados por color: eudicotiledóneas verdes, las monocotiledóneas rojas, los Piperales azules con Hydnoraceae en morado y señalado con un corchete, y las demás Magnólicas y el grado ANA turquesas. Grupos externos en negro.

DISCUSIÓN

Secuenciación Masiva, Ensamblaje y Secuenciación de Gaps

La secuenciación a partir de ADN de material vegetal preservado por largo tiempo, como el material de herbario, puede presentar complicaciones. El ADN puede llegar a tener una alta fragmentación si la muestra fue preservada por secado al horno, ya que las altas temperaturas generan que el 90% del ADN sufra de fragmentación de bandas dobles lo que inhabilita el acceso de la polimerasa para la amplificación. Al mismo tiempo, los materiales almacenados pueden presentar contaminación por microorganismos endófitos o microorganismos que colonizaron post-mortem al espécimen. En cuanto al tiempo de almacenaje de las muestras, el mismo no es determinante para la calidad del ADN que se puede obtener según estudios que concluyen que el uso del calor en la preservación es más perjudicial que el tiempo de almacenamiento para el uso del ADN en futuros análisis moleculares (Staats et al., 2011, 2013). La mejor forma de almacenar muestras vegetales, que permite la obtención de un ADN adecuado para la secuenciación, es la preservación a temperaturas congelantes (Bakker et al., 2016).

El material usado en esta investigación se ha preservado durante 21 años a -12 a -15 °C. Según Bakker et al., (2016) y Staats et al., (2011, 2013), almacenamiento en temperaturas de congelación es la más beneficiosa para realizar posteriores ensayos moleculares. Esto se confirma con las características del ADN extraído del material, además, el plastoma al ser un ADN circular podría estar sujeto a menos efectos de degradación. Este tuvo altos rendimientos que cumplen con los requerimientos para la secuenciación con Illumina, entre 50-100 ng/ul de concentración (Meyer & Kircher, 2010) y una pureza adecuada con una A260/280 de 1.8 a 2.0 (Habibi et al., 2022). Además, la calidad del ADN enviado a CD Genomics (Estados Unidos, Nueva York), confirmó la calidad del material. Este presentó supero los valores los mínimos requeridos para secuenciar, y de este material lograron obtener lecturas R1 y R2 con la calidad suficiente para proseguir con el proceso de ensamblado de acuerdo con los resultados obtenidos con el análisis hecho con FastQC.

A pesar de esta alta calidad obtenida en las lecturas, inicialmente solo se logró obtener un ensamblaje por referencia con gaps en la secuencia. Twyford & Ness, (2017) mencionan que la

estrategia empleada aquí, que implica la secuenciación del genoma completo para el posterior ensamblaje del plastoma, es una de las más usadas ya que es costo-efectiva y reduce la cantidad de trabajo de laboratorio por lo que es más rápida. No obstante, presenta la desventaja de que las bibliotecas de ADN genómico de plantas contienen una alta variabilidad de la concentración de ADN plastidial dependiendo del tejido de origen y el tamaño del ADN nuclear. Esto hace que diseñar correctamente las bibliotecas en la secuenciación para obtener una buena cobertura de lecturas del plastidio sea más difícil. Esto a su vez se complica más cuando el plastoma analizado presenta estructuras atípicas, re-arreglos, deleciones de genes, o pseudogenes, como ocurre con los plastomas de plantas parásitas. Dichos factores dificultan la escogencia de los análisis bioinformáticos para lograr el ensamblaje apropiado del genoma, y a esto se le añade la falta de genomas de referencia que pueden ser usados en estos procesos. Por lo tanto, Twyford & Ness, (2017) mencionan que es común que los ensamblajes de plastomas de plantas parásitas o de materiales de herbario, sean fragmentados o incompletos.

En el caso del protocolo de ensamblaje para plastomas no convencionales, Twyford & Ness, (2017) mencionan que existen dos abordajes diferentes que pueden ser utilizados. El primero es el proceso de alineamiento de lecturas contra un genoma de referencia por medio de un alienador de lecturas cortas como Bowtie2 (Langmead & Salzberg, 2012). Estrategia que puede llegar a presentar regiones con gaps debido a la divergencia de las lecturas con respecto al genoma de referencia. Los gaps son difíciles de solucionar ya que no se recomienda aumentar la dureza del filtrado por similitud en el caso de plastomas con estructuras atípicas como el de las plantas parásitas ya que podría generar aún más gaps (Jost et al., 2020; Straub et al., 2013; Sun et al., 2018) y esto mismo ocurre con genomas mitocondriales (Yu et al., 2022, 2023).

Debido a estas dificultades con los plastomas atípicos se ha generado la necesidad usar otros abordajes para el ensamblaje completo de estos genomas. Entonces, Twyford & Ness, (2017) mencionan como segundo abordaje el filtrado de lecturas. Este busca separar las lecturas del genoma plastidial de las lecturas del genoma completo por medio de las características exclusivas del genoma plastidial, como el contenido GC y la cobertura de secuenciación; con un filtrado por similitud posterior. Una de las herramientas que permite realizar este abordaje es el pipeline de extensión de lectura GetOrganelle (Kim et al., 2023; Twyford & Ness, 2017; Wen et

al., 2022). Con este pipeline se determina un tamaño de lectura a usar para recuperar un set inicial de lecturas que tienen una alta similitud con el o los plastomas de referencias que se elijan y a estos se les denomina semilla. Esta semilla se utiliza para seguir recuperando las lecturas que presenten una superposición con las lecturas de la semilla mediante rondas sucesivas de extensión y con estas se procede a generar el ensamblaje (Jin et al., 2021; Kim et al., 2023; Twyford & Ness, 2017).

Como se evidenció en los resultados, previo al uso del segundo abordaje, se intentó rellenar los fragmentos desconocidos por secuenciación Sanger. Sin embargo, estos tuvieron numerosas limitantes como se ve en las figuras de la 1 a la 4. Como obstáculos se tuvo: una baja concentración de los fragmentos amplificados, una alta producción de múltiples bandas por fragmento y una baja efectividad de secuenciación. Es posible que estos protocolos tuvieron dichas limitantes debido al diseño de imprimadores para PCR. Para diseñar imprimadores adecuados usualmente se prefiere oligos con un contenido GC entre 40-60% (Bustin et al., 2020; Hishamuddin et al., 2023), con una temperatura de fusión (T_m) entre 60 y 65 °C, sin estructuras secundarias y sin complementariedad en sus extremos 3' (Bustin et al., 2020). En esta investigación los contenidos de GC de los imprimadores diseñados fue entre 24% a 50%, donde 40 de 68 oligos tenían un contenido menor a 40%, lo cual rindió imprimadores con T_m con valores menores de 60 °C. Estos se tuvieron que diseñar de esta manera por que el contenido GC del plastoma como tal ya es bajo, lo cual limita de entrada el diseño de oligos. Adicionalmente, las medidas de diseño para aumentar el GC de lo oligos no eran factibles para estos imprimadores. Por ejemplo el alargar los imprimadores para aumentar su T_m (Bustin et al., 2020) generaba estructuras secundarias. Por lo tanto, se tuvo que utilizar estos imprimados con baja T_m y recurrir a otras estrategias que mejoraran la amplificación y especificidad de los fragmentos.

Para promover una alta especificidad (Strien et al., 2013) menciona que se puede variar los protocolos de amplificación utilizando metodologías con temperaturas de amplificación variadas, o se pueden usar reactivos que promuevan la alta especificidad como BSA (Albúmina de suero bovino), DMSO o $MgCl_2$. No obstante, al implementar un protocolo de amplificación con touch up para promover la hibridación de los primer sin tener que usar T_m muy bajas, y al añadir DMSO 10% y $MgCl_2$ para promover la especificidad, no se logró superar el obstáculo de

baja especificidad, sino que más bien se obtuvo una muy baja concentración de las bandas. Entonces se recurrió a probar la amplificación sin touch up, pero con temperaturas de amplificación más bajas para promover la hibridación de los imprimados, pero adicionalmente mantener los reactivos para favorecer especificidad (DMSO y OneTaq® High GC Enhancer). Esta opción propició una mejor concentración de bandas, pero generó múltiples bandas en muchos de los fragmentos. Por lo que, no logró superar de manera correcta este obstáculo para todos los fragmentos, sino que solo se logró optimizar el protocolo de amplificación para 18 fragmentos (cuadro 9). Aunado a esto se encontraron las limitaciones de la tecnología Sanger. Con esta tecnología se facilita la secuenciación de 800 a 1000 pb, secuencias mayores presentan una baja calidad (Crossley et al., 2020). Entonces, es probable que tanto las limitaciones de los imprimadores como las de la tecnología de secuenciación fueran las que impidieron la obtención de todos los fragmentos necesarios para rellenar los fragmentos desconocidos del primer ensamblaje.

Adicionalmente los tres imprimadores (figura 8) que si lograron amplificar secuencias a pesar de que no contaban con secuencias adecuadas para calzar con el ensamblaje final lo lograron debido a el uso de técnicas que relajaban la astringencia de hibridación de los imprimadores en el proceso de amplificación específicamente el uso de temperaturas más bajas de hibridación sin touch up. Pero a la hora de secuenciar por Sanger, las condiciones de hibridación son más exigentes por lo que no se lograron secuenciar correctamente, además de que en esta técnica una de las limitaciones son las secuencias de baja calidad dentro de los primeros 15 a 40 pb debido a la hibridación del imprimador (Crossley et al., 2020), por lo que en este caso para estos fragmentos los imprimadores no lograban hibridar en la secuenciación.

Entonces, debido a todas las limitantes mencionadas, se optó por utilizar al segundo abordaje propuesto por Twyford & Ness, (2017). En el caso del ensamblaje por referencia mediante Bowtie2 y algoritmo de Geneious, se logró obtener un ensamblaje parcial con gaps con tamaño de hasta 3000 pb. Con GetOrganelle se logró obtener un ensamblaje con dos gaps uno de menos de 100 pb y el segundo de ~1000 pb. Al combinar el uso de estas dos técnicas, más el proceso de secuenciación individual de fragmentos se logró obtener el plastoma completo. De esta forma, se confirmó lo mencionado por Twyford & Ness, (2017), respecto a que ambas

técnicas no son mutuamente excluyentes y que se pueden usar para ensamblar plastomas con estructuras inusuales. Adicionalmente, se reafirmó lo experimentado en procesos de ensamblaje de otros plastomas donde es usual la necesidad de secuenciar ciertos fragmentos por Sanger para rellenar gaps o confirmar la existencia de repeticiones invertidas y sus uniones (Ma et al., 2015; Schneider, Chun, et al., 2018; Xu et al., 2020).

Características de los plastomas de Hydnoraceae enfocado en *Prosopanche costaricensis*

Tamaños y genes

Los plastomas de la familia Hydnoraceae presentan una estructura altamente compacta por su reducido tamaño y bajo número de genes, lo que los posiciona dentro de los plastomas más pequeños reportados hasta la fecha (Jost et al., 2022; Naumann et al., 2016). En general los plastomas del género *Hydnora* son más pequeños que los del género *Prosopanche*. En *Hydnora* el tamaño de los plastomas varía, el rango va de *H. esculenta* que tiene el plastoma más corto con 24,479 pb, a *H. triceps* que tiene el más largo con 28 658 pb. En el caso de *Prosopanche* los tamaños varían menos, dentro los plastomas reportados (sin contar a *P. costaricensis*) el más grande lo tiene *P. panguanensis* con 28 262 pb y el más reducido lo tiene *P. americana* con 28 191 pb (Jost et al., 2022). El plastoma de *P. costaricensis* no es la excepción a esta tendencia con un tamaño de 28 705 pb, convirtiéndose así en el plastoma más largo en toda la familia reportado hasta la fecha.

En el caso de los genes, Hydnoraceae como familia presenta 15 genes de proteínas ribosomales (*rps* y *rpl*), el gen de la subunidad beta de la acetil CoA carboxilasa (*accD*), los dos genes de función desconocida como lo menciona (Jost et al., 2020) o con función de traslocación de proteínas por la membrana plastídica como lo indica (Jost et al., 2022) (*ycf1* y *ycf2*), y los tres ARN de transferencia (*trnE-UUC*, *trnI-CAU*, *trnM-CAU*) y exclusivamente en *Prosopanche* se incluyen 2 genes más, *trnW-CCA* and *trnY-GUA*, para un total de 27 genes.

El plastoma de *P. costaricensis* cumple con estas características, pero dos de estos 27 genes no son putativamente funcionales. Existía la posibilidad de que fueran tres genes los no funcionales ya que el ARN de transferencia *trnW* podía ser un pseudogén como sucede en *Prosopanche americana*. En *P. americana* la estructura secundaria de *trnW* presenta como

anticodón la triada ACA en vez de la triada CCA y dicho anticodón no se encuentra presente para este ARN de transferencia en los linajes de Piperales (Jost et al., 2020). Pero en *Prosopanche costaricensis* el *trnW* sí presentó el anticodón tradicional CCA (figura 9) de acuerdo a la predicción de su estructura mediante tRNAscan(ver. 2.0) (Chan et al., 2019). Entonces, solo quedaron dos genes que no son putativamente funcionales en *P. costaricensis*. El gen *ycf2* que es un pseudogén por la falta de un codón de inicio en su marco abierto de lectura y el *trnI* de la repetición directa A. De acuerdo con la predicción de su estructura y anticodón mediante tRNAscan(ver. 2.0) (Chan et al., 2019), este es el ARN de transferencia *Ile-AAU* (Ile2) (Figura 9). En los plastomas de las especies de los linajes de Piperales los ARN de transferencia para isoleucina (Ile) e isoleucina 2 (Ile2) presentan los codones CAU (Ile2) o GAU (Ile). No se logró encontrar un reporte de alguna especie de Piperales con el ARN de transferencia *Ile-AAU* (Ile2). Además, Mohanta et al (2023) mencionan que el anticodón usado en cloroplastos para Ile2, es CAU y los anticodones para Ile en el código genético universal son AAU, UAU y GAU (Rodin et al., 2011). Adicionalmente, AAU para Ile2 tampoco es un codón usual en los genomas de plantas (Mohanta et al., 2020) ni de bacterias (Tomikawa et al., 2018). Por lo que este *trnI-AAU* (Ile2) en *P. costaricensis* es probablemente un pseudogén. En el caso del *trnI-CAU* de la repetición B, este sí cuenta con el codón tradicional CAU (Ile2) de cloroplastos (Mohanta et al., 2023) y de los Piperales, por lo que se considera un gen putativamente funcional.

Ahora bien, la retención de genes en la familia de Hydnoraceae es un tema que Naumann et al., (2016) ha analizado previamente con el plastoma de *Hydnora visseri*. Los genes codificantes de proteínas ribosomales, en los genomas plastidiales de plantas fotosintéticas, se ha observado que usualmente presentan 21, mientras que en las plantas no-fotosintéticas o con capacidad fotosintética reducida, algunos de estos genes se han perdido (Barrett et al., 2014; Pan et al., 2023; Cristina Roquet et al., 2016). En Hydnoraceae las especies retienen 14 de estos genes (Jost et al., 2022) siendo *Prosopanche costaricensis* un ejemplo de esta tendencia. Naumann et al., (2016) menciona que la falta de estos genes en Hydnoraceae puede implicar que los genes se transfirieron a otro genoma (mitocondrial o nuclear). En el proceso de ensamblaje del plastoma de *H. visseri*., Naumann et al., (2016) encontraron genes plastidiales que residen como

pseudogenes en el genoma mitocondrial de *H. visseri* y de ahí parten para mencionar esta transferencia de genes.

Por otro lado se encuentran los genes codificantes para ARNs de transferencia. En los plastomas de *Hydnora* se retienen solo tres, mientras que en los de *Prosopanche* retienen cinco (Jost et al., 2020, 2022; Naumann et al., 2016). Por lo que ambos géneros retienen una menor cantidad de ARNs de transferencia de lo esperado para mantener una función mínima del plastoma, ya que el total de ARNt retenidos en los plastomas de plantas fotosintéticas es de 30. Naumann et al., (2016) indica que probablemente, en plastomas altamente reducidos de familias que han perdido la capacidad fotosintética recientemente, la pérdida de ARN de transferencia sea menor debido a que se ha demostrado mediante simulaciones previas (Lohan & Wolfe, 1998) que es probable que secuencias cortas escapen la delección por casualidad debido a su tamaño y estructura secundaria altamente conservada. No obstante, el holoparasitismo en Hydnoraceae evolucionó hace al menos 55 Ma (Naumann et al., 2013; Sanchez-Puerta et al., 2023), lo que ha proporcionado tiempo suficiente para poder superar la probabilidad de evitar la delección por casualidad. Entonces, Naumann et al., (2016) menciona que es probable que estos ARNs de transferencia conservados (*trnM-CAU*, *trnW-CCA*, *trnI-CAU*) sean esenciales en *Hydnora*.

La situación anterior es extrapolable a *Prosopanche* y su retención de cinco ARNs de transferencia únicos (*trnE-UUC*, *trnM-CAU*, *trnW-CCA*, *trnY-GUA*, *trnI-CAU*). No obstante, en *P. costaricensis*, se logró identificar un ARN de transferencia con signos de pseudogenización. Una de sus copias del *trnI-CAU* (repetición directa A) presenta el codón AAU en vez del CAU, para el aminoácido isoleucina 2 (ver Figura 9). A pesar de que este ARNt con este codón ha sido encontrado en plastomas de otras plantas (López et al., 2023) para isoleucina, este no se ha evidenciado para isoleucina 2 (Mohanta et al, 2023). Además, este codón no aparece en ninguna de las especies Piperales incluidas en este estudio (cuadro 8), por lo que su forma se considera una pseudogenización del gen. Estos hallazgos pueden demostrar una relajación de las presiones selectivas sobre el plastoma (Xin Li et al., 2021).

Posteriormente, se encuentran los genes con funciones inicialmente desconocidas (Jost et al., 2020), ahora también denominados genes de traslocación de proteínas por la membrana plastídica (Jost et al., 2022). Entre estos están *ycf1* y *ycf2*, así como el gen de la subunidad beta

de la acetil CoA carboxilasa (*accD*). Estos genes son retenidos en la mayoría de plastomas de plantas holoparásitas utilizadas en este estudio (ver Figuras de 10 a15). Sin embargo, hay excepciones como los plastomas de *Pilostyles aethiopica* y *Pilostyles hamiltonii* (figura 12) (Bellot & Renner, 2016) que han perdido tanto *ycf1* como *ycf2* y el de *Cytinus hypocistis* (figura 14) (Cristina Roquet et al., 2016) que ha perdido los tres genes. En todos los plastomas reportados hasta la fecha para Hydnoraceae, los tres genes se reportan como putativamente funcionales (Jost et al., 2020, 2022; Naumann et al., 2016). La diferencia la introduce el plastoma de *Prosopanche costaricensis*, que presenta un marco abierto de lectura más corto para *ycf1* y le hace falta el codón de inicio al marco abierto de lectura de *ycf2* siendo este último un pseudogén. Eventos que, de acuerdo a Jost et al., (2022), pueden ser eventos aislados de pseudogenización en especies de este género.

Finalmente, es importante mencionar la pérdida de intrones del grupo IIa en los plastomas de esta familia. La ausencia de estos intrones no es exclusiva de plantas que han perdido la capacidad fotosintética; más bien, se trata de eventos esporádicos poco usuales en plastidios tanto fotosintéticos como no fotosintéticos (Jost et al., 2022; McNeal et al., 2009; Su et al., 2019) En Hydnoraceae los plastomas han conservado todos los genes que tradicionalmente presentan intrones (*rpl2*, *rpl16*, y *rps12*) pero solo uno ha conservado su intrón (*rps12*). Todos estos genes que han perdido los intrones no han sufrido problemas con sus marcos de lectura (Jost et al., 2020, 2022; Naumann et al., 2016). Lo mismo sucede con *P. costaricensis*, los tres genes mencionados anteriormente se han retenido y solo *rps12* retiene su intrón. Esta pérdida de intrones del grupo IIa se alinea con la pérdida del gen maturasa K (*matK*) o intrón del *tRNK* que es necesario para el proceso de empalmado de estos intrones (McNeal et al., 2009).

Estructura y repeticiones invertidas o directas

En el caso de la estructura cuatripartita tradicional de los plastomas, la mayoría de los plastomas de plantas con plastidios altamente reducidos como los de plantas holoparásitas normalmente no presentan este arreglo tradicional debido a la pérdida de las repeticiones invertidas (Jost et al., 2022; Naumann et al., 2016). Ejemplos de este fenómeno son *Balanophora laxiflora* y *B. reflexa* (figura 13) (Su et al., 2019), *Cytinus hypocistis* (figura 14) (Roquet et al., 2016), *Mitrastemon*

kanehirai (figura 14) (Shyu, 2013), y *Pilostyles hamiltonii* y *P. aethiopica* (figura 12) (Bellot & Renner, 2016).

Jost et al., (2020) y Naumann et al., (2016) mencionan que muchos de los genes retenidos en los plastomas de Hydnoraceae son genes encontrados en las repeticiones invertidas de especies autotróficas cercanas. Pero, en la estructura de los plastomas de Hydnoraceae aún hay especies que retienen estas repeticiones (directas o invertidas) por lo que Jost et al., (2022) indican que estas repeticiones en Hydnoraceae son versiones altamente condensadas de las de sus contrapartes autótrofas. Además, Naumann et al., (2016) en su análisis del plastoma de *H. visseri*, mencionan que el reducido tamaño de las repeticiones invertidas de *Hydnora* ha propiciado una relocalización de los genes encontrados en la repetición invertida de especies fotosintéticas cercanas a la copia única larga. En el plastoma de *P. costaricensis*, se han retenido a lo largo de todo el plastoma, diez de los genes de la repetición invertida de *Aristolochia contorta* (figura 16) confirmando así lo mencionado por las investigaciones anteriores.

Las estructuras de las repeticiones encontradas en la familia Hydnoraceae se dividen en dos organizaciones según sus dos géneros (Jost et al., 2022). En *Hydnora* la mayoría de sus especies presentan la estructura cuadripartita tradicional con las repeticiones invertidas, una copia única corta y un copia única larga (a excepción de *H. esculenta* que no tiene repeticiones) (Jost et al., 2020, 2022; Naumann et al., 2016). Para *Prosopanche*, Jost et al., (2020) fueron los primeros en sospechar la existencia de repeticiones directas en vez de invertidas, con la secuenciación y ensamblado del plastoma de *P. americana*. Repeticiones que posteriormente fueron confirmadas por primera vez por Jost et al., (2022) en *P. americana* y en las otras especies disponibles de este género (*P. bonacinae* y *P. panguanensis*). Añadiendo a la evidencia de Jost et al., (2022), en esta investigación se determinó que *P. costaricensis* también presenta dichas repeticiones directas.

Jost et al., (2022) dividen las repeticiones invertidas y directas de esta familia de acuerdo con sus diferencias estructurales y a su relaciones filogenéticas, en tres clados y dos especies solitarias. El clado I *Hydnora* incluye las repeticiones invertidas *H. africana*, *H. triceps*, *H. visseri*, y *H. longicollis* donde estas repeticiones están compuestas por el fragmento 5' del gen *ycf2*, el *trnI-CAU* y el fragmento 5' del gen *rpl2*. El clado II *Hydnora* incluye las repeticiones

invertidas de *H. arabica* y *H. abyssinica*, y estas repeticiones incluyen los genes que van del fragmento 5' del gen *rpl16* al fragmento 3' del gen *rps3*, con diferentes cantidades de genes entre ambos fragmentos. Aquí se incluyen las estructuras de las repeticiones descubiertas para *H. abyssinica* por Mkala et al., (2023). *H. esculenta* no fue asignado a ninguno de los clados de *Hydnora* debido a que no presenta repeticiones invertidas ni directas. Para *Prosopanche* Jost et al., (2022) solo mencionan el clado I *Prosopanche*. Este incluye las repeticiones directas de *P. americana* y *P. panguanensis*, que están compuestas por el gen *trnI-CAU* y el fragmento 5' del gen *rpl2*. En el caso de la especie *P. bonacinae*, esta queda por fuera de los clados ya que sus repeticiones directas están compuestas por los mismos genes que las del clado I *Hydnora*, pero ambas en la misma dirección y más cortas siendo apenas un tercio del tamaño de las repeticiones de las especies *Hydnora*. Las repeticiones directas de *P. costaricensis* sufren una situación similar a las de *P. bonacinae*, tienen la misma composición pero su tamaño es menor con 330 pb. Estas similitudes entre *P. bonacinae* y *P. costaricensis* indican que esta última puede ser un paso intermedio entre el clado y la especie solitaria si nos limitamos a las características de sus repeticiones directas.

Patrones de evolución

Naumann et al., (2016) en su primera investigación sobre los plastomas de Hydnoraceae con el género *Hydnora*, propusieron un modelo evolutivo de pérdida de genes y contracción plastidial por la pérdida de capacidad fotosintética. En la primera etapa ellos establecen una degeneración de los genes que ya no son esenciales y los genes fotosintéticos convirtiéndose estos en pseudogenes para posteriormente perderse del todo. Dicho patrón inicial se ha visto en linajes como: el de la micoheterótrofa *Corallorhiza striata* (Orchidaceae), donde su plastidio se encuentra en sus primeros estadios de degradación (Barrett & Davis, 2012); los plastidios de las plantas hemiparásita *Krameria* (Zygophyllales), que están relativamente intactos (Banerjee et al., 2022); y en Orobanchaceae, en especies del género *Orobanche*, donde los genes que se han perdido son los no esenciales y los fotosintéticos (Cusimano & Wicke, 2016).

Naumann et al., (2016), luego menciona que la segunda etapa se caracteriza por la retención de genes con funciones no fotosintéticas y un ritmo de pérdida de genes más bajo, por

lo que hay una baja generación de pseudogenes. Aquí la pérdida de genes depende de la habilidad del organismo de importar proteínas al plastidio que le permitan funcionar normalmente. Esta etapa puede durar mucho, pero según Naumann et al., (2016) depende de cada linaje y las presiones específicas que este sufra. Ya cuando los genes retenidos son reemplazados por copias de genes nucleares o mitocondriales o son migrados a estos recintos genómicos, el plastoma deja de ser necesario y puede entrar a la tercera etapa. En esta tercera etapa, los genes que se han retenido en el plastoma por ser genes esenciales se van a pseudogenizar y luego perder. En esta etapa la detección del plastoma es mucho más complicado que en la etapa dos. Finalmente está la etapa 4, caracterizada por la ausencia del plastoma, caso que es posible evidenciar en la familia Rafflesiaceae. No obstante, aún no se puede estar seguro ya que puede ser que su plastoma sea difícil de detectar y se encuentre en etapa 3 o ya no esté presente del todo (Krause, 2015). Fuera de las angiospermas ya hay eventos reportados donde se indica la posible pérdida del plastoma en organismos que aun presentan cloroplastos como en el alga verde no fotosintéticas y de vida libre *Polytomella* (Smith & Lee, 2014), en el alga parda (crisófitas) no fotosintética *Paraphysomonas* (Dorrell et al., 2019), y en el protozooario parasito *Perkinsus marinus* del clado Apicomplexa que normalmente contiene plastidios reliquia (Fernández Robledo et al., 2011).

De acuerdo con Naumann et al., (2016), partiendo del plastoma de *Hydnora*, Hydnoraceae se encuentra en la segunda etapa, posiblemente entrando en la tercera, de pérdida de plastoma. Esta hipótesis se refuerza con la evidencia presentada por Jost et al., (2022) sobre 10 plastomas de Hydnoraceae (7 de *Hydnora* y 3 de *Prosopanche*), y se complementa con el plastoma de *P. costaricensis* determinado en esta investigación. Estas 11 especies han perdido todos sus genes fotosintéticos y algunas especies incluso presentan genes con funciones esenciales pseudogenizados (*H. visseri rps7*, *P. americana trnW* y *rsp19* y *P. costaricensis trnI* y *ycf2*). Si a esto se le añade el hecho de que los plastomas de Hydnoraceae presentan más genes y mayor tamaño que los de Balanophoraceae y que Rafflesiaceae (partiendo de los reportes de pérdida del plastoma en especies de esta familia: *Rafflesia lagascae* (Molina et al., 2014) y *Sapria himalayana* (Cai et al., 2021)), sustenta la propuesta de que es probable que los plastomas de Hydnoraceae están en la segunda-tercera etapa de pérdida plastidial, mientras que los de Balanophoraceae y Rafflesiaceae podrían estar más avanzados en el proceso. Esto apoya la

propuesta anterior debido a que el holoparasitismo en Hydnoraceae es el segundo más viejo (hace 55 Ma), mientras que el de Balanophoraceae (entre 110 a 115 Ma) (Kim et al., 2023) y el de Rafflesiaceae (entre 60 a 94 Ma según Naumann et al., (2013) y Wikström et al., (2001) y entre 83 a 109 Ma según Bendiksby et al., (2010)) son más viejos. Entonces el nivel de reducción de los plastomas de estos linajes apoyan que el nivel de pérdida de Hydnoraceae es menor, por lo que está un paso antes que estos dos en el proceso de pérdida.

Adicionalmente, Jost et al., (2022) expande el modelo de Naumann et al., (2016) proponiendo tres modelos evolutivos donde se puede explicar el orden de como estos eventos de cambio o pérdida sucedieron de acuerdo al acomodo filogenético de las 10 especies antes mencionadas. Los tres modelos se diferencian en cómo se perdieron o ganaron las repeticiones directas (RD) o indirectas (RI) en las especies. El primer modelo se denomina modelo de contracción. Este es el menos parsimonioso de los tres modelos ya que requiere de tres eventos independientes de cambios (traslocación e inversión y contracción) de las repeticiones directas (RD) o indirectas (RI). En este modelo se propone que en *Prosopanche* la RI1 se contrajo primero y luego sufrió una translocación e inversión que la convirtió en la RD de estas especies. En el caso de *Hydnora*, *H. esculenta* pierde la RI, mientras que en el clado II la RI1 sufre una contracción y una traslocación convirtiéndose en una RI diferente denominada RI2. En el caso del clado I de *Hydnora*, la RI1 también sufre una contracción y una translocación pero se convierte una RI diferente denominada RI3. Este modelo parte del supuesto de que las RI y RD de Hydnoraceae son una versión condensada de la RI de sus parientes autótrofos cercanos y que estas sufrieron eventos independientes de translocación e inversión. Según Jost et al., (2022) dichas modificación se han discutido poco. Pero ya hay variados reportes de fuertes contracciones de RI en varios plastomas como los resume Mower & Vickrey, (2018). Un ejemplo puntual es el linaje de Papaveraceae donde han reportado importantes re-arreglos en sus plastomas, incluyendo contracciones y expansiones relevantes de sus RI (Raman et al., 2022). Por lo tanto, a pesar de su baja parsimonia, también existe la posibilidad de que este modelo sea correcto.

En el segundo modelo, denominado de pérdida ancestral, los cambios son más parsimoniosos que en el de contracción, pero también presenta tres eventos independientes para explicar las RI y RD de las especies de esta familia. Este propone una pérdida de la RI1 previo a

la separación de *Hydnora* y *Prosopanche*, y luego propone la ganancia de la RD en *Prosopanche* y la ganancia de dos RI diferentes RI2 (clado II) y RI3 (clado I) en *Hydnora*. Pero la ganancia de una IR o de una DR luego de una pérdida ancestral es un fenómeno raro y poco reportado (Choi et al., 2019; Jost et al., 2022). Por lo tanto, hay una baja probabilidad de que este modelo sea correcto.

Finalmente, el tercer modelo también es el más parsimonioso por ser el que tiene menor número de pasos. A este se le denomina el modelo de pérdidas repetidas y propone otros tres eventos importantes. En *Prosopanche* la inversión y traslocación del RI1 que se convierte en una DR y en *Hydnora* propone que *H. esculenta* pierde la RI aisladamente y que solo el clado II pierde su RI1 y gana la RI2, mientras que el clado I no sufre cambios en su IR partiendo del hecho de que esta RI sea una versión condensada o contraída de su pariente fotosintético más cercano (Jost et al., 2022).

Prosopanche costaricensis calza relativamente bien en los tres modelos propuestos por Jost et al., (2022) para clado I de *Prosopanche*. En los tres modelos se predice la ocurrencia de RD (por diferentes medios) y contraídas (menor tamaño aunado a una pérdida del fragmento *ycf2*). Entonces, *P. costaricensis* calza relativamente bien en estos modelos pero el paso de contracción de la RD no fue tan fuerte en esta especie como si lo fue en *P. americana* y *P. panguanensis*. *P. costaricensis* aún conserva el fragmento 5' del gen *ycf2* en su RD, igual que *P. bonacinae* pero con un menor tamaño (la de *P. bonacinae* es de 516 pb mientras que la de *P. costaricensis* es de 330 pb). Además, *P. costaricensis* presenta una mayor similitud filogenética con *P. panguanensis* que con *P. bonacinae*, a pesar de que las RD de que *P. bonacinae* y *P. costaricensis* se parecen más en estructura (ver Figura 18). Debido a esto se puede decir que *P. costaricensis* pertenece al clado I de *Prosopanche* pero tiene una contracción intermedia de la DR, con un tamaño de su RD de 330 pb, mientras que las RD de *P. americana* y *P. panguanensis* contienen 188 pb y 195 pb respectivamente. Adicionalmente, *P. costaricensis* presenta dos pseudogenes antes mencionados, *ycf2* y *trnI*, además de un *ycf1* con un codón de terminación temprano. Estos eventos que pueden ser añadidos sin problema a los tres modelos propuestos por Jost et al., (2022) como eventos aislados. Con la evidencia anterior se puede decir que es probable

que tanto *P. costaricensis* como el resto de las especies disponibles de Hydnoraceae se encuentren en la segunda etapa de degradación entrando en la tercera etapa de la pérdida del plastoma.

Selección sobre genes

Además de examinar cuáles y cuántos genes se retuvieron en el plastoma de *P. costaricensis* se analizó la divergencia de las secuencias de los genes retenidos y la tasa de variación evolutiva de los mismos (Figura 17). En estudios realizados para evaluar esto en los genomas de otras especies heterótrofas (*Cuscuta*, Convolvulaceae), se ha encontrado que la intensidad de la selección positiva o purificadora ha sido elevada por la acelerada evolución de estos plastomas (Banerjee et al., 2022; Barrett et al., 2019).

Dentro de los genes esenciales que codifican proteínas ribosomales analizados aquí *rpl14*, *rpl16*, *rpl2*, *rpl36*, *rps11*, *rps12*, *rps14*, *rps18*, *rps19*, *rps4*, y *rps8* parecen estar bajo una selección purificadora moderada ($0.25 < dN/dS < 0.7$). Los genes restantes que codifican proteínas ribosomales *rps2*, *rps3* y *rps7* presentaron una selección purificadora más débil ($0.7 > dN/dS$). En otras especies heterótrofas y autótrofas genes de las proteínas ribosomales pequeñas han presentado altas sustituciones de nucleótidos (Barrett et al., 2019) (Barnard-Kubow et al., 2014). Por lo que estos valores moderados no están fuera de lo ya encontrado en otras especies. No obstante, para *P. costaricensis* y *P. americana*, *rps7* tuvo valores $dN/dS > 1$ indicando una selección positiva. Aquí es importante mencionar que los alineamientos se realizaron contra los genes de *Hydnora visseri*, genoma que tiene al gen *rps7* con un marco abierto de lectura que le hace falta el codón de inicio pero en el cual Jost et al., (2022) identificó un sitio donde la posible edición postranscripcional podría crear un codón de inicio en el marco de lectura de *rps7*, por lo que podría ser funcional. Por lo tanto esta particularidad podría introducir diferencias que generan altos valores de la tasa dN/dS .

En el caso de los genes con otras funciones *accD* y *ycf1*, estos presentaron una selección purificadora más débil ($0.8 > dN/dS$). El gen *accD* involucrado en la biosíntesis de ácidos grasos, ha presentado, en el plastoma de especies del género *Cuscuta*, valores débiles de selección purificadora, por lo que muestran una divergencia de secuencias mayor a la de otros genes (Banerjee & Stefanović, 2019). Proceso similar se ha encontrado con los genes *ycf1* y *ycf2* en

Cuscuta donde estos han tendido a acumular mayores mutaciones no sinónimas (Banerjee & Stefanović, 2019). Otro ejemplos de estos son el de la especie *Campanulastrum americanum* en el estudio de Barnard-Kubow et al., (2014) donde estos genes presentaron una débil fuerza de selección purificadora; y el caso de la perdida de estos genes en los plastomas de Poaceae estudiados por Guisinger et al., (2010). El *ycf1* de *P. costaricensis* tuvo un valor de dN/dS mayor a 1 señalando una selección positiva, y este valor es probablemente ocasionado por el codón de terminación temprano que presento en su marco abierto de lectura.

Arboles filogenéticos

De acuerdo con (Nickrent, 2002), taxónomos han posicionado a Hydnoraceae junto con Aristolochiaceae. Dicho posicionamiento inicial se basó en ADN nuclear y mitocondrial. En la última actualización realizada en el APG IV (APG et al., 2016) Hydnoraceae se incluyó dentro de las Magnólidas anidada dentro de Aristolochiaceae sensu APG IV (Hydnoraceae, Lactoridaceae y Asaraceae). No obstante, el posicionamiento de la familia junto a su taxón fotosintético más cercano en Aristolochiaceae aún no se ha logrado de manera definitiva. Esta falta de certeza se debe primordialmente a relaciones no resueltas dentro de Piperales (Bolin, 2009). Existen variados posicionamientos de Hydnoraceae con familias y géneros dentro Piperales. Jost et al., (2020) en su análisis con 82 genes posicionaron a Hydnoraceae dentro de Piperales, pero no lo lograron posicionar como hermana de Aristolochiaceae. En el estudio de Naumann et al., (2013) Hydnoraceae se posiciona como hermana de *Aristolochia* y *Thottea* (Aristolochioideae), junto a *Lactoris* (Lactoridaceae). Mientras que Nickrent, (2002) la posiciona como hermana de *Lactoris* (Lactoridaceae) y Massoni et al., (2014) concuerdan con esta posición como hermana de Lactoridaceae pero le añaden a su cercanía como hermana de Asaroideae y Aristolochioideae.

En esta investigación la topología del árbol filogenéticos por máxima verosimilitud (Figura 18) verificó, ahora con más taxones de la familia Hydnoraceae, el posicionamiento de *Hydnora* y *Prosopanche* como géneros hermanos mediante solo secuencias plastidiales, con un soporte máximo. Por lo que se obtuvo el mismo resultado presentado por Jost et al., (2020) y Jost

et al., (2022) en sus reconstrucciones de árboles filogenéticos. En el caso del posicionamiento de toda la familia, esta se posicionó dentro de los Piperales como lo han hecho los reportes anteriores de AGP IV, (2016); Naumann et al., (2013); Nickrent, (2002). Pero no se logró determinar a cuál es la familia más cercana a Hydnoraceae, situación similar reportada previamente en la investigación de Jost et al., (2021), donde los árboles con solo información plastidial arrojan a Hydnoraceae como hermana de Aristolochiaceae + Lactoridaceae. En este caso el árbol realizado con solo los genes retenidos en las holoparasitas (26 genes) arrojó con un soporte de BS de 51 a Piperales, donde Hydnoraceae fue obtenida como hermana del resto. Además, las ramas de las especies de Hydnoraceae se fueron sustancialmente largas, esto puede deberse a la alta modificación de los plastomas de plantas holoparasitas lo que genera que tengan menos información que se puede usar en el árbol para determinar su posicionamiento, demostrando así largas ramas (APG IV, 2016; Jost et al., 2020). Entonces, el posicionamiento específico de Hydnoraceae dentro de Piperales sigue siendo incierto.

Nueva evidencia para considerar *Prosopanche americana* y *Prosopanche costaricensis* como dos especies distintas

Prosopanche se distribuye desde Centro América hasta Sur América (Hatt et al., 2023). *Prosopanche americana* y *P. bonacinae* se encuentran en Argentina y es menos frecuente en Bolivia y Paraguay (Cocucci & Cocucci, 1996). *Prosopanche cocucci* y *P. demogorgoni* se distribuyen de manera restringida en una zona pequeña boscosa del sur de Brasil, siendo este bosque tropical con condiciones más secas que un bosque tropical lluvioso (de Carvalho et al., 2021; Funez et al., 2019). *P. caatingicola* se encuentra en la zona de bosques estacionalmente secos del norte de Brasil (Ferreira-Machado & Paganucci-De queiroz, 2012). Mientras que *P. panguanensis* y *P. costaricensis* son las únicas dos reportadas en zonas de bosque tropical húmedo (Hatt et al., 2023), la primera del bosque lluvioso central de Perú (Martel et al., 2018) y la especie de este estudio en el bosque lluvioso de Costa Rica (Gómez-Pignataro & Gómez-Laurito, 1981). Por lo tanto. *P. costaricensis* evidencia una distribución en parches del género *Prosopanche*, ya que es la única fuera de Sur América (Bolin, 2009). No obstante, la ausencia de especies entre Perú y Costa Rica puede deberse a dificultades en el muestreo o una dificultad de

realizarlo, ya que la planta puede ser confundida con un hongo o raíces de otras plantas, lo que dificulta su observación y recolección (Hatt et al., 2023), a parte, que solo puede ser reconocida cuando emerge el cuerpo floral.

Adicionalmente, Nickrent (2002), menciona que la distribución geográfica de la familia Hydnoraceae en general, puede ser explicada la formación y movimiento de Gondwana (*Hydnora* principalmente en África y *Prosopanche* predominantemente en América del Sur). Pero dado que *Prosopanche* e *Hydnora* parecen haber divergido entre 86.89 y 29.54 Ma, es decir que ocurrió después de la separación África-Sudamérica, hace ca. 100 Ma, se necesitaría también en este caso dispersión a larga distancia, ya que en el medio solo había mar o tal vez islas. En el caso de *Prosopanche*, Hatt et al., (2023) indica que la existencia de una sola especie reportada en Centroamérica y la falta de especies reportadas en regiones intermedias como Panamá se debe mayoritariamente a una dificultad de muestreo, pero que la probabilidad de encontrar especies en estas regiones en el futuro es alta. Partiendo de esto, la formación del istmo centroamericano podría apoyar la dispersión y especiación de *Prosopanche* desde América del Sur hasta Centroamérica mediante algún tipo de dispersión de larga distancia. De acuerdo con, Bacon et al., (2015) y Coates & Obando, (1996) el istmo centroamericano de terminó de cerrar hace 3.5 millones de años, pero Bacon et al., (2015) menciona que evidencia reciente sugiere que la formación y finalización del istmo ocurrió por medio de una formación geológica más compleja, y han evidenciado que el movimiento biótico entre América del Sur y el istmo empezó a darse inclusive antes de que este se cerrara. Ellos encontraron eventos de movimiento de organismos terrestres hace 20 a 6 millones de años. Apoyando esta evidencia, Erkens et al, (2007) mencionan que para el género *Guatteria* de la familia Annonaceae (uno de los linajes más importantes, diversos y abundantes de la Amazonía), su diversificación y dispersión entre América del sur y Centroamérica se dio, no por un movimiento a través del istmo cerrado o atravesando una gran masa de agua, sino por medio de una migración por peldaños durante la formación compleja del istmo. Entonces, es probable que en el caso de *Prosopanche costaricensis*, su dispersión ocurriera mediante pequeños vertebrados que consumieron sus frutos y movieron las semillas de sus ancestros hasta Centroamérica, y ya se sabe que para *Prosopanche costaricensis* esta es la dispersión reportada (Cocucci & Cocucci, 1996).

Esta situación de dispersión en parchones de *Prosopanche* llevó a investigadores como Cocucci & Cocucci, (1996) a no aceptar a *P. costaricensis* como una especie diferente de *P. americana*. Sin embargo, actualmente *P. costaricensis* se acepta como una especie diferente a base de su morfología y filogenias con genes nucleares (Bolin, 2009; Ferreira-Machado & Paganucci-De queiroz, 2012; Funez et al., 2019; Hatt et al., 2023; Martel et al., 2018). No obstante, Mkala et al., (2023) menciona que en muchas holoparásitas la caracterización morfológica y genética es complicada, más aún en especies de familias como Hydnoraceae que presentan una gran modificación física que las hace vivir principalmente bajo tierra haciendo su recolección difícil, lo que genera que la identificación dependa de especímenes de herbario que son principalmente estructuras florales. Por lo tanto, es crítico añadir más información para apoyar la descripción de una especie de *Prosopanche*. En el caso de *P. costaricensis* aún se carecía de evidencia plastidal para respaldar la distinción morfológica de las especies.

En esta investigación, mediante el análisis del plastoma completo de *P. costaricensis*, se obtuvo suficiente evidencia para afirmar lo siguiente: i) los plastomas de *P. costaricensis* y *P. americana* muestran diferencias estructurales en cuanto al contenido de genes, genes pseudogenizados, tamaño del plastoma y la estructura de la repetición directa, y ii) que *P. costaricensis* es genéticamente más similar a *P. panguanensis* que a *P. americana*, según lo evidencia el árbol filogenético (Figura 18) y lo corroborado por Hatt et al., (2023) respecto a los hábitats que estas dos ocupan, siendo las únicas del género encontradas en bosques tropicales lluviosos y las más cercanas geográficamente.

CONCLUSIONES

- A pesar de no encontrar tejidos frescos, se logró extraer con éxito ADN de una muestra preservada por más de 20 años a temperaturas congelantes de entre -12 a -15 °C. Por lo que, en casos en los que no se cuenta con material fresco, este tipo de almacenamiento es una alternativa viable para tejidos que se utilizarán posteriormente en ensayos moleculares.
- El uso combinado de distintos métodos de ensamblaje, más la secuenciación de fragmentos específicos por Sanger permiten superar las dificultades que presentan los plastomas con re-arreglos importantes, los cuales que limitan el ensamblaje tradicional por referencia.
- El plastoma de *Prosopanche costaricensis* apoya la hipótesis de Jost et al., (2022) que destaca a Hydnoraceae como uno de los pocos grupos con plastomas holoparásitos que aún conservan repeticiones invertidas o directas. Además, demuestran que los genes presentes en el plastoma representan una retención de la mayoría de los genes de las repeticiones invertidas de la especie fotosintética más cercana.
- *Prosopanche costaricensis* calza con los modelos de Naumann et al., (2016) y Jost et al., (2022). En el modelo de Naumann et al., (2016) añade a la evidencia existente de que Hydnoraceae está en una etapa 2 a 3 de reducción plastidial, caracterizada por la falta de genes fotosintéticos y el inicio de pseudogenización de genes esenciales. Según Jost et al., (2022), encaja dentro del clado I *Prosopanche* en los tres modelos propuestos, pero con una reducción leve de la DR conservando aun un fragmento de *ycf2* pero con este gen completo pseudogenizado, al igual que el *trnI* de la repetición A. No obstante, es importante no generalizar estos resultados obtenidos para Hydnoraceae con todas las especies de plantas parásitas más aún cuando existe una limitante de información.
- *Prosopanche costaricensis* difiere de *P. americana* en términos de la estructura del plastoma, contenido de genes putativamente funcionales, posición en el análisis filogenético y por el uso de hábitat mencionado en literatura.
- Los análisis filogenéticos indican que *Prosopanche costaricensis* está más cercanamente relacionada a *P. panguanensis*, coincidiendo con su uso similar de hábitat. Ambas especies son las únicas del género registradas en bosques húmedos hasta la fecha.

- El análisis filogenético y las características de estructurales de los plastomas proporcionan evidencia adicional que sugiere que los géneros *Hydnora* y *Prosopanche* son hermanos.
- Hydnoraceae, al agregar los datos del genoma plastidial de todas las especies disponibles de esta familia, se posiciona dentro del clado de los Piperales, pero la familia más cercana a Hydnoraceae no se pudo determinar.

RECOMENDACIONES

Los espaciadores intergénicos (IGS) de los plastidios de plantas donde el genoma de estas organelas sufre una lenta evolución, se han utilizado para obtener mejor resolución en los análisis filogenéticos, debido a que estas regiones tienen una tasa evolutiva más alta (Prince, 2015). No obstante, en los plastomas de plantas parásitas estos espaciadores son muy reducidos y cortos debido a la velocidad evolutiva incrementada que sufren los genomas de estas organelas por la reducción de la presión selectiva. En el caso de *Prosopanche* ocurre lo mismo que en *Balanophora* donde los IGS son escasos y muy cortos (Su et al., 2019). Por lo tanto para intentar determinar cuáles son las familias más cercanas a Hydnoraceae dentro del orden Piperales y evitar inferir relaciones artificiales que se generan entre los plastomas de plantas parásitas con plastomas altamente reducidos de linajes diferentes, una alternativa es incluir genes nucleares, específicamente los espaciadores intergénicos nucleares (IGS). Estos incluyen el espaciador transcrito interno (ITS), el espaciador transcrito externo (ETS), y el espaciador no transcrito (NTS) (Zimmer & Wen, 2015).

Por lo tanto, para mejorar los análisis filogenéticos se recomienda utilizar la metodología propuesta por Krawczyk et al., (2017), donde a partir de los datos de secuenciación completa del genoma se ensamblan los contigs de genes de ARN ribosomal nuclear, se extraen las regiones IGS completas de estos contigs y se usan estas secuencias para la reconstrucción de árboles filogenéticos. Sin embargo, para Hydnoraceae, esto sería posible sólo si se logra tener acceso a los IGS o a los datos nucleares que permitan extraerlos para el resto de taxones incluidos en el análisis filogenético realizado en esta investigación

Apéndice 1.1 Primer ensamblaje a partir de la secuenciación masiva de *Prosopanche costaricensis* (30 192 pb).

```

1 ANTAATTACC ACCCTTAAGG GAAATCGAAT CCCTGTTATT TCCTTGAAAA AGAAATGTCC
61 TAAACCTCTA GACGATAAGG GCAATATTA TAATATAAAC TTAATAATGT NNNNNNNNNN
121 NNNNNNNNTAG TGCGAAGGTG GGAATTGAAC CCACGACCTC AAGGTCATGA GCCTTGCGAG
181 CTACCAAACT GCTCCACTTC GCGCTATANT NNNANNNNAA NNACCAACKN GNTATTTTAA
241 TTACCAACTT GATATTTTTG CTCCGGTAA TTTACATGAA TTAAATGTGT CATAAATTA
301 ATTTCTAGAT TGTCACAAAAT AACGATAATA ACCTCTAGGA CGACCTTTTA TCGAACATCG
361 TTTTCTAATA CGTATTAATA TACTATTTTC AGGCAATTTT TGTACTACTT TATTTATTTT
421 AAATTTAATT TTTAATAAAAT AAACNNNNNN NNNTTTTCTT ATCTTTTTTT TTAAGTTTTG
481 TCTAATAATA CGATATTTTT TTTCTAAAT TATTTTCTTT TTATCTCTAA AAAATATACT
541 TTTTTNNNNN NNNNNNNNNA TTGTATTAAT ATAAATTTGT ATTAATATAA CATAATNAAA
601 TAAAAATAAC AAAAAATAA AATTATAAAT NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
661 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNTGAATAA TNTTTATTTT TTTATTTTTT
721 TATTTAAATA ATATTCAC TAATTTATGTT CATTAAATAA AAAATCTAAT CTTTTTTTGT
781 CTAATTTTTT TGGTATTTTT TTTGGTATTT TTATAGTTTT TTTATTTAAA TTTAGTAAAG
841 TAATTTTTATC TTCAGTTTTA CAAATAAAAC TGGGAATATT TACTGTATAA TTGTTAACTA
901 AAATATGCTT ATGATTTATA AATTGTCTAG CTTCTGAAAT AGTTTTAGAC AAACCAATT
961 TAAATATAAT ATTATCTAAT CGCATTCTA AAAAACTACA TAATTTGTGC ATTTTAAATC
1021 CTTTAATTTT TATATATATA TATAAATAAT TAGATAATTG TTTATTTGTT AAACCATAAT
1081 AAAAGCGCAT TTTTTGTTTT TCTTTTAAAA GAGGAAGATA TTTTGATTCA CGTATAATTT
1141 TAAATTTTTT TTTTTTCTA TTTGTTTTTT TTTTTTTATA TTTTACAATT TCAAACTAC
1201 GCTTTCTTTT TAACTTAGGT CCTAAATATC TAGACNNNAT AATTGTTTAA ATTTTAAGTT
1261 TTAATATTA TAAATATAA AATAATATN NNNANAACNA NAACNAANAC NAANANNNNN
1321 AGNAAANNA AGNNANNNNA AANNNANNAG NANAANNNNN NNNNNNAAAN NNAAGNNANN
1381 NNAANNAAN ANAAAAAAN AAANNAANN AACNNGNNA NNNNNNAANNA AAGNNAANC
1441 ANAGNANGAA AAAAAAAAT AATAATAA TAGATAATAA AAAATATAA GTAAGTTATA
1501 AATATAATA ATTTAAACTT TACTTCAAAG ATGATATAAG TGGTAAAACA TTTGTGTTTT
1561 ATATACAAAT TAATGTAAA TATTTTAA TAATTTTAA AAAAAATTA TTTTATTAGT TTAATAATTAG
1621 AAGAGATTGG TTTTACATCA TTAGAAGATT TAGTATATAA TTATTTTGT TFCAGAAATTG
1681 AATCAAAATT TAGTGATTTT AATATAAATT TTTTCATATGT AAATAAATAGT AAAGAGGAAG
1741 AAANNNNGA AGAAAAAAA NANAAAAA ACAAAAAAA NGAAAAANAN NANNRWTWTN
1801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
1861 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
1921 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
1981 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNATGATGA TTATAAAAA GAACGTAGAA
2041 AAAAAAAAT ATTAATAA TATAAATAA AAAAATAA TAAATACAGA ANNNNNNNNN
2101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2161 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2221 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2281 TTTTATTATC AAAAAAAG AATTTTAAGT GAAGTATTAG ATAAAAAGA AAATTTTGAT
2341 TTTACAAAAT TATGGGTTTT TTGTGAAAAC TGTAATTTCT TAAATATAA AAAAGATTTT
2401 AAGTCTAAAT TATACGTTTG TGAAGAATGT GGTATCAT TTAATAATTAG TAGTTCAGAT
2461 AGAATAAAAA TGATAATAGA TAAAGATACT TGGAAATCCA TAAATGAAGA TATTTTTTCT
2521 ATAGATCCTC TAGAGATTGA TGATACTTTA GATGATAATT ATATAGATGA TGATAATACA
2581 GANGANAANN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2641 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2761 NNNNNNNNNN NNNNNNNNNN NNCNNNCACC AAATAATAAT TAGAAAAAGA ATGACAATGC
2821 ACAAATAA AACAAAAAG ATAATTACTT TGACCGCGAT AATAATTACA AAAATATTCN
2881 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

```

2941 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3061 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3121 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3181 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3241 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3361 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3421 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3481 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3541 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3661 NNNNNNNAAA AAGAAAAAA AAAAAGAAA AAACAACTA AAAAAATNN ANGNNAGNN
 3721 NGAAGAAGAA GNAGNAAAAC NAAANANGAAA AAAACAAAN NANNAAAAAN NNNNNNNNNN
 3781 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3841 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 3961 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNAAAAAA NGAAATAAAA AATAATAAAC
 4021 AAAAAATAC ATAAANANN NAANGAAAA TTTGATGAAA TAGAGAAAA CAAATCATAT
 4081 ATTAATCGTT TAAATTTTTA TCAAGAAAA ACAAGTTTAC CTGAAGCTGT AAAAACAGGT
 4141 ATAGGTAAAA TAAATGGTAT TAAAGTTGCA CTTGGTATAA TGGATTGAGA ATTTATGGGT
 4201 GGAAGCATGG GATCTGCTGT AGGAGAACT TTAACCTCGTT TAATCGAATA TGCCTACTAA
 4261 GAAAAATTAC CTCTTATAAT TTTCTGTGCT TCTGGAGGAG CACGCATACA AGAGGGGATT
 4321 TGTAGTTTAA TACAAATGGC TAAAAATATCT GCTGCATTAC ATTTATTATAA ATCAAACAAA
 4381 AAATCAGTAT TTATATCAAT ACTTACATCT CCAACAGCAG GGGGAGTTAT TGCAAGTTTT
 4441 GGAATGTTGG GTGATATTAT TATAGCTGAA CCAAATACTA CTATTGCTTT TGCAAGGACAA
 4501 ATAATAATTG AACAAATTTT AAGAAAAGAA GTACCTAAAG GTTTTCAAAC AGCTGAATAT
 4561 CTATTTAAAC AAGGTGCTTT TGATTTAATT TTACCACGAC AACTTTTAAA AACAATTGTT
 4621 ACTGTATTAC TTAATTTTCA CGGGTTTTTAT ACATGTTCTT AGTTAATAAT TTAATATTTA
 4681 ATATTTAATA TTAATAAAAT TATTATTTAA ATAAAAATTA TAAANNAAA ACGCGCTCTG
 4741 TAGGATTTKA ACCTACAACA TTAGGTTTTG GAGACCTAAA TTATTACTAA TTGAATTAAG
 4801 AGCGCATATA ANNNAAAAATA TTAAAAATGA AAATAAAAAAT AAGAAATAAA AAAACTATAA
 4861 AAAATTTAAA TTTTTTAAAA AATTTAAATA TAAATAAAAT AAATATAAAA GAAATAAAAT
 4921 TACTTTACAA ATTTTTAAGT ATAAAAGGAA AAATATTATC AAGAAAAGTA ACTAAATTA
 4981 ATTTAAAAAA ACAAAAAAGC ATTACTATAG CTATAAAGAG AGCTCGTATA TTATCTTTAT
 5041 TACCTTTTAA TAAATCAATT TACTTAAATA AAAAATAAAC GATGCAATAA AAAATTAATA
 5101 TATAATAAAT TATTTAATTT TAAAAAATT ATAAAAAAT ATTAATCTTT TTATCCAATA
 5161 TATAATAATA AATAAATTAA ACTTAAATAA CCTTCCTTTT TTTTATTAGT TNNNNNNNNN
 5221 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 5281 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 5341 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 5401 NNNNNNNNNN NAATTA AAC TTTTACTC TATAAAAAATA AAATTAAGT AACGATATTA
 5461 TAAAAATTAG TTTACTGAAC CTTTAATCTG TTAAATTTAT ATAATATAA AATATACATT
 5521 TATAAATTC CGCTTATTTA ATAAAAATAA AATTAAAAAA TGGAATTTAA NTTTTTAAAA
 5581 CAATTAATAA AATTATAATT AAACATAAAC ATACGAATCA CACATACAAT CTCATACACA
 5641 TTCCTAACCG TTGAGGACAT TTTTTAAGAG GAGCTGATGA TTTTGTAATT ATCTTATTTA
 5701 ATTTCTTGT TTTTTTCTA AAAATTTGTT TAATGTTTGT CATAATTTAC TTTTATTTTN
 5761 TATATTATNT NTATAATAAA AAANNNNNNN NNNNNNNNNN NNANANATAG TNTAGATNTA
 5821 TCAAAACCTA TATAACAACA AANNNNNNAN NNGNATTTGT ATATGTAATA TAKKTATTTT
 5881 ATATTTTTTN NANNNACTTA TTTACATATT ATCTTACTTT TTTTAAAGTA GATTTACAAC
 5941 CATTATGAGA AAAAGGAGTA ATATCTAATA TAGAATAAAT TTTTATTTTT TTACTTTTAA
 6001 TAACTTGTA TACTGAAGTT CTTCCTGCAC CTATACCTTT AATACGTATT TCAGCAGAA
 6061 GTATTTTTTAT TATATTAATA ACATTATTTG CTGTAGTTTT AGCTGCAAAA GGAGTACCTT
 6121 TTTGTGTATT TTTAAATTCA TACATACCAG TAGAAGCTGT AAAAATAAGT ATATTTTTTT

6181 CATCCATAAT TGTGATTCTA GTATTATTAA AACTAGCTAA AATAATAATA ATTCCTTTTT
6241 CAAGGAAATT TATTTTTGGT TGGCGCTTAT GACGATAAAG AACTTGTATA GGTTTTATTT
6301 TAATTCCTTT TCTCATATAT NNNNNNNNNN NNNNNNANNN NNKNNAATNTT TTTATTTTTTG
6361 TTTTTGTTTA TGTCTTTTAT TATTACAAAT AATTTAAAATA CGCCTATATC TTTTAACTAA
6421 TTTACATTCT TGGCAAATTT TTCGAATAGA TGTTCTTTTK TTCATTTTTN TNNTCATNTT
6481 TATACAACAT AAAATAAAAT TTCCCCTCCA ATTTTATTTA ATTTAGCTTC TATATCTGTT
6541 ATTATTCCTT GAGATGTAGA AATAAAATAA ACTCCATTTG CATTTAATTT TTTAGGAATT
6601 TTTTTATAAG TTACATAAAAT ACGTTGACTA GGTTTACTTA TTATTTTTTAC AAATATTTTT
6661 TTTTTTAAAA TTATACTAAA ATTATTTTTA TTTTTTATAA AATTTTCAAT AAAACCTTCT
6721 TTTAAAAGTA ATTTTAAAAT ATTATTTGAA GTATTATTAA AAGGTATTGT TGTATCTTT
6781 TTATTTGATA AACTTGCATT TTTTATATGA ACAATTATAG TTATAAAAT ATTTAACATA
6841 ATATTTATAT TTACAATATT TCAGAAGATA ATGAAATTAT TTTATTAAAA TTTAATTTGC
6901 GTAATTCCTG TGTAAATTATA CCAAAAATTC TAGATCCTTT AGGATTGTTT AAATTATCAA
6961 TTACAACAGC AGCATTATCG TCAAATTTAA ATATAATACC ATCAGAACGT TTAATTTCTT
7021 TTCGTGTACG TACAATAACA GCTCTTATTA TTTCTGATTT TTCTATTGAA AACTAGGTA
7081 AAGCTTCTTT AATCACAGAT ATAACAATAT CACCTATATA AGCAGATTTA AATTTACTAA
7141 TACCTAATTT TTGAATACAC ATTAACCTAA TAGCACCAC TTTATCTGTT ACATTAAGAT
7201 AAGTTTGTGTT GTGAATCATA TTTTNNNTAAC CATAAAAANA ATTAGGTTTT AATNNNAANA
7261 ANAAATTTTG TTTTTACAGG TAATTTAGAT GAAACTATTT TAAAAGTATT TTTTGCAAGT
7321 ATTTCCGATA CGTTATCTAC TTCATAAAGT ATTCGTCCAG GTTTAATAAC AGCAACCCAA
7381 TTTTCAGTGG AACCTTTACC AGACCCATA CGTTTTCCTA CAGATCTTTT TGTAATAGGT
7441 TTATCTGTAA ATATTCTAAT CCAGAAATTA CCACCTCGTT TAAAATATCG TGTCAATACT
7501 TTTCTACCAG CTTCTATTTG TTTAGCTGTG ATCCAAGTTG GTTCAAGGGT TTTAATTTCCA
7561 AATTTACCAA AACAAAATAA ATTACCTTTT ACTGAAATTC CTTTCATCCT ACCCCTATGT
7621 TGTTTATCAA ATTTAGTTTG TTTAGGTAAC ATTTTATAAT AAATTTTTAA AAATTTAATT
7681 AAAATAAAT AAANAAAAA TTAATAANA GTAATTAATA GTTAAAGAAA AGGTCTAGGT
7741 TCGCAATATT CACTTATAAT ATAATCTATA AGAATGCTG TACCAAAAT TACTACTATT
7801 CTAACATAAG TTATAACTTT TCTAATGATT TTTTTTTTTG TAATTATCAT AANTTTTTTT
7861 TAATGTTTTT ATTATTATNG TTATATAANT AANTATAATA ATGATATAGA ACTATCACAA
7921 AAAAAAATAA AAAAAGTTAT KTGTATAAAA AAAAATTTAA ATTTAAANAA NTAKTAGNTA
7981 ATATTAATTT ATATATAAAT CCAAATTTTG ATTCCTAAAA GTCCATTATT TGTTTTAACT
8041 GAATATGAAC AAAAACGTAT ATTTTCTTTA ATTGTTTTGTA ATGGTAAAAG ACCTTCTTTA
8101 ATCCAAGTTT TATTTTTTATG TACAATAGCT TTACCTGATG TTCGTCCAGA TATCTTTACT
8161 TTAATTCCTT TAATTTTTTTT TTCATTGTAA ATTTTATATA AAACTTTTTC TTTTATTTTT
8221 TTTAAAGTAA TTCCTTTTTT TAATTGAAAT AATATATATT TAGCAATAAC TTTAGGATTT
8281 TTAAAAGGGT ATTTAATTC TTTGAACTTTT ATATTAAAA TTAACCTTTT AAAATTTATA
8341 AATTGATTTT TAAAAATTAGA ATCTTTATTA AACATTTTTA GTAATTTAAA GTTTAAGAAA
8401 TCTAAAATTA TAATTTTTTAT TTTAATTTTG TTATTTTGTT TTGTTATTCT AATATGTGAA
8461 ATAACATTAT TTACATTAAT ATTTTTTTTTT TTAATAAAT TAATAAACA GTCTTGTATT
8521 TTATTAGTTT CGCATAAAAT TATAGAATAA TTTTATTAT TATAAAACCA AAAATACAAA
8581 TAGTTTGGTA TAATATTAAG TCTAAAACCT ATTTGGATTTA TTTTGTTCAT AAATTTATTA
8641 TTAATTTTTT TTATTTANNT TTGATTTATA ATAATATTTT TTTTTTTTTG ATTTAACATC
8701 TATTTTTTTT TGACGTGGAA AAGTTCTAGT TGGTACAAA TCACCTAATT TATGACCTAT
8761 TTTTTTTTTT TTAATACAAA TAGGTATATG TTCTTTTCCA TTGTGTATTG CTAAAAATTTG
8821 ACCAACCAAT AAAGGTATAA TGGTTGATGC TCTAGACCAA GTTTTAATAA TTTCTTTTTT
8881 TTTTCTATT TTATTTTTAG TTATCTTGTC TATTAAATGA TTATGTACAA AAAGTTTTTT
8941 TTTATTCTTA TTTATAATAN TTATAAKANA TTTTNANAAA ATTTTATATT ATTTTTTAGG
9001 ATAAAAATTA AATCTAAACA TACTAAAGCG GCTACCATTA CTGTATTTTA AGCAATAACG
9061 ATTTATGCAA GCTAAGTCTT CTAAAGAATA ATTAGGCCAG AGTAATTTTA TGAAAAATTT
9121 ATTATTATCA NANNNNACTT AATACATAAT CAAANTAAAT TNATTAATTT TATATTTATT
9181 TTTGTAAATTA TTTTTAAAAG TTTTAAAAAT TCTTAATTGT CTACGAAAT TATAAGTTAA
9241 AATCGTTTCA GGAATAATAA TATTAGTTTT TTCTTTATTT ATTTTATTAT TTAATGAAAG
9301 ACTTAAAAAT TGATGTATAA TAAATTTTTT ATTAATAAAC GAATTTGGTA TTATGCCAAT
9361 AGAAAGCCAT GTAAAGCGTT GTCTATCAGC ATATAAACAG TCTTCATCTT TTATAATCTT

9421 TTTAAACTCA TACTTTTGTC CAAAAAATAA ATGAAGTACA ACAAGATTTT TATGAACTTT
 9481 TTTAAATTTT ATTGATTGAA ATAAAAGTTT ATAAAAAATT TCTTTTTTTT TTTTTTTTAA
 9541 ATCATAATTA TCTTTGAAAT CTATTTTATC TATTAATCTT AGTGTTAAAG GAAGTCTTGT
 9601 TAATAGGAAT TTAAAAAAT ATCTAACCTC GCTAGTTAAT TCTATTTGAT AATTATTATA
 9661 TTTTTTATTA ATTATATACA AAGCTTTGCT AAATTCCTCT TTTCTACTAT CAACTTTATT
 9721 TCCACCANCA ACAKAAATTC TACTATCTAC ATTTTCATCC NNATAAGAAA TATTATTATT
 9781 TTCATCTTGA TCAGAAAAC TATTATTCGT ATCNNNNNNN NNNNNNNNNN TATTTTTATC
 9841 TTTATTTTCA TCTTTATTAA ATTCAGATAA TTTTGTAAA TATATCTGTT CATGTATGTC
 9901 TTCAAATTTA GTATTAAAAA ATTGATTTAT AGGAAAAAGC CAAGGTTTTT GAAGATATTT
 9961 TCTTTGTTTT AATCTAAGAT TTAATAAAAT ATTATTTGGT GCTGTTATAA AAATATTATA
 10021 TAAATWTAAT TTATTTTTAT TAAATTTAAA TTCTGGAATA ATTTTTTTTT CTAAATTTATG
 10081 TAAATTTTTT TTAACATAAT GTCTAAGGAA ATATTTTTTT CTTTATTTTA TTATTAATTT
 10141 TGCTTGCTCT TTTTTCTTTT GTATTTCTCT TTTTATCCTT TCTGCTTCTT CTTTTTTTTT
 10201 ATTGATTTCT TCTTCTTCTT TTCTCTTTTC TTCTTCACTT TTTATTTTAC CTTTTTTTTCT
 10261 TCTTCTTCTT TCTTCTTTT CTTCTTTTTC TTTTTCTTTT CTTTCTTTT CTTCTTTACT
 10321 TTTTCTTTCT TTTTCTTGTT GTTTTCTTTT TTTTATAATA ATTTTTATTA ACTTTTCTGT
 10381 TTCATTATTT TGTTTAAATA ATTCGTCAGA ATCAAAATGT TTTTTTCTT TTTTTAAAAG
 10441 TATATTTTGA ATAGATAAAT TTTCTTTATT GTNNNNNNN NNTTATTTT TAATAGTTAA
 10501 CTTAATMGTT AACTTAATCG GTAATATTTT TTCATTTAAA CTATATTTT TTAATAAATT
 10561 TATAGATTGA TAATTTTTTT TTATATTTTT TTTATTTAAA AAATTTTCTG GTATGAAATT
 10621 ATTAGAACTT ATTATAAAAC TTTCCCAAGA AAAACCAATA TTACTAAAAA GTATTCTTAA
 10681 TTTATTTAAA TTTATATAAT TGTCATGTTT GAATAAAATA CTATATAGTA AATATGCTTG
 10741 AGATATTTTA ATATTTTGTG GATAATTTTT TATACTTAAA TCTATTATAT TTGAAGATAT
 10801 ATCTAATCTT TCTTCTACAC TANNKNGNAA NANANANANA NAAANANNN NNNNNNNAGN
 10861 NNNAANAAAA NNAAANANN CNNCANNNN NNCAANNANA AAANNAANAN NANNAACNNN
 10921 AACANAANAA ANAAAAATAT TCAACTTATA AACATATAA ATTATTATAT ATANNNNNNN
 10981 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 11041 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 11101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 11161 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NATGATAATA TTAATATTAT TAGATATTTT
 11221 TTTTATTTCT ATATCATAAT TTATAATTTT TGTTTTTATA TCAATCATAT TTTTATTTT
 11281 ATACTTTTTT ATTTTATAAT TTTTCAAGATA TAAATTTTTT TKTAATTKAC TTTTTTTAAA
 11341 AATTTTAAAA AGCTTATAAA TATAAATAAT TTTTATTAAT TTACTTTTAC TAATATATTC
 11401 AATTAATTTA NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 11461 NNNNNNNNNA ATAAAANNAA NNAGNANNNT AAAAAWATAA ATAAATANAA NAATATAAAT
 11521 ANAAANANAW ATAAATAAAT ATATATATAT ATATATATAT ATATATATAT ATATATATAT
 11581 ATATATATAT ANTATKTKTT GTATATTATA ATTTAATANA ATTATTAGTC CAATTATATT
 11641 GTAATCTAGT TCATATCTTT TTTTATATAN TTCTATATTA ACAAGTTTAT AAAAAGTTGA
 11701 ATTTAACAAA TATTCAAATT TTTCACTAGA AATGATTCTA GTTCTTGTG CTTTCATGTC
 11761 TAAAAAATA TAATCTCTC TACTTTTAAAT TTTAACTTCT TCTTCTTTTT CAGATTTTTT
 11821 GTCGTACCAA GGCTTAAATC TAAAAGGATT CATTATTTTA ATTTGAATAC CTTCTGACAT
 11881 CCAATCTTCT GGAAGTTTAT TTTCAACTAA TTCAATACCA TTATAAGTAC ACATAATATA
 11941 TTGTCTTTTT TTAAAAATAA ACATATCTTT ATGATATTTA ATAGTTGAA ATAAAAAAT
 12001 ACGAACTAAA GTTTTAACCA TAATTAATAA AGGTATTTTT ATCCACATTC TTATATATAA
 12061 TTGAGAAAATA AGACAACAAG TTCTGAAAAA ATTTACAAAA CCATAAGTAC TATCCCAAGC
 12121 AACATTTCTG CCTGCTCTTT TCTTAGACTC AATTAATTTA AAGAAATCTT GTTGTTTTTT
 12181 TATTAATCTT TTTTTTACAA TATCTGCATT TTCAGGACCA TAAAATTCAA GTAAATCATT
 12241 GTCATTTAAA ATTAAAAGCC TATCTTCTTC GTCTCTCTCA AGAAGAAAAA TGGGAGATGA
 12301 AGGAGATCTA TAAGTTACAA AATCAGGTAA GTTAAGAGGA ATCTCTTTAC GTTTTTGATT
 12361 ACGTAAAGAC CCCCTTACAG AAAATCTATT TAAATCATAA AATCTGAAA ATCTAATATC
 12421 ATGCAATCTC CTTTTTTTTT TTTTCGTCTT TTCAATCTT TTTTCAAAAT TATAACCAA
 12481 AAAGTTATTT TTATCCACTT TTTTATATTC TTTATCTTCA ATTTTTTTTAT TTACATTTAA
 12541 AGAGTCAGAT ATACAAGCTA ATTTTAAATTT ACTTAAANNNT ACTTTATATT CAACTGTAAT
 12601 TTCTTTTATCT GTAACCGTNN NNTCTATANC TTANNNTANC TANTCTTTAN NTTATCTGTT

12661 ATTTCTTTAT CTTTATCTTT TNNWNTTCT TTATCTTTAT CTTTCGGAGG TTCTGTATAA
12721 AGTATCAAAG GTCTTAATAT TCTACATTTT AAACCTGTAA TTTTTTTTAGA TCTCAACAAA
12781 AGAAATCTTA TATCATCTTT TAAGCTACAA ATCCATCTAG GAATATTTTTT TTCTTCGATA
12841 ATAAAAATTT GAGTTTCTTC ATATTTTTTTT TTATTTTTTCA AAAATTTTTAA AATAGGTGGT
12901 AGATTTTCAT ATNATATATT TTTTTTATTGC TNNNNNCATT TTCATTAAAT TTATTTATGA
12961 CTGATTCATC TATATATTAT TNTATTTAAN ATATAAAATTT CATTAAAAAT AATTAGGTAT
13021 GACTTNATAA AAATTACCTT TTAAAAATATC ATCTAAATTTT ATTAAAGCTT TTTTTTCATT
13081 ATCTTTATCA AAATAAATAA AAAATCTATT TTTTATTTTCA AATAAAATTTA AATTTTTTGA
13141 TTGATCAAAA GATTCTATAC GTTTTTTTAAA TTCATTTTCA AAATTATTTAA GTTTTTTTTTT
13201 TAAATTTAAAT CTGTCATAAG GATATATTTT TTTTAAAAAC CTTGTGTTTTT TTTTTATTAA
13261 TTTAAAAAAG GCATACAAAT TTGGTGAAAA TGTATATAAT AATCGTTTTT TATTATCAAT
13321 TTTATAGAAA CTAAAAATA ATTGTGACAT TTCAGTTCGT ATAAATTTAT CAAATTGTTT
13381 ATTTTTTATA TATCTTATAG GACGATTTCCA ACGGTAGTA TAATAAAAAA TATCTACTAA
13441 ATCTTTATTG TCTTCTTCTG ATATATTAAG TAAGTTTAAAC TCTTCTTCGT CTCTACATC
13501 ACTTCTTCT ATTTCTGATT CACCTCCTTT TAGTTTTTTTT GCTCTTCTT TTTTATTGTC
13561 TTTTTCTACG TCATTTATGT CCTTATATTT TTCTTCTTTT TCAACGTTAG AAAATATAGA
13621 TCTTTTTAAA AGAATCTTTC GGTCATTTAA AGTTATTGAT TTTAAACCTG TTCTTTGAAA
13681 ATTTAAAAGTA AATAAAGGTA TAGGTGTTCT AGACAAAAAT ACTAAAAATA ATATTAGATA
13741 AATATATAAA AAATATATAT CTAAACTGTC TTTGTTTTTTA AAACATTTTA TTACTTTATT
13801 TAGAAATTTA TACATAACAC TTTCGCTTTT TTTTAAAAAA AATTTAAACT TAAACCAACT
13861 TATCAGTTTA TAAATATATT TTATTATAAT AAAATATGAA CAAATACTAC CTAAAAATTT
13921 AAATACATAA AATAATATTT TATTAGTATT AGAATAATTC TTAAATAAAA AAGATATACT
13981 CAATCTCTTA AATATTGGGT TTGGTAAATTT AAAATTAATA ATTTGAAAATA TAAAAAGTTT
14041 TATAATTATA TCTAAAAATA AATTTAAATTT AAAATTAATA TCTTTATAAT TTTTAAATTA
14101 ATATATTATA AAATTTTCGA TATATTTTTT TAATTTTATT TTCATTTCAA TATAGTATGA
14161 TTTAAATACA TATAATAAGT TTATTAGTTT AAATTTTNN TAGTATTTTTT TTCATAAAA
14221 TATTTACCWW NAKAAAAANAA AGNNTTATT ATCGGCTAA TTTTAAAAAT AATTTTAAGA
14281 TGNNGNNTTT TTNGNTTTTT AAATATAAAT GTTCTAAAT TATCAAAAT TCTTAATCTA
14341 AAACCATATA AAAGTTAAT TAAAAAATAT ATCCAGTATA TACCCTGTTG TAATATACCT
14401 TTATAAAAAAG TATAATATAA AAAAATGCAT AATAAAGTGT GAGGTCTTTG TAAAAAAGA
14461 AACAAAGGCG TATAATATAT AGATAAACCTC ATTATTAATTT GACCTAAAC TGATGAGCTT
14521 ATTATCACAA ATATTTTTTTC TTGTTTTTTTT TCAAAATCTC TTTTAAATAA GTGCTCTATA
14581 AAAAAAATAT AAGAAGGTAT CACACAGAAT GAACTAAAAA AACCAAAATA AAAACCTATA
14641 AAAGAATAAA AAAAAAGTTT ATTATTAATTT AAANNNNNT TTCTTAATAT TAATAAAAAA
14701 TGTATTATTA CWWWNTGTAT TATTACTTTT ACCATAATTT TTTATATTTG AATGNNTTT
14761 ATATTTAAGT TTGTATATTT TTAAANNNA TTAaaaaaagg TCCCCACCAC TAGTTTTTTTT
14821 TATTCGGCAT AAAATTTATT TTCCATAAGA TTCCCTTAA AGTATCGTCA CTAATTAAGA
14881 CTTTACCAC CTAATTCGAA AAGGTGTTG GTGTAGATCA TCTTAGTTAA AACACCGAAA
14941 GTTTATGAAT ATATTAAATTT TAATTTTTTTA AATTTATTATT ATTCAAATCA TCTTTCGGTC
15001 TTTAGAAAATA TTTAAGCTAA ATATATTACT ATATTTACAC ATAATACCTA TTAACAATTT
15061 ATCTAATTTGT AACCTTAATTT TTANTTATCA TATATATATT TATATAATTA TAAATTTATA
15121 AATAATAAAG GGAGCACTCA TCTTTGTTTA AACTTACTAC TTATATGCTT TCAGCAGTTG
15181 TTCATTACGC ACATAGCTAC CTAGCGTTTA CTATTGGAAT AATAACTAGT ACACAAGTGG
15241 TCGTTCGTT CTGGTCTCT CTGACTAAGA ACGATTAAAC TCAATACTCC AACGTTTACA
15301 CCGATATGG ACCGAATGT CTGCGAGCT TCTGAACCCA GCTCACGTAC CGTTTAAATG
15361 CGCGAACAGC CCAACCCCTG GAACCTCCTT CAGCTCCAGG TGGCGATGAG CCGACATCGA
15421 GGTGCCAAAC CTTCCCCTCG ATGTGGACTC TTGGGGAAGA TCAGCCTGTT ATCCCTAGAG
15481 TAACTTTTAT CCGATAAGCG ATGGCCTTAC CACTCAGTAC TACCATCGGA TCACTAAGAC
15541 CAACTTTTCGT TCCTGCTTGA CTTGTTTGTG TCACAGTTAG GCTTTCCTTT GCCTTTACAC
15601 TTNTAAAAAT CATAAAATAG ATTTAAGAAA ACCTTTATTG CACGCCCTCC TTACTTTTTTT
15661 GGAGGCCTAC GCCCATATA AACTGTCTAC CAAAGACTGT CCTTGCNTAA AAAACAAGTT
15721 AGAAGTTAAA CTTAATTAGG GTGGTATTT ACTGATGACT TAANANNKMM TATTGCCTCC
15781 CACCTATATT TTACAAAAAA AAGACCAACA TCAATCACAG GGAACAGTAA AGCTTCATAG
15841 GGCTTTTCTG TCCTGGTGTG AATAGTCCGC ATCTTCACAG ACATTTTTTAT TTCACCGAGT

15901 CTCTCTCCGA GACAGTGCCT AGATCGTTAC GCCTTTCATG CGGGTCGGAA CTTACCCGAC
15961 AAGGAATTTT GCTACCTTAG GACCGTTATA GTTACGGCCG CCGTTTACTG AGATTTCCGGT
16021 TGCTAGCTTT CCTGTATTTT TCAGGTCAAC AACTTCGTTT ACCTTACAGC ACTGGGCAGG
16081 CGTCAGCCCC CATAATGGT CTTACGACTT TGC GGAGACT TGTGTTTTTTG TTAACAGTC
16141 GTCTGGGCCCT GGTATTGCA ATCCTTTTTA AGGACACCC TTTTCCCGAA GTTACAGGGC
16201 TATTTTGCCG AGTTCCTTAG AGAGAGTTAT CTCGATCCCC TTAGTATTCT CTACTCATCT
16261 ACCTGTGTAG GTTTACGGTA CAGGTATTTA TTTATTTNNN NNNTTTTTTC AATTTAGTTT
16321 TTCCCTGGAAG TATGACCTCA GTTATATTA TACAAATAAC ANNNNNNNNN NNNNNNNNNN
16381 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
16441 NNNNNNNNNN NNNNNNNNNN AGACTTTACA AAATTGTTTA AATTTTTCGCT TATATCTTTC
16501 GACTGATAAC CATTATCAG CTAACCTAGC CTTCTTCGTC CCTAATAATG AAAATAAATA
16561 AATAGTACAG GAATATAAAC CTGTAATCCA TCAACTACGC CTTTCGGCCT TATCTTAGGT
16621 CCTGACTAAA CTTTCGCATA TAAATATTTT GAAAGAATCC TTAGATTGTC GGGGCATAGG
16681 ATTCACACCT ATGTTTTTCGT TACTCAAGCC AACATTCTCA CTTCGACAT AADANNANAA
16741 AATTAAAGCA GAACGCCCC CTACCGANNT TGTTAATCTC ATAGTTTCGG CATATTAATT
16801 TAGACCCGTG TATTTTCGGC AAAAAACGC TTTATCAGTG AGCTTTTACG CACTCTTTAA
16861 AGGATAGCTG CTTCCAAGCA AACCTTCTGA CTGTCTATGC ACTTTCACAT CCTTGTCCAC
16921 TTAATAATAA TTTATGGGCC TTAAC TAATG ATCTGGGCTG TTTCCCTCTT GACAATGAAG
16981 CTTATCCCCC ATTGTCTCAC TAGTTAATTT ATAATTGTAC AGTATTATTA GTTTGTATAA
17041 AATTTAAAAT TTAACAGTG CTTTACCCCT GTTNNNATCA TAATTAAC TG GTCTTCG
17101 ACACATTTTCG GGGGAACCA GCTAGCTCTG GGTTCGAGTG GCATTTTACC GCTATCCACA
17161 ATTCATCCGC TGATTCTGCA TTATCAGTCG GTTCGGACCT CCACTTAGTT TTAAC TAAAT
17221 TTCATCCTGA CCATAGATAG ATCATCCAGT TTCGGGTACA TAAGTAATGA CTTATANATA
17281 TATNANATAT ATATATTTAT TGCCGTTATA GACTTGCTTT CGCTATGGCT ACGACTTATA
17341 TTATTTCTTAA CCAAAGCCAT TAATTATGAG TCGCTGGCTC CTTC TTCAAC AAGCAGCAG
17401 TCAGAGATAA TAAAAATCT CCTCCTACAA TTTGACAGCT TTTGGTTTCA TTTAAATATA
17461 TCTTTTAAAA GTTATATTAT ACCTTTCCCT CACGGTACTA CTTTCGCTATC GTTACTCAG
17521 GAGTATTTAG TCTTACAAGG TGGTCTTGT AGATTCACAC GGAATTTAAC TTGTTCCAGG
17581 TACTTAGAT AAGTTAATTA NTTTTANNNAT TANNNATAAT TAAGTATTAA ATATAAANT
17641 ATAAAAAAA TATGTTTTTA TATTTTGACT ACTGGACTTT AACCATCTAA GGTACAGTAT
17701 TTTTANCGT TTTATCTAAT AATATTTAT AAATTACAN AAAATTTAGT TTTATATAAT
17761 AACTTACCTA CAACTTCATT TTAANNNNTG ATAATGATTT AGACTATTTT CATTTTCGCTC
17821 ACCACTACTA AGGAAATTAC TTTTGTTTTA TTTTCCCTCT GATACTAAGA TGTTCAGTT
17881 CTCCAGGTTA TCTCTTACCT TATGAAGCA GTTTTAAAAA TTAATTTATT CGGGAATCTT
17941 TAATTTTAT AAATTTTAT TAAAGCTTTT TCGTACTTTA TTACGCCCTT CTTCGTCTCT
18001 GAGTACCAAG GTATTTACCA AAAGCTTAT ATTAAGTTTG AATATTATAA AAATAANNAA
18061 AAGATTTAAA TTATAATTAT ATNNATATAT ANNATGTATA TTANNNNNNN NNNNNNNNNN
18121 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18181 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNAA NNNNNNNNAA ANAANTTAAT
18241 ATAAC TTTTA GTACTTTTAA TAAAATTCNT AGNACAAANA NAGANNCANN AANAANNACA
18301 AANAAAAAAG ANAANNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18361 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18421 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18481 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18541 NNNNNNNNNN ANANAGAANA NAAANKTATG TCAATAATA TTTTATTAT TNNNNNNNNN
18601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18661 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18721 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18781 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18841 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18961 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19021 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
19081 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

19141 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNAAAAT AATAAAAAAA AAAGAAANNA
 19201 ANAAAAANA AAAAAAAA AGTTAACCTA TTAGATTNAA NANKNANANA NAANNANANA
 19261 NAANANANAN ATATAAACA TAAAAATTTT ATAATGTAAT AAAAAATTC AGCTGCACCT
 19321 TCCAGTACAG CTACCTTGTT ACGACTTCAT TCTAGTTACC AATTTTGTCT TCGTCATTTA
 19381 TTAAATGATT CTAGACATAA CCAATTTCCA GACTGTGACG GGCAGTGTGT ACAAGGTCCA
 19441 AGAACTTATT CACCGCTATA TTGCTGATTA GCGATTACTA GCGATTTCTA CTTTATGAGT
 19501 GCGAGTTGCA GCCCTCAATC CGAACTGAGA TTGGGTTTTT TGATTAGCTC TCCTTTACAG
 19561 GATTGCAACA TTTTGTCCCA ACCATTGTAG CACGTGTGTA GCCCAGGATT TAAAGAGCAT
 19621 GATGACTAGA CGTCGTCCCT TCCTTCCTCT AGCTTACACC AGCTGATTC TTAAAGATAA
 19681 ATTTAAATTT AATATATATT ATTATAGGGA TAGGGACAAA TCCATAATAT ATATTATTTA
 19741 TAAAAATAAA GATAAGTATA GGGGTTACGC TCGTTAGGGG ACTTAAACCTA ATACCTAAAA
 19801 GCACGAGCTG ATGACAGCCA TGCACCACCT GTATCTTTAT ACTGGATAGT ATAAATTTTA
 19861 TTTTTCATAA AAAATTTCTA CATGTCAAAT CCTGGTAAGG TTCATCGCTT TGCATCGAAT
 19921 TAAACCACAT GCTCCACCGC TTGTGTGGAC CCCCCTCAAT TCCTTTGAGT TTCATTCCTG
 19981 CGAACGTACT TCCCAGGCGG GATACTTTAC GCGTTAGCTT TAGTATAACA TAGGCATATT
 20041 TTTTTTATTA TTTATGCTTA ATATGCTTAG TATCCATTGT TTACAGCTTA TGACTATCGG
 20101 GGTATCTAAT CCCACTCGCT ACCATAGCTT TCGTTACTCA GTGTCAGTAT TAATCTAGCA
 20161 AGGTGCTTTC GCCATTGGTG TTCTTTCCAA TATCAACGTA TTTACCACCT CCATTGAAAA
 20221 TTCCCTTAC TTTTAGTATA CTCAAGTTTT TTAGTTTTCTA ATGCAGGTCA GAGTTGAGCT
 20281 CTAATATTTT ACAATAGACT TAAAAACGA CCTACGAACC CTTTACGCC AATCATTCTG
 20341 GATAATGCTT GCATCCTCTG TATTACCGCG GCGGCTGGCA CAGAGTTGGC CGATGCTTAT
 20401 TCCTGGAATA ACGTCATTTT TTCTTCTACC ATAAAAGAAG TTTACAACCT TTTCAGGCCT
 20461 TAATCCTCCA CGCGACATTG CTCCGTGAGG CTTTCGCCCC TTGCGGAATA TTCTCACTG
 20521 CTGCCTCCCG TAAGAGTCTG GGCCGTGTCT CAGTCCAGC GTGACTGATC ATCCTCTCAG
 20581 ACCAGTTACT GATTATAGCC TTTGGAGGCT GTTACCCTAC AAAC TAGTAC ATCAGACATA
 20641 AATCCAGCCT AGAGCGATTT TTAATAATAA ATATTAATCT TTTACTCCTC AGNTTATAGG
 20701 AGAATTAGCA ACTGTTTCCA ATTTGTTATA CTCTCNTCT AGGATAGGTT TTAATGTATT
 20761 ACTCACTCGT TCGTACTAT AAAAAATAN NNNNNNTCGT TCAACTTGCA TGTGTTAAGC
 20821 ATGTCGCTAG CATTCACTCT GAGCCAGGAT CGAACTCGTT ATATTTTTTTN NNNNNNNNNN
 20881 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 20941 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNTAAT TATTTAGAAA
 21001 GTTATTTTAC TCATAGCTTC ANNNNNNNNC NNNNANNNA NANNAAAAA NNCNANGNNN
 21061 ANGNANNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 21121 NNNNNNNNNN NNNNNNNNTA TACATCTTTA AAAAAATTA TTTNATCTAT AGCTTTTTTA
 21181 GTAGTTTCTT CTTTTATATT TATTGCATTT CCTTTAAAAG CATCTAATAA TTCAAAACCT
 21241 AACTTAGAAA CCATGTTATA TCCAAAACGT TTTT TAGATG CTGATATTA TAATTTGCGA
 21301 GCATACATAT ATTTTTTATC TTTAGGTAATA TTATAATATA TTTTACGTAT TGATGCTTTT
 21361 CCTTTTTTAC GTTTTTTTTT TAATATTTGT GTTTTTACTA TTTTTTTTTT AGTTATTTTA
 21421 TTTAGTGCTT TTTGTACAAC ATAAATGGA TTTAATTTTA TTTTTTTTTT TATTTCTAGT
 21481 AAAGATTTAT AAATATTTT ATATGATATT GATTTTTTTC CTTTCTTAA AATGTAATTA
 21541 GATAGATTAT AAATGATTT TTTTAAAAA ATTGGATCAT ATTTTACTTT ATAAGTCATT
 21601 TNTANNATTA TNTNTTAAT TATTNAATTT TTTGTTTTTT AGCTCCATAT TTTGAACGTC
 21661 TTTGAAATCG TTCTTTTGCT CTACCAGCAT CTTTACTGCC TAAAATAACT TTATGTTTTA
 21721 CACCAGGTAA ATCTTTTACT AAGCCACCT TTACTAAAAC TCTTGAATGT ACTTGAACCTA
 21781 AATTAATACC AATACCTGGT ATTAATGCAG TTACTTCAAT TCCAGAAGTC AATCGTATTC
 21841 GGGCAACTTT TCTTGAGCCA GAATTAGGTT TTTTTGGATG AATAATGTAG TTAACAGATA
 21901 AAAAAGTCTC ACTTTACAGA ACCTATACAT GAAAGTTTCC TTTCATATAG CTCTCATT
 21961 AAATATTTAT TAAACCYTTN AAATTTGTTT TTATTAATAA TTATTAATAA ANNNNNNNNN
 22021 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 22081 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN AAANNNAAN NAAANAGNNN
 22141 ANAANNNNA NNAANAANN AAANGGNNNA ANGNAAGNAA ANNAANNCA NANCNNAANA
 22201 NNNNNNNANN ANNNNNNNNN NNNNNNNNNN NANNACAGAA ATTATATAA AAAAAATACT
 22261 ACATTTANTA AAATAAANA TTTATTTGCT TTAATAATAA ATNTATATA TATNTANNNN
 22321 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

22381 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 22441 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 22501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 22561 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 22621 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNCATTA TTATATAATA TTTTGNTTAA
 22681 ATAATCAATT GCTTCATAAT GAATTATAAT TGAAATATTA TGTATTCT TATATTTATA
 22741 ATTANNNNN NANNNANAAN NAGATTTGTA AGAATCANN TTATTATTAA ACATAAAAT
 22801 TTTATATAGT AGAGGTTTTA ATGTATTATA TACAAAAAAT CTTAGTTGGT CAGAGAATCT
 22861 ATAATCTTCA TACATAGCAT ACATAAAATTT TCATAAAGGA ATATAAAAT CTTTTTTTTNT
 22921 TATTTTATTC GATGGATACN NNNAAATAGTA AACGATAAAG TATCAGATTA GTAATAAAAA
 22981 GCACTAATAA CATGTCNTAA GAATAAAAAAT ATCTTTTTTTN AATTTTTTTAT TTGATGGGTT
 23041 ATTAATGGGT AAAGTATCAG TTTCGTAATT ATAAGCATTT ATAACAGAAA AATCTAATTC
 23101 TAATATATTT GATTTAATTA AATTATATTT ATTAATATTT AATATTTCTT TGTCATTTTT
 23161 TATTAACTTT TCTATAAAAT TTATTAGTCT AAAATTGTCA AAATAAAAAA CTTTTTCTAA
 23221 ATTTAATAAA TTATTAAATTA ACTTTAAATC ATGATTAAAA TATTCATGTT TTATATTTTT
 23281 AATCTCTGAT GTGTTAAAAA TTTTATAAAG AATTAATCCA GATAAACATA TCAATAAATA
 23341 AATAATTACA GTTATTATTT TTATATTTGT TCCGAGTTCA TAATAATTTT TATATAAAGG
 23401 ATATTTTTTTT GTTTTGAAT CTATTAAATA TACAGACATA GGATCAACTA TTAAACTATT
 23461 TAATCTAATT TGTAATAAAAA AACGTCCAAT TTGATAAAAA ACAATTATAA TAGGAACTTT
 23521 GTATCCTTTT TTTATGTATT GATTAAATCT AAAAAAATAA AAAATAAAT TATTAATTAT
 23581 TTTTGAATTT ATTACATCTT GATTTTGTGT AAGACTTAGT GATAAAGTTT GATTTGTCAA
 23641 TATAGATAAA TCTCTTAGAT TTTTATATTT TAAATGTTCA TGTAATAAAT CTTTTTTTTTC
 23701 AATATAAAAA CCTTTTGTTT TTAATAAAT CAAAATATAT TTTTTTTTTTA TTTCATAGT
 23761 AGTTTTTTTT AATAAAAAA ATTTACTTAA TTTATCATAA CCTATTAATA TAGGTTCTAC
 23821 TTCTTTTGGT TTATTTGTAT AAGCTAAAA TAAATTATTT TTATTAAAA TAGAAATTA
 23881 ATTACTTAAT TCATAAAAT GTTTAAAAAT TAAATTATAA TGACATAAAT GAATATTTGG
 23941 AATAATTATA ATACATGAAG AAATAATTC TATTAGTTTA AATTGAGTTA CAATATTAAA
 24001 TACATCATAT ATACTTTTTT TTCTTTTTTG CATTAGTTTA TCTAATATTT TTATTTTTTTC
 24061 GAATTCAAAT GTTCTTTTAT CAGTATTATA AATTTTTTTTA AACATTTTTT TAATTTGTTT
 24121 AGTAGAAGTA TTACTTTTAT TAATATCAA TTTATCAACA AAAAGAAAAT CAGATTCATC
 24181 ATAACATAT TTTGAAATTT CATAGTCACT TAATTCCTTA TTAATTGGGT TTAATATAAT
 24241 AAAAGGCAA TCAGCATAAT CTTTTATAAT TTTTCTTGT AAAACAAAA CGTCTTCAGA
 24301 CTGACCAACT AATAATATAC CTTTTGAAGC TGACGGAATC ATTGTTAATG AAAAAAATA
 24361 AGGTTTTTGT CTTATTTTAT ATCTGTTTTT TTTTAAAAAT AAATATCTG TTATTTTTTT
 24421 ATAATTTGTA AAAAAAATAC GATTTTTTTC TTTTATATTT AATATATTAT TTGNNNCTTT
 24481 TGTATTAATT CTTAACATAT CTAAAAAATA TATTAAATAT AAATAACCAT ATNCMNMNKM
 24541 CACAGAGTTT TCATACTTTT TTATAGATAT ATTTTTAAAA ACATATTTAA TAAGTTTTTG
 24601 TTCATTTTTT ATATTTAAAA AAATATTTAT AAATTGAATT AAAAATTTTC TTTCTTCATC
 24661 AAAAATTAAT TCATTAGTAT TAATCCAAGA TTCTAAATCA AATTGTGTTT CATTTTTTTTT
 24721 NAATATANNG GTTTAGTATA TGTAGTTTCT TTTATAGATG TATATCTTG ATCTTTATTG
 24781 TTAAGAATAC ATTGATAAAA ATCTTCAGTT ATATCTTGTG AAAACATTT AAAAAATCT
 24841 TTACTTATGG AGTTTCTTAA ATAATAAAAA TCATAAATAA AADATTTATA CCAATTTTTT
 24901 AAATTTGATT CAGTATAAAA ATATAAATAT NCNTNATAT TAANTTTNAT ATTTATATAT
 24961 ATATATATA ANTTTAAAT TACTTTATGA NAATAAATAG TAGTTTTTAC TTTGTCTATT
 25021 TCTAAAAAT TAAAAAAT GTCAACAATA AAAAAATCGAA TAATATATTT ATTTATTTCA
 25081 GTTAGTAAAT TAAAAATTTT ATTTATAAAA TTAATAAAAA TATTGTTATA AAGATAAAT
 25141 GTGTTTTTCA TGCATGCTTC TAAATAACTA TATTTACATA ACTTTACATA TAGATCTATA
 25201 CTAAATTTTT TTGTCATTAA TAATAAATTT ATTTCATATT CATTTTTTAA TAATTTAATA
 25261 TATTTAAAAA TTAATTTAAA ATATTTTAAA ATAAAAAATA TAAAAATATA AATAAAAAAT
 25321 ATTTTAAATA AATAATTATA TTTTATAAAA TTTAATTGGT CATATCTAAA AATTTCAATA
 25381 ATATCTTTTT TAATTGGGTG TAATTTAAAC TGATTTTTTAA NAAANNNNN NNNATTTACT
 25441 AATAAATTTT AATTATTTTT ATNAGNTTTT GATAATAAAA TAATTATTTG AATAAAMTTA
 25501 ATAAAAATNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25561 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

25621 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25681 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25741 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25861 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25921 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25981 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26041 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26161 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26221 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26281 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26341 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26461 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26521 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26581 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26641 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26761 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26821 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26881 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26941 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27061 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27121 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27181 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27241 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27361 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27421 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27481 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27541 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27661 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27721 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27781 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27841 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 27961 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 28021 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 28081 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 28141 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 28201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNAATCCCT
 28261 GCTGGATACA AAAAAAAGAT TTATAATCCT TAATAAAAAA AAATGACATT AAATAACATA
 28321 TATAAAAAAA AAGTTAAAAG AAAATTATTA ATTTATGGAT ATAATTTTAA TAAAGGTAGA
 28381 AATTATAAAG GTACAATAAC TGTA AACAT AGAGGAGGGG GTCATAAGCG TTTATATAGA
 28441 AAGATAGATT TTAACGTAA TGAAAAAAT ACTGGAAAAA TATTA AAAAT TGAATATGAT
 28501 CCAATCGGA ATACAACAAT TTGTTTAGTT TATTATAAAA TAAATAATAT ATACAAATAT
 28561 ATAATACATC CACAAAATGT TGGTATTGGA GATACTATTT TATCTAGTCA TGAACATCT
 28621 ATTTATGCTG GAAATACATT ACCTTTAATT AATATACCTT TTGGTACTTT TATACATAAC
 28681 ATAGAAATTA AACCTGAAA GGGCGACAA TTAGTTAGGG CTGCAGGTAC TGCAGCAAAA
 28741 GTAAATTACAA AAAAAA AAAAGT TTTA TTAAAGTTAC CATCAGGAAA ATTATGTTCA
 28801 ATACCTTTAA CATGCTTAGC AACTATAGGA CAAGTTAATA AAAC TGCATT AAAAAA

```

28861 TTAAAAGCTG GGACAAATCG TTGGTTAGGA AAAAGACCTA AAGTACGAGG TACTGCTATG
28921 AACCCGTGTAG ATCATCCTCA CGGAGGAGGA GAAGGTAAAG CTCCTATAGG TAGAAAATTT
28981 CCTGCAATC CTTGGGGATA TCCTGCTTTT GGACGAAGAA AAAAAAGAAA AAAAAATTNAG
29041 ATTAAAAAAA AAAAAATATT AAAAAATAAT AAAAAATAA TAAAGAATAT AATTAAANNN
29101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNTT TTAATTTTTT TGATAAAATC
29161 CTGTTAAAAAT AGCATAAGAT AACTTATTTA AAATATAAAA TATTGAAGAC AAATTAAATG
29221 TATTTATAGG TATAACAATA TCTGCTAAAT CTATATTAAT ATCAGTATNN NNNNTTATAT
29281 TAATTAACA AATTGTTGAA ATTTTTAAAG TAAAACATTC TCTGATAATT ATATCATTTT
29341 GTGGATCAAT AAGTATTACA ATATCAGGTC TCTTAGTCAT ATATTTTATT CCTTCTAAAT
29401 TATTATATAA ACTAATNNNT TKTTGTTTTT TTTTTTTTAT AAAATTAATT TTTGTTTTAC
29461 TAATTAATAA ATTTTTTAAT TTAATTTTTG TAGTTAACCA GTTAGTAAGT AAACCTGCTC
29521 GTATATTTCT ATTAATATAA TGACATCTAA TTCTTTCAGA TATTAATTTA AAATAATTTA
29581 ATTTTAAGGT TTTATATTTA TTATTTATTC GTTTATTTTT TCTTTTTAAT CTCAATTTAG
29641 TTTTTATTTT ATTCCTTAAT TTTTTATNTT TTATTTATTTA TTAAAAACCT TTTAGTTAAA
29701 AAATCTTAG TTTTAATTAT TATAGGTTTT ATTTCTTTTT TATTTTTGTT TTGACTAATT
29761 ATTA AAAATG TTTTTTTTATT TTTTGCTGCA TAAAAAACTA AATTACAAGC TATTGTTAAT
29821 AAAGAAGCTA TTTTGTAAAT ATCTATAATA NTAATATTTT AAATTAATAT TAAAAATATA
29881 ATATATCATT TTAAATTTAA TATTTTTTGT ATTTAAAAAA TGAATTTTTT CTGNNNNNNN
29941 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN AACANANCAN
30001 NNAATTTTAA AATACAATCC ATTKATACTA TCATTTTAAA AATTCAWAAY ACAATCATAA
30061 ACTTGATGAC TATCATTTAT AATTATATAA ATGGATTGTG GATCGAGTTG GATTTGAACC
30121 AACGTAGACA CATAAGTCAA AGGATTTACA GTCCATTCC TTTAACCTCT CGGGCATCGA
30181 TCCAAAATAA NN

```

//

Apéndice 1.2 Secuencia intermedia de *Prosopanche costaricensis* con la inclusión de gap secuenciados (28489 pb).

```

1 ANTAATTACC ACCCTTAAGG GAAATCGAAT CCCTGTTATT TCCTTGAAAA AGAAATGTCC
61 TAAACCTCTA GACGATAAGG GCAATATTAA TAATATAAAC TTAATAAATG TATTGTGCAA
121 TGTTGCC TAG TGC GAAGGTG GGAATTGAAC CCACGACCTC AAGGTCATGA GCCTTGCGAG
181 CTACCAAAC T GCTCCACTTC GCGCTATATT TTTATTTTAA TTACCAACTT GTTATTTTAA
241 TTACCAACTT GATATTTTTG CTCCTGGTAA TTTACATGAA TTAAATGTTT CATAAATTA
301 ATTTCTAGAT TGTCCAAAAT AACGATAATA ACCTCTAGGA CGACCTTTTA TCGAACATCG
361 TTTTCTAATA CGTATTAATA TACTATTTTC AGGCAATTTT TGTACTACTT TATTTATTTT
421 AAATTTAATT TTTAATAAAT AAACTTTTCT TATTCTTTTT TTTAAGTTTT GTCTAATAAT
481 ACGATATTTT TTTTCTAAAT TTATTTTCTT TTTATCTCTA AAAATTATAC TTTTTTTTGA
541 CATATTGTAT TAATATAAAT TTGTATTAAT ATAACATAAT NAAATAAAT AAACAAAAA
601 ATAAAAATAT AAATNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
661 NNNNNNNNNN NNNNNNNNTGA ATAATNTTTA TTTTTTTATT TTTTATTTA AATAATATTC
721 ACTAATTTTA TGTTCATTAA ATAAAAAATC TAATCTTTTT TTGTCTAATT TTTTGGGTAT
781 TTTTTTTGGT ATTTTTATAG TTTTTTTATT TAAATTTAGT AAAGTAATTT TATCTTCAGT
841 TTTACAAAATA AAAC TGGGAA TATTTACTGT ATAATTGTTA ACTAAAAAT GCTTATGATT
901 TATAAATGT CTAGCTTCTG AAATAGTTTT AGACAAACTC AATTTAAATA TAATATTATC
961 TAATCGCATT TCTAAAAAAC TACATAATTT GTGCATTTTT AATCTTTTAA TTTTATATA
1021 TATATATAAA TAATTAGATA ATTGTTTTATT TGTTAAACCA TAATAAAAGC GCATTTTTTTG
1081 TTTTCTTTTT AAAAGAGGAA GATATTTTGA TTCACGTATA ATTTTAAATT TTTTTTTTTT
1141 TCTATTTGTT TTTTTTTTTT TATATTTTAC AATTTCAAAA CTACGCTTTC TTTTAACTT
1201 AGGTCCATAA TATCTAGACA TATAATTGTT TAAATTTTAA GTTTTAATAT TAATAAATAT
1261 AAAAAATAA TNNNNNANAA CNANAACNAA NACNAANANN NNNAGNAAAN NNAAGNNANN
1321 NNAAANNAN NAGNANAANN NNNNNNNNNA AANNNAAGNN ANNNNAAANN AANANAAAA

```

1381 AANAAANNAA ANNAACNNGN NAANNNNNAA NNAAAGNNAA ANCANAGNAN GAAAAAAAAA
1441 AATAATAATA ATATAGATAT AAAAAAATAT AAAGTAAGTT ATAAATATAA TAAATTTTAA
1501 CTTTACTTCA AAGATGATAT AAGTGGTAAA ACATTTGTGT TTTATATACA ATTTAATGTA
1561 AAATATTTTA ATAAAAAAT TAATTTTATT AGTTTAAAA TAGAAGAGAT TGGTTTTTACA
1621 TCATTAGAAG ATTTAGTATA TAATTTATTT TGTTCAGAA TTGAATCAAA ATTTAGTGAT
1681 TTTAATATAA ATTTTTCATA TGTAAATAAT AGTAAAGAGG AAGAAATTTT TGAAGAAAAA
1741 AAANANAAAA NAAACAAAA AAANGAAAA TATTATGATT ATTATATAGA TGAATGTTTA
1801 AAAAAAACT TATTAAATTA TTTATATATA AATAAATAA ATGAATATAT AAAATATTTT
1861 TTAAATGAAA AATTTGTTAA TAATAAATTT TATAATTTTA AATCATTTAA ATCATATGAA
1921 AAAATTTTAA AAGAGCTTAT TATTTTCAGAA TGTACTTATC TAACAAATGA TGATTATAAA
1981 AATGAACGTA GAAAAAAAAA AATATTAAAA AAATTATAAA TAAAAAATA TAATAAATAC
2041 AGAAAGACGA ATTTTATATTA TCAAAAAAAA AGAATTTTAA GTGAAGTATT AGATAAAAAA
2101 GAAAATTTTG ATTTTACAAA ATTATGGGTT TTTTGTGAAA ACTGTAATTT CTTAAATTAT
2161 AAAAAAGATT TTAAGCTTAA ATTATACGTT TGTGAAGAAT GTGGTTATCA TTTTAAAAATG
2221 AGTAGTTCAG ATAGAATAAA AATGATAATA GATAAAGATA CTTGGAATCC CATAAATGAA
2281 GATATTTTTT CTATAGATCC TCTAGAGATT GATGACTT TAGATGATAA TTATATAGAT
2341 GATGATAATA CAGACACCAA ATAATAATTA GAAAAAGAA GACAATGCAC AAAATAAAAA
2401 CAAAAAGAT AATTACTTTG ACCGCGATAA TAATTACAAA AATATTCNNN NNNNNNNNNN
2461 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2521 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2581 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2641 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2761 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2821 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2881 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
2941 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3061 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3121 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
3181 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNAAAAA
3241 GAAAAAAAAA AAAAGAAAA ACAAACTAAA AAAAAATNNAN GNNAGNNNG AAGAAGAAGN
3301 AGNAAAACNA ANANGAAAA AACAAATTC ATAGAAATTA TGGGCAAATG AAATAAAAAA
3361 TAATAAACAA AAAAATACAT AAAAAATGAA AAATTTGATG AAATAGAGAA AACAAATCA
3421 TATATTAATC GTTTAAATTT TTATCAAGAA AAAACAAGTT TACCTGAAGC TGTAAAAACA
3481 GGTATAGGTA AAATAAATGG TATTAAAGTT GCACTTGGTA TAATGGATTC AGAATTTATG
3541 GGTGGAAGCA TGGGATCTGC TGTAGGAGAA ACTTTAATC GTTTAATCGA ATATGCTACT
3601 AAAGAAAAAT TACCTCTTAT AATTTTCTGT GCTTCTGGAG GAGCACGCAT ACAAGAGGGG
3661 ATTTGTAGTT TAATACAAAT GGCTAAAATA TCTGCTGCAT TACATTATTA TAAATCAAC
3721 AAAAAATCAG TATTTATATC AATACTTACA TCTCCAACAG CAGGGGGAGT TATTGCAAGT
3781 TTTGGAATGT TGGGTGATAT TATTATAGCT GAACCAAATA CTACTATTGC TTTTGCAGGA
3841 CAAATAATAA TTGAACAAAT TTTAAGAAAA GAAGTACCTA AAGGTTTTCA AACAGCTGAA
3901 TATCTATTTA AACAAGGTGC TTTTGATTTA ATTTTACCAC GACAACTTTT AAAACAATT
3961 GTTACTGTAT TACTTAATTT TCACGGGTTT TAGACATGTT CTTAGTTAAT AATTTAATAT
4021 TTAATATTTA ATATTAATAA AATTATTTAT TAAATAAAAT TTATAAANN AAAACGCGCT
4081 CTGTAGGATT TTAACCTACA ACATTAGGTT TTGGAGACCT AAATTATTAC TAATTGAATT
4141 AAGAGCGCAT ATAANNAAA ATATTAAAA TGAAAAATAA AATAAGAAAT AAAAAACTA
4201 TAAAAAATTT AAATTTTSTA AAAAAATTA ATATAAATAA AATAAATAT AAAGAAATAA
4261 AATTACTTTA CAAATTTSTA AGTATAAAG GAAAAATAT ATCAAGAAAA GTAACTAAT
4321 TAAATTTAAA AAAACAAAA AGCATTACTA TAGCTATAAA GAGAGCTCGT ATATTATCTT
4381 TATTACCTTT TAATAAATCA ATTTACTTAA ATAAAAATA AACGATGCAA TAAAAATTA
4441 AAATATAATA AATTATTTAA TTTTAAAAAA ATTATAAAAA ACTATTAATC TTTTATCCA
4501 ATATATAATA ATAAATAAAT TAAACTTAAA TAACCTTCCT TTTTTTTATT AGTTAATTAA
4561 AACTTTTTTA CTCTATAAAA ATAAATTTAA AGTAACGATA TTATAAAAA TAGTTTACTG

4621 AACCTTTAAT CTGTTAAATT TATATAATAT ATAAATATAC ATTTATAAAT TTCCGCTTAT
 4681 TTAATAAAAA TAAAAATAAA AAAATGGAATT TAATTTTTTAA AACAATTTAAA ATAATTATAA
 4741 TTAACATAAA ACATACGAAT CACACATACA ATCTCATACA CATTCCATAA CGTTGAGGAC
 4801 ATTTTTTAAAG AGGAGCTGAT GATTTTGTAA TTATCTTATT TAATTTTCCTT GTTTTTTTTTT
 4861 TAAAAATTTG TTTAATATAA ACCATAATAT ATATATGGTT TAGATTTATC AAAACCTATA
 4921 TCACACTACA TTTTGTATTT GTATATGTAA TATATTTATT TTATATTTTTT TTTATTTACT
 4981 TATTTACATA TTATCTTACT TTTTTTAAAG TAGATTTACA ACCATTATGA GAAAAAGGAG
 5041 TAATATCTAA TATAGAATAA ATTTTTATTT TTTTACTTTT AATAACTTGT AATACTGAAG
 5101 TTCTTCCTGC ACCTATACCT TTAATACGTA TTTTCAGCAGA ATGTATTTTTT ATTATATTAA
 5161 AAACATTTATT TGCTGTAGTT TTAGCTGCAA AAGGAGTACC TTTTGTGTGA TTTTTAAATT
 5221 CATAACATACC AGTAGAAGCT GTAAAAATAA GTATATTTTTT TTCATCCATA ATTGTGATTC
 5281 TAGTATTATT AAAACTAGCT AAAATAATAA TAATTCCTTT TTCAAGGAAA TTTATTTTTTG
 5341 GTTGGCGCTT ATGACGATAA AGAACTTGTA TAGGTTTTTAT TTTAATTCCTT TTTCTCATAT
 5401 ATATNTTTTT ATTTTTGTTT TTGTTTATGT CTTTTATTAT TACAAAATAA TAAAATACGC
 5461 CTATATCTTT TAACTAATTT ACATTCCTGG CAAATTTTTT GAATAGATGT TCTTTTTTTC
 5521 ATTTTTTTTCA TTTTATACAA CATAAAATAA AATTTCCCTT CCAATTTTAT TTAATTTAGC
 5581 TTCTATATCT GTTATTATTC CTTGAGATGT AGAAAATAAA TAAACTCCAT TTGCATTTAA
 5641 TTTTTTAGGA ATTATTTTTAT AAGTTACATA AATACGTTGA CTAGGTTTAC TTATTATTTT
 5701 TACAAAATTT TTTTTTTTTA AAAATTATACT AAAATTATTT TTATTTTTTTA TAAAATTTTC
 5761 AATAAAACCT TCTTTTAAAA GTAATTTTTAA AATATTATTT GAAGTATTAT TAAAAGGTAT
 5821 TGTTGTTATC TTTTTATTTG ATAAACTTGC ATTTTTTTATA TGAACAATTA TAGTTATAAA
 5881 ATTATTTAAC ATAATATTTA TATTTACAAT ATTTTCAGAAG ATAATGAAAT TATTTTATTA
 5941 AAATTTAATT TGCCTAATTC TTGTGTAATT ATACCAAAAA TTCATAGATCC TTTAGGATTG
 6001 TTTAAATTTT CAATTACAAC AGCAGCATTA TCGTCAAAT TAAAATAAAT ACCATCAGAA
 6061 CGTTTAAATTT CTTTTCGTGT ACGTACAATA ACAGCTCTTA TTATTTCTGA TTTTTCTATT
 6121 GAAACACTAG GTAAAGCTTC TTTAATCACA GATATAACAA TATCACCTAT ATAAGCAGAT
 6181 TTAATTTTAC TAATACCTAA TTTTTGAATA CACATTAAC TAATAGCACC ACTGTTATCT
 6241 GTTACATTTA GATAAGTTTG TTTGTGAATC ATATTTTTTAA CCATAAAAAA ATTAGGTTTT
 6301 AATAAAAAAT TTTGTTTTTA CAGGTAATTT AGATGAAACT ATTTTTAAAAG TATTTTTTGC
 6361 AAGTATTTCC GATACGTTAT CTACTTCATA AAGTATTCGT CCAGGTTTAA TAACAGCAAC
 6421 CCAATTTTCA GTGGAACCTT TACCAGACCC CATACGTTTTT CCTACAGATC TTTTTGTAAT
 6481 AGGTTTATCT GTAAATATTC TAATCCAGAA TTTACCACCT CGTTTAAAAA ATCGTGTGCAT
 6541 TACTTTTCTA CCAGCTTCTA TTTGTTTAGC TGTGATCCAA GTTGGTTCAA GGGTTTTAAT
 6601 TCCAAATTTA CCAAAACAAA TAAAATTACC TTTTACTGAA ATTCCTTTTCA TCCTACCCCT
 6661 ATGTTGTTTA TCAAATTTAG TTTGTTTAGG TAACATTTTA TAATAAATTT TTAATAAATTT
 6721 AATTAAATAA AACTAAANAA AAAATTAAAT AANAGTAATT AATAGTTAAA GAAAAGGTCT
 6781 AGGTTGCGAA TATTCACCTA TAATATAATC TATAAGAATT GCTGTACCAA ATATAAGTAC
 6841 TATTCTAACA TAAGTTATAA CTTTTCTAAT GATTTTTTTTT TTTGTAATTA TCATAANTTT
 6901 TTTTTAATGT TTTTATTATT ATNGTTATAT AANTAANTAT AATAATGATA TAGAACTATC
 6961 ACAACAAAAA ATAAAAAAG TTATKTGTTA TAAAAAAAAT TAAATTTTAA ANAANTAKTA
 7021 GNTAATATTA AATTATATAT AAATCCAAAT TTTGATTCCCT AAAAGTCCAT TATTTGTTTT
 7081 AACTGAATAT GAACAAAAAC GTATATTTTC TTTAATTGTT TGTAATGGTA AAAGACCTTC
 7141 TTTAATCCAA GTTTTATTTT TATGTACAAT AGCTTTACCT GATGTTCTGC CAGATATCTT
 7201 TACTTTAATT CCTTTAATTT TTTTTTCATT GTAAATTTTA TATAAAACTT TTTCTTTTAT
 7261 TTTTTTTTAAA GTAATTCCTT TTTTTAATTG AAATAATATA TATTTAGCAA TAACTTTAGG
 7321 ATTTTTTAAA GGGTATTTAA TTTCTTGAAC TTTTATATTA AAATTTAACC TTTTAAATTT
 7381 TATAAATTGA TTTCTAAAAT TAGAATCTTT ATTAAACATT TTTAGTAATT TAAAGTTTTAA
 7441 GAAATCTAAA ATTATAATTT TTATTTTAAAT TTTGTTATTT TGTTTTGTGA TTCTAATATG
 7501 TGAAATAACA TTATTTACAT TAATATTTTT TTTTTTAAAA TAATTAATAA AACAGTCTTG
 7561 TATTTTATTA GTTTCGCATA AATTTATAGA ATAATTTTTT TTATTATAAA ACCAAAAATA
 7621 CAAATAGTTT GGTATAATAT TAAGTCTAAA ACTTATTGGA TTTATTTTTTG TCATAAATTT
 7681 ATTATTAATT TTTTTTATTT ANNTTTGATT TTTAATAATA TTTTTTTTTT TTTGATTTAA
 7741 CATCTATTTT TTTATGACGT GGAAAAGTTC TAGTTGGTAC AAATTCACCT AATTTATGAC
 7801 CTATTTTTTT TTTATTAATA CAAATAGGTA TATGTTCTTT TCCATTGTGT ATTGCTAAAA

7861 TTTGACCAAC CATTAAAGGT ATAATGGTTG ATGCTCTAGA CCAAGTTTTA ATAATTTCTT
 7921 TTTTTTTTTT TATTTTATTT TTAGTTATCT TGTCTATTAA ATGATTATGT AAAAAAGTT
 7981 TTTTTTTTATT CATATTTATA ATANTTATAA KANATTTTNA NAAAATTTTA TATTTATTTTT
 8041 TAGGATAAAA ATTAAATCTA AACATACTAA AGCGGCTACC ATTACTTGTA TTTAAGCAAT
 8101 AACGATTTAT GCAAGCTAAG TCTTCTAAAG AATAATTAGG CCAGAGTAAT TTTATGAAAA
 8161 TTTTATTATTT ATCTATTTTA TTTAATTTAT GATCTTGATA CTTAATTAAA TTTAATTTAT
 8221 TAAATTTATA TTTATTTTTG TAATTATTTT TAAAAGTTTT AAAAAATCTT AATTGTCTAC
 8281 GAAATTTATA AGTTAAAATC GTTTCAGGAC TAATAATATT AGTTTTTCTT TTATTTATTT
 8341 TATTATTTAA TGAAAGACTT AAAATTTGAT GTATAAATAA TTTTTTATTA ATAAACGAAT
 8401 TTGGTATTAT GCCAATAGAA AGCCATGTAA AGCGTTGTCT ATCAGCATAT AAACAGTCTT
 8461 CATCTTTTAT AATCTTTTTA AACTCATACT TTTGTCCAAA AAATAAATGA AGTACAACAA
 8521 GATTTTTATG AACTTTTTTA AATTTTATTT ATTGAAATAA AAGTTTATAA AAAATTTCTT
 8581 TTTTTTTTTT TTTTAAATCA TAATTATCTT TGAAATCTAT TTTATCTATT AATCTTAGTG
 8641 TTAAAGGAAG TCTTGTTAAT AGGAATTTAA AAAAAATATCT AACCTCGCTA GTTAATTTCTA
 8701 TTTGATAAAT ATTATATTTT TTATTAATTA TATACAAAGC TTTGCTAAAT TCTCCTTTTC
 8761 TACTATCAAC TTTATTTCCA CCACAACCAT AATTTCTACTA TCTACATTTT CATCCCGATA
 8821 AGAAATATTA TTATTTTCAT CTTGATCAGA AAAC TTATTA TTCGTATCTA TTTTATCTT
 8881 TATTTTCATC TTTATTTAAAT TCAGATAAAT TTTGTAAATA TATCTGTTCA TGTATGTCTT
 8941 CAATTTTAGT ATTTAAAAAT TGATTTATAG GAAAAAGCCA AGGGTTTTGA AGATATTTTC
 9001 TTTGTTTTAA TCTAAGATTT AATAAAAATAT TATTTGGTGC TGTTATAAAA ATATTATATA
 9061 AATWTAATTT ATTTTTATTA AATTTAAATTT CTGGAATAAT TTTTTTTTTCT AATTTATGTA
 9121 AATTTTTTTTT AACTAAATGT CTAAGGAAAT ATTTTTTTCC TTTATTTATTT ATTAATTTTG
 9181 CTTGTCTTTT TTTCTTTTGT ATTTCTCTTT TTATCCTTTT TGCTTCTTCT TTTTTTTTTAT
 9241 TGTATTTCTT TTTCTTTTCT CTCTTTTCTT CTTCACTTTT TATTTTACCCT TTTTTTCTTC
 9301 TTTCTTCTTC TTTCTTTTCC TTTCTTTTCT TTTCTTTTCT TTTCTTTTCT TTTCTTTTCT TCTTTACTTT
 9361 TTTCTTTCTT TTTCTTTTCT TTTCTTTTCT TTTCTTTTCT TTTCTTTTCT TTTCTTTTCT TCTTTACTTT
 9421 CATTATTTTG TTTAATTAAA TTTCTGAGAA TCAAAATGTT TTTTTTCTTT TTTTAAAGT
 9481 ATATTTTGAA TAGATAAATTT TTTCTTTATTT TTTTATTTT TAATAGTTAA CTTAATAGTT
 9541 AACTTAATCG GTAATATTTT TTTCAATTTAA CTATATTTT TTAATAAAT TATAGATTGA
 9601 TAATTTTTTT TTATATTTT TTTATTTAAA AAATTTTCTG GTATGAAAT ATTAGAATTT
 9661 ATTATAAAAC TTTCCCAAGA AAAACCAATA TTACTAAAA GTATTTCTAA TTTATTTAAA
 9721 TTTATATAAT TGTCATGTTT GAATAAAATA CTATATAGTA AATATGCTTG AGATATTTTA
 9781 ATATTTTGTG GATAATTTT TATACTTAAA TCTATTATAT TTGAAGATAT ATCTAATTTCT
 9841 TCTTCTACAC TANNNGNAA NANANANANA NAAANANNN NNNNNNNAGN NNNNAANAAA
 9901 NNAAANANN CNNCANNNN NNCAANNANA AAANNAANAN NANNACNNN AACANAANAA
 9961 ANAAAATATT TCAACTTATA AACATATAAA ATTATTATAT ATANNNNNNN NNNNNNNNNN
 10021 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 10081 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 10141 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 10201 NNNNNNNNNN NNNNNNNNNN NATGATAATA TTAATATTAT TAGATATTTT TTTTATTTCT
 10261 ATATCATAAT TTATAATTTT TGTTTTTATA TCATCATATT TTTTATTTT ATACTTTTTT
 10321 ATTTTATAAT TTTTCAGATA TAAATTTTTT TKTAATTKAC TTTTTTTAAA AATTTTAAAA
 10381 AGCTTATAAA TATAAATAAT TTTTATTAAT TTACTTTTAC TAATATATTC AATTAATTTA
 10441 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 10501 ATAAAANNAA NNAGNANNNT AAAAAWATAA ATAAATANAA NAATATAAAT ANAAANANAW
 10561 ATAAATAAAT ATATATATAT ATATATATAT ATATATATAT ATATATATAT ATATATATAT
 10621 ANTATKTKTT GTATATTATA ATTTAATANA ATTATTAGTC CAATTATATT GTAATCTAGT
 10681 TCATATCTTT TTTTATATAN TTCTATATTA ACAAGTTTAT AAAAAGTTGA ATTTAACAAA
 10741 TATTCAAATTT TTTCACTAGA AATGATTTCTA GTTTCCTTGT CTTTCATGTC TAAAAAATA
 10801 TAATTTCTCT TACTTTTAAAT TTTAACTTCT TCTTCTTTTT CAGATTTTTT GTCGTACCAA
 10861 GGTCTTAATC TAAAAGGATT CATTATTTTA ATTTGAATAC CTTCTGACAT CCAATCTTCT
 10921 GGAAGTTTAT TTTCAACTAA TTCAATACCA TTATAAGTAC ACATAATATA TTGTTCTTTT
 10981 TTAAAATAAA ACATATCTTT ATGATATTCA ATAGGTTGAA ATAAAAAAT ACGAACTAAA
 11041 GTTTTAAACCA TAATTAATAA AGGTATTTTT ATCCACATTC TTATATATAA TTGAGAAATA

11101 AGACAACAAG TTCTGAAAAT ATTTACAAAA CCATAAGTAC TATCCCAAGC AACATTTCTG
11161 CCTGCTCTTT TCTTAGACTC AATTAATTTA AAGAAATCTT GTTGTTTTTTT TATTAATCTT
11221 TTTTTTACAA TATCTGCATT TTCAGGACCA TAAAATTCOA GTAAATCATT GTCATTTAAA
11281 ATTAAAAGCC TATCTTCTTC GTCTCTCTCA AGAAGAAAAA TGGGAGATGA AGGAGATCTA
11341 TAAGTTACAA AATCAGGTAA GTTAAGAGGA ATCTCTTTAC GTTTTTGATT ACGTAAAAGC
11401 CCCCTTACAG AAAATCTATT TAAATCATAA AATTCTGAAA ATCTAATATC ATGCAATTCT
11461 CCTTTTTTTTT TTTTCGTCTTC TTCAATTCTT TTTTCAAAAAT TATAACCAA AAAGTTATTT
11521 TTATCCACTT TTTTATATTC TTTATCTTCA ATTTTTTTTAT TTACATTTAA AGAGTCAGAT
11581 ATACAAGCTA ATTTTAATTT ACTTAAANNNT ACTTTATATT CAATTTTAAAT TTTTTTATCT
11641 GTTATTTCTT TATCTATATC TGTATTCTT CTTCCTTATC TTTATTATCT GTTATTTCTT
11701 TATCTTTATC TGTTTATTTT TTTATCTTCT TTATCTTTTAT CTGTTATTTT TTTATCTTTA
11761 TCTTTTATTT CTTTATCTTT ATCTTTCCGA GGTTCTGTAT AAAGTATCAA AGGTCTTAAT
11821 ATTCACATT TTAAACCTGT AATTTTTTTTA GATCTCAACA AAAGAAATTC TATATCATCT
11881 TTTAAGCTAC AAATCCATCT AGGAATATTT TTTTCTTCGA TAATAAAATTT TTGAGTTTCT
11941 TCATATTTTT TTTTATTTTT CAAAAATTTT AAAATAGGTG GTAGATTTTC ATATNATATA
12001 TTTTTTTTAT GCTNNNNNCA TTTTCATTAA ATTTATTTTAT GACTGATTCA TCTATATATT
12061 ATTNATTTTA ANATATAAAT TTCATTAAAA TTAATTAGGT ATGACTTNAT AAAAATTACC
12121 TTTTAAAATA TCATCTAATT TTATTTAAAGC TTTTTTTTTCA TTATCTTTTAT CAAAATAAAT
12181 AAAAAATCTA TTTTTTATTT CAAATAAATT TAAATTTTTT GATTGATCAA AAGATTCTAT
12241 ACGTTTTTTA AATTCATTTT CAAAATTTAT AAGTTTTTTTT TTTAAATTTA ATCTGTCTATA
12301 AGGATATATT TTTTTTAAAA ACCTTGTTT TTTTTTTTATT AATTTAAAAA AGGCATACAA
12361 ATTTGGTGAA AATGTATATA ATAATCGTTT TTTATTATCA ATTTTATAGA AACTAACAAA
12421 TAATGTGAC ATTTTCAGTTC GTATAAATTT ATCAAATTTGT TTATTTTTTTA TATATCTTAT
12481 AGGACGATTC CAACGGTTAG TATAATAAAA AATATCTACT AAATCTTTAT TGTCTTCTTC
12541 TGATATATTA AGTAAGTTTA ACTCTTCTTC GTCTTCTACA TCACTTTCTT CTATTTCTGA
12601 TTCACCTCCT TTTAGTTTTT TTGCTCTTTC TTTTTTATTG TCTTTTTCTA CGCATTTTAT
12661 TCCTTATAT TTTTCTTCTT TTTCAACGTT AGAAAAATA GATCTTTTTTA AAAGAATTTCT
12721 TCGGTCATTA AAAGTTATTG ATTTTAAACC TGTTCTTTGA AAATTAAGG TAAATAAGG
12781 TATAGGTGTT CTAGACAAAA ATACTAAAAA TAATATTAGA TAAATATATA AAAAATATAT
12841 ATCTAACTG TCTTTGTTTT TAAAACATTT TATTACTTTA TTTAGAAAT TATACATAAC
12901 ACTTTCGCTT TTTTTTAAAA AAAATTTAAA CTTAAACCAA CTTATCAGTT TATAAATATA
12961 TTTTATTATA ATAAAAATG AACAAATACT ACCTAACAAA TTAAATACATA AAAATAATAT
13021 TTTATTAGTA TTAGAATAAT TCTTAAATAA AAAAGATATA CTCAATCTCT TAAATATTGG
13081 GTTTGGTAAA TTAAAAATTA TAATTTGAAA TATAAAAAAGT TTTATAATTA TATCTAAAAAT
13141 AAAATTAAT TTAATTTTAA TATCTTTATA ATTTTTAATTT AAATATATTA TAAAAATTTT
13201 GATATATTTT TTTAATTTTA TTTTCATTTT AATATAGTAT GATTTAAATA CATATAATAA
13261 GTTTATTAGT TTAATTTTTA TAGTATTTTT TTCATAAATA TATTTACCTT TATAAAAAAG
13321 AAAATAAAGG TTTTATTTTTC GGTCTAATTT TAAAAATATT TTTAATATGT GTTTTTTTTTT
13381 GTTTTTTAAA TATAATTGGT TCTAAATTT CAAAATTTCT TAATCTAAAA CCATATAAAA
13441 GTTTAATTA AAAATATATC CAGTATATAC CTTGTTGTAA TATACCTTTA TAAAAAGTAT
13501 AATATAAAAA AATGCATAAT AAAGTGTGAG GTCCTTGTAA AAAAAGAAAC AAAGGCGTAT
13561 AATATATAGA TAAACTCATT ATTAATTGAC CTAAAACTGA TGAGCTTATT ATCACAAATA
13621 TTTTTTCTTG TTTTTTTTCA AAATCTCTTT TAAATAAGTG CTCTATAAAA AAAATATAAG
13681 AAGGTATCAC ACAGAAATGAA CTAAAAAAAC CAAAAATAAA ACCTATAAAA GAAATAAAAA
13741 AAAGTTTATT ATTAATAAAA TTTCTTAATA TTAATAAAAA ATGTATTATT ACTTTTACCA
13801 TAATTTTTTTA TATTTGAATG TTATATTTAA GTTTGTATAT TTTTAAAAAT TAAAAAAGGT
13861 CCCCACTACT AGTTTTTTTT ATTCGGCATA AAATTATTTT TCCATAAGAT TCCTCTTAAA
13921 GTATCGTCAC TAATTAAGAC TTTTACCACC TAATTCGAAA AGGTTGTTGG TGTAGATCAT
13981 CTTAGTTAAA ACACCGAAAG TTTATGAATA TATTAATTTT AATTTTTTTAA ATTTATTATTA
14041 TTCAAATCAT CTTTCGGTCT TTAGAAATAT TTAAGCTAAA TATATTACTA TATTTACACA
14101 TAATACCTAT TAACAATTTA TCTAATTGTA ACCTTAATTT TANTTATCAT ATATATATTT
14161 ATATAATTTA AAATTAATAA ATAATAAAGG GAGCACTCAT CTTTGTTTTAA ACTTACTACT
14221 TATATGCTTT CAGCAGTTGT TCATTACGCA CATAGCTACC TAGCGTTTAC TATTGGAATA
14281 ATAAC TAGTA CACAAGTGGT GCGTTCGTTT TGGTCTCTC GTACTAAGAA CGATTAACAT

14341 CAATACTCCA ACGTTCACAC CGGATATGGA CCGAACTGTC TCGCGACGTT CTGAACCCAG
14401 CTCACGTACC GCTTTAATGG GCGAACAGCC CAACCCTTGG AACCTCCTTC AGCTCCAGGT
14461 GCGGATGAGC CGACATCGAG GTGCCAAACC TTCCCCTCGA TGTGGACTCT TGGGGGAAGAT
14521 CAGCCTGTTA TCCCTAGAGT AACTTTTATC CGATAAGCGA TGGCCTTACC ACTCAGTACT
14581 ACCATCGGAT CACTAAGACC AACTTTTCGT CCTGCCTGAC TTGTTTTGTCT CACAGTTAGG
14641 CTTTCTTTTG CCTTTACACT TTAAAAATCA TAAAAATAGAT TTAAGAAAAAC CTTTATTGCA
14701 CGCCTCCGTT ACTTTTTTGG AGGCCTACGC CCCATATAAA CTGTCTACCA AAGACTGTCC
14761 TTGCTAAAAA AACAAGTTAG AAGTTAAACT TAATTAGGGT GGTATTTTAC TGATGACTTA
14821 ATATTGCCTC CCACCTATAT TTTACAAAAA AAAGACCAAC ATCAATCACA GGAACAGTA
14881 AAGCTTCATA GGGTCTTCT GTCCCTGGTGT GAATAGTCCG CATCTTCACA GACATTTTTTA
14941 TTTACCCGAG TCTCTCTCCG AGACAGTGCC TAGATCGTTA CGCCTTTCAT GCGGGTCGGA
15001 ACTTACCCGA CAAGGAATTT TGCTACCTTA GGACCGTTAT AGTTACGGCC GCCGTTTACT
15061 GAGATTTTCG TTGCTAGCTT TCCTGTATTT TTCAGGTCAC CAACTTCGTT TACCTTACAG
15121 CACTGGGCAG GCGTCAGCCC CCATACATGG TCTTACGACT TTGCGGAGAC TTGTGTTTTTT
15181 GTTAAACAGT CGTCTGGGCC TGGTTATTGC AATCCTTTTT AAGGACACCC TTTTTCCCGA
15241 AGTTACAGGG CTATTTTGCC GAGTTCCTTA GAGAGAGTTA TCTCGATCCC CTTAGTATTC
15301 TCTACTCATC TACCTGTGTA GGTTTACGGT ACAGGTATTT ATTTATTTTT TTTTCAATTT
15361 AGTTTTTCCCT GGAAGTATGA CCTCAGTTAT ATTAATACAA ATAACAATAC ATATAAACGT
15421 ATTTAATATA CTAGTATATT TTTAAGACTT TACAAAATTG TTTAAATTTT TGCTTATATC
15481 TTTCGACTGA TAACCATTTA TCAGCTAACC TAGCCTTCTT CGTCCCCTAAT AATGAAAATA
15541 AATAAATAGT ACAGGAATAT AAACCTGTAA TCCATCAACT ACGCCTTTCG GCCTTATCTT
15601 AGGTCCGAC TAAACTTTCG CATATAAATA TTTCGAAAGA ATCCTTAGAT TGTCCGGGCA
15661 TAGGATTCCT ACCTATGTTT TCGTTACTCA AGCCAACATT CTCACTCTG ACATAAAATT
15721 AAAAAATTAA AGCAGAACGC CCCCCTACCG ATTGTTAATC TCATAGTTTC GGCATATTA
15781 TTTAGACCCG TGTATTTTCC GCACAAAAAC GCTTTATCAG TGAGCTTTTA CGCACTCTTT
15841 AAAGGATAGC TGCTTCCAAG CAAACCTTCT GACTGTCTAT GCACTTTCAC ATCCTTGTCC
15901 ACTTAATAAT AATTTATGGG CCTTAACTAA TGATCTGGGC TGTTCCTCCT TTGACAATGA
15961 AGCTTATCCC CCATTGTCTC ACTAGTTAAT TTATAATTGT ACAGTATTAT TAGTTTGTAT
16021 AAAATTTAAA ATTTAAACAG TGCTTTACCC CTGTTATCAT AATTAAGTGC TGTGCTTCGA
16081 CACATTTTCG GGGGAACCAG CTAGCTCTGG GTTCGAGTGG CATTTCACCG CTATCCACAA
16141 TTCATCCGCT GATTCTGCAT TATCAGTCGG TTCGGACCTC CACTTAGTTT TAACATAATT
16201 TCATCCGAC CATAGATAGA TCATCCAGTT TCGGGTACAT AAGTAATGAC TTATAATATA
16261 TAATATATAT ATTTATTGCC GTTATAGACT TGCTTTCGCT ATGGCTACGA CTTATATTAT
16321 TCTTAACCAA AGCCATTAAT TATGAGTCGG TGGCTCCTTC TTCAACAAGC ACGCAGTCAG
16381 AGATAATAAA AATTCTCCTC CTACAATTTG ACAGCTTTTG GTTTCATTTA AATATATCTT
16441 TTAAAAGTTA TATTATACCT TTCCCTCACG GTACTAGTTC GCTATCGGTT ACTCAGGAGT
16501 ATTTAGTCTT ACAAGGTGGT TCTTGTAGAT TCACACGGAA TTTAAGTTGT TCCAGGTTAC
16561 TTAGATAAGT TAATTANTTT ANNNATTANN NATAATTAAG TATTAAATAT AAAANTATAA
16621 AAAAAATATG TTTTATATTT TTGACTACTG GACTTTAACC ATCTAAGGTA CAGTATTTTT
16681 ATCTGTTTTA TCTAATAATA TTATTTAAAT TTACANAAAA TTTAGTTTTA TATAATAACT
16741 TACCTACAAC TTCATTTTAA TTAATGATAA TGATTTAGAC TATTTCCATT TCGCTCACCA
16801 CTACTAAGGA AATTACTTTT GTTTTATTTT CCTCTGGATA CTAAGATGTT TCAGTTCTCC
16861 AGGTATATCT TTACCTTATG AAGGCAGTTT TAAAAATTAA TTTATTCGGG AATCTTTAAT
16921 TTTTATAAAAT TTTATTTAAA GCTTTTTCGT ACTTTATTAC GCCCTTCTC GTCTCTGAGT
16981 ACCAAGGTAT TTACCAAAAAG CTTATTATTA AGTTTGAATA TTATAAAAAAT AANNAAAAGA
17041 TTTAAATTAT AATTATATNN ATATATANNA TGTATATTAN NNNNNNNNNN NNNNNNNNNN
17101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17161 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNAANANN NNNNAAANAA NTAAATATAA
17221 CTTTATAGTAC TTTTAAATAA ATTCNTAGNA CAAANANAGA NNCANNAANA ANNACAAANA
17281 AAAAAGANAA NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17341 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17461 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17521 NNNNNNNANAN AGAANANAAA NKATATGTTCA ATAATATTTT TATTATTNNN NNNNNNNNNN

17581 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17641 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17761 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17821 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17881 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
17941 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18061 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18121 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
18181 AAAANAAAAA AAAAAAAGTT AACCTATTAG ATTNAANANK NANANANAAN NANANANAAN
18241 ANANANATAT AAAACATAAA ATATTTATAA TGTAATAAAA AATTCCAGCT GCACCTTCCA
18301 GTACAGCTAC CTTGTTACGA CTTCATTCTA GTTACCAATT TTGTCTTCGT CATTTATTAA
18361 ATGATTCTAG ACATAACCAA TTTCCAGACT GTGACGGGCA GTGTGTACAA GGTCCAAGAA
18421 CTTATTCCACC GCTATATTGC TGATTAGCGA TTACTAGCGA TTTCGACTTC ATGAGTGCGA
18481 GTTGCAGCCC TCAATCCGAA CTGAGATTGG GTTTTTTGTAT TAGCTCTCCT TTACAGGATT
18541 GCAACATTTT GTCCCAACCA TTGTAGCACG TGTGTAGCCC AGGATTTAAA GAGCATGATG
18601 ACTAGACGTC GTCCCTTCCT TCCTCTAGCT TACACCAGCT GTATTCCTAA AGATAAATTT
18661 AAATTTAATA TATATTATTA TAGGGATAGG GACAAATCCA TAATATATAT TATTTATAAA
18721 AATAAAGATA AGTATAGGGG TTACGCTCGT TAGGGGACTT AACCTAATAC CTAAGGCAC
18781 GAGCTGATGA CAGCCATGCA CCACCTGTAT CTTTATACTG GATAGTATAA ATTTTATTTT
18841 TCAATAAAAA TTTTCGACATG TCAAATCCCTG GTAAGGTTCA TCGCTTTGCA TCGAATTTAA
18901 CCACATGCTC CACCGCTTGT GTGGACCCCC GTCAATTCCT TTGAGTTTCA TTTCTGCGAA
18961 CGTACTTCCC AGGCGGGATA CTTTACGCGT TAGCTTTAGT ATAACATAGG CATATTTTTT
19021 TTATTATTTA TGCTTAATAT GCTTAGTATC CATTGTTTAC AGCTTATGAC TATCGGGGTA
19081 TCTAATCCCA CTCGCTACCA TAGCTTTCTG TACTCAGTGT CAGTATTAAT CTAGCAAGT
19141 GCTTTTCGCCA TTGGTGTCTC TTCCAATATC AACGTATTTT ACCACTCCAT TGAAAAATTC
19201 CTTTACTTTT AGTATACTCA AGTTTTTTAG TTTCTAATGC AGGTCAGAGT TGAGCTCTAA
19261 TATTTTACAA TAGACTTAAA AAACGACCTA CGAACCCCTT ACGCCCAATC ATTCTGGATA
19321 ATGCTTGCAAT CCTCTGTATT ACCGCGGCGG CTGGCACAGA GTTGGCCGAT GCTTATTCCCT
19381 GGAATAACGT CATTTTTTCT TCTACCATAA AAGAAGTTTA CAACCTTTTC AGGCCTTAAT
19441 CCTCCACGCG ACATTGCTCC GTCAGGCTTT CGCCCATTCG GGAATATTCC TCACTGCTGC
19501 CTCCCGTAAG AGTCTGGGCC GTGTCTCAGT CCCAGCGTGA CTGATCATCC TCTCAGACCA
19561 GTTACTGATT ATAGCCTTTG GAGGCTGTTA CCCTACAAAC TAGCTAATCA GACATAAATC
19621 CAGCCTAGAG CGATTTTTTAA TAATAAATAT TAATCTTTTA CTCCTCAGTT TATGGAGAAT
19681 TAGCAACTGT TTCCAATTTG TTATACTCCT CTCTAGGATA GGTTTTAATG TATTACTCAC
19741 TCGTTCGCTA CTATAAAAAT ATATCGTTCA ACTTGCATGT GTTAAGCATG TCGCTAGCAT
19801 TCATCCTGAG CCAGGATCGA ACTCTTAAGA GAAATACCCG CCCCCCCCCA AAGAAAACCC
19861 CGCCATTATC TCAACCTACT ACTTCTAAAC AATTTATTAT ATTTTTTTAT AACTTTTGAA
19921 AGTTATTTTA CTCATAGCTT CATATACATC TTTAAAAAAT TTATTTTTAT CTATAGCTTT
19981 TTTAGTAGTT TCCCTTTTTA TATTTATTGC ATTTCCTTTT AAAGCATCTA ATAATTCAAA
20041 ACTTAACCTA GAAACCATGT TATATCCAAA ACGTTTTTTA GATGCTGATA TTAATAAATTT
20101 GCGAGCATA ATATATTTTT TATCTTTAGG TAAATATAA TATATTTTAC GTATTGATGC
20161 TTTTCCTTTT TTACGTTTTT TTTTAAATAT TTGTGTTTTT ACTATTTTTT TTTTAGTTAT
20221 TTTATTTAGT GCTTTTTGTA CAACATAAAT TGGATTTAAT TTTATTTTTT TTTTATTTT
20281 TAGTAAAGAT TTATAAATTA TTTTATATGA TATTGATTTT TTTCCTTTTT TTAATAATGTA
20341 ATTAGATAGA TTATAAATTG ATTTTTTTTAA AAAAAATTGGA TCATATTTTA CTTTATAAGG
20401 CATTTTTTAA ATTTAATTTT TAAGTTTTTG TTTTTTAGCC CCATATTTTG AACGCTTTTG
20461 AAATCGTTCT TTTGCTCTAC CAGCATCTTT ACTGCCAAA ATAACCTTAT GTTTTACACC
20521 AGGTAAATCT TTTACTAAGC CACCTCTTAC TAAAACCTT GAATGTACTT GAACTAAATTT
20581 AATACCAATA CCTGGTATTA ATGCAGTTAC TTCAATTTCA GAAGTCAATC GTATTCGGGC
20641 AACTTTTCTT GAGCCAGAA TAGGTTTTTT TGGATGAATA ATGTAGTTAA CAGATAAAAA
20701 ACTGTCACCT TACAGAACCT ATACATGAAA GTTTCCTTTC ATATAGCTCC TCATTTAAAT
20761 ATTTATTAAA CCYTTNAAAT TTGTTTTTAT TAAAAATTAT TAAAAATTG TCATGGACTC

20821 TCGTCGAAGA TGCACCTAACG GACTCTAAGG TAGAAGGACC ATGTCTCAAA GCATGATCAC
 20881 TCAAAAATCTC AAGCCCTATAC TGGGCTCCTT GCCTAGAAGA CATAATGAAA TCCACTCGAC
 20941 TTAAAGAAGG CAAAGTAAGC GAATAGGCAC GAAATAGACT CGCATGGTAA AGACCCGCTAG
 21001 AAACAAACCC ACACCTCTAAA ACAAAGAGAC TCACCTAAAC AACTGCTAAG TGACGGGACA
 21061 AAACAACAAA CGCGCTCAAA AGACACCCTAA GCAAACACAA ACACGAACAA AGATAACTAA
 21121 ACTACCTACA TAGGTGAGCG GACAGCCAAA TAAGGGAACA CATCAAAAAT GACATGGAGC
 21181 AAAGGAGAAA AAAGCTCAAG TGAGGTACTC GAAATCCCAA AGACCTAGGC TCTGAGGCAA
 21241 CTCAATCGCC TAGCGAACCC TAAAGCTTTG ATACCAAGCT GGTGCAAGAC CAACCCGGGA
 21301 ATGCACAGAA CCTCGCTAAG AGGCCCTTC AGCCACCCGG AAAATATATTT TCAAAATTTTT
 21361 TATTTGATGG GTTATTAATT GGTAAAGTAT CAGTTTCGTA ATTATAAGCA TTTATAACAG
 21421 AAAAACTAA TTCTAATATA TTTGATTTAA TTAAATTATA TTTATTAATA TTTAATATTT
 21481 CTTTGTCAAT TTTTATTAAC TTTTCTATAA AATTTATTAG TCTAAAATTTG TCAAAAATAAA
 21541 AAACTTTTTC TAAATTTAAT AAATTTATTA TTAACTTTAA ATCATGATTA AAATATTCAT
 21601 GTTTTATATTT TTTAATCTCT GATGTGTTAA AAATTTTATA AGCAATTAAT CCAGATAAAC
 21661 ATATCAATAA ATAAATAATT ACAGTTATTA TTTTTATATTT TGTCCGAGT TCATAATAAT
 21721 TTTTATATAA AGGATATTTT TTTGTTTTGA AATCTATTAA ATATACAGAC ATAGGATCAA
 21781 CTATTAAACT ATTTAATCTA ATTTGTAAAA AAAACGTCC AATTTGATAA AAAACAATTA
 21841 TAATAGGAAC TTTGTATCCT TTTTTTATGT ATTGATTAAA TCTAAAAAAA TAAAAAATAA
 21901 ATTTATTAAT TATTTTTGAA TTTATTACAT CTTGATTTTG TGTAAGACTT AGTGATAAAG
 21961 TTTGATTTGT CAATATAGAT AAATCTCTTA GATTTTTTAT ATTTAAATGT TCATGTAATA
 22021 ATTCTTTTTT TTCAATATAA AAACCTTTTG TTTTAAATAA ATTCAAAATA TATTTTTTTTT
 22081 TTATTTCTAT AGTAGGTTTT TTTAATAAAA TAAATTTACT TAATTTATCA TAACCTATTA
 22141 ATATAGGTTT TACTTCTTTT GGTTTATTTG TATAAGCTAA AATTAAATTA TTTTTATTAA
 22201 AATTAGAAAT TAAATTAATT AATTCATAAA ATTGTTTTAAA ATTTAAATTA TAATGACATA
 22261 AATGAATATTT TGAATAAAT ATAATACATG AAGAAATATA TTCTATTAGT TTAAATTGAG
 22321 TTACAATATTT AAATACATCA TATATACTTT TTTTCTTTTT TGGCATTAGT TTATCAATAA
 22381 TTTTTATTTTT TTCGAATTCA AATGTTTCTTT TATCAGTATTT ATAAATTTTT TTAACAATTT
 22441 TTTTAAATTTG TTTAGTAGAA GTATTACTTT TATTAATATC AAATTTATCA ACAAAGAA
 22501 AATCAGATTC ATCATAACTA TATTTTGAAA TTTCATAGTC ACTTAATTTCT TTATTAATTG
 22561 GGTTTAAATAT AATAAAAGGC AAATCAGCAT AATCTTTTAT AATTTTTTCT TGTAACAA
 22621 AAACGTCCTC AGACTGACCA ACTAATAATA TACCCTTTGA AGCTGACGGA ATCATTTGTTA
 22681 ATGAAAAAAA ATAAGGTTTT TGCTTATTTT TATATCTGTT TTTTTTTAAA AATAAATATTT
 22741 CTGTTATTTTT TTTATAAATTT GTAATAAAAA TACGATTTTT TTCTTTTATA TTTAATATAT
 22801 TATTTGCTTT TGTATTAATTT CTTAACATAT CTAATAAATA TATTAAATAT AAATAACCAT
 22861 ATTCACAGA GTTTTCATAC TTTTTTATAG ATATATTTTT AAAACATAT TTAATAAGTT
 22921 TTTGTTCAAT TTTTATATTT AAAAAAATAT TTATAAATTG AATTAAAAAT TTTCTTTCTT
 22981 CATCAAAAAT TAAATCATTG ATTTAATCC AAGATTCTAA ATCAAATTTG GTTTCATTTT
 23041 TTTTAAATAT ATTTTLAGTA TATGTAGTTT CTTTTATAGA TGTATATTCT TGATCTTTAT
 23101 TGTTAAGAAT ACATTGATAA AAATCTTCAG TTATATCTTG TGAAAAACAT TTAATAAATTT
 23161 CTTTACTTAT GGAGTTTCTT AAATAATAAA AATCATAAAT AAAAGATTTA TACCAATTTT
 23221 TTAAATTTGA TTCAGTATAA AAATATAAAT ATTCCTTATA TTTAAGTTTT ATATTTATAT
 23281 ATATATATAT AAAATTTAAA TTTACTTTAT GANAATAATA AGTAGTTTTT ACTTTGTCTA
 23341 TTTCTAAAAA TTTAAATAAA ATGTCAACAA TAAAAAATCG AATAATATAT TTATTTATTT
 23401 CAGTTAGTAA ATTAATAAAT TCATTTATAA AATTAAAAAA AATATTGTTA TAAAGATAAA
 23461 TTGTGTTTTT CATGCATGCT TCTAAATAAC TATATTTACA TAACTTTACA TATAGATCTA
 23521 TACTAAATTT TTTTGTCAAT AATAATAAAT TTATTTTATA TTCATTTTTT AATAATTTAA
 23581 TATATTTAAA AATTAATTTA AAATATTTTA AAATAAAAAA TATAAAAAATA TAAATAAAAA
 23641 ATATTTTAAA TAATAAATTA TATTTTATAA AATTTAATTG GTCATATCTA AAAATTTCAA
 23701 TAATATCTTT TTTAATTGGG TGTAATTTAA ACTGATTTTT AATAAATGAA AATTGATTTA
 23761 CTAATAAATTT ATAATTTATTT TTATTAGTTT TTGATAATAA AATAATTTAT TGAATAAACT
 23821 TAATAAAATN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 23881 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 23941 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN

24061 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24121 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24181 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24241 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24361 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24421 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24481 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24541 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24661 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24721 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24781 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24841 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 24961 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25021 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25081 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25141 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25261 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25321 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25381 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25441 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25561 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25621 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25681 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25741 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25861 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25921 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 25981 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26041 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26161 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26221 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26281 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26341 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26461 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 26521 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNAATTC
 26581 CTGCTGGATA CAAAAAAAAG ATTTATAATC CTTAATAAAA AAAAATGACA TTAATAACA
 26641 TATATAAAAA AAAAGTTAAA AGAAAAATTAT TAATTTATGG ATATAATTTT AATAAAGGTA
 26701 GAAATTATAA AGGTACAATA ACTGTAAAAC ATAGAGGAGG GGTTCATAAG CGTTTATATA
 26761 GAAAGATAGA TTTTAAACGT AATGAAAAAA ATACTGGAAA AATATTAAAA ATTGAATATG
 26821 ATCCCAATCG GAATACAACA ATTTGTTT TAGTTATTATAA AATAAATAAT ATATACAAAT
 26881 ATATAATACA TCCACAAAAT GTTGGTATTG GAGATACTAT TTTATCTAGT CATGAAACAT
 26941 CTATTTATGC TGGAAATACA TTACCTTTAA TTAATATACC TTTTGGTACT TTTATACATA
 27001 ACATAGAAAT TAAACCTGGA AAGGGCGGAC AATTAGTTAG GGCTGCAGGT ACTGCAGCAA
 27061 AAGTAATTAC AAAAAAAAAA AAAAGAGTTT TATTAAAGTT ACCATCAGGA AAATTATGTT
 27121 CAATACCTTT AACATGCTTA GCAACTATAG GACAAGTTAA TAAAACGCA TTAATAAAAA
 27181 AATTAAAAGC TGGGACAAAT CGTTGGTTAG GAAAAAGACC TAAAGTACGA GGTACTGCTA
 27241 TGAACCTGT AGATCATCCT CACGGAGGAG GAGAAGGTAA AGCTCCTATA GGTAGAAAAAT

27301 TTCCTGCAAC TCCTTGGGGA TATCCTGCCT TTGGACGAAG AAAAAAAGA AAAAAATTT
27361 TTTAAAAAAA AAAAATATTA AAAATAAATA AAATAAATAT AAAGAATATA ATTAANNNN
27421 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNTT TAATTTTTTTT GATAAAATCC
27481 TGTAAAAATA GCATAAGATA ACTTATTTAA AATATAAAAT ATTGAAGACA AATTAAATGT
27541 ATTTATAGGT ATAACAATAT CTGCTAAATC TATATTAATA TCAGTATTTA TATTAATTAA
27601 ACAAATTGTT GAAATTTTTA AAGTAAAACA TTCTCTGATA ATTATATCAT TTTGTGGATC
27661 AATAAGTATT ACAATATCAG GTCTCTTAGT CATATATTTT ATTCCTTCTA AATTATTATA
27721 TAAACTAATT ATTTTTGTTT TTTTTTTACT AATTAATAAA TTTTTTAATT TAATTTTTGT
27781 AGTTAACCAG TTAGTAAGTA AACCTGCTCG TATATTTCTA TTAATATAAT GACATCTAAT
27841 TCTTCAGAT ATTAATTTAA AATAATTTAA TTTTAAGGTT TTATATTTAT TATTTATTCG
27901 TTTATTTTTT CTTTTTAATC TCAATTTAGT TTTTATTTTA TTCCTTAATT TTTTATTATT
27961 TATTAAAAAC CTTTTAGTTA AAAAATTCCT AGTTTTAATT ATTATAGGTT TTATTTCTTT
28021 TTTATTTTTG TTTTGACTAA TTATTAAAAA TGTTTTTTTA TTTTTTGCTG CATAAAAAAC
28081 TAAATTACAA GCTATTGTTA ATAAAGAAGC TATTTTTGTA ATATCTAAAA TATAAATTTT
28141 AGATTTTAAA TTAATATTAA AAATATAATA TATCATTTTA AATTTAATAT TTTTTGTATT
28201 TAAAAAATGA ATTTTTTCTG TTAACATATC ATTTAATTTT ATTTTCCAAT CTATTTTTTT
28261 TTTCATTTTA ACATTTTTAA TTTTAACATT TTTCATTTTA ACATTTTAAA ATACAATCCA
28321 TTTATACTAT CATTTTAAAA ATTCAAAAATA CAATCATAAA CTTGATGACT ATCATTTATA
28381 ATTATAATAA TGGATTGTGG ATCGAGTTGG ATTTGAACCA ACGTAGACAC ATAAGTCAAA
28441 GGATTTACAG TCCATTCCTT TTAACCTCTC GGGCATCGAT CCAAATAA

Apéndice 1.3 *Anotación final del plastoma de Prosopanche costaricensis.*

```

LOCUS      Prosopanche_costaricensis      28705 bp DNA      circular UNA 11-MAR-2024
DEFINITION Prosopanche_costaricensis plastid.
ACCESSION  .
VERSION    .
KEYWORDS   .
SOURCE     plastid Prosopanche costaricensis
  ORGANISM  plastid Prosopanche costaricensis
REFERENCE  1 (bases 1 to 28705)
  AUTHORS   Obando-Coronado M, Albertazzi, F.
  TITLE     El genoma plastidial de la planta holoparasita Prosopanche
            costaricensis: Reduccion del plastoma en la familia Hydnoraceae

  JOURNAL   Unpublished
REFERENCE  2 (bases 1 to 28705)
  AUTHORS   Obando-Coronado M, Albertazzi, F.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-MAR-2024) Your Department, Your Institute, Your
            Address 1, City, State 12345, Country
REFERENCE  3 (bases 1 to 28705)
  AUTHORS   .
  TITLE     Direct Submission
  JOURNAL   Exported Mar 23, 2024 from SnapGene Viewer 7.1.2
            https://www.snapgene.com
COMMENT    SGRef: number: 1; type: "Journal Article"; journalName:
            "Unpublished"
COMMENT    SGRef: number: 2; type: "Journal Article"; journalName: "Submitted
            (11-MAR-2024) Your Department, Your Institute, Your Address 1, City,
            State 12345, Country"
FEATURES   Location/Qualifiers
  source    1..28705
            /organelle="plastid"
            /mol_type="genomic DNA"
            /organism="Prosopanche costaricensis"
  gene      complement(10..82)
            /gene="trnE-UUC"
            /label=trnE-UUC
  tRNA      complement(10..82)
            /gene="trnE-UUC"
            /product="tRNA-Glu"
            /label=trnE-UUC
  gene      complement(131..203)
            /gene="trnFM-CAU"
            /label=trnFM-CAU
  tRNA      complement(131..203)
            /gene="trnFM-CAU"
            /product="tRNA-fMet"
            /label=trnFM-CAU
  gene      complement(241..543)
            /gene="rps14"
            /label=rps14
  CDS       complement(241..543)
            /codon_start=1
            /transl_table=11
            /gene="rps14"
            /product="ribosomal protein S14"
            /label=rps14
            /translation="MSKKSIIFRDKKKINLEKKYRIIRQNLKKRIRKVVYLLKIKFKINK
            VVQKLPNRSILIRIRKRCISIKGRPRGYRYFGQSRNLIYEQFNSCKLPGAKISSW"

```

gene complement (839..1372)
/gene="rps4"
/label=rps4

CDS complement (839..1372)
/codon_start=1
/transl_table=11
/gene="rps4"
/product="ribosomal protein S4"
/label=rps4
/translation="MSRYLGPKLKRKRSFEIVKYKKKKTNRKKKKFKI IRESKYLPLLK
EKQKMRFYGLTNKQLSNLYIYIKIKGLKMHKLCFLEMRLDNIIFKLSLTKTISEAR
QFINHKKHILVNNYTVNIPSFICKTEDKITLLNLNKKTIKIPKKIPKLDKRLDFLFNE
HKISEYYLNKKIKK"

gene 2320..4341
/gene="accD"
/label=accD

CDS 2320..4341
/codon_start=1
/transl_table=11
/gene="accD"
/product="acetyl-CoA carboxylase beta subunit"
/label=accD
/translation="MSSSDRIKMIIDKDTWNPINEDIFSIDPLEIDDTLYDNYIDDDNT
VNDNYIYDDNVDDIYKISNYFEWFFNYEIFLDYIDLRLNLENLIEELPQEQQEEYLIKK
ELKGSPLPNKEEIQKYLQEYSTLIKEELENYLPQEQQELEEYLIKKELKKSPLTKEE
IQKYLQEYSPLIKEELENYLRLTKEEIQKYLQEYSPLSKEELENYLPQEEDLEKRQRQ
REFDLYIGDKLIKLDIKRYFFLYLGDKLVKLNKIRKKKKEKKEKELYVVSFEFEEV
VKLIENQKQIEKELNNENESKFKFYALHETFDILIKEDKKKYEKEHYNLTDIHKDI
EFRTSLFERLNKEVKEYEENNEKEEYEEENNEKELKEENNEKEEYEEENNEKELKEEN
EKEEYEEENNEKELKEENNEKEEYEEENNEKELKEENNEKEEYEEENNEKEEYEEENNE
TKKILKILNEKLDDEEKFDEIEKNKSYINRLNFYRETTSLVEAVKTGIGKINGIKVALGI
MDSEFMGGSMGSAVGETLRLIEYATKEKLP LIIFCASSGGARIQEGICSLIQMAKISAA
LHYKSNKKS VFISILTSPTAGGVIASFGMLGDI IIAEPNTTIAFAGQIIIEQILRKEV
PKGFQTAEYLFKQGAFDLILPRQLLKTIVTVLLNFHGF"

gene complement (4416..4489)
/gene="trnW-CCA"
/label=trnW-CCA

tRNA complement (4416..4489)
/gene="trnW-CCA"
/product="tRNA-Trp"
/label=trnW-CCA

gene 4508..4759
/gene="rps18"
/label=rps18

CDS 4508..4759
/codon_start=1
/transl_table=11
/gene="rps18"
/product="ribosomal protein S18"
/label=rps18
/translation="MKIQIRNKTIKLNFLKLNLNINKINYKEIKLLYKFLSIKGIKLS
RKVTKLNLKQKSITIAIKRARILSLLPFNKSIYLNKK"

exon complement (5246..5365)
/gene="rps12"
/number=1

gene complement (5511..5918)
/gene="rps11"
/label=rps11

CDS complement (5511..5918)
/codon_start=1

```

/transl_table=11
/gene="rps11"
/product="ribosomal protein S11"
/label=rps11
/translation="MRKRKIKPIQVLYRHKRQPKINFLEKGI I I I L A S F N N T R I T I M D
EKNILIFTASTGMIEFKNTQKGT P F A A K T T A N N V F N I I K I H S A E I R I K G I G A G R T S V L Q
VIKSKKIKIYSILDITPFSHNGCKSTLKKVR"
gene      complement(5927..6040)
          /gene="rpl36"
          /label=rpl36
CDS       complement(5927..6040)
          /codon_start=1
          /transl_table=11
          /gene="rpl36"
          /product="ribosomal protein L36"
          /label=rpl36
          /translation="MKKRTSIRKICQECKLVKRYRRIL I C N N K R H K Q K Q K"
gene      complement(6043..6402)
          /gene="rps8"
          /label=rps8
CDS       complement(6043..6402)
          /codon_start=1
          /transl_table=11
          /gene="rps8"
          /product="ribosomal protein S8"
          /label=rps8
          /translation="MLNNFIT I I V H I K N A S L S N K K I T T I P F N N T S N N I L K L L L K E G F I E
NFIKKNNF S I I L K K K I F V K I I S K P S Q R I Y V T Y K I I P K K L N A N G V Y F I S T S Q G I I T D I E
AKLNKIGGEILFYVV"
gene      complement(6429..6797)
          /gene="rpl14"
          /label=rpl14
CDS       complement(6429..6797)
          /codon_start=1
          /transl_table=11
          /gene="rpl14"
          /product="ribosomal protein L14"
          /label=rpl14
          /translation="MIHKQTYLNVDN S G A I K L M C I Q K L G I S K F K S A Y I G D I V I S V I K E
ALPSV S I E K S E I R A V I V R T R K E I K R S D G I I F K F D D N A A V V I D N L N N P K G S R I F G I I T Q
ELRKLNFNK I I S L S S E I L"
gene      complement(6817..7221)
          /gene="rpl16"
          /label=rpl16
CDS       complement(6817..7221)
          /codon_start=1
          /transl_table=11
          /gene="rpl16"
          /product="ribosomal protein L16"
          /label=rpl16
          /translation="MLPKQTKFDKQHRGRMKG I S V K G N F I C F G K F G I K T L E P T W I T A K Q
IEAGRKVMTRYFKRGGKFWIRIFTDKPITKRSV G K R M G S G K G S T E N W V A V I K P G R I L Y E
VDNVSEILAKNTFKIVSSKLPVKT K F F I K T"
gene      complement(7680..8321)
          /gene="rps3"
          /label=rps3
CDS       complement(7680..8321)
          /codon_start=1
          /transl_table=11
          /gene="rps3"

```

```

/product="ribosomal protein S3"
/label=rps3
/translation="MTKINPISFRLNIIPNYLYFWFYNNKNYSINLCETNKIQDCFINY
EKKKNINVNVNIVSHIRITKQNNKIKIKIIILDFLNFKLLKMFNKDSNFRNQFINFKRLN
FNIKVQRIKYPFKNPKVIKYLQKGGITLKKIKEKVLKYIYNEKKIKGKVKISGR
TSGKAIHVHKNKTWIKEGILPLQTIKENIRFCSYSVKTNNGLLGIKIWIYI"
gene      complement (8357..8638)
          /gene="rps19"
          /label=rps19
CDS       complement (8357..8638)
          /codon_start=1
          /transl_table=11
          /gene="rps19"
          /product="ribosomal protein S19"
          /label=rps19
          /translation="MNKKLFLVHNHLIDKITKNKIEKKKEIKTWSRASTIIPLMVGQI
LAIHNGKEHIPICINKKKIGHKLGFEVPTRTFFPRHKKIDVKSKKKKYY"
gene      complement (11018..14470)
          /gene="ycf1"
          /label=ycf1
CDS       complement (11018..14470)
          /codon_start=1
          /transl_table=11
          /gene="ycf1"
          /product="hypothetical protein RF1"
          /label=ycf1
          /translation="MSLSIYYTPLFLFLQRPHTLLCIFYLYTFYKGIQQGIYWIYFLI
KLLYGFRLINFDNLEPIIFKKQKKTTHIKNIKMRPKIKPLFSFYKGYIYEKNTIKFKL
INLLYVFKSYYIEMKIKLKKYIENFIIYLIKNYKDIKFKNFILDIKLFIFQIINFN
LPNPIFKRLSISFLFKNYSNTNKILFLVFNFLGSICSFYIIKIYKLLISWFKFKFFLK
KSESESLMYIFLNRVINSFKNKDSLDIYFLYIYLILFLVFLSRTPIPLFTFNFQRTGLK
SITFNDRRILLKRSIFSVEKEEKYKDINDVEKDNKKERAKKLGGESEIEESDVEDEE
ELNLLNISEEDNKDLVDIFYYTNRWNRPIRYIKNKQFDKFI RTEMSQLFFSFYKIDNKK
RLLYTFSPNLYAFFKLIKNTFRFLKKIYPYDRFNLKKLNNFENEFKKRIESFDQSKNL
NLFEIKNRFIYFDKDNEKKALIKLDDILKGNFYKEKTNENKEKININKSYINNFNEID
KLNKININKSDINKFNENEKAIKKYIYENLPPILKFLKNKKYEETQNFIEEKNIPRW
ICSLKDDIEFLLRSKKITGLKCRILRPLILYTEPPKDKDKEIKDKDKEITDKDKEDKE
ITDKDKEITDKDKEDKEITDKDKEITDKKIKIEYKVPPLSKLKLACISDSLVNKKIEDK
EYKKVDKNNFFGYNFEKRIEEDKKGELHDIRFSEFYDLNRFVSRGSLRNQKRKEIPL
NLPDFVTYRSPSSPIFLLERDEEDRLLILNDNDLLEFYGPENADIVKRLIKKQQDFK
LIESKKRAGRNAVDSTYGFVNI FRTCCCLISQLYIRMWIKIPFLIMVKTIVRIFLFQPI
EYHKDMFYFKKEQYIMCTYNGIELVENKLPEDWMSEGIQIKIMNPFRLRPWYDEKSEKE
EEVKIKSRENYIFLDMKGQETRIISSEKFEYLLNSTFYKLVNIEIYKKRYELDYNIGL
KIIVNYNINKIFYTFIYFLFFYKYIININFLIFRFIYLFKILKILINFIKICYVI
SKFRELSSIFSKSLIEYISKSLIKIYIYKLFKIFKKSCLKKNLLYENYKIKKYKIK
KYDDIKTKIINYDIEIKKISNNINILKKKISNINTKNKKRLMP"
gene      complement (14770..14888)
          /gene="rrn5"
          /label=rrn5
rRNA     complement (14770..14888)
          /gene="rrn5"
          /product="5S ribosomal RNA"
          /label=rrn5
gene      complement (14929..15025)
          /gene="rrn4.5"
          /label=rrn4.5
rRNA     complement (14929..15025)
          /gene="rrn4.55"
          /product="4.5S ribosomal RNA"
          /label=rrn4.55

```

```

gene      complement (15095..17909)
          /gene="rrn23"
          /label=rrn23
rRNA     complement (15095..17909)
          /gene="rrn23"
          /product="23S ribosomal RNA"
          /label=rrn23
misc_feature 19223..19552
          /label=direct repeat A
          /note="direct repeat A"
gene      complement (19223..19398)
          /pseudo
          /gene="ycf2"
          /label=ycf2
gene      19402..19474
          /gene="trnI-AAU"
          /label=trnI-AAU
tRNA     19402..19474
          /gene="trnI-AAU"
          /product="tRNA-Ile"
          /label=trnI-AAU
gene      19517..19552
          /pseudo
          /gene="rpl2"
          /label=rpl2
gene      complement (19706..21248)
          /gene="rrn16"
          /label=rrn16
rRNA     complement (19706..21248)
          /gene="rrn16"
          /product="16S ribosomal RNA"
          /label=rrn16
gene      complement (21346..21921)
          /gene="rps7"
          /label=rps7
CDS      complement (21346..21921)
          /codon_start=1
          /transl_table=11
          /gene="rps7"
          /product="ribosomal protein S7"
          /label=rps7
          /translation="MTYKVKYDPIFLKKS IYNLSNYILRKGKKSISYKIIYKSLLEIKK
KIKLNP IYVQKALNKITKKKIVKTQILKKRKKKGKASIRKIYYNLPKDKKMYARKLL
ISASKRFRGYNMVSKLSFELLD AFKGNAINIKEETTKKAIDKNKYFKKVYKTKYRKKLY
KNKKVYKKKYRKKVKEKNYKYKELKKKD"
gene      complement (join (21931..22200,5246..5365))
          /trans_splicing
          /gene="rps12"
          /label=rps12
CDS      complement (join (21931..22200,5246..5365))
          /codon_start=1
          /transl_table=11
          /trans_splicing
          /gene="rps12"
          /product="ribosomal protein S12"
          /label=rps12
          /translation="MTNIKQIFRKKTRKLNKIITKSSAPLKKCPQRLGMCMLRYIIHPK
KPNSGSRKVARIRLTSGIEVTALIPGIGINLVQVHSRVLVRGGLVKDLPGVKHKVILGS
KDAGRAKERFQRRSKYGAKKQKIKN"
exon     complement (21931..22200)

```

```

        /gene="rps12"
        /number=2
gene      complement (22732..26805)
        /gene="ycf2"
        /label=ycf2
CDS       complement (22732..26805)
        /pseudo
        /gene="ycf2"
        /label=ycf2
misc_feature 26630..26947
        /label=direct repeat B
        /note="direct repeat B"
gene      26807..26879
        /gene="trnI-CAU"
        /label=trnI-CAU
tRNA      26807..26879
        /gene="trnI-CAU"
        /product="tRNA-Ile"
        /label=trnI-CAU
gene      26912..27642
        /gene="rpl2"
        /label=rpl2
CDS       26912..27642
        /codon_start=1
        /transl_table=11
        /gene="rpl2"
        /product="ribosomal protein L2"
        /label=rpl2
        /translation="MTLNNIYKKKVKRKLIIYGYNFNKGRNYKGTITVKHRGGGKHRLY
        RKIDFKRNEKNTGKILKIEYDPNRNTTICLVYYKINNIYKYIIHPQNVGIGDTILSSHE
        TSIYAGNTLPLINIPFGTFIHNIIEIKPGKGGQLVRAAGTAAKVITKKKKRVLKLPSPGK
        LCSIPTCLATIGQVNKTALKKKLKAGTNRWLGKRPKVVRGTAMNPVDHPHGGGEGKAPI
        GRKFPATPWGYPAFGRRKKNK"
gene      complement (27709..28530)
        /gene="rps2"
        /label=rps2
CDS       complement (27709..28530)
        /codon_start=1
        /transl_table=11
        /gene="rps2"
        /product="ribosomal protein S2"
        /label=rps2
        /translation="MKKKIDWKIKLNDMLTEKIHFLNNTKNIKFKMIYYIFNINLKYYII
        DITKIASLLTIACNLV FYAAKNKKTFLII SQNKNKKEIKPIIIKTKNFKRFLINNK
        LRNKIKTKLRLKRKNKRINNKYKTLKLN YFKLISERIRCHYINRNIRAGLLTNWLTTKI
        KLKNLLISKTKINFIKKKKQKISLYNNLEGIKYMTRKRPDIVILIDPQNDIIIRECFTLK
        ISTICLININTDINIDLADIVIPINTFNLSIFYILNKLSYAILTG FYQKN"
gene      complement (28614..28698)
        /gene="trnY-GUA"
        /label=trnY-GUA
tRNA      complement (28614..28698)
        /gene="trnY-GUA"
        /product="tRNA-Tyr"
        /label=trnY-GUA
ORIGIN
1 ATTAATTACC ACCCTTAAGG GAAATCGAAT CCCTGTTATT TCCTTGAAAA AGAAATGTCC
61 TAAACCTCTA GACGATAAGG GCAATATTAA TAATATAAAC TTAATAAATG TATTGTGCAA
121 TGTTGCCTAG TGCGAAGGTG GGAATTGAAC CCACGACCTC AAGGTCATGA GCCTTGCGAG
181 CTACCAAAC TCTCCACTTC GCGCTATATT TTTATTTTAA TTACCAACTT GTTATTTTAA
241 TTACCAACTT GATATTTTTG CTCCTGGTAA TTTACATGAA TTAAATTGTT CATAAATTAA

```

301 ATTTCTAGAT TGTCACAAAAC AACGATAATA ACCTCTAGGA CGACCTTTTA TCGAACATCG
361 TTTTCTAATA CGTATTAATA TACTATTTTCG AGGCAATTTT TGTACTACTT TATTTATTTT
421 AAATTTAATT TTTAATAAAT AAACCTTTCTT TATTCTTTTT TTTAAGTTTT GTCTAATAAT
481 ACGATATTTT TTTTCTAAAT TTATTTTCTT TTTATCTCTA AAAATTATAC TTTTTTTTGA
541 CATAATACAT AATATTAACA AAAAAATAAA ATTATTTATTT AATATTTTAA ATTTATATTT
601 GAATATAAAT ATAATATTAT TATTATTTTT TTTTATTTTT ATATTAATTA AATTATTAAT
661 ATATTTATAA ATAATATATA AATAATATAT TTATAAAAATA AAAATATATA TTATATATA
721 TTGTATTAAT ATAAATTTGT ATTAATATAA CATCATATAA TATAATAAAT ATATAAGTGT
781 TATTTTAAAT TATAAATTTA TAAATTAATT TTTAGTAAAT AAATAATATG AATAATTTTT
841 ATTTTGTAT TTTTATAAAT AAATAATATT CACTAATTTT ATGTTTCATTA AATAAAAAAT
901 CTAATCTTTT TTTGTCTAAT TTTTTTGGTA TTTTTTTTGG TATTTTATA GTTTTTTTAT
961 TAAATTTTAG TAAAGTAATT TTATCTTCAG TTTTACAAAT AAAACTGGGA ATATTTACTG
1021 TATAATTGTT AACTAAAATA TGCTTATGAT TTATAAATG TCTAGCTTCT GAAATAGTTT
1081 TAGACAAACT CAATTTAAAT ATAATATTAT CTAATCGCAT TTCTAAAAAA CTACATAAAT
1141 TGTGCATTTT TAATCCTTTA ATTTTTTATAT ATATATATAA ATAATTAGAT AATTGTTTAT
1201 TTGTTAAACC ATAATAAAG CGCATTTTTT GTTTTTCTTT TAAAAGAGGA AGATATTTT
1261 ATTCACGTAT AATTTTAAAT TTTTTTTTTT TTCTATTGTT TTTTTTTTTT TTATATTTTA
1321 CAATTTCAAA ACTACGCTTT CTTTTTAACT TAGGTCCTAA ATATCTAGAC ATATAAATTG
1381 TTTAAATTTT AAGTTTTAAT ATTAATAAAT ATAAAAATA TATTATAACT ATAACATAA
1441 CTAATATTTT TAGTAAATTT AAGTTATTTT AAATTAATAT AAAAAATAA ATTAATTA
1501 CTTGTTAATT TTTAATTAAG GTTAAATCAT AGTATGAAA AAAAAATAA TAATAATATA
1561 GATATAAAAA AATATAAAGT AAGTTATAAA TATAATAAAT TAAACTTTA CTTCAAAGAT
1621 GATATAAGTG GTAAAACATT TGTGTTTTAT ATACAATTTA ATGTAAAAA TTTTAATAA
1681 AAATTTAATT TTATTAGTTT AAAATTAGAA GAGATTGGTT TTACATCATT AGAAGATTTA
1741 GTATATAAAT ATTTTTGTTC AGAATTTGAA TCAAAATTTA GTGATTTTAA TATAAATTTT
1801 TCATATGTAA ATAATAGTAA AGAGGAAGAA ATTTTTGAAG AAAAAATAA TAAAATAAAC
1861 AAAAAAATG AAAAAATATTA TGATTATTAT ATAGATGAAT GTTTAAAAAA AAACTTATTA
1921 AATTATTTAT ATATAAATAA ATATAATGAA TATATAAAT ATTTTTTAAA TGAAAAATTT
1981 GTTAATAATA AATTTTATAA TTTTAAATCA TTTAAATCAT ATGAAAAAT TTTAAAAGAG
2041 CTTATTATTT CAGAATGTAC TTACTAACA ATTGATGATT TTATATTTGA CTTAGACAAA
2101 GAAAAATTTT ATAAAGATTT TAAATATAAA AATATTATAA CTTAGATGA ATATTTTAT
2161 TATCAAAAAA AAAGAATTTT AAGTGAAGTA TTAGATAAAA AAGAAAATTT TGATTTTACA
2221 AAATTATGGG TTTTTTGTGA AAAGTGAAT TTCTTAAAT ATAAAAAGA TTTTAAGTCT
2281 AAATTATACG TTTGTGAAGA ATGTGGTTAT CTTTTTAAA TGAGTAGTTC AGATAGAATA
2341 AAAATGATAA TAGATAAAGA TACTTGGAAT CCCATAAATG AAGATATTTT TTCTATAGAT
2401 CCTCTAGAGA TTGATGATAC TTTATATGAT AATTATATAG ATGATGATA TACTGTAAAT
2461 GATAATATA TATATGATGA TAATGTAGAT TAATGTAGAT GATATATATA AAATATCTAA TTATTTTGGAG
2521 TGGTTTTTTA ATTATGAAAT TTTTTTGGAT TATATAGATT TAAGAAATTT AGAAAATTTA
2581 ATAGAAGAGT TACCTCAAGA ACAAGAAGAA GAAGAATATT TAATTAATAA AGAACTAAAA
2641 GGATCTTTAC CTTTAAATAA AGAAGAAAT CAAAAATATT TACAAGAATA TTCAACATTA
2701 ATTAAGAAG AACTAGAAAA CTATTTACCT CAAGAACAAG AAGAAC TAGA AGAAGAATAT
2761 TTAATTAATA AAGAACTAAA AAAATCATTA CCTTTAACTA AAGAAGAAAT TCAAAAATAT
2821 TTACAAGAAT ATTCACCTTT AATTAAGAA GAATTAGAAA ACTATTTACG TTTAACTAAA
2881 GAAGAAATTC AAAAAATTTT ACAAGAATAC TCACCATTAA GTAAAAGA AACTAGAAAA
2941 TCTTTACCTC AAGAACAAGA AGACCTAGAA AAAAGACAAA GACAACGAGA ATTTGATCTT
3001 TATATTGGAG ATAAATTAAT AAAACTAAAA GATATAAAAA GATACTTTTT TTATCTTTAT
3061 CTTGGAGATA AATTAGTAAA ACTAATAAAT ATAAAAATAC GAAAAAATAA AGAAAAATAA
3121 ATAAAAGAAA AAGAACTTTA TGTTAGTTTT GAAGAAGAAG TAGTAAACT AATAGAAAAT
3181 CAAAAACAAA TTATTGAAAA AGAATTAAT AATGAAATG AAAGTAAAT TAAATTTTAT
3241 GAGCACCTT ATGAGACTTT TGATATTTTA ATAAAAGAA ATAAAAATA ATATGAAAA
3301 GAACATTATA ATCTAACTGA TGAAATTTAA AAAGATATTG AATTTAGAAC ATCATTATTT
3361 GAAAGACTTA ATAAAGAAGT AAAAGAATAT GAAGAAAACA ATGAAAAAGA AGAAGAATAT
3421 GAAGAAAACA ATGAAAAAGA ATTAAGAA GAAAACAATG AAAAAGAAGA AGAATATGAA
3481 GAAAACAATG AAAAAAGATT AAAAGAAGAA AACAATGAAA AAGAAGAAGA ATATGAAGAA
3541 AACAATGAAA AAGAATTAATA AGAAGAAAAC AATGAAAAAG AAGAAGAATA TGAAGAAAA
3601 AACAATGAAA AAGAATTAATA AGAAGAAAAC AATGAAAAAG AAGAAGAATA TGAAGAAAAC
3661 AATGAAAAAG AAGAAGAAAA GGATAAACAA CAACTAAGA AAATATTAATA AATATTAAT
3721 GAAAAATTTG ATGAAGAAAA ATTTGATGAA ATAGAGAAAA ACAAATCATA TATTAATCGT
3781 TAAATTTTTT ATCGAGAAAC AACAAGTTTA GTTGAAGCTG TAAAAACAGG TATAGGTAAA
3841 ATAAATGGTA TAAAGTTGC ACTTGGTATA ATGGATTGAG AATTTATGGG TGGAAGCATG

3901 GGATCTGCTG TAGGAGAAAC TTTAACTCGT TTAATCGAAT ATGCTACTAA AGAAAAATTA
3961 CCTCTTATAA TTTTCTGTGC TTCTGGAGGA GCACGCATAC AAGAGGGGAT TTGTAGTTTA
4021 ATACAAATGG CTAAAATATC TGCTGCATTA CATTATTATA AATCAAACAA AAAATCAGTA
4081 TTTATATCAA TACTTACATC TCCAACAGCA GGGGGAGTTA TTGCAAGTTT TGGAATGTTG
4141 GGTGATATTA TTATAGCTGA ACCAAATACT ACTATTGCTT TTGCAGGACA AATAATAATT
4201 GAACAAATTT TAAGAAAAGA AGTACCATAA GGTTTTCAAA CAGCTGAATA TCTATTTAAA
4261 CAAGGTGCTT TTGATTTAAT TTTACCACGA CAACTTTTAA AAACAATTGT TACTGTATTA
4321 CTTAATTTTC ACGGGTTTTA GACATGTTCT TAGTTAATAA TTTAATATTT AATATTAATA
4381 AAATTATTAT TTAATTAAAT TTTATAAAAA AAAAACGCGC TCTGTAGGAT TTTAACCTAC
4441 AACTTAAGGT TTTGGAGACC TAAATCTTA CTAATTGAAT TAAGAGCGCG TATAAAAAAT
4501 ATTA AAAAATG AAAATACAAA TAAGAAATAA AAAAACTATA AAAAATTTAA ATTTTTTAAA
4561 AAATTTAAAT ATAAATAAAA TAAATTATAA AGAAATAAAA TTACTTTACA AATTTTTAAG
4621 TATAAAAGGA AAAATATTAT CAAGAAAAGT AACTAAATTA AATTTAAAAA AACAAAAAAG
4681 CATTACTATA GCTATAAAGA GAGCTCGTAT ATTATCTTTA TTACCTTTTA ATAAATCAAT
4741 TTACTTAAAT AAAAAATAAA CGATGCAATA AAAAATTTAA ATATAATAAA TTATTTAATT
4801 TAAAAAAAT TATAAAAAAC TATTAATCTT TTTATCCAAT ATATAATAAT AAATAAATTA
4861 AACTTAAATA ACCTTCCTTT TTTTATTAG TTTTAAGTTT ATAGTATAAA CCAATACTAT
4921 ATTTTTTTAT ATCGTTTTATA TGTTTTAATTA TTTAATTTAA ACAGTTAGTT CTTTTTTTGT
4981 ATAAAATTTT TGTATAAAAT ACCGGAGTTT ATTAATTTTA GTATTAECTT GTATAATTAA
5041 AACTTTTTTA CTCTATAAAA ATAAAATTA AGTAACGATA TTATAAAAAAT TAGTTTACTG
5101 AACCTTTAAT CTGTAAATTT TATATAATAT ATAAATATAC ATTTATAAAT TTCCGCTTAT
5161 TTAATAAAAA TAAAATTTAA AAATGGAATT TAATTTTTTA AACAAATTTAA ATAATTATAA
5221 TTAACATAAA ACATACGAAT CACACATACA ATCTCATACA CATTCCATAA CGTTGAGGAC
5281 ATTTTTTAAG AGGAGCTGAT GATTTTGTAA TTATCTTATT TAATTTTCTT GTTTTTTTTT
5341 TAAAAATTTG TTTAATGTTT GTCATAATTT ACTTTTTATT TTTATATTAT TTTTATATTA
5401 TATATTATTA TAATTATATA TAAATTTATA TATAAATTAT AATATATTAA TAATATATTA
5461 ATTTATATAT ATTAATTATA TATTATATTT TTTATTTACT TATTTACATA TTATCTTACT
5521 TTTTTTAAAG TAGATTTTACA ACCATTATGA GAAAAAGGAG TAATATCTAA TATAGAATAA
5581 ATTTTTATTT TTTTACTTTT AATAACTTGT AACTACTGAAG TTCTTCCTGC ACCTATACCT
5641 TTAACATGTA TTTCAGCAGA ATGTATTTT ATTATATTAA AAACATTATT TGCTGTAGTT
5701 TTAGCTGCAA AAGGAGTACC TTTTTGTGTA TTTTAAATTT CATACATACC AGTAGAAGCT
5761 GTAAAAATAA GTATATTTTT TTCATCCATA ATTGATGATC TAGTATTATT AAAACTAGCT
5821 AAAATAATAA TAATTCCTTT TTCAAGGAAA TTTATTTTTG GTTGGCGCTT ATGACGATAA
5881 AGAACTTGTA TAGGTTTTAT TTTAATCTT TTTCTCATAT ATTTTTTTAT TTTTGTTTTT
5941 GTTTATGCT TTTATTATTA CAAATAATTA AAATACGCCT ATATCTTTTA ACTAATTTAC
6001 ATCTTTGGCA AATTTTTCGA ATAGATGTTT TTTTTTTCAT TTTTATACAA CATAAAAAAT
6061 AATTTCCCTT CCAATTTTAT TTAATTTAGC TTCTATATCT TTTATATTCT CTTGAGATGT
6121 AGAAATAAAA TAAACTCCAT TTGCATTTAA TTTTTTAGGA ATTATTTTAT AAGTTACATA
6181 AATACGTTGA CTAGGTTTAC TTATTATTTT TACAAATATT TTTTTTTTTA AAATTATACT
6241 AAAATTATTT TTATTTTTTTA TAAAATTTTC AATAAAACCT TCTTTTTAAA GTAATTTTAA
6301 AATATTATTT GAAGTATTAT TAAAAGGAT TGTTGTTATC TTTTTATTTG ATAAACTTGC
6361 ATTTTTTATA TGAACAATTA TAGTTATAAA ATTATTTAAC ATAATAATAT TTAACATAAT
6421 ATTTATATTT ACAATATTTT AGAAGATAAT GAAATTTATT TATTAATAAT TAATTTGCGT
6481 AATTCCTGTT TAATTATACC AAAAAATCTA GATCCTTTAG GATTGTTTAA ATTATCAATT
6541 ACAACAGCAG CATTATCGTC AAATTTAAAT ATAATACCAT CAGAACGTTT AATTTCTTTT
6601 CGTGTACGTA CAATAACAGC TCTTATTATT TCTGATTTTT CTATTGAAAC ACTAGGTAAT
6661 GCTTCTTTAA TCACAGATAT AACAATATCA CCTATATAAG CAGATTTAAA TTTACTAATA
6721 CCTAATTTTT GAATACACAT TAACTTAATA GCACCCTGT TATCTGTTAC ATTAAGATAA
6781 GTTTGTTTGT GAATCATATT TTTAACCATA AAAAAATTAG GTTTTAATAA AAAATTTTGT
6841 TTTTACAGGT AATTTAGATG AAACTATTTT AAAAGTATT TTTGCAAGTA TTTCCGATAC
6901 GTTATCTACT TCATAAAGTA TTCTGCCAGG TTTAATAACA GCAACCCAAT TTTTCAGTGA
6961 ACCTTTACCA GACCCCATAC GTTTTCTTAC AGATCTTTTT GTAATAGGTT TATCTGTAAA
7021 TATTCTAATC CAGAATTTAC CACCTCGTTT AAAATATCGT GTCATTACTT TTCTACCAGC
7081 TTCTATTTGT TTAGCTGTGA TCCAAGTTGG TTCAAGGTTT TTAATTTCAA ATTTACCAAA
7141 ACAAAATAAAA TTACCTTTTA CTGAAATTTCC TTTTCATCCTA CCCCTATGTT GTTTATCAAA
7201 TTTAGTTTGT TTAGGTAACA TTTTATAATA AATTTTTAAA AATTTAATTA AATAAACTA
7261 AAAACTAAAT TTAATTTTAT TTTTATAATA AATAAAAAAC AGCTACAATT ATTCAAAAAA
7321 ATATAATTAA ACTTGTAGCA AATTTTTTTTT TATTTAATTA TTAATTTAAT TAATAGTTAT
7381 TAATAGTTAA AGAAAAGGTC TAGGTTCCGA ATATTCACTT ATAATATAAT CTATAAGAAT
7441 TGCTGTACCA AATATAAGTA CTATTTCTAAC ATAAGTTATA ACTTTTCTAA TGATTTTTTT

7501 TTTTGTAATT ATCATAATTT TTTTAAATGT TTTTATTATT ATTGTTATAT AATAATTATA
7561 ATTATTATAT ATAATTATTA TATATAAATA TAATAGTGGT TTAGATCTAT CACAAACCTA
7621 ATTTTATAAA GTTATTTGTT ATAAAAAAA TTTAAATTTA AATATTAGTT AATATTAAT
7681 TATATATAAA TCCAAATTTT GATTCCTAAA AGTCCATTAT TTGTTTTAAC TGAATATGAA
7741 CAAAAACGTA TATTTTCTTT AATTGTTTGT AATGGTAAAA TACCTTCTTT AATCCAAGTT
7801 TTATTTTTTAT GTACAATAGC TTTACCTGAT GTTCGTCCAG ATATCTTTAC TTTAATTCCT
7861 TTAATTTTTTT TTTCATTGTA AATTTTATAT AAAACTTTTT CTTTTATTTT TTTTAAAGTA
7921 ATTCTTTTTT TTAATTGAAA TAATATATAT TTAGCAATAA CTTTLAGGATT TTTAAAGGG
7981 TATTTAATCT TTTGAACCTT TATATTAATA TTTAACCTTT TAAAATTTAT AAATTGATTT
8041 CTAATAATTAG AATCTTTTAT AAACATTTT AGTAATTTAA AGTTTAAAGAA ATCTAAAATT
8101 ATAATTTTTTA TTTTAAATTT GTTATTTTGT TTTGTTATTC TAATATGTGA AATAACATTA
8161 TTTACATTAA TATTTTTTTT TTTAAAATAA TTAATAAAAC AGTCTTGTAT TTTATTAGTT
8221 TCGCATAAAT TTATAGAATA ATTTTTATTA TTATAAAACC AAAAATACAA ATAGTTTGGT
8281 ATAATATTAA GTCTAAAAC TATTGGATTT ATTTTTGTCA TAAATTTATT ATTAATTTTT
8341 TTTATTTTAT TGATTTTTTA TAATATTTT TTTTTTTTGA TTTAACATCT ATTTTTTTTAT
8401 GACGTGGAAA AGTTCCTAGT GGTACAAAAT CACCTAATTT ATGACCTATT TTTTTTTTAT
8461 TAATACAAAT AGGTATATGT TCTTTTCCAT TGTGTATTGC TAAAATTTGA CCAACCATTA
8521 AAGGTATAAT GGTTGATGCT CTAGACCAAG TTTTAAATAAT TTCTTTTTTT TTTTCTATTT
8581 TATTTTTTAGT TATCTTGCTC ATTTAAATGAT TATGTACAAA AAGTTTTTTT TTATTCCTAT
8641 TTATAATATT TATAATATAT TTATAATATT TATATATAAT AATATATAAT AATTAGTTAT
8701 TTTAAAAATT TTATATTATT TTTTAGGATA AAAATTTAAT CTAAACATAC TAAAGCGGCT
8761 ACCATTACTT GTATTTAAGC AATAACGATT TATGCAAGCT AAGTCTTCTA AAGAATAATT
8821 AGGCCAGAGT AATTTTATGA AAATTTTATT ATTATCTATT TTATTTAATT TATGATCTTG
8881 ATACTTAATT AAATTAATTT TATTAATTT ATATTTTATT TTGTAATTAT TTTTAAAAGT
8941 TTTAAAAATT CTTAATTGTC TACGAAATTT ATAAGTTAAA ATCGTTTCAG GACTAATAAT
9001 ATTAGTTTTT TCTTTATTTA TTTTATTATT TAATGAAAGA CTTAAAATTT GATGTATAAT
9061 AAATTTTTTTA TTAATAAACG AATTTGGTAT TATGCCAATA GAAAGCCATG TAAAGCGTTG
9121 TCTATCAGCA TATAAACAGT CTTTCATCTTT TATAATCTTT TTAAACTCAT ACTTTTGTCC
9181 AAAAAATAAA TGAAGTACAA CAAGATTTTT ATGAACTTTT TTAAATTTTA TTGATTGAAA
9241 TAAAAGTTTA TAAAAAATTT CTTTTTTTTT TTTTTTTAAA TCATAAATTAT CTTTGAATC
9301 TATTTTATCT ATTAATCTTA GTGTTAAAGG AAGTCTTGTT AATAGGAATT TAAAAAATA
9361 TCTAACCTCG CTAGTTAATT CTATTTGATA ATTATTATAT TTTTTATTAA TTATATACAA
9421 AGCTTTGCTA AATTCTCCTT TTCTACTATC AACTTTTTTT CTACTATCAA CTTTTTTTCT
9481 ACTATCTACT TTTTTTTCAC TATCAACTTT ATCAGAAATA TTATTATTTT TATAATTATC
9541 AGTAATATTA TTAATTTTCA TGTGATCAGC AGAGTTATCA GCAGAAATTT TTTTATTTTC
9601 CTCATTATCA GAAATCTTAT TATTTTCCCT CTTGTAAGAA ATATTATFAT TTTTATTTT
9661 GTCATAAAAT TTATTTTTTT TATCGTTATG AGAAATATTA TTATTTTTAT TATTATCTC
9721 TTTATTTAAT TCAGATAAGT TTTGATCAGC AGAAATTTTT TTATTTTCCT CATTATCAGA
9781 AATCGTCTTA TTTTCTCCTT TGTAAGAAAT ATTATTCCTT TCATTTTGGT CAGAAAATTT
9841 ATTTTTTGTA TCCTTATTAT AATTTGTATT ATTTTTATCT TTATTTTCAT CTTTATTTAA
9901 TTCAGATAAT TTTTGTAAAT ATATCTGTTC ATGTATGTCT TCAATTTTAG TATTAATAAA
9961 TTGATTTATA GGAAAAAGCC AAGGGTTTTG AAGATATTTT CTTTGTTTTA ATCTAAGATT
10021 TAATAAAATA TTATTTGGTG CTGTTATAAA AATATTATAT AAATATAATT TTTTATTATT
10081 AAATTTAAAT TCTGGAATAA TTTTTTTTTT TAATTTATGT AAATTTTTTT TAACATAATG
10141 TCTAAGGAAA TATTTTTTTC CTTTATTTAT TATTAATTTT GCTTGTCTTT TTTTCTTTTG
10201 TATTTCTCCT TTTATCCTTT CTGCTTCTTC TTTTTTTTTA TTGTATCTTT CTTCTTCTTT
10261 TCTCTTTTCT TCTTCACTTT TTATTTTACC TTTTTTCTT CTTTCTTCTT CTTCTTTTTC
10321 CTTCTTTTCT TTTTCTTTTC TTTCTTTTTC TTCTTTACTT TTTCTTTCTT TTTCTTGTG
10381 TTTTCTTTT TTTATAATAA TTTTATTATA CTTTCTGTGTT TCATTATTTT GTTTAATTA
10441 ATTCGTCAGA ATCAAAATGT TTTTTTCTT TTTTTTAAAG TATATTTTGA ATAGATAAAT
10501 TTTCTTTTATT GTTTTTATTT TTAATAGTTA ACTTAATCGG TAATATTTT TCAATTTAAC
10561 TATATTTTTT TAATAAATTT ATAGATTGAT AATTTTTTTT TATATTTTTT TTATTTAAAA
10621 AATTTTCTGG TATGAAATTA TTAGAATTTA TTATAAACT TTCCAAGAA AAACCAATAT
10681 TACTAAAAAG TATTCTTAAT TTATTTAAAT TTATATAAAT GTCATGTTTG AATAAAATAC
10741 TATATAGTAA ATATGCTTGA GATATTTTAA TATTTTGTGG ATAATTTTTT ATACTTAAAT
10801 CTATTATATT TGAAGATATA TCTAATCTT CTTCTTCTTT TATTTCTTCT TTTTACTTA
10861 TACTATTATT ATATATAATA TTATATATAT ATATAAATAT AGTTTTAATA AATTTAAATA
10921 TTTCTTCATT ATTTTCAATT ATAAATTTAA ATTTATTAACT TTTAACATAA TAAATAAAAT
10981 TAGACAACCT TTATGCAAGT GTGTATCAAT TTTGATTCTA GGCATCAAT CTTTCTTGT
11041 TTTTCGTATT TATATTACTT ATTTTTTTTT TTAATATATT AATATTATTA GATATTTTTT

11101 TTATTTCTAT ATCATAATTT ATAATTTTTG TTTTTATATC ATCATATTTT TTTATTTTAT
11161 ACTTTTTTAT TTTATAATTT TCATATAATA AATTTTTTTT TAATTTACTT TTTTTAAAAA
11221 TTTAAAAAAG CTTATAAATA TAAATAATTT TTATTAATTT ACTTTTACTA ATATATFCAA
11281 TTAATTTTACT TTTACTAAAA ATACTACTTA ATTCTCTAAA TTTACTAATA ACATAACATA
11341 TTTAAAAAAT AAAATTAATT AGTATTTTTA AAATTTTTTT AAATAAATAA ATTATAAATC
11401 TAAATATTA AAATTAATA TTTATAATAT ATTTATAAAA AAAAAATAAA AAATATATAA
11461 AAGTATAGAA GATTTTGTTT ATATTATAAT TTACTATAAT TTTTAGTCCA ATTATATTGT
11521 AATCTAGTTC ATATCTTTTT TTATATATTT CTATATTAAC AAGTTTATAA AAAGTTGAAT
11581 TTAACAAATA TTCAAATTTT TCACTAGAAA TGATTCTAGT TTCTTGTCCT TTCATGTCTA
11641 AAAAAATATA ATTTCACTA CTTTTAATTT TAACCTCTC TTCTTTTTCA GAATTTTCGT
11701 CGTACCAAGG TCTTAATCTA AAAGGATTTA TTATTTAAT TTGAATACCT TCTGACATCC
11761 AATCTTCTGG AAGTTTATTT TCAACTAATT CAATACCATT ATAAGTACAC ATAATATATT
11821 GTTCTTTTTT AAAATAAAC ATATCTTTAT GATATTCAAT AGGTTGAAAT AAAAAATAC
11881 GAACTAAAGT TTTAACCAT AATAAAAAAG GTATTTTTTAT CCACATCTT ATATATAATT
11941 GAGAAATAAG ACAACAAGTT CTGAAAATAT TTACAAAACC ATAAGTACTA TCCCAAGCAA
12001 CATTCTGTC TGCTCTTTTC TTAGACTCAA TTAATTTAAA GAAATCTTGT TGTTTTTTTA
12061 TTAATCTTTT TTTTACAATA TCTGCATTTT CAGGACCATA AAATTCAGT GAATCATTTGT
12121 CATTAAAAAT TAAAAGCCTA TCTTCTTCGT CTCTCTCAAG AAGAAAAATG GGAGATGAAG
12181 GAGATCTATA AGTTACAAAA TCAGGTAAGT TAAGAGGAAT CTCTTTACGT TTTTGATTAC
12241 GTAAAGACCC CCTTACAGAA AATCTATTTA AATCATAAAA TTCTGAAAAT CTAATATCAT
12301 GCAATTCTCC TTTTTTTTTT TCGTCTTCTT CAATCTTTT TTCAAAATTA TAACCAAAAA
12361 AGTTATTTTT ATCCACTTTT TTATATCTT TATCTTCAAT TTTTTTATTT ACATTTAAAG
12421 AGTCAGATAT ACAAGCTAAT TTTAATTTAC TTAAAGGTAC TTTATATCTT ATTTTAAATT
12481 TTTTATCTGT TATTTCTTTA TCTTTATCTG TTATTTCTTT ATCTTCTTTA TCTTTATCTG
12541 TTATTTCTTT ATCTTTATCT GTTATTTCTT TATCTTCTTT ATCTTTATCT GTTATTTCTT
12601 TATCTTTATC TTTTATTTCT TTATCTTTAT CTTTCGGAGG TTCTGTATAA AGTATCAAAG
12661 GTCTTAATAT TCTACATTTT AAACCTGTAA TTTTTTTAGA TCTCAACAAA AGAAATCTA
12721 TATCATCTTT TAAGCTACAA ATCCATCTAG GAATATTTTT TTCTTCGATA ATAAAAATTT
12781 GAGTTTCTTC ATATTTTTTT TTATTTTTCA AAAATTTTTA AATAGGTGGT AGATTTTCAT
12841 ATATATATTT TTTTATTGCT TTTTCTTTT CATTAAATTT ATTTATGTCT GATTTATTTA
12901 TATTTATTTT ATTTAATTTA TCAATTTCTT TAAAATTTAT TATGTATGAT TTATTTATAT
12961 TTATTTTTTC TTTATTTTCA TTAGTTTTTT CTTTATAAAA ATTACCTTTT AAAATATCAT
13021 CTAATTTTAT TAAAGCTTTT TTTTCATTAT CTTTATCAAA ATAAATAAAA AATCTATTTT
13081 TTATTTCAAA TAAATTTAAA TTTTTTGATT GATCAAAAAGA TTCTATACGT TTTTTAAATT
13141 CATTTTCAAA ATTATTAAGT TTTTTTTTTA AATTAATCT GTCATAAGGA TATATTTTTT
13201 TTAAAAACCT TGTGTTTTTT TTTATTAATT TAAAAAGGC ATACAAATTT GGTGAAAATG
13261 TATAATAATA TCGTTTTTTA TTATCAATTT TATAGAAACT AAAAAATAAT TGTGACATTT
13321 CAGTTCGTAT AAATTTATCA AATTGTTTTT TTTTTATATA TCTTATAGGA CGATTTCCAAC
13381 GGTTAGTATA ATAAAAATA TCTACTAAAT CTTTATTGTC TTCTTCTGAT ATATTAAGTA
13441 AGTTTAACTC TTCTTCGTCT TCTACATCAC TTTCTTCTAT TTCTGATCA CCTCCTTTA
13501 GTTTTTTTGC TCTTTCTTTT TTATTGTCTT TTTCTACGTC ATTTATGTCC TTATATTTTT
13561 CTTCTTTTTC AACGTTAGAA AATATAGATC TTTTTAAAAG AATCTTCCG TCATTTAAAAG
13621 TTATTGATTT TAAACCTGTT CTTTGAAAA TAAAAGTAAA TAAAGGTATA GGTGTTCTAG
13681 ACAAAAATAC TAAAAATAAT ATTAGATAAA TATATAAAAA ATATATATCT AAACGTCTT
13741 TGTTTTTAAA ACTATTTAT ACTCGATTTA AAAATATATA CATAAGACTT TCGCTTTCGC
13801 TTTTTTTTAA AAAAAATTTA AACTTAAACC AACTTATCAG TTTATAAATA TATTTTATTA
13861 TAATAAAAATA TGAACAAATA CTACCTAAAA AATTAATAC TAAAAATAAT ATTTTATTAG
13921 TATTAGAATA ATTCCTAAAT AAAAAAGATA TACTCAATCT CTTAAATATT GGGTTTGGTA
13981 AATTAANAAT AATAATTTGA AATATAAAA GTTTTATAAT TATATCTAAA ATAAAAATA
14041 ATTTAAATTT AATATCTTTA TAATTTTTTA TAAATATAT TATAAAAATTT TCGATATATT
14101 TTTTTAATTT TATTTTCATT TCAATATAGT ATGATTTAAA TACATATAAT AAGTTTATTA
14161 GTTTAAATTT TATAGTATTT TTTTCATAAA TATATTTACC TTTATAAAAA GAAAATAAAG
14221 GTTTTATTTT CGGTCTCATT TTAATAATAT TTTTAATATG TGTTTTTTTT TGTTTTTTAA
14281 ATATAATTGG TTCTAAATTA TCAAAATTTA TTAATCTAAA ACCGTATAAA AGTTTAAATTA
14341 AAAAAATATAT CCAGTATATA CCTTGTGTA ATATACCTTT ATAAAAAGTA TAATATAAAA
14401 AAATGCATAA TAAAGTGTGA GGTCTTTGTA AAAAAAGAAA CAAAGGCGTA TAATATATAG
14461 ATAAACTCAT TATTAATTGA CTAAAAACGT ATGAGCTTAT TATCACAAT ATTTTTCTT
14521 GTTTTTTTTC AAAATCTCTT TTAAATAAGT GCTCTATAAA AAAAAATAA GAAGGTATCA
14581 CACAGAATGA ACTAAAAAAA CCAAAATAAA AACCTATAAA AGAATAAAAA AAAAGTTTAT
14641 TATTAAATAA ATTTCTTAAT ATTAATAAAA AATGTATTAT TACTTTTACC ATAATTTTTT

14701 ATATTTGAAT GTTATATTTA AGTTTGTATA TTTTAAATT AAAAAAGGTC CCCACCACTA
14761 GTTTTTTTTA TTCGGCATAA AATTATTTTT CCATAAGATT CCTCTTAAAG TATCGTCACT
14821 AATTAAGACT TTTACCACCT AATTCGAAAA GGTGTTGGT GTAGATCATC TTAGTTAAAA
14881 CACCGAAAGT TTATGAATAT ATTAATTTA ATTTTTTAAA TTATTATTAT TCAAAATCATC
14941 TTTCGGTCTT TAGAAATATT TAAGCTAAAT ATATTACTAT ATTTACACAT AATACCTATT
15001 AACAAATTTAT CTAATTGTAA CCTTAATTTA TTATAATAAT TTTATTATCA TATATATATT
15061 TATATAATTA TAAATTATAA AATAATAAAG GGAGCACTCA TCTTTGTTTA AACTTACTAC
15121 TTATATGCTT TCAGCAGTTG TTCATTACGC ACATAGCTAC CTAGCGTTTA CTATTGGAAT
15181 AATAACTAGT ACACAAGTGG TCGGTTCCGT CTGGTCCTCT CGTACTAAGA ACGATTAACA
15241 TCAATACTCC AACGTTTACA CCGATATTGG ACCGAACTGT CTCGCGACGT TCTGAACCCA
15301 GCTCACGTAC CGCTTTAATG GCGAACACGC CCAACCCCTG GAACCTCCTT CAGCTCCAGG
15361 TGGCGATGAG CCGACATCGA GGTGCCAAC CTTCCCGTCG ATGTGGACTC TTGGGGGAAGA
15421 TCAGCCTGTT ATCCCTAGAG TAACTTTTAT CCGATAAGCG ATGGCCTTAC CACTCAGTAC
15481 TACCATCGGA TCACTAAGAC CAACTTTCGT TCCTGCTTGA CTTGTTTGTG TCACAGTTAG
15541 GCTTCTTTTT GCCTTTACAC TTTAAAAATC ATAAAATAGA TTTAAGAAAA CCTTTATTGC
15601 ACGCCTCCGT TACTTTTTTG GAGGCCTACG CCCCATATAA ACTGTCTACC AAAGACTGTC
15661 CTTGCTAAAA AACAAAGTTAG AAGTTAAACT TAATTAGGGT GGTATTTTAC TGATGACTTA
15721 ATATTGCCTC CCACCTATAT TTTACAAAAA AAAGACCAAC ATCAATCACA GGGAACAGTA
15781 AAGCTTCATA GGGTCTTTCT GTCCTGGTGT GAATAGTCCG CATCTTCACA GACATTTTTA
15841 TTTCACCGAG TCTCTCTCCG AGACAGTGCC TAGATCGTTA CGCCTTTCAT GCGGGTCGGA
15901 ACTTACCCGA CAAGGAATTT TGCTACCTTA GGACCGTTAT AGTTACGCC GCCGTTTACT
15961 GAGATTTCCG TTGCTAGCTT TCCTGTATTT TTCAGGTCAC CAACTTCGTT TACCTTACAG
16021 CACTGGGCG GCGTCAGCCC CCATACATGG TCTTACGACT TTGCGGAGAC TTGTGTTTTT
16081 GTTAAACAGT CGTCTGGGCC TGGTTATTGC AATCCTTTTT AAGGACACCC TTTTTCCCGA
16141 AGTTACAGGG CTATTTTGCC GAGTTCCTTA GAGAGAGTTA TCTCGATCCC CTTAGTATTC
16201 TCTACTCATC TACCTGTGTA GGTTTACGGT ACAGGTATTT ATTTATTTTT TTTTCAATTT
16261 AGTTTTTCTT GGAAGTATGA CCTCAGTTAT ATTAATACAA ATAACAATAC ATATAAACGT
16321 ATTTAATATA CTAGTATATT TTTAAGACTT TACAAAATG TTTAAATTTT TGCTTATATC
16381 TTTCGACTGA TAACCATTTA TCAGCTAACC TAGCCTTCTT CGTCCCATAA AATGAAAAATA
16441 AATAAATAGT ACAGGAATAT AAACCTGTAA TCCATCAACT ACGCCTTTCG GCCTTATCTT
16501 AGGTCTGAC TAAACTTTTCG CATATAATAA TTTTCGAAAGA ATCCTTAGAT TGTCGGGGCA
16561 TAGGATTCTC ACCTATGTTT TCGTTACTCA AGCCAACATT CTCACTTCTG ACATAAAATT
16621 AAAGCAGAAC GCCCCCCTAC CGATTGTAA TCTCATAGTT TCGGCATATT AATTTAGACC
16681 CGTGTATTTT CCGCACAAAA ACGCTTTTATC AGTGAGCTTT TACGCACTCT TTAAAGGATA
16741 GCTGCTTCCA AGCAAACCTT CTGACTGTCT ATGCACCTT ACATCCTTGT CCACTTAATA
16801 ATAATTTATG GGCCTTAACT AATGATCTGG GCTGTTTCCC TCTTGACAAT GAAGCTTATC
16861 CCCCATTTGCT TCACCTAGTTA ATTTATAAAT GTACAGTATT ATTAGTTTGT ATAAAAATTA
16921 AAATTTAAAC AGTGCTTTTAC CCCTGTTTATC ATAATTAACT GCTGTGCTTC GACACATTTT
16981 GGGGGGAACC AGCTAGCTCT GGGTTCGAGT GGCATTTTAC CGCTATCCAC AATTCATCCG
17041 CTGATTCTGC ATTATCAGTC GGTTCGGACC TCCACTTAGT TTTAACTAAA TTTCATCCTG
17101 ACCATAGATA GATCATCCAG TTTTCGGGTAC ATAAGTAATG ACTTATATAT ATATTATATA
17161 TATATATTTA TTGCCGTTAT AGACTTGCTT TCGCTATGGC TACGACTTAT ATTATTCTTA
17221 ACCAAAAGCCA TTAATTATGA GTCGCTGGCT CTTTCTTCAA CAAGCACGCA GTCAGAGATA
17281 ATAAAAATTC TCCTCCTACA ATTTGACAGC TTTTGGTTTC ATTTAAATAT ATCTTTTAAA
17341 AGTTATATTA TACCTTTCCC TCACGGTACT AGTTTCGCTAT CGGTTACTCA GGAGTATTTA
17401 GTCTTACAAG GTGGTTCTTG TAGATTACACA CGGAATTTAA CTTGTTCCAG GTTACTTAGA
17461 TAAGTTAATT AATTTAAGTA TTAAATATAA TTAAGTATTA AATATAAAAT TATAAAAAAA
17521 ATATGTTTTT ATATTTTGAC TACTGGACTT TAACCATCTA AGGTACAGTA TTTTATCTG
17581 TTTTATCTAA TAATATTATT TAAATTTTACA TAAAATTTAG TTTTATATAA TAACCTACCT
17641 ACAACTTCAT TTTAATGATA ATGATTTAGA CTATTTCCAT TTCGCTCACC ACTACTAAGG
17701 AAATTACTTT TGTTTTATTT TCCTCTGGAT ACTAAGATGT TTCAGTTCTC CAGGTTATCT
17761 CTTACCTTAT GAAGGCAGTT TTA AAAATTA ATTTATTCGG GAATCTTTAA TTTTATAAAA
17821 TTTTATTTAA AGCTTTTTTC TACTTTATTA CGCCCTTCTT CGTCTCTGAG TACCAAGGTA
17881 TTTACAAAA GCTTATTATT AAGTTTGAAT ATTATAAAAA TAAAAAAGAT TTAATTTATA
17941 ATTATATATA TATATAACGT TTATATTATT TATATAACTT TTATTACTTT TATTAAAAAT
18001 CTAAGTTTTT TCATTTGTAT TTTTATTATT ATTTTCTAAT AAATCATFAA TTTCTTCTGT
18061 ATTTTATTA TTATTTTCTA ATAAATCATT AATTTCTTCT GTATTTTAT TATTATTTT
18121 TAATAAATCA TTAATTTCTT CTGTATTTTT ATTATTATTT TCTAATAAAT TTAATTTTTT
18181 ATTTTACCT CTTTTATCTT TTTTATTTTC TAATTTTACG TTTTTTTTAC TATCTTTTTT
18241 ATTTTGAAT TTGTTAATAC GTTCTTTAAT ATTTTTTAAA TAATTTAATA TTTTTTTTAG

18301 TTTTTTTGAT AAAATTTGTA GTACAAATAT AGATTCATTA ATAATTACAA ATAAAAAAGA
18361 TAATATTAATA AATAAAATAT CTAATATGAA AATTACAAAA TCAATTAGTA AATTAAATTT
18421 AAGTATTATA AGTTTTTTTA TAACTTTTAT AAATTCATT GTTAAATATT TTGATATTTT
18481 AAATTTAAGC AAATATTTTA CTAATTTTAT AAAAAATAAA AAAATAATAA TAAAAATACA
18541 AATTTTATAG TAAACATTTA AATAATTGTT TAAATACACT AAATTAACCT TATATATTTT
18601 TAGTTTATTA TTTTTTAATA AATTAATTTG TGTTTTTTAT AAATTTAGTA AAATTTCTTT
18661 ATATATATAA ATATAAAAATA ATTTAAATGT TTTAAGTAAA TTTATAATTT GATCATAAAA
18721 CGATGAATAA AAAAAACAAA ATATTTTGT TTTAAGTAAA TTTATAATTT GATCATAAAA
18781 AAAATTATTA AAAAATTTTT TATATACATT TACAGCTTTG TAATCTAAAA TTTGTTCAAC
18841 ATTATTTTGT TTTAACTCTT CTTTTATTTT TTTAAACAAA TGTTCTTTGT TTACTAAATT
18901 TACTAACGTG TTTTTTTTTT TTTTTATAAA AATATTTTTA ATGTAATATT TCATAAAAAG
18961 AGTTGTAACA CCGACTAATA ATAAAATTAC TTTATTTATA TTTCTCATAT TAATATTAAT
19021 AAATTATTAT TTTTTTTTAT TTTATTTAAT ATTATTATTA ATAATTTAAT ATTATTATTA
19081 ATATTATTTA ATTATTTAAT ATTATATATA ATTATATATA TAAATATAAT TAAAATATA
19141 GTTAAATTA TTAATATAAT TATGTATATA TATTTATATA TAAAATAAAT ATGTATTTTT
19201 CTTAATAAAT GTTTTTTATA AAAATGATAC ATTTTTTATA TATTTACCAT ATTTTATTTT
19261 TTTATCTTTT TTTTTATTAA AATTTATTTA TAATAATATT ATTATTTTFT ATTTTFTTTA
19321 TTTTAGGCTT AATCCCATAC AGCTAATTTA AAAAAAATTA TTTATAAATT AATGTTTAAA
19381 AAATTAATAC AATATCTAAG TGTATCCATG GCTGATTGGT GAAAGCATCC AACTAATAAT
19441 TGGAGATATG TAGGTTCAAT TCCTGCTGTA TACAAAAAAA AATATTTTATA ATCATTAAATA
19501 AAAAAATAATA AAAAAAATGA CATTAAATAA CAGATATAAA AAAAAAGTT AACCTATTAG
19561 ATTTATAAAA CATAAAATAT TTATAATCAT TAAATTTATA ATCATTAAATA AAAAAATAA
19621 AAAAAAATG ACATTAATA ACAGATATAA AAAAAAGTT AACCTATTAG ATTTATAAAA
19681 CATAAAATAT TTATAATGTA ATAAAAAAT CCAGCTGCAC CTTCCAGTAC AGCTACCTTG
19741 TTACGACTTC ATTCTAGTTA CCAATTTTGT CTTCTGTCATT TATTAATGA TTCTAGACAT
19801 AACCAATTTT CAGACTGTGA CGGGCAGTGT GTACAAGGTC CAAGAACTTA TTCACCGTA
19861 TATTGCTGAT TAGCGATTAC TAGCGATTTC GACTTCATGA GTGCGAGTTG CAGCCCTCAA
19921 TCCGAAGTGA GATTGGGTTT TTTGATTAGC TCTCCTTTAC AGGATTGCAA CATTTTGTCC
19981 CAACCATTTG AGCACGTGTG TAGCCCAGGA TTTAAAGAGC ATGATGACTA GACGTCGTCC
20041 CTTCCCTTCT CTAGCTTACA CCAGCTGTAT TCTTAAAGAT AAATTTAAT TTAATATATA
20101 TTATTATAGG GATAGGGACA AATCCATAT ATATATTATT TATAAAAATA GATAAAGTA
20161 TAGGGGTTAC GCTCGTTAGG GGACTTAACC TAATACCCTA AAGCACGAGC TGATGACAGC
20221 CATGCACCAC CTGTATCTTT ATACTGGATA GTATAAATTT TATTTTCAA TAAAAATTTT
20281 GACATGTCAA ATCCTGGTAA GGTTCATCGC TTTGCATCGA ATTAACCAC ATGCTCCACC
20341 GCTTGTGTGG ACCCCCGTCA ATTCCTTTGA GTTTCATTCT TGCGAACGTA CTTCCCAGGC
20401 GGGATACTTT ACGCGTTAGC TTTAGTATAA CATAGGCATA TTTTTTTTAT TATTTATGCT
20461 TAATATGCTT AGTATCCATT GTTTACAGT TATGACTATC TATGACTATC ATCCCACCTG
20521 CTACCATAGC TTTCTGTTACT CAGTGTGCTG ATTAATCTAG CAAGGTGCTT TCGCCATTGG
20581 TGTTCTTTCC AATATCAACG TATTTACCA CTCCATTGAA AATTCCTT ACTTTTAGTA
20641 TACTCAAGTT TTTTAGTTTC TAATGCAGGT CAGAGTTGAG CTCTAATATT TTACAATAGA
20701 CTTAAAAAAC GACCTACGAA CCCTTTACGC CCAATCATTC TGGATAATGC TTGCATCCTC
20761 TGTATTACCG CGGCGGCTGG CACAGAGTTG GCCGATGCTT ATTCCTGGAA TAACGTCATT
20821 TTTTCTTCTA CCATAAAAGA AGTTTACAAC CTTTTCAGGC CTTAATCCTC CACGCGACAT
20881 TGCTCCGTC GGCCTTCGCC CATTGCGGAA TATTCCTCAC TGCTGCCTCC CGTAAGAGTC
20941 TGGGCCGTGT CTCAGTCCCA GCGTACTGTA TCATCCTCTC AGACCAGTTA CTGATTATAG
21001 CCTTGGAGG CTGTTACCCT ACAAAGTAGC TAATCAGACA TAAATCCAGC CTAGAGCGAT
21061 TTTAATAAAT AAATATTAAT CTTTTACTCC TCAGTTTATG GAGAATTAGC AACTGTTTCC
21121 AATTTGTTAT ACTCCTCTCT AGGATAGGTT TTAATGTATT ACTCACTCGT TCGCTACTAT
21181 AAAAAATCGT TTCAACTTGC ATGTGTTAAG CATGTCGCTA GCATTCATCC TGAGCCAGGA
21241 TCGAATCTG TATATTTTTT AATTATTAG CAAGTTATT TACTCATAGC TTCACTTTTA
21301 TTTATATTA AATATTATAT TTTTATATAT AATTATATAA TTATTTAAT CTTTTTTTTT
21361 TAATTCCTTA TATTTATAAT TTTTTCTTT AACTTTTTTA CGGTACTTTT TTTTATATAC
21421 TTTTTTATTT TTATATAAAT TTTTACGGTA TTTAGTTTTA TATACTTTTT TAAAAATTTT
21481 ATTTTTATCT ATAGCTTTTT TAGTAGTTTC CTCTTTTATA TTTATTGCAT TTCTTTAAA
21541 AGCATCTAAT AATTCAAAAC TTAAGTTAGA AACCATGTTA TATCCAAAAC GTTTTTTAGA
21601 TGCTGATATT AATAATTTGC GAGCATACAT ATATTTTTTA TCTTTAGGTA AATTATAATA
21661 TATTTTACGT ATTGATGCTT TTCCTTTTTT ACGTTTTTTT TTTAATATT TTTTATTAC
21721 TATTTTTTTT TTAGTTATTT TATTTAGTGC TTTTGTACA ACATAAATG GATTTAATTT
21781 TATTTTTTTT TTTATTTCTA GTAAAGATTT ATAAATTATT TTATATGATA TTGATTTTTT
21841 TCCTTTTCTT AAAATGTAAT TAGATAGATT ATAAATGAT TTTTTTAAA AAATTGATC

21901 ATATTTTACT TTATAAGTCA TTTTATTTT TTAATTTTTA ATTTTTTGTT TTTTAGCTCC
 21961 ATATTTTGAA CGTCTTTGAA ATCGTCTCTT TGCTCTACCA GCATCTTTAC TGCCTAAAAA
 22021 AACTTTATGT TTTACACCAG GTAAATCTTT TACTAAGCCA CCTCTTACTA AACTCTTGA
 22081 ATGTACTTGA ACTAAATTA TACCAATACC TGGTATTAAT GCAGTTACTT CAATTCCAGA
 22141 AGTCAATCGT ATTCGGGCAA CTTTTCTTGA GCCAGAATTA GGTTTTTTTG GATGAATAAT
 22201 GTAGTTAACA GATAAAAAAC TGTCACTTTA CAGAACCTAT ACATGAAAGT TTCCTTTCAT
 22261 ATAGCTCCTC ATTTAAATAT TTATTAACC TTTAAATTTG TTTTATATA AAATTATATA
 22321 AAAAGCATAA ATTATATATG TAAAATATTT ATATTTATTA TATTTATATT TATTTTTTTT
 22381 TATTATAAAT TATATAATTT ATTTATATTT ATAATATTTA TTTTATATAT ATTTTATAT
 22441 TAATTTTAA TTAATTAAT AGTTAATAA TTTAAACGGG TTAATGTAAG TAATTTAAAT
 22501 TTATATCTTA ATATTTTTT ATTATTTAAT TTTATTTTTT AAAATTAAT ATTTTATATT
 22561 TTATTTGCTT TAATACTTAA TTTTATATAT AATTATTTTT TTAGTTTATA TTATTATTTA
 22621 ATTTTAAATT AAAATATTA GATTAATAAT GAATATAAAT ATAAAAAAT ATATAACAAT
 22681 ATAACAAATA ATATTATAAT TATAAATTA AAAATTTAAA TTTAATAAA TTTAAATAAG
 22741 GTCTTGAAAA TTATTTGGCC ATAAAAATTC TTCTTTAATT AATTTTTTAA AAAGTTTATT
 22801 AATAAGTTTA CTATTATTTA ATAATATTTT TGTTAAATA TCAATTGCTT CATAATGAAT
 22861 TATAATTGAA ATATTATGTT TATCTTTATA TTTATAAATA GATTTGTAAG AATCAGATTT
 22921 ATTATTAAC ATAAAAATTT TATATAGTAG AGGTTTTAAT GTATTATATA CAAAAATCT
 22981 TAGTTGGTCA GAGAATCTAT AATCTTCATA CATAGCATA ATAAATTTTT CATAAGGAAT
 23041 ATAAATATCT TTTTTTTTTT TTTTATTTGA TGGGTTATTA ATGGGTAAG TATCAGTTTC
 23101 GTAATTATAA GCATTTATAA CAGAAAAATC TAATTCATTA ATATTTGATT TAATTTAAAT
 23161 ATATTTATTA ATATTTAATA TTTCTTTGTC ATTTTTTATT AACTTTTCTA TAAAATTTAT
 23221 TAGTCTAAAA TTGTCAAAA AAAAACTTT TTCTAAATTT AATAAATTA TAATTAACTT
 23281 TAAATCATGA TTAAAATATT CATGTTTTAT ATTTTTAATC TCTGATGTGT TAAAAATTTT
 23341 ATAAGCAATT AATCCAGATA AACATATCAA TAAATAAATA ATTACAGTTA TTATTTTAT
 23401 ATTTGTTCCG AGTTCATAAT AATTTTTATA TAAAGGATAT TTTTTGTTT TGAAATCTAT
 23461 TAAATATACA GACATAGGAT CAACTATTA ACTATTTAAT CTAATTTGTA AAAAAAACC
 23521 TCCAATTTGA TAAAAACAA TTATAATAGG AACTTTGTAT CCTTTTTTTA TGTATTGATT
 23581 AAATCTAAAA AAATAAAAA TAAATTTATT AATTATTTTT GAATTTATTA CATCTTGATT
 23641 TTGTGATAAGA CTTAGTGATA AAGTTTGATT TGTCATATA GATAAATCTC TTAGATTTT
 23701 TATATTTAAA TGTTTCATGTA ATAATCTTT TTTTCAATA TAAAACTTT TTGTTTTTAA
 23761 TAAATTCAAA ATATATTTTT TTTTATTTT TATAGTAGGT TTTTAAATA AAATAAATTT
 23821 ACTTAATTTA TCATAACCTA TTAATATAGG TTCTACTTCT TTTGGTTTAT TTGTATAAGC
 23881 TAAATTTAAA TTATTTTTT TAAAATTAGA AATTAATTA CTTAATTCAT AAAATGTTT
 23941 AAAATTTAAA TTATAATGAC ATAAATGAAT ATTTGGAATA ATTATAATAC ATGAAGAAAT
 24001 ATATTTCTAT AGTTTAAAT GAGTTACAAT ATTAATACA TCATATATAC TTTTTTTCT
 24061 TTTTGGCATT AGTTTATCTA ATATTTTTT TTTTTCGAAT TCAAATGTTT CTTTATCAGT
 24121 ATTATAAATT TTTTAAACA TTTTTTTAAT TTGTTTAGTA GAAGTATTAC TTTTATTAAT
 24181 ATCAAATTTA TCAACAAAA GAAATCAGA TTCATCATA CTATATTTG AAATTCATA
 24241 GTCACTTAAT TCTTTATTAA TTGGGTTTAA TATAATAAAA GGCAAATCAG CATAATCTTT
 24301 TATAATTTTT TCTTGTA AAAAACGTC TTCAGACTGA CCAACTAATA ATATACCTTT
 24361 TGAAGCTGAC GGAATCATTG TTAATGAAAA AAAATAAGGT TTTTGTCTTA TTTTATATCT
 24421 GTTTTTTTTT AAAATAAAT ATTCTGTTAT TTTTATAA TTTGTAATA AAATACGATT
 24481 TTTTCTTTT ATATTTAATA TATTATTTG TTTGTATTA ATTCTTAACA TATCTAAAAA
 24541 ATATATTA TAATAAATAC CATATTCAC AGAGTTTCA TACTTTTTTA TAGATATATT
 24601 TTTAAAAACA TATTTAATA GTTTTTGTT ATTTTTTATA TTTAAAAAA TATTTATAA
 24661 TTGAATTA AATTTCTTT CTTCATCAA AATTAATCA TTAGTATTA TCCAAGATTC
 24721 TAAATCAAT TGTGTTTCAT TTTTTTTTAA TATAGTTTGA GTATATGTAG TTTCTTTTAT
 24781 AGATGTATAT TCTTGATCTT TATTGTTAAG AATACATTGA TAAAACTCT CAGTTATATC
 24841 TTGTGAAAAA CATTTAAAA ATTCTTTACT TATGGAGTTT CTTAAATAA AAAATCATA
 24901 AATAAAATAT TTATACCAAT TTTTAAATTT TGATTCAGTA TAAAAATA AATATCTTT
 24961 ATATTTAATT TTTATATTTA TATATATATT TTTTAAATTT ACTTTATGAT TATAAGTAGT
 25021 TTTTACTTTG TCTATTTCTA AAAATTTAAA TAAAATGTCA ACAATAAAA ATCGAATAAT
 25081 ATATTTATTT ATTTTCAGTTA GTAAATTA AATTTCAATTT ATAAAAATA AAAAAATATT
 25141 GTTATAAAGA TAAATTGTGT TTTTCATGCA TGCTTCTAAA TAACTATATT TACATAACTT
 25201 TACATATAGA TCTATACTAA ATTTTTTTGT CATTAATAAT AAATTTATT CATATTCATT
 25261 TTTAATAAAT TTAATATATT TAAAAATTA TTTAAAAAT TTTAAAAATA AAAATATAA
 25321 AATATAAATA AAAAAATTT TAAATAATA ATTATATTT ATAAAAATTA ATTGGTCATA
 25381 TCTAAAAATT TCAATAATAT CTTTTTTAAT TGGGTGTAAT TTAACCTGAT TTTTATAAAA
 25441 TGCAAGTTGA TTTACTAATA AATTATAATT ATTTTTATTA GTTTTTGATA ATAAAAAAT

25501 TATTTGAATA ATTTTAATAA AATTTAACTC ATTTAACTCA TTTAAAAAAT AGTCAATTTG
 25561 ATCACCACATA TATTTAAAAA AAATAAATTT TATTATATTT ATAATTTTAT ATTTTAAATA
 25621 TTTTTTAATT TCAAAAATTT TTTTAAATAT ATAAATATAA ATAATTTTAT TTATAATAA
 25681 TTTTTCAATA TTTTTAAAAA TATTTTTTAT ATTATTATAA TAATTATAAT ATTTTTTTGC
 25741 TTTAAAAAAA AAAAGAAAAA GTATATATAA TCCAATAAAC TTTTAAATTT ATTTATATCT
 25801 TTACTTATAT TTAAATCTTT ACTTATATTA ACTATATAAG AATTTATAT TAATAAAAAA
 25861 ATATTTCTAA ATTTATAAAA AATATTATTA TTAAAAAAT TATTATAATA TTTATTTTGA
 25921 ATTAAAAATT TTTTTTTTTT TTCTAAATAT TTATTATATT TATTTTTAAT TTTGTTTTTT
 25981 TTATATTTAT AATTTATATT ATTATTATTA TTTTATAAAA TATATTTTGT TTCTCTGAA
 26041 ATTAATAAAA AATTAATAA AATATCTTTA AGAGAATTTT TGTATTCATC AAAAATATAT
 26101 TCAATTTTTA AATTATCTTT ATTAATATTA ATAAAACTAA AATCTTCTGT AATTTTGT
 26161 ATAGACATAG TTATTCCTAT TATATATATA TATTTATCAT AAAATGTAAT TTTTGATATA
 26221 TTTTTTTTTG ATAAAAAAC TTTATATTTT AAGTTTTTTT TTTTCTTAA AATAATTTTA
 26281 TTTTCACTTA TATTTAATGG AATTAAGAG GTAAAAATCAC CTATAGTAAA ATTAAAAAAT
 26341 AAATATTTAT TAAAAAATGA AACAAAAAAA ATAGAAAAATA TTATAATCTT AATTATATTA
 26401 TTTAAAAATA ATAAAAAAA CTTTTTTAT TTTAAAAATA AATATTTTTT AAATTTATTT
 26461 ATTAGGTCAT TAGATTCAGT ATAAAAAAG AACTAAATTT TGTAGAAATT AAAGGTTTCT
 26521 TTACTGTATA TAGTATTTTT ATTAATATTA TAAACCATAC TATCAAATTT ATAAAAAAGT
 26581 TTATATTTTT TTAAATTTTT TTTTGTTTTT TGTAATTTAT TATTATTATA ATTATATATT
 26641 TTTATAATAT TTACCATATT TTATTTTTTT ATCTTTTTTT TTTATTAAAA TATTATTAGA
 26701 ATAATATTAT TATTTTTTAT TTTTTTATTT TAGCTTAATC CCATACAGCT AATTTTAAAA
 26761 TAATTTATTT ATAAATTAAT TTTTAAAAAA TTAATACAAT ATATAAGTAT CCATGGCTGA
 26821 TTGGTGAAAG CACCCAATC ATAATTGGAG ATATGTAGGT TCAATTCCTG CTGGATACAA
 26881 AAAAAAGATT TATAATCCTT AATAAAAAAA AATGACATTA AATAACATAT ATAAAAAAA
 26941 AGTTAAAAGA AAATTATTA TTTATGGATA TAATTTTAAAT AAAGGTAGAA ATTATAAAGG
 27001 TACAATAACT GTAAAACATA GAGGAGGGGG TCATAAGCGT TTATATAGAA AGATAGATTT
 27061 TAAACGTAAT GAAAAAATA CTGGAAAAAT ATTAAAAAAT GAATATGATC CCAATCGGAA
 27121 TACAACAATT TGTTTTAGTTT ATTATAAAAT AAATAATATA TACAAATATA TAATACATCC
 27181 ACAAAATGTT GGTATTGGAG ATACTATTTT ATCTAGTCAT GAAACATCTA TTTATGCTGG
 27241 AAATACATTA CCTTAAATTA ATATACCTTT TGGTACTTTT ATACATAACA TAGAAATTA
 27301 ACCTGGAAAG GGCGGACAAAT TAGTTAGGCG TGCAGGTAAT GCAGCAAAAG TAATTTACAA
 27361 AAAAAAATA AGAGTTTTAT TAAAGTTACC ATCAGGAAAA TTATGTTCAA TACCTTAAAC
 27421 ATGCTTAGCA ACTATAGGAC AAGTTAATAA AACTGCATTA AAAAAAAT TAAAAGCTGG
 27481 GACAAATCGT TGGTTAGGAA AAAGACCTAA AGTACGAGGT ACTGCTATGA ACCCTGTAGA
 27541 TCATCCTCAC GGAGGAGGAG AAGGTAAAGC TCCTATAGGT AGAAAATTTT CTGCAACTCC
 27601 TTGGGGATAT CCTGCTTTTG GACGAAGAAA AAATAAAAAA TAAATGAGA TTA AAAAGAA
 27661 AACATAAACA TATTTAAAAA AAATATAAAG AATATAATTA AATTTCTTTT AATTTTTTTG
 27721 ATAAAATCCT GTTAAAATAG CATAAGATAA CTTATTTAAA ATATAAAATA TTGAAGACAA
 27781 ATTAATGTA TTTATAGGTA TAACAATATC TGCTAAATCT ATATTAATAT CAGTATTTAT
 27841 ATTAATTAATA CAAATTGTTG AAATTTTTTA AGTAAAACAT TCTCTGATAA TTATATCATT
 27901 TTGTGGATCA ATAAGTATTA CAATATCAGG TCTCTTAGTC ATATATTTTA TTCCTTCTAA
 27961 ATTATATAT AAATAATTT TTTGTTTTTT TTTTTTTATA AAATTAATTT TTGTTTTACT
 28021 AATTAATAA TTTTTTAATT TAATTTTTGT AGTTAACAG TTAGTAAGTA AACCTGCTCG
 28081 TATATTTCTA TTAATATAAT GACATCTAAT TCTTTCAGAT ATTAATTTAA AATAATTTAA
 28141 TTTAAGGTT TTATATTTAT TATTTATTCG TTTATTTTTT CTTTTTAAAT TCAATTTAGT
 28201 TTTATTTTTA TTCCTTAATT TTTTATTAT TATTAAAAAC CTTTTAGTTA AAAAATCTCT
 28261 AGTTTTAATT ATTATAGGTT TTATTTCTTT TTTATTTTTG TTTTGACTAA TTATTA AAAA
 28321 TGTTTTTTTT TTTTTTGCTG CATAAAAAAC TAAATTACAA GCTATTGTTA ATAAAGAAAGC
 28381 TATTTTTGTA ATATCTATAA TATAATATTT TAAATTAATA TTA AAAATAT AATATATCAT
 28441 TTTAAATTTA ATATTTTTTT TATTTAAAAA ATGAATTTTT TCTGTTAACA TATCATTTAA
 28501 TTTTATTTTC CAATCTATTT TTTTTTCTAT TTTAACATTT TTAATTTTAA CTTTTTTCAT
 28561 TTTAACATTT TAAAATACAA TCCATTTATA CTATCTTTTA TATAAATGGA TTGTGGATCG
 28621 AGTTGGATTT GAACCAACGT AGACACATAA GTCAAAGGAT TTACAGTCCA TTCCTTTTAA
 28681 CCTCTCGGGC ATCGATCCAA AATAA

REFERENCIAS

- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410.
[https://doi.org/https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/https://doi.org/10.1016/S0022-2836(05)80360-2)
- Arias-Agudelo, L. M., González, F., Isaza, J. P., Alzate, J. F., & Pabón-Mora, N. (2019). Plastome reduction and gene content in New World Pilostyles (Apodanthaceae) unveils high similarities to African and Australian congeners. *Molecular Phylogenetics and Evolution*, *135*, 193–202. <https://doi.org/https://doi.org/10.1016/j.ympev.2019.03.014>
- Bacon, C. D., Silvestro, D., Jaramillo, C., Smith, B. T., Chakrabarty, P., & Antonelli, A. (2015). Biological evidence supports an early and complex emergence of the Isthmus of Panama. *Proceedings of the National Academy of Sciences*, *112*(19), 6110–6115.
<https://doi.org/10.1073/pnas.1423853112>
- Bakker, F. T., Lei, D., Yu, J., Mohammadin, S., Wei, Z., van de Kerke, S., Gravendeel, B., Nieuwenhuis, M., Staats, M., Alquezar-Planas, D. E., & Holmer, R. (2016). Herbarium genomics: plastome sequence assembly from a range of herbarium specimens using an Iterative Organelle Genome Assembly pipeline. *Biological Journal of the Linnean Society*, *117*(1), 33–43. <https://doi.org/https://doi.org/10.1111/bij.12642>
- Banerjee, A., Schneider, A. C., & Stefanović, S. (2022). Plastid genomes of the hemiparasitic genus *Krameria* (Zygophyllales) are intact and exhibit little relaxation in selection. *International Journal of Plant Sciences*, *183*(5), 393–403.
- Banerjee, A., & Stefanović, S. (2019). Caught in action: fine-scale plastome evolution in the parasitic plants of *Cuscuta* section *Ceratophorae* (Convolvulaceae). *Plant Molecular Biology*, *100*(6), 621–634. <https://doi.org/10.1007/s11103-019-00884-0>
- Banerjee, A., & Stefanović, S. (2020). Reconstructing plastome evolution across the phylogenetic backbone of the parasitic plant genus *Cuscuta* (Convolvulaceae). *Botanical Journal of the Linnean Society*. <https://doi.org/10.1093/botlinnean/boaa056>
- Barbrook, A. C., Howe, C. J., & Purton, S. (2006). Why are plastid genomes retained in non-photosynthetic organisms? *Trends in Plant Science*, *11*(2), 101–108.
<https://doi.org/https://doi.org/10.1016/j.tplants.2005.12.004>

- Barnard-Kubow, K. B., Sloan, D. B., & Galloway, L. F. (2014). Correlation between sequence divergence and polymorphism reveals similar evolutionary mechanisms acting across multiple timescales in a rapidly evolving plastid genome. *BMC Evolutionary Biology*, *14*(1), 268. <https://doi.org/10.1186/s12862-014-0268-y>
- Barrett, C. F., & Davis, J. I. (2012). The plastid genome of the mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *American Journal of Botany*, *99*(9), 1513–1523. <https://doi.org/10.3732/ajb.1200256>
- Barrett, C. F., Freudenstein, J. V., Li, J., Mayfield-Jones, D. R., Perez, L., Pires, J. C., & Santos, C. (2014). Investigating the Path of Plastid Genome Degradation in an Early-Transitional Clade of Heterotrophic Orchids, and Implications for Heterotrophic Angiosperms. *Molecular Biology and Evolution*, *31*(12), 3095–3112. <https://doi.org/10.1093/molbev/msu252>
- Barrett, C. F., Sinn, B. T., & Kennedy, A. H. (2019). Unprecedented Parallel Photosynthetic Losses in a Heterotrophic Orchid Genus. *Molecular Biology and Evolution*, *36*(9), 1884–1901. <https://doi.org/10.1093/molbev/msz111>
- Bellot, S., Cusimano, N., Luo, S., Sun, G., Zarre, S., Gröger, A., Tensch, E., & Renner, S. S. (2016). Assembled Plastid and Mitochondrial Genomes, as well as Nuclear Genes, Place the Parasite Family Cynomoriaceae in the Saxifragales. *Genome Biology and Evolution*, *8*(7), 2214–2230. <https://doi.org/10.1093/gbe/evw147>
- Bellot, S., & Renner, S. S. (2016). The plastomes of two species in the endoparasite genus *pilostyles* (apodanthaceae) each retain just five or six possibly functional genes. *Genome Biology and Evolution*, *8*(1), 189–201. <https://doi.org/10.1093/gbe/evv251>
- Bendiksby, M., Schumacher, T., Gussarova, G., Nais, J., Mat-Salleh, K., Sofiyanti, N., Madulid, D., Smith, S. A., & Barkman, T. (2010). Elucidating the evolutionary history of the Southeast Asian, holoparasitic, giant-flowered Rafflesiaceae: Pliocene vicariance, morphological convergence and character displacement. *Molecular Phylogenetics and Evolution*, *57*(2), 620–633. <https://doi.org/https://doi.org/10.1016/j.ympev.2010.08.005>
- Benore-Parsons, M., & Ayoub, M. A. (1997). Presence of RNase A causes aberrant DNA band shifts. *BioTechniques*, *23*(1), 128–131. <https://doi.org/10.2144/97231st04>

- Bolin, J. F. (2009). *Ecology and Molecular Phylogenetics of Hydnora (Hydnoraceae) in Southern Africa* by [Old Dominion University]. <https://doi.org/10.25777/e44d-9n09>
- Burger, G., Lavrov, D. V., Forget, L., & Lang, B. F. (2007). Sequencing complete mitochondrial and plastid genomes. *Nature Protocols*, 2(3), 603–614. <https://doi.org/10.1038/nprot.2007.59>
- Bustin, S. A., Mueller, R., & Nolan, T. (2020). *Parameters for Successful PCR Primer Design BT - Quantitative Real-Time PCR: Methods and Protocols* (R. Biassoni & A. Raso (eds.); pp. 5–22). Springer New York. https://doi.org/10.1007/978-1-4939-9833-3_2
- Cai, L., Arnold, B. J., Xi, Z., Khost, D. E., Patel, N., Hartmann, C. B., Manickam, S., Sasirat, S., Nikolov, L. A., Mathews, S., Sackton, T. B., & Davis, C. C. (2021). Deeply Altered Genome Architecture in the Endoparasitic Flowering Plant *Sapria himalayana* Griff. (Rafflesiaceae). *Current Biology*, 31(5), 1002-1011.e9. <https://doi.org/10.1016/j.cub.2020.12.045>
- Chan, P. P., Lin, B. Y., Mak, A. J., & Lowe, T. M. (2019). tRNAscan-SE 2.0: Improved Detection and Functional Classification of Transfer RNA Genes. *BioRxiv*, 614032. <https://doi.org/10.1101/614032>
- Choi, I.-S., Jansen, R., & 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. <https://doi.org/10.1093/gbe/evz076>
- Coates, G. ., & Obando, J. A. (1996). The geological evolution of the Central American Isthmus. *Evolution and Environment in Tropical America*, 21–56. <https://cir.nii.ac.jp/crid/1570572700876583168.bib?lang=en>
- Cocucci, A. E., & Cocucci, A. A. (1996). Prosopanche (Hydnoraceae): somatic and reproductive structures, biology, systematics, phylogeny and potentialities as a parasitic weed. *Congresos y Jornadas-Junta de Andalucia (Espana)*, 36.
- Crossley, B. M., Bai, J., Glaser, A., Maes, R., Porter, E., Killian, M. L., Clement, T., & Toohey-Kurth, K. (2020). Guidelines for Sanger sequencing and molecular assay monitoring. *Journal of Veterinary Diagnostic Investigation*, 32(6), 767–775. <https://doi.org/10.1177/1040638720905833>

- Cusimano, N., & Wicke, S. (2016). Massive intracellular gene transfer during plastid genome reduction in nongreen Orobanchaceae. *New Phytologist*, *210*(2), 680–693.
<https://doi.org/https://doi.org/10.1111/nph.13784>
- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2). <https://doi.org/10.1093/gigascience/giab008>
- Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*, *9*(8), 772.
<https://doi.org/10.1038/nmeth.2109>
- Davis, C. C., & Xi, Z. (2015). Horizontal gene transfer in parasitic plants. *Current Opinion in Plant Biology*, *26*, 14–19. <https://doi.org/https://doi.org/10.1016/j.pbi.2015.05.008>
- de Carvalho, J., Markus, C., Merotto, A., Záchia, R., Schuch, M., Sulzbach, E., & Mariath, J. (2021). *Prosopanche cocuccii* (Hydnoraceae): a new species from Southern Brazil. *Phytotaxa*, *521*, 177–192. <https://doi.org/10.11646/phytotaxa.521.3.3>
- Dexter, K. G., Lavin, M., Torke, B. M., Twyford, A. D., Kursar, T. A., Coley, P. D., Drake, C., Hollands, R., & Pennington, R. T. (2017). Dispersal assembly of rain forest tree communities across the Amazon basin. *Proceedings of the National Academy of Sciences*, *114*(10), 2645–2650. <https://doi.org/10.1073/pnas.1613655114>
- Dorrell, R. G., Azuma, T., Nomura, M., Audren de Kerdel, G., Paoli, L., Yang, S., Bowler, C., Ishii, K., Miyashita, H., Gile, G. H., & Kamikawa, R. (2019). Principles of plastid reductive evolution illuminated by nonphotosynthetic chrysophytes. *Proceedings of the National Academy of Sciences*, *116*(14), 6914–6923.
<https://doi.org/10.1073/pnas.1819976116>
- Doyle, JJ; Doyle, J. (1990). Isolation of plant DNA from fresh tissue. *Focus*, *12*(1), 13–15.
- Erkens, R. H. J., Chatrou, L. W., Maas, J. W., van der Niet, T., & Savolainen, V. (2007). A rapid diversification of rainforest trees (Guttaria; Annonaceae) following dispersal from Central into South America. *Molecular Phylogenetics and Evolution*, *44*(1), 399–411.
<https://doi.org/https://doi.org/10.1016/j.ympev.2007.02.017>
- Fan, W., Zhu, A., Kozaczek, M., Shah, N., Pabón-Mora, N., González, F., & Mower, J. P.

- (2016). Limited mitogenomic degradation in response to a parasitic lifestyle in Orobanchaceae. *Scientific Reports*, 6(1), 36285. <https://doi.org/10.1038/srep36285>
- Fernández Robledo, J. A., Caler, E., Matsuzaki, M., Keeling, P. J., Shanmugam, D., Roos, D. S., & Vasta, G. R. (2011). The search for the missing link: a relic plastid in *Perkinsus*? *International Journal for Parasitology*, 41(12), 1217–1229. <https://doi.org/10.1016/j.ijpara.2011.07.008>
- Ferreira-Machado, R., & Paganucci-De queiroz, L. (2012). A new species of *Prosopanche* (Hydnoraceae) from northeastern Brazil. *Phytotaxa*, 75, 58–64. <https://doi.org/10.11646/phytotaxa.75.1.5>
- Funez, L. A., Ribeiro-Nardes, W., Kossmann, T., Peroni, N., & Drechsler-Santos, E. R. (2019). *Prosopanche demogorgoni*: A new species of *Prosopanche* (Aristolochiaceae: Hydnoroideae) from southern Brazil. *Phytotaxa*, 422(1), 93–100. <https://doi.org/10.11646/phytotaxa.422.1.6>
- Funk, H. T., Berg, S., Krupinska, K., Maier, U. G., & Krause, K. (2007). Complete DNA sequences of the plastid genomes of two parasitic flowering plant species, *Cuscuta reflexa* and *Cuscuta gronovii*. *BMC Plant Biology*, 7(1), 45. <https://doi.org/10.1186/1471-2229-7-45>
- Gitzendanner, M. A., Soltis, P. S., Wong, G. K.-S., Ruhfel, B. R., & Soltis, D. E. (2018). Plastid phylogenomic analysis of green plants: A billion years of evolutionary history. *American Journal of Botany*, 105(3), 291–301. <https://doi.org/https://doi.org/10.1002/ajb2.1048>
- Gómez-Pignataro, L. D., & Gómez-Laurito, J. (1981). A new species of *Prosopanche* (Hydnoraceae) from Costa Rica. Una especie nueva de *Prosopanche* (Hydnoraceae) de Costa Rica. *Phytologia.*, 49(1), 53–55.
- Gonçalves, D. J. P., Simpson, B. B., Ortiz, E. M., Shimizu, G. H., & Jansen, R. K. (2019). Incongruence between gene trees and species trees and phylogenetic signal variation in plastid genes. *Molecular Phylogenetics and Evolution*, 138, 219–232. <https://doi.org/https://doi.org/10.1016/j.ympev.2019.05.022>
- Greiner, S., Lehwark, P., & Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic*

- Acids Research*, 47(W1), W59–W64. <https://doi.org/10.1093/nar/gkz238>
- Group, T. A. P., Chase, M. W., Christenhusz, M. J. M., Fay, M. F., Byng, J. W., Judd, W. S., Soltis, D. E., Mabberley, D. J., Sennikov, A. N., Soltis, P. S., & Stevens, P. F. (2016). An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG IV. *Botanical Journal of the Linnean Society*, 181(1), 1–20. <https://doi.org/10.1111/boj.12385>
- Gruzdev, E. V., Kadnikov, V. V., Beletsky, A. V., Mardanov, A. V., & Ravin, N. V. (2019). Extensive plastome reduction and loss of photosynthesis genes in *Diphelypaea coccinea*, a holoparasitic plant of the family Orobanchaceae. *PeerJ*, 7, e7830. <https://doi.org/10.7717/peerj.7830>
- Guindon, S., & Gascuel, O. (2003). A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies by Maximum Likelihood. *Systematic Biology*, 52(5), 696–704. <https://doi.org/10.1080/10635150390235520>
- Guisinger, M. M., Chumley, T. W., Kuehl, J. V., Boore, J. L., & Jansen, R. K. (2010). Implications of the plastid genome sequence of typha (typhaceae, poales) for understanding genome evolution in poaceae. *Journal of Molecular Evolution*, 70(2), 149–166. <https://doi.org/10.1007/s00239-009-9317-3>
- Habibi, N., Al Salameen, F., Rahman, M., Shajan, A., Zakir, F., & Abdulrazzack, N. (2022). Comparison and optimization of DNA Isolation protocols for high throughput genomic studies of *Acacia pachyceras* Schwartz. *MethodsX*, 9, 101799. <https://doi.org/10.1016/j.mex.2022.101799>
- Hasenstab-Lehman, K. E. (2017). Phylogenetics of the Borage Family: Delimiting Boraginales and Assessing Closest Relatives. *Aliso: A Journal of Systematic and Evolutionary Botany*, 35(1), 41–49.
- Hatt, S. A., Cameron, D. D., Grace, O. M., Rocamundi, N., Cocucci, A. A., Martel, C., & Thorogood, C. J. (2023). Prosopanche: A remarkable genus of parasitic plants. *PLANTS, PEOPLE, PLANET*, 5(2), 163–168. <https://doi.org/https://doi.org/10.1002/ppp3.10340>
- Hedenäs, L., Larsson, P., Cronholm, B., & Bisang, I. (2021). Evidence of horizontal gene transfer between land plant plastids has surprising conservation implications. *Annals of*

- Botany*, 127(7), 903–908. <https://doi.org/10.1093/aob/mcab021>
- Hishamuddin, M. S., Lee, S. Y., Syazwan, S. A., Ramlee, S. I., Lamasudin, D. U., & Mohamed, R. (2023). Highly divergent regions in the complete plastome sequences of *Aquilaria* are suitable for DNA barcoding applications including identifying species origin of agarwood products. *3 Biotech*, 13(3), 78. <https://doi.org/10.1007/s13205-023-03479-1>
- Jansen, R. K., Cai, Z., Raubeson, L. A., Daniell, H., dePamphilis, C. W., Leebens-Mack, J., Müller, K. F., Guisinger-Bellian, M., Haberle, R. C., Hansen, A. K., Chumley, T. W., Lee, S.-B., Peery, R., McNeal, J. R., Kuehl, J. V., & Boore, J. L. (2007). Analysis of 81 genes from 64 plastid genomes resolves relationships in angiosperms and identifies genome-scale evolutionary patterns. *Proceedings of the National Academy of Sciences*, 104(49), 19369 LP – 19374. <https://doi.org/10.1073/pnas.0709121104>
- Jin, J. J., Yu, W. B., Yang, J. B., Song, Y., DePamphilis, C. W., Yi, T. S., & GetOrganelle, D. Z. L. (2021). A fast and versatile toolkit for accurate de novo assembly of organelle genomes. *DOI: <https://doi.org/10.1186/S13059-020-02154-5>*, 241.
- Jost, M., Naumann, J., Bolin, J. F., Martel, C., Rocamundi, N., Cocucci, A. A., Lupton, D., Neinhuis, C., & Wanke, S. (2022). Structural Plastome Evolution in Holoparasitic Hydnoraceae with Special Focus on Inverted and Direct Repeats. *Genome Biology and Evolution*, 14(6), evac077. <https://doi.org/10.1093/gbe/evac077>
- Jost, M., Naumann, J., Rocamundi, N., Cocucci, A. A., & Wanke, S. (2020). The first plastid genome of the holoparasitic genus *Prosopanche* (Hydnoraceae). *Plants*, 9(3), 7–15. <https://doi.org/10.3390/plants9030306>
- Jost, M., Samain, M.-S., Marques, I., Graham, S. W., & Wanke, S. (2021). Discordant Phylogenomic Placement of Hydnoraceae and Lactoridaceae Within Piperales Using Data From All Three Genomes. *Frontiers in Plant Science*, 12, 642598. <https://doi.org/10.3389/fpls.2021.642598>
- Kapli, P., Yang, Z., & Telford, M. J. (2020). Phylogenetic tree building in the genomic age. *Nature Reviews Genetics*, 21(7), 428–444. <https://doi.org/10.1038/s41576-020-0233-0>
- Katoh, K., Misawa, K., Kuma, K., & Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research*,

- 30(14), 3059–3066. <https://doi.org/10.1093/nar/gkf436>
- Katoh, K., Rozewicki, J., & Yamada, K. D. (2019). MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, 20(4), 1160–1166.
- Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4), 772–780. <https://doi.org/10.1093/molbev/mst010>
- Kim, W., Lautenschläger, T., Bolin, J. F., Rees, M., Nzuzi, A., Zhou, R., Wanke, S., & Jost, M. (2023). Extreme plastomes in holoparasitic Balanophoraceae are not the norm. *BMC Genomics*, 24(1), 330. <https://doi.org/10.1186/s12864-023-09422-1>
- Knapp, S., & Peña-Chocarro, M. (2015). *Tropicos* | Name - *Prosopanche costaricensis* L.D. Gómez & Gómez-Laur. <http://legacy.tropicos.org/Name/15900004?projectid=3>
- Kozlov, A. M., Darriba, D., Flouri, T., Morel, B., & Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics*, 35(21), 4453–4455. <https://doi.org/10.1093/bioinformatics/btz305>
- Krause, K. (2008). From chloroplasts to “cryptic” plastids: evolution of plastid genomes in parasitic plants. *Current Genetics*, 54(3), 111. <https://doi.org/10.1007/s00294-008-0208-8>
- Krause, K. (2012). Plastid genomes of parasitic plants: a trail of reductions and losses. In C. E. Bullerwell (Ed.), *Organelle genetics* (pp. 79–103). Springer. https://doi.org/https://doi.org/10.1007/978-3-642-22380-8_4
- Krause, K. (2015). Grand-scale theft: Kleptoplasty in parasitic plants? *Trends in Plant Science*, 20(4), 196–198. <https://doi.org/https://doi.org/10.1016/j.tplants.2015.03.005>
- Krawczyk, K., Nobis, M., Nowak, A., Szczecińska, M., & Sawicki, J. (2017). Phylogenetic implications of nuclear rRNA IGS variation in *Stipa* L. (Poaceae). *Scientific Reports*, 7(1), 11506. <https://doi.org/10.1038/s41598-017-11804-x>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- Lee, C., Choi, I.-S., Cardoso, D., de Lima, H. C., de Queiroz, L. P., Wojciechowski, M. F., Jansen, R. K., & Ruhlman, T. A. (2021). The chicken or the egg? Plastome evolution and

- an independent loss of the inverted repeat in papilionoid legumes. *The Plant Journal*, *107*(3), 861–875. <https://doi.org/https://doi.org/10.1111/tpj.15351>
- Li, Xi, Zhang, T.-C., Qiao, Q., Ren, Z., Zhao, J., Yonezawa, T., Hasegawa, M., Crabbe, M. J. C., Li, J., & Zhong, Y. (2013). Complete chloroplast genome sequence of holoparasite *Cistanche deserticola* (Orobanchaceae) reveals gene loss and horizontal gene transfer from its host *Haloxylon ammodendron* (Chenopodiaceae). *PloS One*, *8*(3), e58747–e58747. <https://doi.org/10.1371/journal.pone.0058747>
- Li, Xin, Yang, J.-B., Wang, H., Song, Y., Corlett, R. T., Yao, X., Li, D.-Z., & Yu, W.-B. (2021). Plastid NDH Pseudogenization and Gene Loss in a Recently Derived Lineage from the Largest Hemiparasitic Plant Genus *Pedicularis* (Orobanchaceae). *Plant and Cell Physiology*, *62*(6), 971–984. <https://doi.org/10.1093/pcp/pcab074>
- Lockhart, J. (2013). Plastid genes that were lost along the road to parasitism. *Plant Cell*, *25*(10), 3636. <https://doi.org/10.1105/tpc.113.251011>
- Lohan, A. J., & Wolfe, K. H. (1998). A subset of conserved tRNA genes in plastid DNA of nongreen plants. *Genetics*, *150*(1), 425–433. <https://doi.org/10.1093/genetics/150.1.425>
- López, K. E. R., Armijos, C. E., Parra, M., & Torres, M. D. (2023). The First Complete Chloroplast Genome Sequence of Mortiño (*Vaccinium floribundum*) and Comparative Analyses with Other *Vaccinium* Species. In *Horticulturae* (Vol. 9, Issue 3). <https://doi.org/10.3390/horticulturae9030302>
- Lyko, P., & Wicke, S. (2021). Genomic reconfiguration in parasitic plants involves considerable gene losses alongside global genome size inflation and gene births. *Plant Physiology*, *186*(3), 1412–1423. <https://doi.org/10.1093/plphys/kiab192>
- Ma, P.-F., Zhang, Y.-X., Guo, Z.-H., & Li, D.-Z. (2015). Evidence for horizontal transfer of mitochondrial DNA to the plastid genome in a bamboo genus. *Scientific Reports*, *5*(1), 11608. <https://doi.org/10.1038/srep11608>
- Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, *50*(W1), W276–W279. <https://doi.org/10.1093/nar/gkac240>

- Martel, C., Fernandez-Hilario, R., Tello, J. A., Arteaga, R. G., & Gerlach, G. (2018). *Prosopanche panguanensis* (Aristolochiaceae), a new species from central Peru. *Phytotaxa*, *364*(3), 241–249.
- Massoni, J., Forest, F., & Sauquet, H. (2014). Increased sampling of both genes and taxa improves resolution of phylogenetic relationships within Magnoliidae, a large and early-diverging clade of angiosperms. *Molecular Phylogenetics and Evolution*, *70*, 84–93. <https://doi.org/https://doi.org/10.1016/j.ympev.2013.09.010>
- McNeal, J. R., Kuehl, J. V., Boore, J. L., Leebens-Mack, J., & dePamphilis, C. W. (2009). Parallel loss of plastid introns and their maturase in the genus *Cuscuta*. *PloS One*, *4*(6), e5982. <https://doi.org/10.1371/journal.pone.0005982>
- Meyer, M., & Kircher, M. (2010). Illumina Sequencing Library Preparation for Highly Multiplexed Target Capture and Sequencing. *Cold Spring Harbor Protocols*, *2010*(6), pdb.prot5448. <https://doi.org/10.1101/pdb.prot5448>
- Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *2010 Gateway Computing Environments Workshop (GCE)*, 1–8. <https://doi.org/10.1109/GCE.2010.5676129>
- Mkala, E. M., Jost, M., Dong, X., Mwachala, G., Musili, P. M., Wanke, S., Hu, G.-W., & Wang, Q.-F. (2023). Phylogenetic and comparative analyses of *Hydnora abyssinica* plastomes provide evidence for hidden diversity within Hydnoraceae. *BMC Ecology and Evolution*, *23*(1), 34. <https://doi.org/10.1186/s12862-023-02142-w>
- Mohanta, T. K., Mishra, A. K., Hashem, A., Abd_Allah, E. F., Khan, A. L., & Al-Harrasi, A. (2020). Construction of anti-codon table of the plant kingdom and evolution of tRNA selenocysteine (tRNA^{Sec}). *BMC Genomics*, *21*(1), 804. <https://doi.org/10.1186/s12864-020-07216-3>
- Mohanta, T. K., Mohanta, Y. K., & Sharma, N. (2023). Anticodon table of the chloroplast genome and identification of putative quadruplet anticodons in chloroplast tRNAs. *Scientific Reports*, *13*(1), 760. <https://doi.org/10.1038/s41598-023-27886-9>
- Molina, J., Hazzouri, K. M., Nickrent, D., Geisler, M., Meyer, R. S., Pentony, M. M., Flowers, J. M., Pelsner, P., Barcelona, J., Inovejas, S. A., Uy, I., Yuan, W., Wilkins, O., Michel, C.-

- I., Locklear, S., Concepcion, G. P., & Purugganan, M. D. (2014). Possible Loss of the Chloroplast Genome in the Parasitic Flowering Plant *Rafflesia lagascae* (Rafflesiaceae). *Molecular Biology and Evolution*, *31*(4), 793–803. <https://doi.org/10.1093/molbev/msu051>
- Mower, J. P., & Vickrey, T. L. (2018). Chapter Nine - Structural Diversity Among Plastid Genomes of Land Plants. In S.-M. Chaw & R. K. B. T.-A. in B. R. Jansen (Eds.), *Plastid Genome Evolution* (Vol. 85, pp. 263–292). Academic Press.
<https://doi.org/https://doi.org/10.1016/bs.abr.2017.11.013>
- Muche, M., Muasya, A. M., & Tsegay, B. A. (2022). Biology and resource acquisition of mistletoes, and the defense responses of host plants. *Ecological Processes*, *11*(1), 24. <https://doi.org/10.1186/s13717-021-00355-9>
- Musselman, L. J., & Visser, J. H. (1989). Taxonomy and natural history of *Hydnora* (Hydnoraceae). *Aliso*, *12*(2), 317–326. <https://doi.org/10.5642/aliso.19891202.09>
- Naumann, J., Der, J. P., Wafula, E. K., Jones, S. S., Wagner, S. T., Honaas, L. A., Ralph, P. E., Bolin, J. F., Maass, E., Neinhuis, C., Wanke, S., DePamphilis, C. W., & De Pamphilis, C. W. (2016). Detecting and Characterizing the Highly Divergent Plastid Genome of the Nonphotosynthetic Parasitic Plant *Hydnora visseri* (Hydnoraceae). *Genome Biology and Evolution*, *8*(2), 345–363. <https://doi.org/10.1093/gbe/evv256>
- Naumann, J., Salomo, K., Der, J. P., Wafula, E. K., Bolin, J. F., Maass, E., Frenze, L., Samain, M.-S., Neinhuis, C., dePamphilis, C. W., & Wanke, S. (2013). Single-Copy Nuclear Genes Place Haustorial Hydnoraceae within Piperales and Reveal a Cretaceous Origin of Multiple Parasitic Angiosperm Lineages. *PLOS ONE*, *8*(11), e79204. <https://doi.org/10.1371/journal.pone.0079204>
- Nickrent, D. L. (2002). Molecular data place Hydnoraceae with Aristolochiaceae. *American Journal of Botany*, *89*(11), 1809–1817. <https://doi.org/10.3732/ajb.89.11.1809>
- Nickrent, D. L. (2020). Parasitic angiosperms: How often and how many? *TAXON*, *69*(1), 5–27. <https://doi.org/10.1002/tax.12195>
- Nickrent, D. L., Blarer, A., Qiu, Y. L., Soltis, D. E., Soltis, P. S., & Zanis, M. (2002). Molecular data place Hydnoraceae with Aristolochiaceae. *American Journal of Botany*, *89*(11), 1809–1817. <https://doi.org/10.3732/ajb.89.11.1809>

- Pan, H., Zagorchev, L., Chen, L., Tao, Y., Cai, C., Jiang, M., Sun, Z., & Li, J. (2023). Complete chloroplast genomes of five *Cuscuta* species and their evolutionary significance in the *Cuscuta* genus. *BMC Genomics*, *24*(1), 310. <https://doi.org/10.1186/s12864-023-09427-w>
- Park, J.-M., Manen, J.-F., & Schneeweiss, G. M. (2007). Horizontal gene transfer of a plastid gene in the non-photosynthetic flowering plants *Orobanche* and *Phelipanche* (Orobanchaceae). *Molecular Phylogenetics and Evolution*, *43*(3), 974–985. <https://doi.org/https://doi.org/10.1016/j.ympev.2006.10.011>
- Piaggio, A. J., Engeman, R. M., Hopken, M. W., Humphrey, J. S., Keacher, K. L., Bruce, W. E., & Avery, M. L. (2014). Detecting an elusive invasive species: a diagnostic PCR to detect Burmese python in Florida waters and an assessment of persistence of environmental DNA. *Molecular Ecology Resources*, *14*(2), 374–380. <https://doi.org/10.1111/1755-0998.12180>
- Prince, L. M. (2015). Plastid primers for angiosperm phylogenetics and phylogeography. *Applications in Plant Sciences*, *3*(6). <https://doi.org/10.3732/apps.1400085>
- Raman, G., Nam, G. H., & Park, S. J. (2022). Extensive reorganization of the chloroplast genome of *Corydalis platycarpa*: A comparative analysis of their organization and evolution with other *Corydalis* plastomes. *Frontiers in Plant Science*, *13*, 1043740. <https://doi.org/10.3389/FPLS.2022.1043740/BIBTEX>
- Renner, S. S., & Bellot, S. (2012). *Horizontal Gene Transfer in Eukaryotes: Fungi-to-Plant and Plant-to-Plant Transfers of Organellar DNA BT - Genomics of Chloroplasts and Mitochondria* (R. Bock & V. Knoop (eds.); pp. 223–235). Springer Netherlands. https://doi.org/10.1007/978-94-007-2920-9_10
- Robinson, J. T., Thorvaldsdottir, H., Turner, D., & Mesirov, J. P. (2023). igv.js: an embeddable JavaScript implementation of the Integrative Genomics Viewer (IGV). *Bioinformatics*, *39*(1), btac830. <https://doi.org/10.1093/bioinformatics/btac830>
- Rodin, A. S., Szathmáry, E., & Rodin, S. N. (2011). On origin of genetic code and tRNA before translation. *Biology Direct*, *6*(1), 14. <https://doi.org/10.1186/1745-6150-6-14>
- Roquet, C., Coissac, É., Cruaud, C., Boleda, M., Boyer, F., Alberti, A., Gielly, L., Taberlet, P., Thuiller, W., Es, J. van, & Lavergne, S. (2016). Understanding the evolution of

- holoparasitic plants: the complete plastid genome of the holoparasite *Cytinus hypocistis* (Cytinaceae). *Annals of Botany*, 118(5), 885–896.
- Roquet, Cristina, Coissac, É., Cruaud, C., Boleda, M., Boyer, F., Alberti, A., Gielly, L., Taberlet, P., Thuiller, W., Van Es, J., & Lavergne, S. (2016). Understanding the evolution of holoparasitic plants: the complete plastid genome of the holoparasite *Cytinus hypocistis* (Cytinaceae). *Annals of Botany*, 118(5), 885–896. <https://doi.org/10.1093/aob/mcw135>
- Rozen, S., & Skaletsky, H. J. (2000). Primer3 on the WWW for general users and for biologist programmers. In S. Krawetz & S. Misener (Eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. (pp. 365–386). Humana Press.
- Sanchez-Puerta, M. V., Ceriotti, L. F., Gatica-Soria, L. M., Roulet, M. E., Garcia, L. E., & Sato, H. A. (2023). Invited Review Beyond parasitic convergence: unravelling the evolution of the organellar genomes in holoparasites. *Annals of Botany*, 132(5), 909–928. <https://doi.org/10.1093/aob/mcad108>
- Schneider, A. C., Braukmann, T., Banerjee, A., & Stefanović, S. (2018). Convergent Plastome Evolution and Gene Loss in Holoparasitic Lennoaceae. *Genome Biology and Evolution*, 10(10), 2663–2670. <https://doi.org/10.1093/gbe/evy190>
- Schneider, A. C., Chun, H., Stefanović, S., & Baldwin, B. G. (2018). Punctuated plastome reduction and host–parasite horizontal gene transfer in the holoparasitic plant genus *Aphyllon*. *Proceedings of the Royal Society B: Biological Sciences*, 285(1887), 20181535. <https://doi.org/10.1098/rspb.2018.1535>
- Shyu, S.-Y. (2013). 全寄生植物菱形奴草質體基因組演化之研究 = Evolutionary studies of plastid genome of holoparasitic *Mitrastemon kanehirai* / 徐馨怡(Shin-Yi Shyu)撰. In 全寄生植物菱形奴草質體基因組演化之研究 = *Evolutionary studies of plastid genome of holoparasitic Mitrastemon kanehirai*. 博士論文--國立臺灣大學生態學與演化生物學研究所.
- Sibbald, S. J., & Archibald, J. M. (2020). Genomic Insights into Plastid Evolution. *Genome Biology and Evolution*, 12(7), 978–990. <https://doi.org/10.1093/gbe/evaa096>
- Sievers, F., & Higgins, D. G. (2014). Clustal Omega, accurate alignment of very large numbers of sequences. *Methods in Molecular Biology (Clifton, N.J.)*, 1079, 105–116.

https://doi.org/10.1007/978-1-62703-646-7_6

- Smith, D. R., & Lee, R. W. (2014). A Plastid without a Genome: Evidence from the Nonphotosynthetic Green Algal Genus *Polytomella*. *Plant Physiology*, *164*(4), 1812–1819. <https://doi.org/10.1104/pp.113.233718>
- Song, Y., Yu, W.-B., Tan, Y., Liu, B., Yao, X., Jin, J., Padmanaba, M., Yang, J.-B., & Corlett, R. T. (2017). Evolutionary Comparisons of the Chloroplast Genome in Lauraceae and Insights into Loss Events in the Magnoliids. *Genome Biology and Evolution*, *9*(9), 2354–2364. <https://doi.org/10.1093/gbe/evx180>
- Staats, M., Cuenca, A., Richardson, J. E., Vrieling-van Ginkel, R., Petersen, G., Seberg, O., & Bakker, F. T. (2011). DNA Damage in Plant Herbarium Tissue. *PLOS ONE*, *6*(12), e28448. <https://doi.org/10.1371/journal.pone.0028448>
- Staats, M., Erkens, R. H. J., van de Vossen, B., Wieringa, J. J., Kraaijeveld, K., Stielow, B., Geml, J., Richardson, J. E., & Bakker, F. T. (2013). Genomic Treasure Troves: Complete Genome Sequencing of Herbarium and Insect Museum Specimens. *PLOS ONE*, *8*(7), e69189. <https://doi.org/10.1371/journal.pone.0069189>
- Straub, S. C. K., Cronn, R. C., Edwards, C., Fishbein, M., & Liston, A. (2013). Horizontal transfer of DNA from the mitochondrial to the plastid genome and its subsequent evolution in milkweeds (apocynaceae). *Genome Biology and Evolution*, *5*(10), 1872–1885. <https://doi.org/10.1093/gbe/evt140>
- Strien, J., Sanft, J., & Mall, G. (2013). Enhancement of PCR Amplification of Moderate GC-Containing and Highly GC-Rich DNA Sequences. *Molecular Biotechnology*, *54*. <https://doi.org/10.1007/s12033-013-9660-x>
- Su, H.-J., Barkman, T. J., Hao, W., Jones, S. S., Naumann, J., Skippington, E., Wafula, E. K., Hu, J.-M., Palmer, J. D., & DePamphilis, C. W. (2019). Novel genetic code and record-setting AT-richness in the highly reduced plastid genome of the holoparasitic plant *Balanophora*. *Proceedings of the National Academy of Sciences*, *116*(3), 934 LP – 943. <https://doi.org/10.1073/pnas.1816822116>
- Sun, G., Xu, Y., Liu, H., Sun, T., Zhang, J., Hettenhausen, C., Shen, G., Qi, J., Qin, Y., Li, J., Wang, L., Chang, W., Guo, Z., Baldwin, I. T., & Wu, J. (2018). Large-scale gene losses

- underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nature Communications*, 9(1), 2683. <https://doi.org/10.1038/s41467-018-04721-8>
- Těšitel, J. (2016). Functional biology of parasitic plants: a review. *Plant Ecology and Evolution*, 149(1), 5–20.
- Thorogood, C. (2019). Hydnora : The strangest plant in the world? *Plants, People, Planet*, 1(1), 5–7. <https://doi.org/10.1002/ppp3.9>
- Tillich, M., Lehwark, P., Pellizzer, T., Ulbricht-Jones, E. S., Fischer, A., Bock, R., & Greiner, S. (2017). GeSeq – versatile and accurate annotation of organelle genomes. *Nucleic Acids Research*, 45(W1), W6–W11. <https://doi.org/10.1093/nar/gkx391>
- Tomikawa, C., Auxilien, S., Guérineau, V., Yoshioka, Y., Miyoshi, K., Hori, H., Fourmy, D., Takai, K., & Yoshizawa, S. (2018). Characterization of redundant tRNAs with CAU and UAU anticodons in *Lactobacillus plantarum*. *The Journal of Biochemistry*, 163(3), 233–241. <https://doi.org/10.1093/jb/mvx075>
- Twyford, A. D., & Ness, R. W. (2017). Strategies for complete plastid genome sequencing. *Molecular Ecology Resources*, 17(5), 858–868. <https://doi.org/10.1111/1755-0998.12626>
- Wen, Y., Qin, Y., Shao, B., Li, J., Ma, C., Liu, Y., Yang, B., & Jin, X. (2022). The extremely reduced, diverged and reconfigured plastomes of the largest mycoheterotrophic orchid lineage. *BMC Plant Biology*, 22(1), 448. <https://doi.org/10.1186/s12870-022-03836-x>
- Werner, J. J., Zhou, D., Caporaso, J. G., Knight, R., & Angenent, L. T. (2012). Comparison of Illumina paired-end and single-direction sequencing for microbial 16S rRNA gene amplicon surveys. *The ISME Journal*, 6(7), 1273–1276. <https://doi.org/10.1038/ismej.2011.186>
- Westwood, J. H., Yoder, J. I., Timko, M. P., & dePamphilis, C. W. (2010). The evolution of parasitism in plants. *Trends in Plant Science*, 15(4), 227–235. <https://doi.org/https://doi.org/10.1016/j.tplants.2010.01.004>
- Wicke, S., Müller, K. F., dePamphilis, C. W., Quandt, D., Bellot, S., & Schneeweiss, G. M. (2016). Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences*, 113(32), 9045 LP – 9050. <https://doi.org/10.1073/pnas.1607576113>

- Wicke, S., & Naumann, J. (2018). Chapter Eleven - Molecular Evolution of Plastid Genomes in Parasitic Flowering Plants. In S.-M. Chaw & R. K. B. T.-A. in B. R. Jansen (Eds.), *Plastid Genome Evolution* (Vol. 85, pp. 315–347). Academic Press.
<https://doi.org/https://doi.org/10.1016/bs.abr.2017.11.014>
- Wikström, N., Savolainen, V., & Chase, M. W. (2001). Evolution of the angiosperms: calibrating the family tree. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1482), 2211–2220. <https://doi.org/10.1098/rspb.2001.1782>
- Wu, C.-S., Wang, T.-J., Wu, C.-W., Wang, Y.-N., & Chaw, S.-M. (2017). Plastome Evolution in the Sole Hemiparasitic Genus Laurel Dodder (*Cassytha*) and Insights into the Plastid Phylogenomics of Lauraceae. *Genome Biology and Evolution*, 9(10), 2604–2614.
<https://doi.org/10.1093/gbe/evx177>
- Xi, Z., Bradley, R. K., Wurdack, K. J., Wong, K. M., Sugumaran, M., Bomblies, K., Rest, J. S., & Davis, C. C. (2012). Horizontal transfer of expressed genes in a parasitic flowering plant. *BMC Genomics*, 13(1), 227. <https://doi.org/10.1186/1471-2164-13-227>
- Xu, W., Chen, H., Tian, L., Jiang, M., Yang, Q., Wang, L., Ahmad, B., & Huang, L. (2020). Extensive gene loss in the plastome of holoparasitic plant *Cistanche tubulosa* (Orobanchaceae). *Mitochondrial DNA Part B*, 5(3), 2679–2681.
<https://doi.org/10.1080/23802359.2020.1787273>
- Yoshida, S., Cui, S., Ichihashi, Y., & Shirasu, K. (2016). The Haustorium, a Specialized Invasive Organ in Parasitic Plants. *Annual Review of Plant Biology*, 67(1), 643–667.
<https://doi.org/10.1146/annurev-arplant-043015-111702>
- Yu, R., Chen, X., Long, L., Jost, M., Zhao, R., Liu, L., Mower, J. P., dePamphilis, C. W., Wanke, S., & Jiao, Y. (2023). De novo Assembly and Comparative Analyses of Mitochondrial Genomes in Piperales. *Genome Biology and Evolution*, 15(3), evad041.
<https://doi.org/10.1093/gbe/evad041>
- Yu, R., Sun, C., Zhong, Y., Liu, Y., Sanchez-Puerta, M. V., Mower, J. P., & Zhou, R. (2022). The minicircular and extremely heteroplasmic mitogenome of the holoparasitic plant *Rhopalocnemis phalloides*. *Current Biology*, 32(2), 470-479.e5.
<https://doi.org/https://doi.org/10.1016/j.cub.2021.11.053>

- Zhang, G., Li, Q., & Sun, S. (2018). Diversity and distribution of parasitic angiosperms in China. *Ecology and Evolution*, 8(9), 4378–4386. <https://doi.org/10.1002/ece3.3992>
- Zimmer, E. A., & Wen, J. (2015). Using nuclear gene data for plant phylogenetics: Progress and prospects II. Next-gen approaches. *Journal of Systematics and Evolution*, 53(5), 371–379. <https://doi.org/https://doi.org/10.1111/jse.12174>