UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA VEGETAL

AMÁLIA IBIAPINO MOURA

EVOLUÇÃO CARIOTÍPICA EM *Cuscuta* L. (CONVOLVULACEAE)

Recife – PE
2022

AMÁLIA IBIAPINO MOURA

EVOLUÇÃO CARIOTÍPICA EM *Cuscuta* L. (CONVOLVULACEAE)

Tese apresentada por Amália Ibiapino Moura ao Prograna de Pós-Graduação em Biologia Vegetal da Universidade Federal de Pernambuco como parte dos requisitos necessários para obtenção do Grau de Doutorado em Biologia Vegetal.

Área de concentração: Sistemática e Evolução

Orientadora: Profa. Dra. Andrea Pedrosa-Harand (Depto. Botânica, UFPE)

Co-orientadora: Mariana Alejandra Baez (Departamento de Melhoramento de Plantas, Universidade de Bonn)

Recife – PE

2022

Catalogação na Fonte:
Bibliotecária Elaine Cristina Barroso, CRB4/1728

Moura, Amália Ibiapino
Evolução cariotípica em *Cuscuta* L. (Convolvulaceae) / Amália Ibiapino Moura – 2022.

106 f. : il., fig., tab.

Orientadora: Andrea Pedrosa-Harandt
Coorientadora: Mariana Alejandra Baez
Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Biologia Vegetal, Recife, 2022.
Inclui referências e apêndice.

1. Convolvulaceae 2. Heterocromatina 3. DNA I. Pedrosa-Harandt, Andrea (orient). II. Baez, Mariana Alejandra (coorient.) III. Título

583.94 CDD (22.ed.) UFPE/CB – 2022-169

AMÁLIA IBIAPINO MOURA

EVOLUÇÃO CARIOTÍPICA EM *Cuscuta* L. (CONVOLVULACEAE)

APROVADA EM: 21/02/2022

BANCA EXAMINADORA:

Profa. Dra. Andrea Pedrosa-Harand (Orientadora) – UFPE

Prof. Dr. Luiz Gustavo Rodrigues Souza – UFPE

Dra. Mariela Analía Sader – IMBIV-CONICET

Profa. Dra. Ana Christina Brasileiro-Vidal – UFPE

Prof. Dr. Reginaldo de Carvalho - UFRPE

Recife – PE

2022

Dedico essa tese às vozes da minha cabeça.
(Ansiosos entenderão)

Agradecimentos

Recentemente eu tive oportunidade de conhecer a história de um orixá chamado Oxóssi, cujo símbolo, Ofá (um arco e uma flecha), foi adotado por Francisco Brennand, artista plástico bem conhecido pelos recifenses. Segundo a mitologia, esse era o Orixá da sabedoria e do ganho de conhecimento, dentre as outras coisas. Reza a lenda que ele salvou sua aldeia de um ataque usando apenas uma única flecha, por isso ficou conhecido como: O Caçador de uma flecha só. Além de força e coragem, o símbolo do Ofá representa a chance única de obtenção de êxito. Mostrando que tudo é possível, por mais que você tenha uma única chance, você só precisa dela para o "tiro certeiro".

Refleti sobre isso indagando as chances que tive na vida. A única chance de ser filha de Ronilda e Laurentino e a eles agradeço pela vida, pelo apoio e pela formação de meu caráter. A chance de ser irmã de Lelia e Laurentino Júnior a quem servem de ponte com meu passado e com as primeiras boas lembranças que tenho. A minha chance de ser tia de Giovana e Ana Vitória, a quem quero cuidar e permitir que um dia olhem pra mim como um bom exemplo. A única chance de fazer parte de uma família e ela foi certeira.

Me tornei bióloga pela chance de ter tido professores inspiradoras durante a vida inteira, então deixo aqui meus agradecimentos a tia Evinha, Maria Carolina e Ana Carolina e Marcelo Guerra por sempre me mostrarem o leques de possibilidades dentro da biologia. A chance de conhecer Andrea Pedrosa-Harand que é de longe uma das pessoas mais competentes que eu conheço. Agradeço a ela pelo cuidado e compreensão, por acreditar em mim e no meu trabalho e por me inspirar a ser uma pessoa e profissional cada vez melhor.

Eu tive também uma única chance de achar outra família na minha vida de pós graduação e nessa chance achei Mariana, Mariela, Yennifer e Paulo. A eles deixo meu agradecimento por sempre estarem presente em momentos bons e ruins, por me guiarem, por me ensinarem e sempre acreditarem que eu poderia ir cada vez mais além. Tivemos a chance de formar o melhor grupo de "CitoGatas" (e agregados né Paulo?!) e conseguimos.

Tive a única chance de ser amiga de Yhanndra, de Brena, Lucas, Amanda Melo, Amanda Santos, Jessica, Gustavo ("inho" e "ão") e Cláudio e espero que eu tenha tornado a vida deles tão alegre e divertida como eles tornaram a minha. E já no fim do doutorado tive a única chance de dividir apartamento com Thiago e a ele agradeço pela companhia e por me ensinar muitas coisas sobre a vida em tão pouco tempo. E agradeço a chance de ter encontrado Bruno que colaborou com meu segundo artigo e me ajudou a manter a fé no meu próprio trabalho.

Agradeço ao Laboratório de Citogenética e Evolução Vegetal e a todas as pessoas que passaram por ele nesses últimos seis anos, em especial Tiago Ribeiro e Tiago Esposito, a quem guardo um enorme carinho mesmo estando tão distantes. Esse é "o melhor laboratório que se pode estar" (Báez, Mariana). Vocês nem imaginam o quanto marcaram minha vida.

Que o Ofá de Oxóssi nos ajude em nossa busca por conhecimento.

RESUMO

O gênero *Cuscuta* L. (Convolvulaceae) está representado por cerca de 200 espécies de distribuição cosmopolita. São plantas holoparasitas que se prendem ao seu hospedeiro por haustórios. Citogeneticamente esse gênero é muito variado e aparentemente possui características citogenéticas especificas em cada um dos seus quatro subgêneros. Apresenta muitos indivíduos poliploides, alguns deles alopoliploides confirmados; além de espécies com cariótipos bimodais, e com cromossomos monocentricos ou holocêntricos. Apresenta também ampla variação quanto aos tamanhos cromossômicos e tamanhos do genoma. Estudos mostrando caracterização da heterocromatina e composição da fração repetitiva revelaram que, em *Cuscuta*, a variação dos tamanhos genômicos está correlacionada com a amplificação de diferentes classes de DNA repetitivo. Esses resultados sugerem a influência da fração repetitiva e da heterocromatina na evolução cariotípica do gênero. A fim de entender o papel das sequências repetitivas na evolução da diversidade cariotipica de *Cuscuta*, foi utilizado o sequenciamento de baixa cobertura do DNA genômico de *C. nitida*, uma espécie com cariótipo bimodal do subgênero *Pachystigma*, com o objetivo de caracterizar e mapear as principais classes de sequências repetitivas envolvidas no surgimento de cariótipos bimodais no subgênero. Além disso, foram avaliadas a distribuição de bandas CMA/DAPI e de sítios de DNAr nessa e em outras duas espécies do mesmo subgênero. Os dados demonstram que *Pachystigma* é caracterizado pela presença de espécies com cariótipos bimodais. Além disso, o acúmulo de DNA repetitivo, predominantemente nos dois maiores pares cromossômicos de *C. nitida*, foi o principal responsável pelo surgimento do cariótipo bimodal dessa espécie. No segundo capítulo, foram realizadas reconstruções de caracteres ancestrais de número cromossômico, tamanho do genoma e distribuição de sítios de DNAr 5S e 35S, para entender os mecanismos envolvidos na evolução cariotípica do gênero *Cuscuta*. Além de um levantamento de dados citogenéticos publicados, foram gerados dados inéditos utilizando técnicas como hibridização *in situ* fluorescente e citometria de fluxo. Os dados suportam $n = 15$ como número ancestral do gênero *Cuscuta* e as mudanças numéricas ocorreram principalmente por poliploidias, mais frequentes no subgênero *Grammica* e disploidias, mais comuns no subgênero *Cuscuta*, comprovando a influência da holocentricidade na evolução do número cromossômico no grupo. Além disso, foi verificada uma expansão do tamanho do genoma em *Cuscuta* em relação às outras Convolvulaceae, o que pode estar correlacionado com o estilo de vida parasita das espécies desse gênero. Por fim, os sítios de DNAr em *Cuscuta* apresentam grande

variação de número e posição, principalmente o DNAr 35S, em parte relacionado à evolução da heterocromatina. *Cuscuta* apresenta uma excepcional diversidade cariotípica dentro das angiospermas. Os dados sugerem que diferentes mecanismos evolutivos são preponderantes nos diferentes subgêneros em *Cuscuta* e a dinâmica do DNA repetitivo desse gênero é a principal responsável pela diversidade reportada nessas espécies.

Palavras-chave: *Cuscuta*, DNA ribossomal, DNA repetitivo, Evolução cariotípica, Heterocromatina

ABSTRACT

Cuscuta L. (Convolvulaceae) is represented by ~200 cosmopolitan species. They have holoparasitic habit and attach to the host by haustoria. This genus is very diverse and apparently has particular cytogenetic features for each of its four subgenera. It has many polyploid individuals and species, often interspecific allopolyploids; besides species with bimodal karyotypes, and with monocentric or holocentric chromosomes. A wide variation of both chromosome and genome sizes are observed. Studies about the heterochromatin characterization and the composition of the repetitive fraction showed that variation of genomic sizes in *Cuscuta* is correlated with the amplification of different classes of repetitive DNA. These suggest an influence of the repetitive fraction and heterochromatin on its karyotypic evolution. In order to understand the role of repetitive sequences in the evolution of *Cuscuta* karyotype diversity, low-coverage sequencing of *C. nitida* genomic DNA, a species with bimodal karyotype of the subgenus *Pachystigma*, was used to characterize and map the main classes of repetitive sequences involved in the origin of the its bimodal karyotypes. In addition, the distribution of CMA/DAPI bands in this and two other species of the same subgenus were evaluated. The data suggest that the *Pachystigma* is characterized by the presence of species with bimodal karyotypes. Furthermore, the repetitive DNA accumulation, predominantly in the two *C. nitida* largest chromosome pairs, was the main responsible for the emergence of the bimodal karyotype of this species. In the second chapter, reconstructions of ancestral characters of chromosome number, genome size and 5S and 35S rDNA sites distribution were performed, in order to understand the mechanisms involved in the karyotypic evolution of the genus *Cuscuta*. In addition to a survey of published cytogenetic data, original data was generated using techniques such as fluorescent *in situ* hybridization and flow cytometry. The data supported $n = 15$ as an ancestral chromosome number of *Cuscuta*. Numerical changes occurred mainly by polyploidies, more frequent in the *Grammica* subgenus, and dysploidies, more common in the *Cuscuta* subgenus, proving the influence of holocentricity in the evolution of the chromosome number in the group. Moreover, an expansion of the genome size in *Cuscuta* compared to other Convolvulaceae was observed, which may be correlated with the parasitic lifestyle of the species of this genus. Finally, the rDNA sites in *Cuscuta* show great variation in number and position, especially the 35S rDNA, in part related to the evolution of heterochromatin. *Cuscuta* has an exceptional karyotypic diversity within angiosperms. Altogether, different

evolutionary mechanisms took place in the different *Cuscuta* subgenera and the repetitive DNA dynamics within this genus is the main responsible for the diversity reported in these species.

Keywords: *Cuscuta*, Heterochromatin, Karyotypic evolution, Repetitive DNA, Ribosomal DNA

SUMÁRIO

1 INTRODUÇÃO..... 11

2 REVISÃO DE LITERATURA..... 13

2.1 EVOLUÇÃO CARIOTÍPICA..... 13

2.1.1 Cariótipos bimodais..... 14

2.1.2 Reconstrução de caracteres ancestrais..... 15

2.2 SEQUÊNCIAS DE DNA..... 16

2.2.1 A fração repetitiva do genoma..... 16

2.2.2 Distribuição cromossômica da fração repetitiva..... 20

2.3 O GÊNERO *Cuscuta* L..... 21

2.3.1 Citogenética de *Cuscuta*..... 22

3 CONSIDERAÇÕES FINAIS..... 25

REFERÊNCIAS..... 26

APÊNDICE A –Karyotype asymmetry in *Cuscuta* L. subgenus *Pachystigma* reflects its repeat DNA composition..... 30

APÊNDICE B –The evolution of cytogenetic traits in *Cuscuta* (Convolvulaceae), the genus with the most diverse chromosomes in angiosperms..... 67

1 INTRODUÇÃO

O gênero *Cuscuta* L. (Convolvulaceae), com cerca de 200 espécies, é composto por plantas holoparasitas de distribuição quase cosmopolita, com cerca de 150 representantes nas Américas. Pode ser subdividido em quatro subgêneros, sendo eles *Cuscuta* Yunck, *Grammica* (Lour.) Peter, Engl. & Prantl, *Monogynella* (Des Moul.) Peter, Engl. & Prantl e *Pachystigma* (Engelm.) Baker & C.H. Wright (GARCÍA et al., 2014; COSTEA et al., 2015). Morfologicamente, são hastes filiformes que se prendem ao hospedeiro por haustórios, possuem folhas reduzidas a escamas e quantidades de clorofila escassas ou inexistentes (COSTEA; TARDIF, 2006). O estilo de vida parasita dessas espécies permite que haja transferência horizontal de genes entre essas plantas e seus hospedeiros (VOEGEL et al., 2018).

Citogeneticamente, as espécies do gênero *Cuscuta* foram mais estudadas quanto ao número cromossômico, que varia de $2n = 8$ a $2n = 150$ (PAZY; PLITMANN, 1995; GARCÍA; CASTROVIEJO, 2003; GARCIA et al., 2019). Possuem grande variação cariotípica, com diversidade de tamanho cromossômico, quantidade de DNA nuclear, presença de cromossomos monocêntricos ou holocêntricos (com centrômero difuso), além de cariótipos bimodais (dois conjuntos de cromossomos morfologicamente distintos no mesmo genoma) (GARCÍA; CASTROVIEJO, 2003; MCNEAL et al., 2007; GARCÍA et al., 2019). Estudos mais recentes mostram que a variação cariotípica encontrada em *Cuscuta* se estende também ao numero e posição de bandas heterocromáticas e sítios de DNA ribossomal, além da diversidade de famílias de DNA repetitivo encontrada nessas espécies (IBIAPINO et al., 2019, 2020; NEUMANN et al., 2020).

Cada um dos quatro subgêneros de *Cuscuta* pode ser caracterizado por particularidades citogenéticas. O subgênero *Grammica*, por exemplo, apresenta a maior variação de número e tamanho cromossômicos, e também de tamanho do genoma. Além disso, possui casos confirmados de hibridização interespecífica e alopoliploidia (MCNEAL et al., 2007; GARCIA et al., 2019; IBIAPINO et al., 2019). O subgênero *Monogynella* é composto por espécies com maior tamanho cromossômico e maior tamanho do genoma (IBIAPINO et al., 2020; NEUMANN et al., 2020). O subgênero *Cuscuta* é representado por espécies exclusivamente holocentricas (GARCIA E CASTROVIEJO, 2003; NEUMANN et al., 2020) e, por fim, o subgênero *Pachystigma* é composto por espécies com cariótipos bimodais (GARCIA et al., 2019). A evolução cariotípica do gênero parece fortemente influenciada pela dinâmica da fração repetitiva.

O aumento do tamanho do genoma e dos cromossomos reportado em *C. monogyna* e *C. indecora*, por exemplo, ocorreu principalmente pela amplificação de diferentes classes de DNA repetitivo (IBIAPINO et al., 2020; NEUMANN et al., 2020). Além disso, estudos utilizando *reads* longos demonstraram que alguns satélites possuem uma estrutura complexa e podem contribuir para a dinâmica evolutiva das sequências repetitivas do gênero, inclusive no surgimento da diversidade do DNA ribossomal (DNAr) dessas espécies (VONDRAK et al., 2021).

Essa diversidade de caracteres torna *Cuscuta* um gênero particularmente interessante para estudos sobre evolução cromossômica. Por isso, este trabalho teve como objetivo entender como ocorre a evolução cariotípica no gênero *Cuscuta* com foco principalmente no surgimento dos cariótipos bimodais do subgênero *Pachystigma* e nos eventos evolutivos que levaram à diversidade de numero cromossômico, tamanho do genoma e DNAr. O primeiro capítulo teve como principal questão: Como as sequências de DNA repetitivo influenciam na formação dos cariótipos bimodais encontrados no subgênero *Pachystigma*? Tendo em vista que os diferentes subgêneros de *Cuscuta* parecem evoluir de forma independente, e não há casos de hibridização interespecífica ou fusões cromossômicas reportadas em *Pachystigma*, a principal hipótese foi que o acúmulo diferencial de sequências seja o principal mecanismo envolvido no surgimento desses cariótipos bimodais. Já o capítulo dois trouxe como questão principal: Como ocorre a evolução dos cariótipos no gênero *Cuscuta*? A hipótese principal foi que o clado holocêntrico (subgênero *Cuscuta*) apresenta uma dinâmica evolutiva diferente dos demais clados, pois sabe-se que eventos de disploidias são mais comuns em cromossomos holocêntricos, onde o centrômero é difuso.

2 REVISÃO DE LITERATURA

2.1 EVOLUÇÃO CARIOTÍPICA

Todo ser vivo pode ser classificado com base em suas características cariotípicas. Os cariótipos são classificados de acordo com vários parâmetros como, por exemplo, número cromossômico. Em plantas, estes números podem variar extensivamente entre as espécies, por exemplo, em eudicotiledôneas a variação é de $2n = 4$ a $2n = 640$. Embora poucas espécies de plantas tenham mais que 200 cromossomos, o maior número cromossômico relatado em plantas é $2n = 1200$ em *Ophioglossum reticulatum* L., uma pteridófita (HESLOP-HARRISON; SCHWARZACHER, 2011). Além da variação em número, o conjunto cromossômico também varia quanto ao tamanho cromossômico e posição do centrômero, sendo chamados de cariótipos assimétricos ou simétricos, quando ao grau de variação no tamanho cromossômico e posição do centrômero. Já os cariótipos bimodais possuem um conjunto de cromossomos maiores e um conjunto de cromossomos menores. (GUERRA, 1988).

As mudanças cariotípicas ocorrem devido a mudanças estruturais que levam a adições e deleções de nucleotídeos, duplicações total do genoma ou duplicação em pequenos blocos cromossômicos e hibridização interespecífica (SCHUBERT, 2007). As alterações cromossômicas podem ocorrer devido a mudanças na composição de DNA, como resultado de eventos de recombinação primária e secundária. As recombinações primárias ocorrem durante o mecanismo de reparo de DNA como inserções, deleções, duplicações, inversões e translocações recíprocas. Já as recombinações secundárias são recombinações homólogas e não homólogas entre regiões cromossômicas rearranjadas, que origina novos cariótipos (SCHUBERT; LYSAK, 2011). Uma forma de estudar essas mudanças é a comparação de números cromossômicos e ordem de fragmentos cromossômicos entre as espécies por métodos como hibridização *in situ* fluorescente (FISH). A FISH permite usar uma sonda do DNA de interesse marcada com um fluorocromo e ao hibridiza-la no cromossomo, é possível verificar a posição da mesma. Essa abordagem fornece conhecimento sobre a presença de rearranjos dentro de um cariótipo, porém, atualmente, com o sequenciamento do genoma completo em larga escala, é possível fazer análises comparativas entre genomas e detectar quebras de macro- e microsintenia de forma ainda mais acurada (EICHLER, 2003; HESLOP-HARRISON; SCHWARZACHER, 2011).

2.1.1 Cariótipos bimodais

A primeira expressão fenotípica do genótipo é o cariótipo. Sua organização fornece uma visão geral da organização do material genético no cromossomo (GUERRA, 2008). A simetria cariotípica está relacionada ao tamanho dos cromossomos e posição do centrômero. Grande parte das angiospermas possui cariótipos simétricos com cromossomos pequenos (STEBBINS, 1971; WEISS-SCHNEEWEISS; SCHNEEWEISS, 2013). Várias espécies, no entanto, possuem cariótipos assimétricos, caracterizados por conterem cromossomos com centrômeros terminais e subterminais, ou cromossomos de tamanhos diferentes. A expansão de sequências de DNA repetitivo, eventos de fusão e fissão cromossômica, podem alterar essa simetria. Além da morfologia e número cromossômico, outras características podem ser verificadas nos cariótipos, como diversidade de bandas heterocromáticas e localização de genes e outras sequências únicas. (WEISS-SCHNEEWEISS; SCHNEEWEISS, 2013). Cariótipos que possuem dois conjuntos cromossômicos com tamanhos diferentes são chamados de bimodais (STEBBINS, 1971; MCKAIN et al., 2012).

O estado extremo de assimetria cariotípica é o cariótipo bimodal. Existem três hipóteses principais que propõem como os cariótipos bimodais surgem. A primeira envolve rearranjos cromossômicos do tipo fusão-fissão através de translocações Robertsonianas. Assim, cromossomos maiores podem ser produtos de fusão de cromossomos pequenos ou cromossomos pequenos são resultados da fissão dos cromossomos maiores (SCHUBERT; LYSAK, 2011). O segundo mecanismo, sugerido para alguns gêneros como *Agave* L., seria por alopoliploidia, envolvendo espécies parentais com diferentes tamanhos cromossômicos (MCKAIN et al., 2012). A terceira possibilidade seria pela amplificação progressiva de sequências repetitivas (DE LA HERRÁN et al., 2001). Os cariótipos bimodais são mais bem estudados em animais. *Lygosoma bowringii* é uma espécie de réptil da família Scincidae que possui variação cariotípica intraespecífica, e cariótipos com $2n = 32$, sendo 18 macrocromossomos e 14 microcromossomos (LISACHOV et al., 2018). As aves também possuem cariótipos bimodais, *Gallus gallus domesticus* apresenta $2n = 78$, seis pares de macrocromossomos e 33 pares de microcromossomos, onde foi constatado que os microcromossomos carregam aproximadamente 30% do tamanho do genoma e estão ligados a maior atividade transcricional (SMITH et al., 2000). Em vegetais, as famílias Agavaceae e Iridaceae (ordem Asparagales) possui um amplo número de espécies com cariótipos bimodais. O gênero *Tigridia* Jussieu (Iridaceae), por exemplo, possui números cromossômicos bem

conservados (2*n* = 28), com quatro cromossomos grandes e 24 cromossomos menores (ARROYO MARTÍNEZ et al., 2017).

2.1.2 Reconstrução de caracteres ancestrais

A utilização de filogenia molecular tem contribuído para o entendimento da direção de mudanças cromossômicas. A interpretação dos dados citogenéticos considerando as relações filogenéticas entre as espécies é fundamental para traçar a origem e evolução das mudanças cariotípicas envolvidas na diversificação das espécies (VAIO et al., 2013; COSTA et al., 2017). Uma diversidade de dados citogenéticos pode ser usada para essas análises, como números cromossômicos, o que permite inferir eventos de poliploidia e disploidia. Essas análises têm demonstrado que disploidias descendentes, ou seja, diminuição do número cromossômico, são mais frequentes que as ascendentes, quando este número aumenta (GUERRA, 2008). Além do número cromossômico, as análises filogenéticas podem ser utilizadas para estudar os padrões do bandeamento cromossômico, número e distribuição de sítios de DNA ribossomal, tamanho do genoma, entre outros (COSTA et al., 2017; VAN-LUME et al., 2017; SADER et al., 2019). Uma análise citogenética revelou que *Oxalis rhombeo-ovata* A.St.-Hil. e *O. psoraleoides* Kunth compartilham algumas características cromossômicas como número de braços cromossômicos, número de sítios de DNAr 5S e bandas CMA positivas nos braços longos, apesar de *O. rhombeo-ovata* ter apresentado apenas metade do tamanho do genoma de *O. psoraleoides*, uma fissão cêntrica, uma banda CMA a mais e sítios de DNAr 35S em todos os cromossomos. As análises filogenéticas dessas duas espécies confirmam os resultados citogenéticos, sugerindo que elas deveriam ser colocadas juntas na seção Psoraleoideae (VAIO et al., 2018).

Algumas dessas análises são feitas simplesmente comparando e mapeando os dados citogenéticos sobre as relações entre as espécies presentes em uma árvore filogenética. Mas uma outra opção é fazer a reconstrução de caracteres ancestrais. Métodos baseados em análises probabilísticas têm permitido testar hipóteses de evolução cromossômica dentro de uma filogenia, reconstruindo os caracteres ancestrais como, por exemplo, número cromossômico (MORAES et al., 2015). A reconstrução de caracteres cromossômicos pode ser feita por meio de ferramentas, como o ChromEvol, que utiliza uma série de modelos de máxima verossimilhança e é capaz de estimar o número cromossômico ancestral ao longo de cada ramo de uma filogenia, levando em consideração fenômenos como poliploidia e disploidia. Além disso, é possível incluir

tamanho genômico nessas análises utilizando acessórios como a ferramenta phytools do softwere R, que possibilita a reconstrução do tamanho do genoma das espécies. E ainda reconstruir número e posição de bandas heterocromáticas e sítios de DNAr 5S e 35S utilizando o programa Mesquite (REVELL, 2012; GLICK; MAYROSE, 2014; MADDISON; MADDISON, 2008). A inclusão de dados de bandas heterocromáticas e sítios de DNAr é muito importante para os estudos de evolução cariotípica, por permitir entender como as sequencias de DNA atuam na transição dos cariótipos (GARCIA et al., 2017).

2.2 SEQUÊNCIAS DE DNA

Apesar da grande variedade de tamanho, número cromossômico e quantidade de DNA presente nas espécies vegetais, características estruturais dos cromossomos como centrômeros, telômeros e compactação da cromatina são em geral bem conservadas. O DNA nuclear de plantas é constituído por sequências codificadoras de cópia única e também por várias classes de sequências repetitivas. Estas constituem grande parte dos genomas e podem ser encontradas tanto organizadas em *tandem* (micro-, mini- e DNA-satélites), quanto dispersas no genoma (HESLOP-HARRISON; SCHWARZACHER, 2011). Estas sequências repetitivas estão associadas a formação de heterocromatina que, no geral, podem estar localizadas nos centrômeros, telômeros e em blocos heterocromáticos intersticiais (GREWAL; JIA, 2007).

As sequências repetitivas evoluem mais rapidamente que as sequências únicas e estão envolvidas com mecanismos de evolução genômica. Os elementos transponíveis, por exemplo, são sequências capazes de se movimentar dentro do genoma e gerar expansão do mesmo e, até mesmo, alterar a função dos genes (TENAILLON; HOLLISTER; GAUT, 2010). Estes elementos também podem facilitar a ocorrência de rearranjos cromossômicos, devido a eventos de recombinação homóloga ectópica entre as várias cópias de elementos transponíveis presentes no genoma (GRAY, 2000; SCHUBERT; VU, 2016).

2.2.1 A fração repetitiva do genoma

A grande maioria dos genomas são compostos por sequências de DNA repetitivo. Em humanos, por exemplo, essa porção corresponde a mais de 50% do genoma total. Essas repetições surgem por mecanismos que geram cópias extras de uma sequência que é inserida no genoma. Essas repetições podem se apresentar intercaladas com sequências

únicas, chamadas de DNA repetitivo disperso, ou alinhadas lado a lado, ou seja, em *tandem* (TREANGEN; SALZBERG, 2012). O DNA repetitivo pode ser espécie específico ou, até mesmo, cromossomo específico. Algumas cópias são altamente conservadas, como o DNAr 5S e 35S (ROA; GUERRA, 2012, 2015); e outras apresentar mudanças rápidas, gerando diferenciação. Por muito tempo o DNA repetitivo foi considerado DNA "lixo" ou "egoísta". Porém, sabe-se hoje que o DNA repetitivo desempenha importante papel na composição e evolução dos genomas. Essas sequencias compõe regiões importantes dos cromossomos como centrômeros e telômeros, além de contribuir para as mudanças dos tamanhos do genoma (BISCOTTI; OLMO; HESLOP-HARRISON, 2015; GARRIDO-RAMOS, 2017).

As sequências repetidas em *tandem* ou DNA satélite são sequências geralmente não codificantes, abundantes e dispostas em uma conformação *head-to-tail*, ou seja, orientadas na mesma direção. Essas sequências são dinâmicas, alterando-se rapidamente também em número e posição de sítios (BISCOTTI; OLMO; HESLOP-HARRISON, 2015; GARRIDO-RAMOS, 2015, 2017). Os DNAs satélite são formados por monômeros que podem variar em comprimento, composição de nucleotídeos, complexidade das sequências e abundância (PLOHL; MEŠTROVIĆ; MRAVINAC, 2012). O *crossing over* desigual entre regiões homólogas é um dos mecanismos que pode levar a expansão e redução dos arranjos de sequências. Além disso, uma mutação que ocorre em um monômero pode se espalhar dentro das unidades de repetição ou serem eliminadas por homogeneização. Os mecanismos envolvem transferência de sequência não reciproca por recombinação desigual, conversão gênica ou transposição, e a fixação de uma variante na população é resultado da recombinação na meiose e segregação cromossômica. Este fenômeno é conhecido como evolução em concerto (FELINER; ROSSELLÓ, 2012; PLOHL; MEŠTROVIĆ; MRAVINAC, 2012), e origina padrões variados de famílias de DNA repetitivo, gerando homogeneidade dentro das espécies e diversidade entre as espécies. Assim, espécies distintas podem apresentar famílias de DNA satélites diferentes ou esses satélites podem ser compartilhados entre espécies relacionadas. Em *Phaseolus* L., diferentes famílias de satélites, como o satélite *khipu* e o satélite *jumper* compõe as regiões subteloméricas e pericentroméricas de algumas espécies analisadas deste gênero (RIBEIRO et al., 2017a).

Além das sequências distribuídas em *tandem*, os organismos eucariontes possuem sequências de DNA repetidas no genoma de maneira dispersa. Conhecidas como elementos transponíveis, essas sequências são capazes de se mover ao longo do genoma

e se inserir de forma autônoma em lugares diferentes. Podendo compor 75% do DNA nuclear, os elementos transponíveis foram caracterizados pela primeira vez em 1940 por Barbara McClintock no genoma de milho (LEE; KIM, 2014; NEUMANN et al., 2019). Os elementos transponíveis podem ser classificados de acordo com seu mecanismo de transposição em duas classes. Os elementos de Classe I, ou retrotransposons, se transpõem por meio de um mecanismo de *copy-and-paste*. Para se transpor, esses elementos são transcritos em RNA mensageiro e, por meio da ação da transcriptase reversa, cópias de DNA complementar são geradas, e posteriormente reintegradas no genoma. Os elementos de Classe I são subdivididos em subclasses de acordo com seu mecanismo de reinserção do genoma, são essas LTR e não LTR, conforme apresentam presença ou ausência do *Long terminal repeat* (LTR) (BOURQUE et al., 2018; WICKER et al., 2018; NEUMANN et al., 2019). Já os elementos de Classe II, ou transposons de DNA, são elementos que não utilizam a transcriptase reversa e se transpõe por um mecanismo de *cut-and-past.* Estes elementos possuem em seus terminais, sequências invertidas que são reconhecidas por uma enzima transposase, que é responsável pelo corte e reinserção destes elementos em outros locais do genoma (WICKER et al;. 2018; BOURQUE et al., 2018; NEUMANN et al., 2019). Além dessas classificações, os elementos transponíveis também podem ser divididos em autônomos e não-autonoms, de acordo com sua capacidade de codificar todas as proteínas envolvidas na transposição. Também são ordenados em superfamílias e famílias de acordo com a similaridade e organização dos seus domínios proteicos. Por exemplo, os retrotransposons do tipo LTR possuem uma estrutura bem conservada e apresentam pelo menos os domínios: GAG, protease (PROT), transcriptase reversa (RT), ribonuclease H (RH) e integrase (INT). A ordem desses domínios classificam os elementos LTRs em Ty1/Copia e Ty3/Gypsy e em diferentes famílias, podem apresentar domínios extras, alguns deles podem estar relacionados ao lugar de inserção desses elementos no genoma, como o CRM, do inglês "centromeric retrotransposon of maize", primeiramente descrito no milho, porém é comumente encontrado em regiões centroméricas (PLOHL; MEŠTROVIĆ; MRAVINAC, 2014; NEUMANN et al., 2019).

Os elementos transponíveis apresentam regiões preferenciais para sua inserção e são extensas fontes de mutações e polimorfismos, estando assim associados a rearranjos genômicos. Podem também desempenhar funções específicas nos cromossomos. Em *Drosophila*, um elemento do tipo LINE é responsável pela manutenção dos telômeros para compensar a perda da telomerase que ocorreu em dípteros (PARDUE;

DEBARYSHE, 2011; BOURQUE et al., 2018). Os efeitos dos elementos transponíveis no genoma podem ser bastante dinâmicos. Eles podem se expandir em massa e gerar genomas grandes ou se acumular em regiões heterocromáticas de genomas pequenos. Em espécies do gênero *Zea* L., o aumento de algumas famílias de elementos resultou em um genoma 2× maior de *Zea luxurians* (Durieu & Asch.) R.M. Bird em relação a *Z. mays* L. e *Z. diploperennis* Iltis, Doebley & R. Guzmánem em menos de 2 milhões de anos (ESTEP; DEBARRY; BENNETZEN, 2013). A transposição desses elementos pode também alterar a expressão gênica ao se inserir perto ou dentro de regiões como éxons, introns e regiões reguladoras. No dendê (*Elaeis guineensis* Jacq.), os níveis de metilação de um elemento transponível que se encontra dentro do gene promotor do florescimento, determina a quantidade de óleo nos frutos (ONG-ABDULLAH et al., 2015). Os elementos transponíveis podem ainda "doar" seus genes para o genoma hospedeiro e ainda servir de matéria prima para o surgimento de genes codificadores de proteínas e RNAs não codificantes que podem assumir função celular importante. Esse fenômeno de domesticação de elementos transponíveis pode ser observado com os genes Rag 1 e Rag 2, que codificam a imunoglobulina e as células T do sistema imunológico dos mamíferos. Estes genes derivaram de transposons de DNA há cerca de 500 milhões de anos (HUANG et al., 2016).

Por se apresentar em grande número de cópias e poderem conter genes próprios, o DNA repetitivo dificulta a montagem e anotação genômicas. Com o avanço das tecnologias de sequenciamento de nova geração (NGS), ficou cada vez mais fácil fazer um sequenciamento de baixa cobertura para identificação das sequências repetitivas (WEISS-SCHNEEWEISS et al., 2015). Dessa forma, também foi possível notar um avanço em programas e metodologias para a análise e montagem dessas sequências. O RepeatExplorer é uma ferramenta cada vez mais utilizada na caracterização das sequências repetitivas. É acessível através da web (https://repeatexplorer-elixir.cerit-sc.cz) e, usando um algoritmo de agrupamento de sequências, facilita a identificação de repetições e compara essas repetições com bancos de dados próprios contendo elementos já conhecidos. Como o algoritmo usa sequências curtas e aleatórias do genoma, é ideal para análise de dados gerados por sequenciamento de nova geração. Algumas ferramentas adicionais auxiliam não só na classificação do DNA repetitivo, mas também na comparação de sequências repetitivas entre múltiplas espécies (NOVÁK et al., 2013; NOVÁK; NEUMANN; MACAS, 2020). Diversos trabalhos recentes demonstram a aplicabilidade desse programa na identificação e caracterização de sequências repetitivas

(MARQUES et al., 2015; RIBEIRO et al., 2017b; UTSUNOMIA et al., 2017, MATA-SUCRE et al., 2020; COSTA et al., 2021)

2.2.2 Distribuição cromossômica da fração repetitiva

O uso de técnicas moleculares como a FISH é particularmente indispensável para o mapeamento das sequências repetitivas nos cromossomos, possibilitando o entendimento da distribuição e dinâmica dessas sequências (WEISS-SCHNEEWEISS et al., 2015). Essas sequencias podem ser organizadas formando blocos ou ainda dispesas em toda a extenção cromossômica. Em muitos eucariontes as regiões centroméricas e teloméricas são compostas por sequências repetitivas que podem influenciar nos rearranjos cromossômicos. Regiões subteloméricas e pericentroméricas são consideradas *hotspots* de inserções ou retenção de sequências repetitivas (EICHLER; SANKOFF, 2003). Os elementos transponíveis, por exemplo, são encontrados em regiões preferenciais dos genomas, isso é, as inserções não ocorrem de maneira aleatória. Enquanto elementos Ty3/Gypsy se insere na heterocromatina, os elementos Ty1/Copia se inserem em regiões próximas a genes (BORQUE et al., 2018). Elementos Rex, do tipo não LTR, são distribuídos ao longo do genoma de peixes. Em espécies de *Characidium*, por exemplo, são encontrados de maneira dispersa, em pequenos aglomerados ao longo dos cromossomos. Essas sequências não ajudam no mapeamento cromossômico das espécies deste gênero, porém estes elementos são os principais causadores da inserção de sequências de DNA em regiões gênicas de *Characidium*, o que pode provocar aumento ou diminuição da expressão gênica (PUCCI et al., 2018).

Uma família do elemento LTR-Ty3/Gypsy-Chromovirus-CRM (retrotransposon centromérico de milho) foram encontradas na mesma região em diversas outras espécies e interagem com a proteína centromérica CENH3 (NAGAKI et al., 2004; LI et al., 2013; ZHANG et al., 2017). Em *Rhynchospora pubera* (Vahl) Boeckeler foram identificadas e confirmadas sequências específicas das regiões centroméricas, pela primeira vez em uma espécie com cromossomos holocêntricos. A família de DNA satélite *Tyba* e retrolementos centroméricos são as principais sequências de DNA associadas com os centrômeros dessa espécie (MARQUES et al., 2015; RIBEIRO et al., 2017b).

2.3 O GÊNERO *Cuscuta* L.

O gênero *Cuscuta* pertence à família Convolvulaceae Juss. é representado por ca. de 200 espécies, separadas em quatro subgêneros: *Grammica* (Lour.) Yunck., *Pachystigma* (Engelm.) Baker & C. H. Wright, *Cuscuta* e *Monogynella* (Des Moul.) (GARCÍA et al., 2014). O subgênero *Grammica*, com cerca de 150 espécies, tem distribuição quase exclusiva nas Américas, enquanto que o subgênero *Cuscuta* e *Pachystigma* são mais geograficamente restritos. *Cuscuta* é nativo da Europa, África e Ásia e *Pachystigma* exclusivamente africano. Por fim, o subgênero *Monogynella* tem como centro de origem a Ásia central e é bem distribuído pela Europa e África, com uma espécie ocorrendo na América do Norte. O gênero *Cuscuta* habita ambientes de clima tropical a temperado, sendo essa preferência possivelmente relacionada ao período de dormência das sementes (YUNCKER, 1932; COSTEA; TARDIF, 2006; GARCÍA et al., 2014).

Conhecidas como "fios de ovos", "videira estranguladora", e ainda "cadarço de bruxa" (YUNCKER, 1932), as plantas desse gênero são parasitas, apresentam folhas reduzidas, em pouca quantidade e aparentemente sem função. Têm hábito herbáceo, com haste filiforme. Expressam pouca ou nenhuma clorofila e se prendem aos seus hospedeiros através de haustórios. O subgênero *Monogynella* é o mais ancestral e, portanto, possui características mais próximas de ancestrais não parasitas, como expressão de maior quantidade de clorofila, pseudoraiz no período embrionário e estrutura vascular menos reduzida (YUNCKER, 1921; YUNCKER, 1932; MCNEAL et al., 2007; COSTEA; GARCÍA; STEFANOVIĆ, 2015). As espécies de *Cuscuta* possuem células epidérmicas com paredes maleáveis, que permitem sua associação à superfície do hospedeiro, e secretam uma camada proteica que permite a fixação como um adesivo. Logo após sua fixação no hospedeiro, as espécies de *Cuscuta* iniciam o desenvolvimento dos seus haustórios, que posteriormente penetram no hospedeiro. Assim como acontece com outras espécies parasitas, em *Cuscuta* também pode haver transmissão horizontal de genes entre essas espécies e suas hospedeiras. Estudos demonstram trocas de três genes mitocondriais entre *Cuscuta* e *Plantago* L. (VAUGHN, 2002; MOWER et al., 2010;VOEGEL et al., 2018). Apresentam pequenas flores e escassez de características morfológicas diferenciais, tornando-se um desafio para a classificação ao nível de espécies (MCNEAL et al., 2007a). É o único gênero da família que apresenta parasitismo e chegou a ser considerada anteriormente uma família separada (*Cuscuta*ceae). A inserção dentro da família Convolvulaceae se deve à sua morfologia reprodutiva e análises de

DNA mitocondrial (YUNCKER, 1932; MCNEAL et al., 2007; STEFANOVIC; KUZMINA; COSTEA, 2007).

Dados moleculares recentes demonstram que cada um dos quatro subgêneros representam clados muito bem suportados, onde *Grammica* e *Monogynela* são monofiléticos. O subgênero *Pachystigma* já foi considerado apenas uma seção do subgênero *Cuscuta*, porém filogenias moleculares recentes classificam esse clado como subgênero irmão de *Grammica* (GARCÍA et al., 2014). *Cuscuta* também possui 15 clados bem suportados e casos de incongruências bem suportados entre os dados de DNA plastidial e nuclear. Esses casos de incongruência podem indicar a origem híbrida de algumas espécies (GARCÍA et al., 2014, 2018; STEFANOVIC; KUZMINA; COSTEA, 2007). Relações de hibridização em plantas holoparasitas não são bem estabelecidas, ou por não serem bem estudadas ou por serem raras. No subgênero *Grammica* há cinco casos de incongruências filogenéticas, sugerindo que hibridização podem influenciar a diversidade de espécies desse gênero (STEFANOVIĆ; COSTEA, 2008). Um desses casos de incongruência é verificado na seção Denticulata. Esta seção é composta por três espécies, *C. denticulata* Engelm. *C. nevadensis* I. M. Johnst. e *C. veatchii* Brandegee que se caracterizam morfologicamente pela extremidade radicular esférica do embrião que aumenta de volume durante a germinação da semente. Este clado apresenta incongruências filogenéticas, onde o DNA plastidial indica *C. veatchii* como espécie irmã de *C. denticulata* e o DNA nuclear, indica *C. veatchii* como espécie irmã de *C. nevadensis* (STEFANOVIĆ; COSTEA, 2008; GARCÍA et al., 2018). Recentemente, uma análise citogenética foi feita nesta seção, confirmando que tanto *C. denticulata* quanto *C. nevadensis* apresentaram números cromossómicos $2n = 30$, enquanto que *C. veatchii* apresenta $2n = 60$. Além disso, este trabalho verificou padrões de bandas heterocromáticas e sítios de DNA ribossomal 5S e 35S e confirmou por hibridização genômica *in situ* (GISH) que *C. veatchii* é um hibrido originado do cruzamento entre *C. denticulata* e *C. nevadensis* (IBIAPINO et al., 2019).

2.3.1 Citogenética de *Cuscuta*

Citogeneticamente o gênero *Cuscuta* é extremamente diversificado. As espécies de *Cuscuta* apresentam números cromossômicos que variam de $2n = 8$ a $2n = 150$ e número básico $x = 15$ (PAZY; PLITMANN, 1995; GARCÍA; CASTROVIEJO, 2003; IBIAPINO et al. submetido). A maioria das espécies são diploides com $2n = 30$, mas algumas espécies são poliploides, originadas tanto por eventos de alopoliploidia, quanto

por eventos de autopoliploidia (GARCÍA et al., 2014, 2018). Os tamanhos cromossômicos e quantidade de DNA das espécies de *Cuscuta* também são bastante variados, o tamanho cromossômico pode variar de 1,66 µm em *C. denticulata* a 21,60 µm em *C. monogyna*. E o tamanho do genoma varia de 1C = 0,27Gbp pg em *C. australis* Engelm. a 1C = 34,73Gbp em *C. reflexa* Vahl (SUN et al., 2018; NEUMANN et al., 2020; IBIAPINO et al., submetido). Diferentes subgêneros de *Cuscuta* apresentam diferentes características cariotípicas. Por exemplo, as espécies do subgênero *Monogynella* apresentam os maiores tamanhos de genoma e os maiores cromossomos do gênero. Já o subgênero *Cuscuta* é o único que possui espécies com cromossomos exclusivamente holocentricos. *Grammica* é o subgênero com maior variação de número e tamanho cromossômico, e tamanho do genoma; com eventos de hibridização e poliploidia (FOGELBERG, 1938; PAZY; PLITMANN, 1994; GARCÍA, 2001; GARCÍA; CASTROVIEJO, 2003; MCNEAL et al., 2007; IBIAPINO et al., 2019). Por fim, três das cinco espécies do subgênero *Pachystigma* possuem cariótipos bimodais. Indicando que essa possa ser a característica citogenética que defina esse subgênero (GARCÍA et al., 2019; IBIAPINO et al., 2021).

No gênero também há relatos de variação intraespecífica, principalmente de número cromossômico. Em espécies como *C. epithymum* (L.) L. e *C. planiflora* Ten., o número cromossômico pode mudar entre populações diferentes da mesma espécie. Essa variação é ainda mais interessante em *C. epithymum*, que apresenta cromossomos holocêntricos e pode variar entre 2*n* = 14, 16, 28, 30 e 32, onde cariótipos com 2*n* = 28 são bimodais, com um par de cromossomos maiores, enquanto que cariótipos com 2*n* = 16 são perfeitamente simétricos, com cromossomos relativamente pequenos (GARCÍA; CASTROVIEJO, 2003). Esses dados sugerem uma alta dinâmica evolutiva nos cariótipos das espécies do gênero Cuscuta.

Outros parâmetros como número e posição de bandas CMA/DAPI e sítios de DNAr 5S e 35S também revelam uma enorme variação. *Cuscuta* possui desde poucas bandas CMA/DAPI e sítios de DNAr, como em *C. denticulata* com um par de bandas CMA/DAPI, um par de DNAr 5S e um par de DNAr 35S; até cariótipos como *C. monogyna*, com pelo menos 90 bandas CMA+, 80 bandas DAPI+, 36 sítios de DNAr 5S e 30 sítios de DNAr 35S (IBIAPINO et al., 2019, 2020). Esses dados indicam que a fração repetitiva do gênero *Cuscuta* tem uma grande influência na evolução cariotípica do gênero. Análises citogenômicas demonstraram que as espécies de *Cuscuta* que possuem maior tamanho do genoma, também possuem a maior proporção de DNA repetitivo em

seu genoma, além de bandas heterocromáticas bem evidentes e numerosas. Mostrando que nesse gênero, a expansão do DNA repetitivo foi o responsável pelo aumento do tamanho do genoma em espécies como *C. indecora* (subgênero *Grammica*), por exemplo (IBIAPINO et al., 2020; NEUMANN et al., 2020). Outro exemplo claro da influencia do DNA repetitivo na evolução desses cariótipos pode ser observado em *C. nitida* (subgênero *Pachystigma*). O acúmulo de diversas classes de DNA repetitivo somente em dois pares do seu conjunto cromossômico $2n = 30$, levou a bimodalidade cariotípica nessa espécies (IBIAPINO et al., 2021).

A dinâmica do DNA repetitivo em *Cuscuta* também está presente a nível molecular. Estudos citogenômicos indicam que alguns DNAs satélites descritos em espécies como *C. nitida* podem ter sido originadas a partir de elementos transponíveis (IBIAPINO et al., 2022, presente tese). Além disso, foi verificado em *C. europaea* (subgênero *Cuscuta*) a presença de satélites complexos. O satélite CUS-TR24, que é colocalizado com a proteína centromérica, possuí insersões de outras sequencias de DNA, principalmente do elemento LINE, em seus monômeros (VONDRAK et al., 2021). Uma outra característica bastante surpreensente em *Cuscuta* pode ser encontrada no subgênero *Cuscuta*. Como mencionado anteriormente, as espécies desse subgênero são holocêntricas, porém a proteína centromérica CENH3 está presente apenas em alguns blocos e não ao longo de todo o cromossomo, como é esperado em cromossomos holocêntricos, apesar deste fato, os microtúbulos se ligam ao longo de todo o cromossomo (OLIVEIRA et al., 2019).

Essa enorme variação cariotípica descrita em *Cuscuta*, além da dinâmica evolutiva de suas sequencias de DNA repetitivo, torna o gênero um ótimo modelo para estudar eventos de poliploidia, evolução genômica e cariotípica em plantas e a transição de cromossomos monocêntricos para holocêntricos.

3 CONSIDERAÇÕES FINAIS

A extensa variação cariotípica presente em *Cuscuta* foi investigada com diferentes abordagens, com a finalidade de entender a origem e a evolução dessa diversidade. A reconstrução de caracteres ancestrais de número cromossômico revelou $n = 15$ como o provável número ancestral desse gênero. Os eventos de duplicação foram os principais responsáveis pela mudança numérica, exceto no subgênero *Cuscuta*, que apresenta cromossomos holocêntricos, onde eventos de disploidia geraram uma grande diversidade numérica inter- e intraespecífica. O trabalho também mostrou que a presença do clado holocêntrico influencia diretamente a reconstrução de número cromossômico ancestral.

Também foi usada a abordagem de sequenciamento de próxima geração para investigar o papel da fração repetitiva do DNA na evolução cariotípica do subgênero *Pachystigma*, que é caracterizado pela presença de cariótipos bimodais. A bimodalidade se mostrou associada ao acúmulo preferencial de heterocromatina nos pares cromossômicos maiores de algumas espécies desse subgêneroe o sequenciamento de baixa cobertura feito em *Cuscuta nitida*, um dos representantes deste subgênero, revelou o acúmulo de sequências repetitivas em tandem e dispersas apenas nos dois pares cromossômicos maiores do complemento. Esses dados corroboram a importância da fração repetitiva, tanto presente na heterocromatina como fora dela, nas mudanças cariotípicas encontradas no gênero *Cuscuta*.

A fração repetitiva também pode ser composta por satélites complexos, como observado em *C. nitida*, e influencia a grande diversidade de DNAr 5S e 35S nesse gênero, pois a associação do DNAr com alguns satélites pode levar à amplificação desses sítios. Além disso, as sequências repetitivas estão ligadas à expansão do genoma reportada em *Cuscuta* em relação a outras Convolvulaceae.

O gênero *Cuscuta* reúne espécies monocêntricas e holocêntricas, cariótipos que vão de simétricos a bimodais, além de espécies alopoliploides e autopoliploides. Sendo assim, além da poliploidia e disploidia, responsáveis pela variação numérica e influenciadas pela organização centromérica, a dinâmica da fração repetitiva das espécies desse gênero contribuiu diretamente para a diversidade encontrada no tamanho, simetria e organização dos seus cariótipos.

REFERÊNCIAS

ARROYO MARTÍNEZ, H. A. et al. Karyotype determination of three *Tigridia* species (Asparagales, Iridaceae). Caryologia, v. 70, n. 3, p. 211–215, 2017.

BISCOTTI, M. A.; OLMO, E.; HESLOP-HARRISON, J. S. Repetitive DNA in eukaryotic genomes. Chromosome Research, v. 23, n. 3, p. 415–420, 2015.

BOURQUE, G. et al. Ten things you should know about transposable elements. Genome Biology, v. 19, n. 1, p. 199, 2018.

COSTA, L. et al. Comparative cytomolecular analyses reveal karyotype variability related to biogeographic and species richness patterns in Bombacoideae (Malvaceae). Plant Systematics and Evolution, v. 303, n. 9, p. 1131–1144, 2017.

COSTEA, M.; GARCÍA, M. A.; STEFANOVIĆ, S. A phylogenetically based infrageneric classification of the parasitic plant genus Cuscuta (Dodders, Convolvulaceae). Systematic Botany, v. 40, n. 1, p. 269–285, 2015.

COSTEA, M.; TARDIF, F. J. The biology of Canadian weeds. 133. *Cuscuta campestris* Yuncker, *C. gronovii* Willd. ex Schult., *C. umbrosa* Beyr. ex Hook., *C. epithymum* (L.) L. and *C. epilinum* Weihe. Canadian Journal of Plant Science, v. 86, n. 1, p. 293–316, 2006.

DE LA HERRÁN, R. et al. A heterochromatic satellite DNA is highly amplified in a single chromosome of *Muscari* (Hyacinthaceae). Chromosoma, v. 110, n. 3, p. 197–202, 2001.

EICHLER, E. E. Structural dynamics of eukaryotic chromosome evolution. Science, v. 301, n. 5634, p. 793–797, 2003.

ESTEP, M. C.; DEBARRY, J. D.; BENNETZEN, J. L. The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. Heredity, v. 110, n. 2, p. 194–204, 2013.

FELINER, G. N.; ROSSELLÓ, J. A. Concerted evolution of multigene families and homoeologous recombination. Plant Genome Diversity Volume 1. Vienna: Springer Vienna. p. 171–193, 2012.

FOGELBERG, S. O. The Cytology of *Cuscuta*. Bulletin of the Torrey Botanical Club, v. 65, n. 9, p. 631, 1938.

GARCÍA, M. A. A new western mediterranean species of *Cuscuta* (Convolvulaceae) confirms the presence of holocentric chromosomes in subgenus Cuscuta. Botanical Journal of the Linnean Society, v. 135, n. 2, p. 169–178, 2001.

GARCÍA, M. A. et al. Phylogeny, character evolution, and biogeography of *Cuscuta* (dodders; Convolvulaceae) inferred from coding plastid and nuclear sequences. American Journal of Botany, v. 101, n. 4, p. 670–690, 2014.

GARCÍA, M. A. et al. Cladogenesis and reticulation in *Cuscuta* sect. *Denticulatae* (Convolvulaceae). Organisms Diversity & Evolution, v. 18, n. 4, p. 383–398, 2018.

GARCÍA, M. Á.; CASTROVIEJO, S. Estudios citotaxonómicos en las especies Ibéricas del género *Cuscuta* (Convolvulaceae). Anales del Jardin Botanico de Madrid, v. 60, n. 1, p. 33-44, 2003.

GARRIDO-RAMOS, M. A. Satellite DNA in plants: More than Just Rubbish. Cytogenetic and Genome Research, v. 146, n. 2, p. 153–170, 2015.

GARRIDO-RAMOS, M. A. Satellite DNA: An Evolving Topic. Genes, v. 8, n. 9, p. 230, 2017.

GLICK, L.; MAYROSE, I. ChromEvol: Assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. Molecular Biology and Evolution, v. 31, n. 7, p. 1914–1922, 2014.

GRAY, Y. H. M. It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends in Genetics, v. 16, n. 10, p. 461–468, 2000.

GREWAL, S. I. S.; JIA, S. Heterochromatin revisited. Nature Reviews Genetics, v. 8, n. 1, p. 35–46, 2007.

GUERRA, M. Chromosome numbers in plant cytotaxonomy: concepts and implications. Cytogenetic and Genome Research, v. 120, n. 3–4, p. 339–350, 2008.

GUERRA, M. DOS S. Introdução à citogenética geral. [s.l.] Guanabara Koogan, 1988.

GUERRA, M.; GARCÍA, M. A. Heterochromatin and rDNA sites distribution in the holocentric chromosomes of *Cuscuta approximata* Bab. (Convolvulaceae). Genome, v. 47, n. 1, p. 134–140, 2004.

HESLOP-HARRISON, J. S. P.; SCHWARZACHER, T. Organisation of the plant genome in chromosomes. The Plant Journal, v. 66, n. 1, p. 18–33, 2011.

HUANG, S. et al. Discovery of an active RAG transposon illuminates the origins of V (D) J recombination. Cell, v. 166, n. 1, p. 102–114, 2016.

IBIAPINO, A. et al. Intense proliferation of rDNA sites and heterochromatic bands in two distantly related *Cuscuta* species (Convolvulaceae) with very large genomes and symmetric karyotypes. Genet etic and Molecular Biology, v. 43, p. e20190068, 2020.

IBIAPINO, A. et al. Allopolyploid origin and genome differentiation of the parasitic species *Cuscuta veatchii* (Convolvulaceae) revealed by genomic in situ hybridization. Genome, n. 999, p. 1–9, 2019.

IBIAPINO, A. et al. Karyotype asymmetry in Cuscuta L. subgenus Pachystigma reflects its repeat DNA composition. Chromosome Research, 2021.

LEE, S.-I.; KIM, N.-S. Transposable elements and genome size variations in plants. Genomics & Informatics, v. 12, n. 3, p. 87, 2014.

LISACHOV, A. et al. New karyotype of *Lygosoma bowringii* suggests cryptic diversity. Herpetology Notes, v. 11, p. 1083-1088, 2018.

MADDISON, W. P.; MADDISON, D. R. Mesquite: a modular system for evolutionary analysis. Evolution, v. 62, p. 1103-1118, 2008.

MATA-SUCRE, Y. et al. How diverse is heterochromatin in the Caesalpinia group? Cytogenomic characterization of *Erythrostemon hughesii* Gagnon & GP Lewis (Leguminosae: Caesalpinioideae). Planta, v. 252, p. 1-14, 2020.

MCKAIN, M. R. et al. Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). American Journal of Botany, v. 99, n. 2, p. 397–406, 2012.

MCNEAL, J. R. et al. Systematics and plastid genome evolution of the cryptically photosynthetic parasitic plant genus *Cuscuta* (Convolvulaceae). BMC Biology, v. 5, n. 1, p. 55, 2007.

MORAES, A. P. et al. Evolutionary trends in Iridaceae: new cytogenetic findings from the New World: Evolutionary trends in Iridoideae. Botanical Journal of the Linnean Society, v. 177, n. 1, p. 27–49, 2015.

MOWER, J. P. et al. Horizontal acquisition of multiple mitochondrial genes from a parasitic plant followed by gene conversion with host mitochondrial genes. BMC Biology, v. 8, n. 1, p. 150, 2010.

NAGAKI, K. et al. Sequencing of a rice centromere uncovers active genes. Nature genetics, v. 36, n. 2, p. 138, 2004.

NEUMANN, P. et al. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mobile DNA, v. 10, n. 1, p. 1, 2019.

NEUMANN, P. et al. Impact of parasitic lifestyle and different types of centromere organization on chromosome and genome evolution in the plant genus Cuscuta. New Phytologyst, v. 4, p. 2365–2377, 2020.

NOVÁK, P.; NEUMANN, P; MACAS, J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. Nature protocols, v. 15, n. 11, p. 3745-3776, 2020.

NOVAK, P. et al. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics, v. 29, n. 6, p. 792–793, 2013.

OLIVEIRA, L. et al. Mitotic spindle attachment to the holocentric chromosomes of *Cuscuta europaea* does not correlate with the distribution of CENH3 chromatin. Frontiers in plant science, v. 10, p. 1799, 2020.

ONG-ABDULLAH, M. et al. Loss of Karma transposon methylation underlies the mantled somaclonal variant of oil palm. Nature, v. 525, n. 7570, p. 533, 2015.

PARDUE, M.-L.; DEBARYSHE, P. G. Retrotransposons that maintain chromosome ends. Proceedings of the National Academy of Sciences, v. 108, n. 51, p. 20317–20324, 2011.

PAZY, B.; PLITMANN, U. Holocentric chromosome behaviour in *Cuscuta* (Cuscutaceae). Plant Systematics and Evolution, v. 191, n. 1, p. 105–109, 1994.

PAZY, B.; PLITMANN, U. Chromosome divergence in the genus *Cuscuta* and its systematic implications. Caryologia, v. 48, n. 2, p. 173–180, 1995.

PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Satellite DNA evolution. In: Repetitive DNA. Karger Publishers, v. 7, p. 126–152, 2012.

PUCCI, M. B. et al. Dispersion of transposable elements and multigene families: Microstructural variation in *Characidium* (Characiformes: Crenuchidae) genomes. Genetics and Molecular Biology, v. 41, n. 3, p. 585–592, 2018.

REVELL, L. J. phytools: an R package for phylogenetic comparative biology (and other things): phytools: R package. Methods in Ecology and Evolution, v. 3, n. 2, p. 217–223, 2012.

RIBEIRO, T. et al. Evolutionary dynamics of satellite DNA repeats from *Phaseolus* beans. Protoplasma, v. 254, n. 2, p. 791–801, 2017a.

SADER, M. A. et al. The role of chromosome changes in the diversification of *Passiflora* L. (Passifloraceae). Systematics and Biodiversity, v. 17, n. 1, p. 7–21, 2019.

SCHUBERT, I. Chromosome evolution. Current Opinion in Plant Biology, v. 10, n. 2, p. 109–115, 2007.

SCHUBERT, I.; LYSAK, M. A. Interpretation of karyotype evolution should consider chromosome structural constraints. Trends in Genetics, v. 27, n. 6, p. 207–216, 2011.

SMITH, J. et al. Differences in gene density on chicken macrochromosomes and microchromosomes. Animal Genetics, v. 31, n. 2, p. 96–103, 2000.

STEBBINS, G. L. Chromosomal evolution in higher plants. Chromosomal evolution in higher plants, 1971.

STEFANOVIĆ, S.; COSTEA, M. Reticulate evolution in the parasitic genus *Cuscuta* (Convolvulaceae): over and over again. Botany, v. 86, n. 8, p. 791–808, 2008.

STEFANOVIC, S.; KUZMINA, M.; COSTEA, M. Delimitation of major lineages within *Cuscuta* subgenus *Grammica* (Convolvulaceae) using plastid and nuclear DNA sequences. American Journal of Botany, v. 94, n. 4, p. 568–589, 2007.

SUN, G. et al. Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. Nature communications, v. 9, n. 1, p. 1-8, 2018.

TENAILLON, M. I.; HOLLISTER, J. D.; GAUT, B. S. A triptych of the evolution of plant transposable elements. Trends in Plant Science, v. 15, n. 8, p. 471–478, ago. 2010.

TREANGEN, T. J.; SALZBERG, S. L. Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nature Reviews Genetics, v. 13, n. 1, p. 36–46, 2012.

UTSUNOMIA, R. et al. A Glimpse into the satellite DNA library in Characidae fish (Teleostei, Characiformes). Frontiers in Genetics, v. 8, p. 103, 2017.

VAIO, M. et al. Molecular phylogeny and chromosome evolution among the creeping herbaceous *Oxalis* species of sections Corniculatae and Ripariae (Oxalidaceae). Molecular Phylogenetics and Evolution, v. 68, n. 2, p. 199–211, 2013.

VAIO, M. et al. Multiple karyotype changes distinguish two closely related species of *Oxalis* (*O. psoraleoides* and *O. rhombeo-ovata*) and suggest an artificial grouping of section Polymorphae (Oxalidaceae). Botanical Journal of the Linnean Society, v. 188, n. 3, p. 269-280, 2018.

VAN-LUME, B. et al. Heterochromatic and cytomolecular diversification in the *Caesalpinia* group (Leguminosae): Relationships between phylogenetic and cytogeographical data. Perspectives in Plant Ecology, Evolution and Systematics, v. 29, p. 51–63, 2017.

VAUGHN, K. C. Attachment of the parasitic weed dodder to the host. Protoplasma, v. 219, n. 3–4, p. 227–237, 2002.

VONDRAK T. et al. Complex sequence organization of heterochromatin in the holocentric plant *Cuscuta europaea* elucidated by the computational analysis of nanopore reads. Computational and Structural Biotechnology Journal, v. 19, p. 2179–2189, 2021.

WEISS-SCHNEEWEISS, H. et al. Employing next generation sequencing to explore the repeat landscape of the plant genome. Next Generation Sequencing in Plant Systematics. Regnum Vegetabile, v. 157, p. 155-179, 2015.

WEISS-SCHNEEWEISS, H.; SCHNEEWEISS, G. M. Karyotype diversity and evolutionary trends in Angiosperms. Plant Genome Diversity Volume 2. Vienna: Springer Vienna, p. 209–230, 2013.

WU, J. Composition and structure of the centromeric region of Rice Chromosome 8. The plant cell online, v. 16, n. 4, p. 967–976, 2004.

YUNCKER, T. G. Revision of the North American and West Indian species of *Cuscuta*. Illinois biological monographs; v. 6, no. 2, p. 1923, 1921.

YUNCKER, T. G. The Genus *Cuscuta*. Memoirs of the Torrey Botanical Club, v. 18, n. 2, p. 109, 111, 113–331, 1932.

ZHANG, Y. et al. The Diversity of sequence and chromosomal distribution of new transposable element-related Segments in the Rye genome revealed by FISH and lineage annotation. Frontiers in Plant Science, v. 8, p. 1706, 2017.

APÊNDICE A – ARTIGO PUBLICADO NA REVISTA CHROMOSOME RESEARCH

Karyotype asymmetry in *Cuscuta* L. subgenus *Pachystigma* reflects its repeat DNA composition

Amalia Ibiapino[1], Mariana Báez[1,2], Miguel A. García[3], Mihai Costea[4], Saša Stefanović[5] and Andrea Pedrosa-Harand[1]*

[1]Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Federal University of Pernambuco, Recife, PE, Brazil
[2]Present address: Plant Breeding Department, University of Bonn, Bonn, Germany
[3]Real Jardín Botanico-CSIC, Madrid, Spain
[4]Department of Biology, University of Wilfrid Laurier, Waterloo, Ontario, Canada
[5]Department of Biology, University of Toronto Mississauga, Mississauga, Ontario, Canada

Short title: Karyotype asymmetry in *Cuscuta* L. subgenus *Pachystigma*

Corresponding author's address:
Andrea Pedrosa-Harand
Laboratório de Citogenética e Evolução Vegetal, Departamento de Botânica, Universidade Federal de Pernambuco – UFPE
R. Prof. Moraes Rego, s/n, CDU, 50670-420, Recife, PE, Brazil
Tel. number: + 55 81 2126 8846
Fax number: + 55 81 2126 8358
E-mail: andrea.harand@ufpe.br

ORCIDs:
Amalia Ibiapino 0000-0002-2613-5259
Mariana Baez 0000-0002-7874-6385
Miguel A. García 0000-0002-0366-043X
Mihai Costea 0000-0003-3049-1763
Saša Stefanović 0000-0001-8290-895X
Andrea Pedrosa-Harand 0000-0001-5213-4770

Key message

Cuscuta subgenus *Pachystigma* contains species with strikingly bimodal karyotypes. The emergence of these karyotypes is linked to the enrichment of varied repetitive sequences in the largest chromosomal pairs.

Abstract *Cuscuta* is a cytogenetically diverse genus, with karyotypes varying 18-fold in chromosome number and 89-fold in genome size. Each of its four subgenera also presents particular chromosomal features, such as bimodal karyotypes in *Pachystigma*. We used low coverage sequencing of the *Cuscuta nitida* genome (subgenus *Pachystigma*), as well as chromosome banding and molecular cytogenetics of three subgenus representatives, to understand the origin of bimodal karyotypes. All three species, *C. nitida*, *C. africana* (2*n* = 28) and *C. angulata* (2*n* = 30), showed heterochromatic bands mainly in the largest chromosome pairs. Eighteen satellite DNAs were identified in *C. nitida* genome, two showing similarity to mobile elements. The most abundant were present at the largest pairs, as well as the highly abundant ribosomal DNAs. The most abundant Ty1/Copia and Ty3/Gypsy elements were also highly enriched in the largest pairs, except for the Ty3/Gypsy CRM, which also labelled the pericentromeric regions of the smallest chromosomes. This accumulation of repetitive DNA in the larger pairs indicates that these sequences are largely responsible for the formation of bimodal karyotypes in the subgenus *Pachystigma*. The repetitive DNA fraction is directly linked to karyotype evolution in *Cuscuta*.

Keywords: bimodal karyotypes; chromosomal evolution; heterochromatin; repetitive DNA; repeat amplification

Declarations:

Funding information: The present study received financial support and fellowships from CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico), CAPES (Coordenação de Pessoal de Nível Superior: Finance Code 001) and FACEPE (Fundação de Amparo a Ciencia e Tecnologia de PE). The field work was supported in part by the Natural Sciences and Engineering Research Council of Canada Discovery grant (no. 326439).

Conflict of interest: The authors declare that they have no conflict of interest.

Ethics approval: Not applicable

Consent to participate: Not applicable

Consent for publication: All authors approved the final version of the manuscript.

Availability of data and material: Raw sequences reads available at NCBI, accession number PRJNA749043

Code availability: Not applicable

Abbreviations:

CMA Chromomycin A3

DAPI 4,6-Diamidino-2-phenylindole

Gbp Giga base pair

FISH Fluorescence in situ hybridization

LTR Long Terminal Repeat

PCR Polymerase chain reaction

rDNA Ribosomal DNA

satDNA Satellite DNA

Introduction

Most angiosperms have symmetric karyotypes (Stebbins 1971; Weiss-Schneeweiss and Schneeweiss 2013), with chromosomes similar in size and morphology. However, several lineages have been known to have asymmetrical karyotypes, characterized by centromeres at different positions along chromosomes, or chromosomes of different sizes. Karyotypes with two sets of chromosomes markedly different in size are called bimodal and represent the extreme of karyotype asymmetry (Stebbins 1971; McKain et al. 2012). Three main hypotheses were suggested for the origin of bimodal karyotypes. The first mechanism involves chromosomal rearrangements such as fusion-fission events. Thus, larger chromosomes can be the product of small chromosome fusions or small chromosomes the result of fission of larger ones (Schubert and Lysak 2011). *Lygosoma bowringii* Günther 1864, a lizard of the Scincidae family, has $2n = 32$, with 18 macrochromosomes and 14 microchromosomes. As in most reptiles, macrochromosomes have originated from the fusion of microchromosomes (Lisachov et al. 2018, 2020). The second mechanism is allopolyploidy involving parental species with different chromosomal sizes, suggested for some genera such as *Agave* L. (McKain et al. 2012). The third possibility is the progressive amplification of repetitive DNA sequences in one set of chromosomes (de la Herrán et al. 2001). In *Eleutherine* Herb. (Iridaceae), for example, an accumulation of different families of repetitive sequences in the larger chromosome pair was suggested to be the cause of the differences between both chromosome sets (Báez et al. 2019).

Repetitive DNA constitutes a large fraction of plant genomes and can be found either organized in tandem (micro-, mini- and satellite DNAs), or dispersed through the genome (transposons and retrotransposons) (Heslop-Harrison and Schwarzacher 2011). Transposable elements are capable of moving within the genome, impacting genome structure and even the function of genes (Bourque et al. 2018). The highly abundant repetitive sequences are frequently associated with heterochromatin formation located at the (peri-) centromeres, subtelomeres and in interstitial heterochromatic blocks (Barros e Silva et al. 2010; Van-Lume et al. 2019). The effects of transposable element and satellite DNA accumulation on genomes are dynamic and can lead to significant increase in genome size. In species of the genus *Zea* L., the accumulation of repetitive DNAs families, mainly LTR retrotransposons like Ty3/Gypsy, resulted in a two-times larger genome in *Zea luxurians* (Durieu) R.M.Bird in relation to *Z. mays* L. and *Z. diploperennis* Iltis, Doebley & R. Guzmán in less than two million years (Estep et al. 2013).

Satellite DNAs are composed by monomers that are oriented head-to-tail and can vary in length, nucleotide composition, sequence complexity, and abundance. These sequences frequently form clusters that can rapidly change in number, position and size (Garrido-Ramos 2015; Biscotti et al. 2015). A mutation that occurs within a monomer can spread among the repeat units or be eliminated by homogenization (Plohl et al. 2012). The mechanism of concerted evolution, for example, can generate varied patterns of repetitive DNA families, producing in general homogeneity within species and diversity between species. Thus, different species can have different families of satellite DNAs or these satellites can be shared between related species (Feliner and Rosselló 2012; Plohl et al. 2012). Furthermore, these tandem repeats can be species-specific or even chromosome specific. Some tandem repeats are highly conserved among species, such as the 5S and 35S ribosomal DNAs (rDNAs), which encode for the ribosomal RNAs. But most repetitive families are usually non-coding sequences evolving rapidly and generating genomic differentiation (Biscotti et al. 2015).

The parasitic genus *Cuscuta* L. (Convolvulaceae Juss.) includes some 200 species, divided into four subgenera: *Grammica* (Lour.) Peter, Engl. & Prantl, *Pachystigma* (Engelm.) Baker & C.H. Wright, *Cuscuta* Yunck, and *Monogynella* (Des Moul.) Peter, Engl. & Prantl (García et al. 2014; Costea et al. 2015a). Subgenus *Grammica*, with about 150 species, has almost exclusive distribution in the Americas. *Pachystigma* includes only five species, all endemic to South Africa, and *Cuscuta* is native to Europe, Africa, and Asia, with a few species introduced and naturalized in the Americas, Australia, and New Zealand. Subgenus *Monogynella* had its origin in Central Asia from where it dispersed to S, E and SE Asia, Europe, Africa, and one species, *C. exaltata* Engelm., is native to south-eastern North America (García et al. 2014; Costea et al. 2015b).

The genus *Cuscuta* shows high cytogenetic variation in chromosome number ($2n = 8$ to $2n = 150$), chromosome size (1.66 µm to 21.60 µm), and genome size (1C = 0.39 Gbp to 1C = 34.73 Gbp). The genus also presents symmetric to bimodal karyotypes, as well as monocentric and holocentric chromosomes (García and Castroviejo 2003; Guerra and García 2004; McNeal et al. 2007; Ibiapino et al. 2019, 2020;

García et al. 2019; Oliveira et al. 2020; Neumann et al. 2020). Each *Cuscuta* subgenus seems to have different karyotypic features. Species of subgenus *Monogynella* have the largest genome sizes and the largest chromosomes. Subgenus *Cuscuta* is the only one that has species with exclusively holocentric chromosomes. Subgenus *Grammica* presents the largest variation in chromosome number and size. This subgenus has at least five cases of interspecific hybridization which can contribute to this chromosome number variation (Fogelberg 1938; Pazy and Plitmann 1994; García 2001; García and Castroviejo 2003; McNeal et al. 2007; Ibiapino et al. 2019; García et al. 2019). A preliminary study of two species of *Pachystigma* revealed bimodal karyotypes and extensive heterochromatic blocks in the larger chromosomes, suggesting the influence of repetitive DNA in the emergence of bimodality in this subgenus (García et al. 2019). An asymmetrical karyotype was also reported for some populations of the holocentric *C. epithymum* (L.) L. (subgenus *Cuscuta*), with $2n = 14$ individuals showing bimodal karyotype while $2n = 16$ individuals having symmetric karyotypes (García and Castroviejo 2003).

Species of the genus *Cuscuta* also vary in heterochromatin content, ranging from species with few bands and few rDNA sites, such as *C. denticulata* (Ibiapino et al. 2019), up to species with numerous bands, where heterochromatin may have contributed to the expansion of the genome size, as in *C. monogyna* and *C. indecora* (Ibiapino et al. 2020; Oliveira et al. 2020). In the latter two species, heterochromatin may have contributed to maintaining karyotype symmetry, since both have similar karyotypes, but belong to different subgenera (Ibiapino et al. 2020). Repeat DNA composition was investigated in 12 *Cuscuta* species, demonstrating that the extensive variation in genome size in species of this genus is caused by the differential accumulation of repetitive sequences (Neumann et al. 2020). However, no representatives of subgenus *Pachystigma* were included in that study.

Our current work investigates heterochromatin distribution in three of the five species of the subgenus *Pachystigma* (*C. nitida* E. Mey. ex Choisy, *C. africana* Thunb. and *C. angulata* Engelm.) and evaluates the repetitive DNA composition of *C. nitida* genome, in order to better understand the role played by repetitive DNA sequences in the emergence of bimodal karyotypes within this subgenus.

Materials and methods

Material

Flower buds of two accessions of *C. africana*, one of *C. angulata* and three of *C. nitida* (subgenus *Pachystigma*) were collected in November 2017 from the Cape region of South Africa, where they are endemic (Table 1). Vouchers were deposited at the herbaria of the University of Toronto Mississauga (TRTE) and Wilfrid Laurier University (WLU), Canada.

Slide preparation and CMA/DAPI double staining

Slides were prepared using flower buds collected and fixed in the field in Carnoy (ethanol: acetic acid, 3:1, v/v). The material was washed in distilled water, digested in an enzymatic solution containing 2% cellulase (Onozuka) and 20% pectinase (Sigma) for 40 minutes. The slides were prepared by air drying, mainly using the ovary wall, as described by De Carvalho and Saraiva (1993), with small modifications. After the material was macerated and dried, the slides were dipped in 60% acetic acid for up to 5 minutes to clear the cytoplasm. Finally, the slides were left at 37° C until completely dry.

For double CMA/DAPI staining, the slides were aged at room temperature for three days, stained with 8 µL of 0.1 mg/µL chromomycin A3 (CMA) for 60 minutes, mounted in 8 uL of 1µg/mL 4', 6-diamidino-2-phenylindole (DAPI) in mounting medium (glycerol:McIlvaine buffer pH 7.0, 1:1, v/v), and aged again for three days at room temperature. The images were captured with a COHU CCD camera attached to a Leica DMLB fluorescence microscope equipped with Leica QFISH software. After image capture, slides were destained for 30 minutes in Carnoy, for one hour in absolute ethanol and stored at -20º C for *in situ* hybridization.

For chromosomal measurements, five metaphases of *C. nitida* were used. Chromosomes were measured with the ruler tool in Adobe Photoshop CS3 version 10.0.

DNA extraction and *in silico* repetitive DNA analysis

Cuscuta nitida genomic DNA was extracted following Doyle and Doyle (1987) protocol. Sequencing of the total genomic DNA generated low coverage (0.01×), 250-bp paired-end reads in an Illumina HiSeq 2500 (BGI, Hong Kong, China). Repetitive DNA analysis was performed by the RepeatExplorer pipeline (https://galaxy-elixir.cerit-sc.cz/; Novak et al. 2013), where reads showing at least 95% similarity in at least 55% of its length were clustered together.

Clusters showing an abundance greater than 0.01% were automatically annotated and manually checked. Clusters similar to plastomes or mitogenomes were considered putative contamination and excluded from the final annotation. All contigs with tandem repetitions identified by TAREAN (Novák et al. 2017), as well as other satellites not identified by this tool, but which presented typical satellite graph layouts after clustering, were confirmed with DOTTER (Sonnhammer and Durbin 1995). High abundance dispersed elements had their integrase domain identified using the NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi). The consensus sequences of the satellites and the integrase domains of the transposable elements were used for primer design using the primer design tool implemented in Geneious version 7.1.9 (Kearse et al. 2012).

The consensus sequences of all identified satellites were compared to verify their homology. The consensus monomers that showed similarity in DOTTER were aligned using Muscle in Geneious. Different satellite families were considered as part of the same superfamily when monomer sequences showed identity between 50% and 80%. Sequences with 80-95% similarity were considered subfamilies of the same family and similarity greater than 95% were considered variants of the same family (Ruiz-Ruano et al. 2016). As two of these satellites showed similarity with transposable elements, alignments were made of the consensus satellite sequence with the most similar transposable element domains indicated by the RepeatExplorer. Two of the satellites that showed similarity with transposable elements also showed *in situ* colocalization with the 35S rDNA cluster. Therefore, a comparison of the satellite consensus sequence with a putative *C. campestris* (GenBank accession number PRJEB19879) 35S rDNA consensus sequence, assembled using the NOVOPlasty algorithm (Dierckxsens et al. 2017) was included. This assembly was made using Illumina reads obtained from Vogel et al. (2018). After assembled, the complete 35S rDNA was aligned with the satellite consensus sequence using Muscle in Geneious. Satellites were named as follows: code referring to the species name (Cn), followed by "Sat", a number referring to the abundance order, and the size of the consensus monomer in base pairs.

Repeat amplification, probe preparation and *in situ* hybridization (FISH)

Polymerase chain reaction (PCR) for repeat amplification was performed in 50 µL reactions containing 1 ng of *C. nitida* genomic DNA, 1× PCR buffer (20 mM Tris-HCl pH 8.4, 50 mM KCl), 2 mM $MgCl_2$, 0.1 mM dNTPs, 0.4 µM of each primer, 0.4× TBT (750 mM trehalose, 1 mg/ml BSA, 1% Tween 20, 8.5 mM Tris hydrochloride) and 0.025 µL of a homemade Taq Polymerase. Amplification program was 1× 94ºC for three minutes, plus 30 cycles of 94 ºC for one minute, 55-65 ºC for one minute (see Supplementary Table 1 for annealing temperatures of each primer pair) and 72 ºC for one minute, followed by a final extension of 7 min at 72 ºC. Primers for amplification of satellite superfamilies SF1 and SF2 annealed in a conserved region shared between the satellite variants from each superfamily. All satellite DNA primers were designed facing outwards, so amplification was additional evidence of a tandem organization. The transposable elements primers were designed using the sequences of the integrase domain of each element (the sequences of the integrase domains can be found in Supplementary Table 2).

The PCR products were sequenced to confirm its identity and labelled in a total volume reaction of 12.5 µL, containing 1 µg of amplified DNA, 1× Nick Translation buffer (0.5 M Tris HCl pH 7.5; 50 mM $MgCl_2$), dNTP mix (0.016 mM each of dATP, dCTP, dGTP), 0.08 mM Cy3-dUTP or Alexa-dUTP, 7,5 U of DNA Polymerase I and 0,006 U of DNase I. The mixture was incubated at 15º C for one hour or longer if needed, until most fragments were under 500 bp, and reactions were stopped using 0.5 M EDTA.

For rDNA probes, the plasmids D2 of *Lotus japonicus* (Regel) K. Larsen (5S rDNA) and p*Ta*71 of wheat (25-28S, 5.8S and 18S rDNA) were used (Pedrosa et al. 2002; Sousa et al. 2011). Probes were labelled by Nick translation with Cy3-dUTP (5S), as described above, and digoxigenin 11-dUTP (35S) with a Nick Translation kit (Invitrogen – Oregon, USA).

Fluorescence *in situ* hybridizations followed Pedrosa et al. (2002). The hybridization mixture, composed of 50% formamide, 10% dextran sulphate, 2× SSC, and 5 ng/µl probe, was denatured at 75ºC for 10 minutes. Slides were denatured for 5 minutes with the hybridization mixture and hybridized for 18-20 hours at 37ºC in a humid chamber. Final stringency was 76% for 5S and 35S rDNA and satellite DNAs. The 35S rDNA probe was detected with anti-digoxigenin produced in sheep, conjugated with FITC (Roche - Basel, Switzerland) and the signal amplified with anti-sheep IgG produced in rabbit conjugated with FITC (Serotec - California, EUA). The slides were mounted as described above. With 76% stringency, transposable element probes did not show strong, clear signals. For this reason, these elements were hybridized with lower stringency (40%), as described in Ribeiro et al. (2017a).

To verify the putative localization of telomeric sequences at interstitial chromosome sites, the ND-FISH protocol described by Cuadrado et al. (2009) was applied. Thirty µL of the hybridization solution containing 2 pmol (25 ng) of the diluted probe ($TTTAGGGTTTAGGGTTTAGGGTTTAGGGT_5$ directly labelled with Cy3, Macrogen, Seoul, Korea) in 2× SSC was added per slide and cover with a coverslip. The slide was incubated 2 h at room temperature protected from light. The coverslip was removed with 2× SSC, washed in 4× SSC/0.2% Tween 20 at room temperature for 10 minutes under agitation, and mounted in DAPI with mounting medium as described above. The satellites and rDNA images were captured as previously described. For transposable elements, images were captured using an epifluorescence Leica DMLB microscope equipped with a COHU 4912-5010 CCD Camera using the Leica QFISH software.

Results

Bimodal karyotypes are a typical feature of the subgenus *Pachystigma*, with large chromosomes enriched in heterochromatin

All three samples of *C. nitida* analysed showed $2n = 28$ and bimodal karyotypes (4L+24S), with two larger chromosomes pairs (average sizes of 12.34 µm and 8.19 µm) and 12 smaller chromosome pairs (average size of 2.67 µm). The second largest pair harbours a proximal nucleolus organizing region (NOR), as evidenced by the large decondensed region between both chromosome arms. The total haploid complement size was 107.6 µm.

For heterochromatin characterization, CMA/DAPI double staining was performed in the three species. *Cuscuta nitida* showed only two pairs of heterochromatic bands, a large $CMA^+/DAPI^-$ band in the short arm of the largest chromosome pair (Fig. 2 a-c) and a second $CMA^+/DAPI^-$ band colocalised with the NOR in the second chromosome pair (Fig. 2 b-c). *In situ* hybridization with 5S and 35S rDNA revealed one large pair of 5S site, co-localized with the CMA^+ band at the short arm of the largest pair (Fig. 3 a-b). One major and one minor pair of 35S rDNA sites were observed, both in the large chromosome pairs. The major 35S site was observed highly decondensed, co-localised with a CMA^+ band in the second largest chromosome pair, while the minor site, not always visible, was observed proximally at the largest pair (Fig. 3 b-d).

Cuscuta africana and *C. angulata* also showed bimodal karyotypes (Fig. 2 d-i). *Cuscuta africana* exhibited karyotype similarities with *C. nitida* ($2n = 28$, 4L +24S), but the heterochromatic band present in the largest pair was $DAPI^+/CMA^-$ (Fig. 2 d-f). On the other hand, *C. angulata* ($2n = 30$, 10L+20S) presented ten large chromosomes in its karyotype. These chromosomes have numerous heterochromatic bands, mainly $DAPI^+/CMA^-$. In addition, the set of smaller chromosomes of this species has pericentromeric bands, mainly $DAPI^+/CMA^-$ (Fig. 2 i). Thus, the heterochromatin characterization evidenced the presence of bimodal karyotypes in the three analysed species of *Pachystigma*, with a high number of heterochromatic bands in the large chromosome pairs, with different compositions, GC or AT rich, depending on the species.

The repetitive fraction of *C. nitida* genome is rich in tandem repeats

To understand the composition of the repetitive DNA fraction and its relation to the heterochromatin content, we performed genome skimming in *C. nitida* and characterized the most abundant DNA repeats. A total of 5,156,846 reads were generated, of which 1,173,600 reads were randomly sampled by RepeatExplorer for analysis. A total of 53,117 clusters were identified and 330 clusters, containing at least 0.01% of genome abundance, were grouped into 322 superclusters (Fig. S1) and annotated. Five and 33 clusters that showed similarity to mitochondrial and plastid sequences, respectively, were excluded from further analysis (Table 2).

The repetitive fraction corresponded to 42.83% of the *C. nitida* genome. It was possible to annotate 198 of the 330 clusters, the rest (6.56% of the total genome) remained unclassified (Table 2). The dispersed repetitive DNA sequences corresponded to 22.01% of the total genome. LTR-retrotransposons from the Ty1/Copia superfamily comprised 5.6% of the genome, while Ty3/Gypsy elements were 1.6 times more abundant (8.86%). Within Ty1/Copia, the SIRE lineage was the most abundant with 3.5%, while the Tekay lineage was the most represented among Ty3/Gypsy with 4.19%. LTR elements without a clear lineage classification corresponded to 5.13% of the total genome. LINEs corresponded to 0.96%, while Class 2 transposable elements corresponded to 1.46%, with CACTA being the most abundant (1.05%).

Among tandem repeats, the 5S rDNA showed a large abundance, comprising 4.75% of the *C. nitida* genome, even larger than the 35S rDNA (3.1%). Other tandem repeats (satDNA) corresponded to 6.4% of the genome. TAREAN identified six clusters with high-confidence satellites and six with low-confidence. Another six clusters, not identified by the TAREAN, showed typical circular graphs and were confirmed as tandem repeats by a self-similarity dot-plot analysis. The comparative dot-plot with the consensus sequences of the 18 identified satDNAs revealed some sequences with similarity to each other and were grouped into three superfamilies (Table 3; Fig. S2 and Fig. S3).

Superfamily 1 (SF1), with 2.44% abundance, is composed of two satellites: CnSat1-213, classified by TAREAN, and CnSat2-295, not identified by TAREAN, showing 68.6% similarity between consensus sequences. The superfamily 2 (SF2) is composed of three satellites classified with high-confidence by TAREAN, CnSat3-111, CnSat4-115 and CnSat8-125, showing 72.3% similarity among consensus sequences and representing together 1.08% of the genome (Fig. 1). The third superfamily is composed of two satellites, one classified with low confidence and one not identified by TAREAN, CnSat11-1357 and CnSat15-990, together corresponding to 0.07% of the genome (Table 3). The consensus sequences are provided in Supplementary Table 2. In addition to these satellites, some microsatellites were identified and described in Supplementary Table 3. Two satellites classified with high confidence by TAREAN showed similarity to retrotransposon domains. The alignment of the consensus sequence of these satellites with these domains showed 32.2% identity between CnSat10-1400 and the Reverse Transcriptase (RT) domain of a LINE element, while CnSat12-1060 showed 39.5% identity to the Ribonuclease (RH) domain of Ty1/Copia Reina lineage.

Mapping repetitive sequences

Different repeats were selected for investigating their chromosomal distribution and putative association with heterochromatin and the largest chromosome pairs. Apart from the 5S and 35S rDNA, four others satDNA were selected: superfamily 1 (SF1), the most abundant among satDNAs, superfamily 2 (SF2), and two tandem repeats that showed similarity with mobile elements. Superfamily SF1 signals colocalized with 35S rDNA in both large chromosome pairs. The SF1 signals, however, were stronger and more extended than the rDNA signal, occupying the proximal region on the long arm of the largest pair (Fig. 3f). Superfamily SF2 is also located in the largest chromosome pair, presenting two signals on each homologue. These signals flanked the 5S rDNA sites (Fig. h). The 5S rDNA and SF2 sites occupy most of the short arm of the largest chromosome pair. The CnSat10-1400 satellite showed a small signal in the pericentromeric region of the short arm of the largest pair and a larger signal in the distended region of the second pair, similar to the 35S rDNA. (Fig. 3g Despite colocalization of satellites SF1 and CnSat10-1400 with the 35S rDNA, these satellites did not show any significant *in silico* similarity with the 35S rDNA assembled from *C. campestris* (2.4% to CnSat1-213, 3.6% to CnSat2-295 and 29.9% to CnSat10-1400, identity to the aligned sequence, Fig. 1). CnSat12-1060, on the other hand, showed no evident chromosome hybridization (data not shown). The ND-FISH with telomeric probe showed terminal signals in all chromosomes of the complement, but no interstitial signals that could suggest previous chromosome fusions (Fig. 3f).

The most abundant LTR retrotransposon lineages, SIRE (Ty1/Copia), Tekay and Retand (Ty3/Gypsy), as well as the putative centromeric CRM lineage (Ty3/Gypsy), were also selected for hybridization in *C. nitida* chromosomes, showing slightly different patterns. The SIRE element showed signals along the entire length of the largest chromosome pair, with a gap in the 5S rDNA site. In addition, it also labelled the distal regions of the second largest pair, not including the proximal NOR (Fig. 4a). The Retand element showed a similar pattern, but it also displayed dispersed signals along the distended region of the NOR and along the 5S rDNA cluster (Fig. 4b). CRM labelled the largest chromosomes pairs, but also showed a weak labelling of the small chromosomes, slightly enriched in the pericentromeric region at least in some of them (Fig. 4c). Tekay element showed scattered proximal signals on the largest chromosome pairs and no signal in the NOR (Fig. 4d).

Combined, these data demonstrate the enrichment of the large chromosome pairs of the *C. nitida* karyotype with tandem (rDNAs and satDNAs) and disperse (LTR retrotransposons) repetitive sequences. Several of these repeats colocalize in the largest pairs, evidencing a complex chromosome organization of this large chromosomes and indicating them as cause for the bimodal karyotype in this species.

Discussion

All three analysed species - *C. nitida*, *C. africana*, and *C. angulata* - presented bimodal karyotypes. Although the remaining two species, *C. gerrardii* Baker and *C. natalensis* Baker, should be analysed in the future for confirmation, the presence of bimodal karyotype is likely a synapomorphy of *Cuscuta* subgenus *Pachystigma*. The phylogenetic relationships within the subgenus resolved *C. nitida* as sister to a clade with *C. natalensis* and *C. gerrardii* (García et al. 2014), with this clade sister to *C. africana* + *C. angulata*. This suggests that this karyotypic feature was maintained in the whole subclades, thus supporting the hypothesis that all species of this subgenus share bimodal karyotypes. In fact, the four subgenera of *Cuscuta* are not only delimited by phylogenetic, biogeographic, and morphological data, but each present unique cytogenetic peculiarities, such as the presence of holocentric chromosomes in the subgenus *Cuscuta* (García et al. 2014; Costea et al. 2015; García et al. 2019).

Much of the cytogenetic studies in the genus *Cuscuta* are restricted to conventional staining techniques with an emphasis on chromosome counting. Fewer studies conducted more detailed cytogenetic analyses, such as CMA/DAPI banding for characterization of heterochromatin. The latter studies revealed a numerical variation of these bands, with species having few bands, like *C. denticulata* Engelm. with only one pair of evident $CMA^+/DAPI^-$ bands, to species like *C. monogyna* Vahl, with approximately 90 $CMA^+/DAPI^-$ bands and 80 $DAPI^+/CMA^-$ bands in pachytene (Ibiapino et al. 2019, 2020). Despite this variation, our previous unpublished data showed that karyotypes with less bands are more frequent, even in polyploid species. In most of those cases, CMA^+ bands are found in pericentromeric regions and are colocalized with 5S and 35S rDNA. In the species with a higher number of $CMA^+/DAPI^-$ or $DAPI^+/CMA^-$ bands, they are localised in interstitial regions. There is no evidence that this banding pattern is different for each subgenus and there may be similar patterns between different subgenera (Ibiapino et al. 2020). In the three species of the subgenus *Pachystigma* analysed here, the multiple CMA/DAPI bands were mainly present at the largest pairs. These bands do not differ much in number and position from those already reported in the genus, however, in these bimodal karyotypes, the bands are larger. In *C. nitida*, for example, the largest $CMA^+/DAPI^-$ band occupies a large part of the short arm of the largest chromosome pair. *Cuscuta africana* showed a similar pattern, however the largest heterochromatic band was $DAPI^+/CMA^-$. The large number of heterochromatic bands on the smallest chromosomes is not observed in the other two species of the *Pachystigma* subgenus. This characteristic may indicate an incipient accumulation of heterochromatin in these chromosomes, which could eventually lead to a less asymmetrical karyotype, such as the amplification observed in other unrelated *Cuscuta* species, such as *C. indecora* Choisy (*Grammica* subgenus) and *C. monogyna* (*Monogynella* subgenus) (Ibiapino et al. 2019), indicating that this character is homoplastic in the genus. Alternatively, karyotype asymmetry in *Pachystigma* may be maintained by an unknown mechanism.

Bimodality in *Pachystigma* is not due to chromosome fusion. Although *C. nitida* has $2n = 28$, lower than the basic number proposed for the genus *Cuscuta*, which is $x = 15$ (Pazy and Plitmann 1995), the sizes of the two largest pairs cannot be explained by a single fusion of two pairs of small chromosomes.

In addition, the ND-FISH with telomeric probe did not provide evidence for any interstitial sites in *C. nitida*, which may indicate that there was no fusion event in the origin of this karyotype. Furthermore, *Cuscuta angulata* presented $2n = 30$, showing no reduction in chromosome number and a bimodal karyotype. Many *Cuscuta* species have $2n = 30$, but there are species with $2n = 8, 10, 14, 16, 18, 20, 28, 30, 32, 34$ and polyploids with $2n = 28, 42, 44, 56, 60, 90, 150$. *Cuscuta epithymum* (L.) L. (subgenus *Cuscuta*), for instance, shows an intraspecific variation which could be attributed to chromosome fusions and polyploidy, with $2n = 14, 16, 28, 30, 32$ and 34. Individuals with $2n = 14$ and $2n = 32$ are bimodal, while $2n = 16$ and $2n = 34$ are symmetric (García and Castroviejo 2003; García et al. 2019). Three species of subgenus *Cuscuta* (holocentric), *C. epithymum* ($2n = 14$), *C. europaea* L. ($2n = 14$), and *C. epilinum* Weihe ($2n = 6x = 42$), had a reduction in the chromosome number. In the case of this subgenus, there may have been chromosomal fusion events, since holocentric chromosomes have diffuse kinetochores, and consequently these chromosomes can stabilize fragments or fused chromosomes favouring rearrangements (Mandrioli and Manicardi 2020).

Bimodal karyotypes may also originate through interspecific hybridization, as proposed for the genus *Agave*, in which allopolyploid species might have chromosomes of different sizes inherited from different parents (McKain et al. 2012). In *Cuscuta*, there are numerous cases of interspecific hybridization and polyploidy (reviewed by García et al. 2014). For example, *C. veatchii* Brandegee is an allopolyploid originated from the hybridization of *C. denticulata* and *C. nevadensis* I.M. Johnst. With $2n = 60$, *C. veatchii* possess 30 smaller chromosomes and 30 slightly larger chromosomes with very evident centromeres, characteristic of *C. denticulata*, and *C. nevadensis*, respectively (Ibiapino et al. 2019). However, molecular phylogenetic analyses have showed that reticulate evolution occurs mainly in the subgenus *Grammica* (e.g., Stefanović and Costea 2008; Costea and Stefanović 2010; García et al. 2014; Costea et al. 2015a). There is also preliminary phylogenetic evidence suggesting that some species of subgenus *Cuscuta* may have a hybrid origin. Different accessions of *C. approximata*, for example, have polymorphism in the ITS, and the location of *C. kurdica* differed between the ITS and *trnL* trees (García and Martín 2007). These contrasting topologies may indicate hybridization events similar to those reported in subgenus *Grammica* (e.g., Stefanović and Costea 2008; García et al. 2014). Evidence such as this has not been observed in *C. africana*, *C angulata*, and *C. nitida* with ITS, 26S, *trnL* nor *rbcL* sequences analyses (García and Martín 2007). Chromosome number and size, as well as the number of rDNA sites in *C. nitida*, were within the range of variation already reported for species of the genus *Cuscuta*. So far, most *Cuscuta* species have shown few rDNA sites, varying from two to 36 sites of 5S rDNA and from two to 30 sites of 35S rDNA (Fogelberg 1938; García 2001; García and Castroviejo 2003; Guerra and García 2004; McNeal et al. 2007; Ibiapino et al. 2019, 2020; García et al. 2019). Thus, our results do not suggest neither hybridization nor polyploidy as the cause of bimodality in *Pachystigma*.

Cuscuta africana presented divergence in chromosome number compared to previous report ($2n = 30$, García et al. 2019). Intraspecific variation is unlikely because samples were plants collected from the same population. It is more likely that the $2n = 30$ reported earlier was a mistake, since conventional staining may leave the proximal, distended NOR unnoticed. NORs are more easily identified as $CMA^{+}/DAPI^{-}$ bands (see, for example Fig. 2d and 2f). Similar miscounts have been registered for *Passiflora foetida* L., which was first described as having $2n = 22$ (Snow and MacDougal 1993) and later corrected to $2n = 20$ (De Melo and Guerra 2003).

Because chromosome fusions and intraspecific hybridization seem less probable, repetitive sequences accumulation in specific chromosome pairs could be a probable mechanism for karyotype asymmetry in subg. *Pachystigma*. Indeed, repetitive DNA in *Cuscuta* is involved in the expansion of the genome, causing an increase in chromosomes, such as in *C. monogyna* and *C. indecora* (Ibiapino et al. 2020; Neumann et al. 2021). In these two cases, however, chromosomes increased proportionally in size, maintaining karyotype symmetry, and resulting in similar karyotypes, although *C. monogyna* and *C. indecora* belong to different subgenera (*Monogynella* and *Grammica*, respectively). All 12 *Cuscuta* species sequenced by Neumann et al. (2021) showed a greater abundance of LTR type elements, with SIRE being the most dominant among Ty1/Copia lineages and Tekay most dominant among Ty3/Gypsy, ranged from 8.5% to 30.8% for Ty1/Copia and 7.3% to 28.5% for Ty3/Gypsy. Many small genome species such as *C. pentagona* Engelm. showed a higher proportion of Class II elements, representing 12.6% of the genome. A large fraction of these genomes was also composed of satellite DNA, reaching up to 18% in *C. europaea*. In this species, the satDNA CUS-TR24 is the major constituent of its heterochromatic bands (Vondrak et

al. 2021). Similar results were observed for *C. nitida*, with 3.5% SIRE and 4.19% Tekay. Satellites also made up a significant percentage of the *C. nitida* genome, 6.4%. However, Class II elements showed a low proportion, 1.46%. These demonstrate that repetitive accumulation is a common mechanism in the evolution of genomes within *Cuscuta* genus, increasing the size of chromosomes within a particular karyotype.

Bimodal karyotypes can maintain a differential composition of DNA sequences between large and small chromosomes, thus creating subgenomes different in structure and function (Baez et al. 2019). In chicken, for example, the recombination rate in microchromosomes is much higher than in macrochromosomes. The high density of genes favours meiotic recombination and could lead to an increase of gene density in small chromosomes (Smith et al. 2000; Rodionov et al. 2002). On the other hand, large chromosomes may be richer in heterochromatin. Within these regions, repetitive sequences are more prone to accumulate because new insertions are not selected against and less likely to be removed by recombination (Biscotti et al., 2015). The evident accumulation of repetitive DNA sequences in the largest chromosomal pairs of *C. nitida* supports the influence of heterochromatin in the karyotype asymmetry of *Cuscuta*. In the bimodal karyotypes of subgenus *Pachystigma*, the most evident heterochromatic bands are restricted to the largest pairs. The 5S rDNA corresponded to 4.75% of the genome of *C. nitida* and was colocalized with the largest CMA^+ band of that species. In addition, all hybridized satellite DNAs, as well as the 35S rDNA sites and most of the transposable elements, are restricted or highly enriched in the largest chromosomal pairs (Fig. 5). In *Muscari* Mill. (Asparagaceae), a massive amplification of the MCSAT satDNA family occurred in only one chromosome pair. This single satDNA family corresponds to 5% of the total genome of *M. comosum* (L.) Mill. and contributed to the progressive increase in the karyotype asymmetry of *Muscari* species (de la Herrán et al. 2001). In *Eleutherine*, two of the *E. bulbosa* satellites, Ebusat1 and Ebusat4, occur in the interstitial region of the largest pair of *E. bulbosa* and *E. latifolia*, both with bimodal karyotypes. In addition, the four most abundant retrotransposons also showed accumulation in the larger pair. This demonstrates that accumulation of repetitive sequences can generate an increase of only part of the chromosomes of a karyotype and lead to a change in karyotype symmetry (Báez et al. 2019). This suggests that the bimodality of *Pachystigma* subgenus could also originate from the asymmetric expansion of multiple repetitive DNA lineages (Fig. 5).

The SF1 and CnSat10-1400 signals colocalized with the 35S rDNA but showed no similarity with the 35S rDNA of *C. campestris*, which is a species of subg. *Grammica*. This may suggest that this satellite DNA unit has originated from tandem duplications of a less-conserved, intergenic region of *Cuscuta* rDNA, such as the IGS, or that it was inserted in *C. nitida* rDNA locus after the divergence between subgenera *Grammica* and *Pachystigma*. In *Phaseolus* L. (Fabaceae), jumper satDNA was inserted into the NTS region of 5S rDNA (Ribeiro et al. 2017a). However, the 35S rDNA and CnSat10-1400 formed independent clusters, suggesting that this satDNA has not become part of the rDNA unit but is rather interspersed along the rDNA site. Nevertheless, no 35S rDNA cluster from *C. nitida* presented a circular graph, indicating its incompleteness. Therefore, the association of this satellite with the 35S of *C. nitida* cannot be excluded and await further investigations using long reads. Furthermore, CnSat10-1400 showed higher similarity with the reverse transcriptase domain of a LINE element, which may indicate a possible origin of this satellite from a TE, and later interspersion within the 35S rDNA loci, or the insertion of LINEs in this satDNA, as observed in *Cuscuta europaea* for CUS-TR24 (Vondrak et al. 2021).

The satellites of *C. nitida* CnSat10-1400 and CnSat12-1060 showed similarity with transposable elements, LINE and Reina, respectively. Some transposable elements and repetitive genes can contribute to the formation and dissemination of satellite DNAs. In *Lathyrus sativus* L., most of the satellites originated from small tandem repetitions present in the 3' untranslated region of the Ogre retrotransposons (Vondrak et al. 2020). MITE transposable elements were appointed as generators of satellite DNA in bivalve molluscs and *Drosophila* (Miller et al. 2000; Pons 2004). Similarly, in ants of the genus *Messor*, a Mariner element gave rise to the expansion of satellite DNA IRE-130 (Palomeque and Lorite 2008). In fishes, copies of the 5S rDNA originated the satellite 5SHindIII. Ancestors of tRNA were probably responsible for the formation of tandemly repeated sequences in higher plants (López-Flores and Garrido-Ramos 2012). In humans, it has been identified that a quarter of all mini/satellites are derived from transposable elements. TE-derived satellites usually have monomers above the standard 500 bp of size and generally occupy pericentromeric regions (Meštrović et al. 2015). This is the case of CnSat10-1400 and CnSat12-1060, with monomers of 1,400 bp and 1,060 bp, respectively. In *Pisum sativum*, variants of the

satellite PisTR-A are incorporated into Ty3/Gypsy Ogre elements. The untranslated region that separates the 3' gag-pol domains from the LTR is highly variable in the pea Ogre elements and carries several other tandem repeats (Macas et al. 2009). In maize, the CRM1TR and CRM4TR tandem repeats are entirely derived from centromeric retrotransopsons (CM) (Sharma et al. 2013). None of the satellites found in *C. nitida* showed similarity to satellites previously described by Oliveira et al. (2020) and Neumann et al. (2021) either between species of the closest subgenus *Cuscuta*, or between the other subgenera„ indicating that these satellites have independent origins and the composition of heterochromatic bands in holocentric and monocentric chromosomes of the genus or between different subgenera are different.

The well-supported clade of subgenus *Pachystigma* is characterized by the presence of bimodal karyotypes in all species analysed. Although the three species had different CMA/DAPI band patterns, these bands were more enriched in the larger chromosomes of the three karyotypes. The genome organization of *C. nitida* repetitive fraction suggested a differential chromosome accumulation of diverse repetitive families, mainly satDNA, rDNA and retrotransposons, as the probable mechanism of origin for the bimodal karyotypes within this subgenus. This shows that the increase in chromosomes, which led to the emergence of bimodality in this clade, was associated to the accumulation of repetitive sequences in heterochromatin. The composition of this heterochromatin may be different among species. In *Cuscuta*, the amount and diversity of repetitive DNA is high and satellite DNAs can originate from transposable elements and potentially be incorporated or interspersed with the rDNA.

Acknowledgments

We thank the Fundação de Amparo a Ciência e Tecnologia de Pernambuco (FACEPE) for the financing of the Postgraduate scholarship; the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) for the financial support for the development of the project. Computational resources for the Galaxy based RepeatExplorer were provided by the ELIXIR-CZ project (LM2015047), part of the international ELIXIR infrastructure. The field work was supported in part by the Natural Sciences and Engineering Research Council of Canada Discovery grant (no. 326439).

References

Báez M, Vaio M, Dreissig S, et al (2019) Together but Different: The subgenomes of the bimodal *Eleutherine* karyotypes are differentially organized. Front Plant Sci 10: 1170. https://doi.org/10.3389/fpls.2019.01170

Barros e Silva AE, Marques A, dos Santos KGB, Guerra M (2010) The evolution of CMA bands in *Citrus* and related genera. Chromosome Res 18: 503–514. https://doi.org/10.1007/s10577-010-9130-2

Biscotti MA, Olmo E, Heslop-Harrison JS (2015) Repetitive DNA in eukaryotic genomes. Chromosome Research 23: 415–420. https://doi.org/10.1007/s10577-015-9499-z

Bourque G, Burns KH, Gehring M, et al (2018) Ten things you should know about transposable elements. Genome Biology 19: 199. https://doi.org/10.1186/s13059-018-1577-z

Costea M, García MA, Baute K, Stefanović S (2015a) Entangled evolutionary history of *Cuscuta pentagona* clade: A story involving hybridization and Darwin in the Galapagos. TAXON 64: 1225–1242. https://doi.org/10.12705/646.7

Costea M, García MA, Stefanović S (2015b) A Phylogenetically based infrageneric classification of the parasitic plant genus *Cuscuta* (Dodders, Convolvulaceae). Systematic Botany 40: 269–285. https://doi.org/10.1600/036364415X686567

Costea M, Stefanović S (2010) Evolutionary history and taxonomy of the *Cuscuta umbellata* complex (Convolvulaceae): Evidence of extensive hybridization from discordant nuclear and plastid phylogenies. Taxon 59: 1783–1800. https://doi.org/10.1002/tax.596011

Cuadrado Á, Golczyk H, Jouve N (2009) A novel, simple and rapid nondenaturing FISH (ND-FISH) technique for the detection of plant telomeres. Potential used and possible target structures detected. Chromosome Res 17: 755–762. https://doi.org/10.1007/s10577-009-9060-z

De Carvalho CR, Saraiva LS (1993) An air drying technique for maize chromosomes without enzymatic maceration. Biotech Histochem 68: 142–5. https://doi.org/10.3109/10520299309104684

de la Herrán R, Robles F, Cuñado N, et al (2001) A heterochromatic satellite DNA is highly amplified in a single chromosome of *Muscari* (Hyacinthaceae). Chromosoma 110: 197–202. https://doi.org/10.1007/s004120000115

De Melo NF, Guerra M (2003) Variability of the 5S and 45S rDNA sites in *Passiflora* L. species with distinct base chromosome numbers. Ann Bot 92: 309–316. https://doi.org/10.1093/aob/mcg138

Dierckxsens N, Mardulyn P, Smits G (2017) NOVOPlasty: de novo assembly of organelle genomes from whole genome data. Nucleic Acids Res 45: e18. https://doi.org/10.1093/nar/gkw955

Doyle J, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical Bulletin 19: 11-15

Estep MC, DeBarry JD, Bennetzen JL (2013) The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. Heredity 110: 194–204. https://doi.org/10.1038/hdy.2012.99

Feliner GN, Rosselló JA (2012) Concerted evolution of multigene families and homoeologous recombination. In: Wendel JF, Greilhuber J, Dolezel J, Leitch IJ (eds) Plant Genome Diversity 1rd edn. Springer Vienna, Vienna, pp 171–193

Fogelberg SO (1938) The cytology of *Cuscuta*. Bulletin of the Torrey Botanical Club 65: 631. https://doi.org/10.2307/2481064

García MA (2001) A new western Mediterranean species of *Cuscuta* (Convolvulaceae) confirms the presence of holocentric chromosomes in subgenus *Cuscuta*. Botanical Journal of the Linnean Society 135: 169–178. https://doi.org/10.1111/j.1095-8339.2001.tb01089.x

García MÁ, Castroviejo S (2003) Estudios citotaxonómicos en las especies Ibéricas del género *Cuscuta* (Convolvulaceae). Anales del Jardín Botánico de Madrid 60: 33-44

García MA, Costea M, Kuzmina M, Stefanović S (2014) Phylogeny, character evolution, and biogeography of *Cuscuta* (dodders; Convolvulaceae) inferred from coding plastid and nuclear sequences. American Journal of Botany 101: 670–690. https://doi.org/10.3732/ajb.1300449

García MA, Martín MP (2007) Phylogeny of *Cuscuta* subgenus *Cuscuta* (Convolvulaceae) based on nrDNA ITS and chloroplast *trnL* intron sequences. Systematic Botany 32: 899–916. https://doi.org/10.1600/036364407783390872

García MÁ, Costea M, Guerra M, García-Ruiz I, Stefanović S (2019) IAPT chromosome data 31. TAXON 68:1374–1380. https://doi.org/10.1002/tax.12176

Garrido-Ramos MA (2015) Satellite DNA in plants: More than just rubbish. Cytogenetic and Genome Research 146: 153–170. https://doi.org/10.1159/000437008

Guerra M, García MA (2004) Heterochromatin and rDNA sites distribution in the holocentric chromosomes of *Cuscuta approximata* Bab. (Convolvulaceae). Genome 47: 134–140. https://doi.org/10.1139/g03-098

Heslop-Harrison JSP, Schwarzacher T (2011) Organisation of the plant genome in chromosomes: Organisation of the plant genome in chromosomes. The Plant Journal 66: 18–33. https://doi.org/10.1111/j.1365-313X.2011.04544.x

Ibiapino A, García MÁ, Costea M, et al (2020) Intense proliferation of rDNA sites and heterochromatic bands in two distantly related *Cuscuta* species (Convolvulaceae) with very large genomes and symmetric karyotypes. Genet Mol Biol 43: e20190068. https://doi.org/10.1590/1678-4685-gmb-2019–0068

Ibiapino A, García MA, Ferraz ME, et al (2019) Allopolyploid origin and genome differentiation of the parasitic species *Cuscuta veatchii* (Convolvulaceae) revealed by genomic in situ hybridization. Genome 62: 467–475. https://doi.org/10.1139/gen-2018-0184

Kearse M, Moir R, Wilson A, et al (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28: 1647–9. https://doi.org/10.1093/bioinformatics/bts199.

Lisachov A, Poyarkov N, Pawangkhanant P, et al New karyotype of *Lygosoma bowringii* suggests cryptic diversity. Herpetology Notes 11: 1083–1088

Lisachov AP, Tishakova KV, Romanenko SA, et al (2020) Whole-chromosome fusions in the karyotype evolution of *Sceloporus* (Iguania, Reptilia) are more intense in sex chromosomes than autosomes. Transactions of the Royal Society B 376: 1833. https://doi.org/10.1098/rstb.2020.0099

López-Flores I, Garrido-Ramos MA (2012) The Repetitive DNA Content of Eukaryotic Genomes. In: Garrido-Ramos MA (ed) Genome Dynamics, 7rd edn. S. KARGER AG, Basel, pp 1–28

Macas J, Koblížková A, Navrátilová A, Neumann P (2009) Hypervariable 3′ UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. Gene 448: 198–206. https://doi.org/10.1016/j.gene.2009.06.014

Mandrioli M, Manicardi GC (2020) Holocentric chromosomes. PLoS Genet 16: e1008918. https://doi.org/10.1371/journal.pgen.1008918

McKain MR, Wickett N, Zhang Y, et al (2012) Phylogenomic analysis of transcriptome data elucidates co-occurrence of a paleopolyploid event and the origin of bimodal karyotypes in Agavoideae (Asparagaceae). American Journal of Botany 99: 397–406. https://doi.org/10.3732/ajb.1100537

McNeal JR, Kuehl JV, Boore JL, de Pamphilis CW (2007) Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. BMC Plant Biology 7: 57. https://doi.org/10.1186/1471-2229-7-57

Miller WJ, Nagel A, Bachmann J, Bachmann L (2000). Evolutionary dynamics of the SGM transposon family in the *Drosophila obscura* species group. Mol Biol Evol 17: 1597–1609. https://doi.org/10.1093/oxfordjournals.molbev.a026259

Neumann P, Oliveira L, Čížková J, et al (2021) Impact of parasitic lifestyle and different types of centromere organization on chromosome and genome evolution in the plant genus *Cuscuta*. New Phytol 229: 2365–2377. https://doi.org/10.1111/nph.17003

Novák P, Ávila Robledillo L, Koblížková A, et al (2017) TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic acids research 45: e111. https://doi.org/10.1093/nar/gkx257

Novak P, Neumann P, Pech J, et al (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29:792–793. https://doi.org/10.1093/bioinformatics/btt054

Oliveira L, Neumann P, Jang T-S, et al (2020) Mitotic spindle attachment to the holocentric chromosomes of *Cuscuta europae*a does not correlate with the distribution of CENH3 chromatin. Front Plant Sci 10: 1799. https://doi.org/10.3389/fpls.2019.01799

Palomeque T, Lorite P (2008) Satellite DNA in insects: a review. Heredity 100: 564–573. https://doi.org/10.1038/hdy.2008.24

Pazy B, Plitmann U (1994) Holocentric chromosome behaviour in *Cuscuta* (Cuscutaceae). Plant Systematics and Evolution 191: 105–109. https://doi.org/10.1007/BF00985345

Pedrosa A, Sandal N, Stougaard J, et al (2002) Chromosomal map of the model legume Lotus japonicus. Genetics 161:1661–1672. https://doi.org/10.1093/genetics/161.4.1661

Plohl M, Meštrović N, Mravinac B (2012) Satellite DNA evolution. In: Garrido-Ramos MA (ed) Genome Dynamics, 7rd edn. pp 126–152Pons J (2004). Cloning and characterization of a transposablelike repeat in the heterochromatin of the darkling beetle *Misolampus goudoti*. Genome 47: 769–774. https://doi.org/10.1139/g04-019

Ribeiro T, dos Santos KGB, Richard MMS, et al (2017a) Evolutionary dynamics of satellite DNA repeats from *Phaseolus* beans. Protoplasma 254: 791–801. https://doi.org/10.1007/s00709-016-0993-8

Ribeiro T, Marques A, Novák P, et al (2017b) Centromeric and non-centromeric satellite DNA organisation differs in holocentric *Rhynchospora* species. Chromosoma 126: 325–335. https://doi.org/10.1007/s00412-016-0616-3

Rodionov AV, Lukina NA, Galkina SA, Solovei I, Saccone S (2002) Crossing over in chicken oogenesis: cytological and chiasma-based genetic maps of the chicken lampbrush chromosome 1 J Hered 93:125–129. doi: 10.1093/jhered/93.2.125

Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM (2016) High-throughput analysis of the satellitome illuminates satellite DNA evolution. Sci Rep 6: 28333. https://doi.org/10.1038/srep28333

Schubert I, Lysak MA (2011) Interpretation of karyotype evolution should consider chromosome structural constraints. Trends in Genetics 27: 207–216. https://doi.org/10.1016/j.tig.2011.03.004

Sharma A, Wolfgruber TK, Presting GG (2013) Tandem repeats derived from centromeric retrotransposons. BMC Genomics 14:142. https://doi.org/10.1186/1471-2164-14-142

Smith J, Bruley CK, Paton IR, Dunn I, Jones CT, Windsor D et al (2000) Differences in gene density on chicken macrochromosomes and microchromosomes. Anim Genet 31:96–103. doi: 10.1046/j.1365-2052.2000.00565.x

Snow N, MacDougal JM (1993) New chromosome reports in *Passiflora* (Passifloraceae). Systematic Botany 261–273. https://doi.org/10.2307/2419402

Sonnhammer EL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene 167: GC1–GC10. https://doi.org/10.1016/0378-1119(95)00714-8

Sousa A, e Silva AB, Cuadrado A, et al (2011) Distribution of 5S and 45S rDNA sites in plants with holokinetic chromosomes and the "chromosome field" hypothesis. Micron 42:625–631. https://doi.org/10.1016/j.micron.2011.03.002

Stebbins GL (1971) Chromosomal evolution in higher plants. Edward Arnold Ltd., London.

Stefanović S, Costea M (2008) Reticulate evolution in the parasitic genus *Cuscuta* (Convolvulaceae): over and over again. Botany 86: 791–808. https://doi.org/10.1139/B08-033

Van-Lume B, Mata-Sucre Y, Báez M, et al (2019) Evolutionary convergence or homology? Comparative cytogenomics of *Caesalpinia* group species (Leguminosae) reveals diversification in the pericentromeric heterochromatic composition. Planta 250: 2173–2186. https://doi.org/10.1007/s00425-019-03287-z

Vondrak T, Ávila Robledillo L, Novák P, et al (2020) Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. Plant J 101: 484–500. https://doi.org/10.1111/tpj.14546

Vondrak T, Oliveira L, Novák P, et al (2021) Complex sequence organization of heterochromatin in the holocentric plant *Cuscuta europaea* elucidated by the computational analysis of nanopore reads. Computational and Structural Biotechnology Journal 19: 2179–2189. https://doi.org/10.1016/j.csbj.2021.04.011

Weiss-Schneeweiss H, Schneeweiss GM (2013) Karyotype Diversity and Evolutionary Trends in Angiosperms. In: Greilhuber J, Dolezel J, Wendel JF (eds) Plant Genome Diversity 2rd edn. Springer Vienna, Vienna, pp 209–230

Table 1: Collection sites for three *Cuscuta* species of the subgenus *Pachystigma*

Species	Voucher	Site	GPS coordinates
Cuscuta nitida	*Stefanović SS-17-126*	West Cape; St. James, M75, Old Mule Path, 300 m from trailhead	34°07'01"S 18° 27'18"E
	Stefanović SS-17-143	West Cape; Stellenbosch, Jonkershoek Nature Reserve, 10 km loop	33°59'03"S 18° 57'03"E
	Stefanović SS-17-144	West Cape; West Coast NP, peninsula, sandy soil above house	33°10'03"S 18° 03'19"E
*C*uscuta *angulata*	*Stefanović SS-17-134*	West Cape; De Hoop Nature Reserve, 500 m on dirt rd S of main road to Infanta	34°25'32"S 20° 46'34"E
*C*uscuta *africana*	*Stefanović SS-17-141*	East Cape; just E of pay booth on N2, S side of the road, just passed red and white radio pole	33°57'19"S 23° 37'36"E
	Stefanović SS-17-138	West Cape; N of George, Outeniekwapass, Hwy N9/12, roadside park, downhill	33°53'27"S 22° 24'12"E

Table 2: General annotation of the Repeat type and genome proportion [%]of *Cuscuta nitida*

Repeat type	%
Unclassified	6.56
5S rDNA	4.75
35S rDNA	3.10
Satellites	6.40
LTR elements	
LTR non classified	5.13
Ty1/Copia	
Ale	0.54
Angela	0.25
Bianca	0.78
Ikeros	0.15
Ivana	0.01
SIRE	3.50
TAR	0.18
Tork	0.20
Ty3/Gypgy	
Athila	0.86
TatIII	0.03
Ogre	0.14
Retand	2.08
Tcn1	0.03
CRM	0.69
Galadriel	0.04
Tekay	4.19
Reina	0.80
Non LTR elements	
LINE	0.96
ClassII	
CACTA	1.05
hAT	0.07
MuDR_Mutator	0.08
Mariner	0.01
Helintron	0.25
Total	42.83

Table 3: Satellite DNA families and superfamilies identified in *C. nitida* genome, showing genome proportion (%), percentage of Guanine and Cytosine (%GC), TAREAN output and similarity to other repeats. First number in the name of the satellite represents its order of abundance, while the second number, the consensus size of its monomer sequence

SatDNA	Cluster	Superfamily	%	%GC	Confidence by TAREAN	Similarity to other repeats
CnSat1-213	CL8	SF1	1.51	49.8	Low	-
CnSat2-295	CL14	SF1	0.93	48.5	-	-
CnSat3-111	CL27	SF2	0.49	39.6	High	-
CnSat4-115	CL32	SF2	0.45	37.4	High	-
CnSat5-159	CL62		0.18	45.2	-	-
CnSat6-289	CL63		0.17	15.2	High	-
CnSat7-3702	CL67		0.15	39.9	Low	-
CnSat8-125	CL72	SF2	0.14	48.8	High	-
CnSat9-1185	CL78		0.12	33.8	Low	-
CnSat10-1400	CL92		0.08	38.4	High	LINE
CnSat11-1357	CL113	SF3	0.05	36.1	Low	-
CnSat12-1060	CL114		0.05	34.3	High	Reina
CnSat13-382	CL117		0.05	48.6	-	-
CnSat14-720	CL124		0.04	40.3	Low	-
CnSat15-990	CL166	SF3	0.02	37.9	-	-
CnSat16-300	CL171		0.02	30.3	-	-
CnSat17-526	CL178		0.02	36.4	-	-
CnSat18-472	CL284		0.01	35.8	Low	-

- Not identified

Figure legends

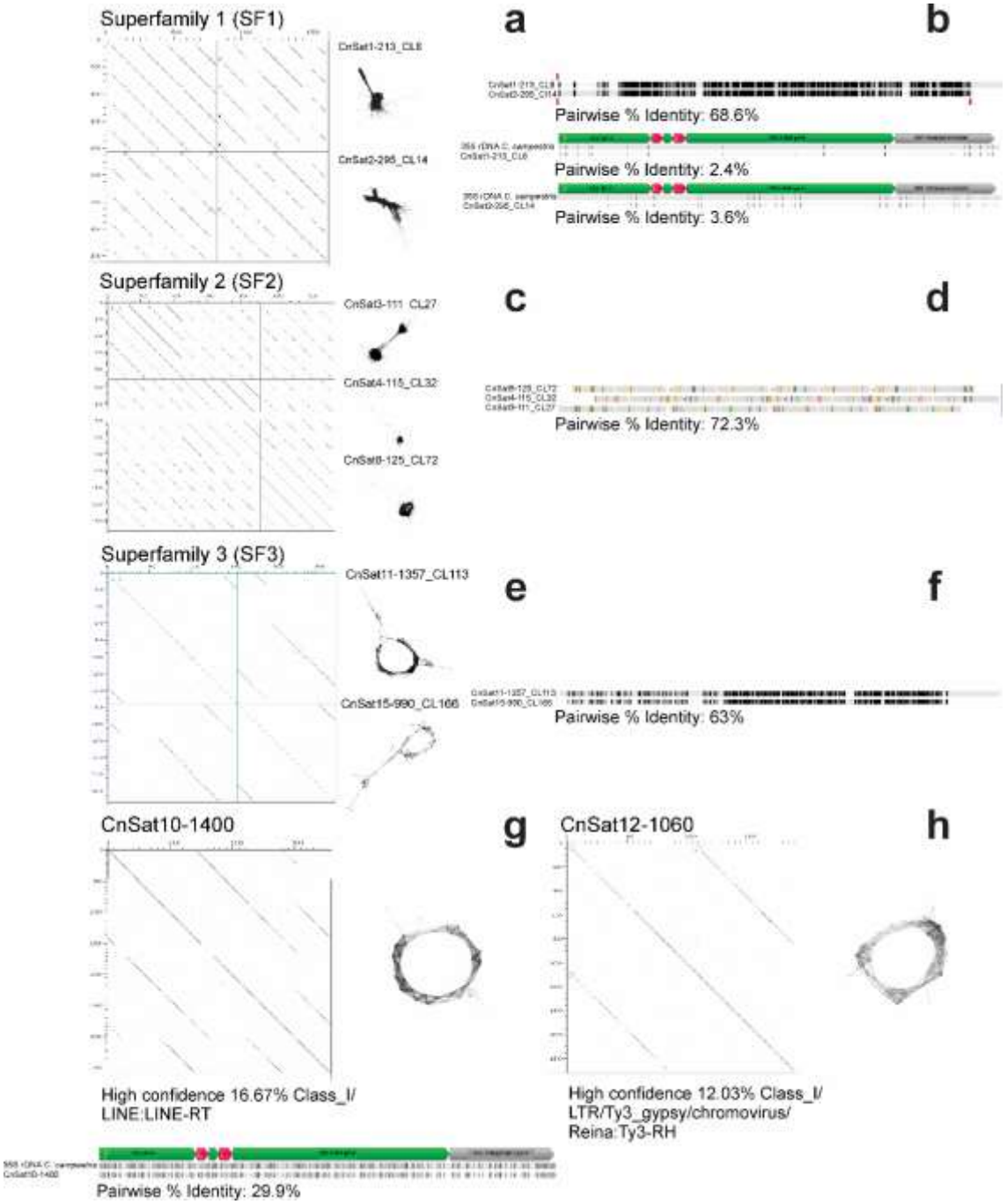


Fig. 1. DOTTER and cluster graphics of hybridized satellites in *C. nitida*. **a** and **b,** Superfamily 1; **c** and **d,** Superfamily 2; **e** and **f,** superfamily 3; **g,** satellite CnSat10-1400; and **h,** CnSat12-1060. **b**, **d** and **f** show the alignments and similarity between the satellites subunits that constitute Superfamilies 1, 2 and 3, respectively. **b** and **g**, similarity of the subunits with the 35S rDNA previously assembled from *C. campestris* for Superfamily 1 and CnSat10-1400, respectively

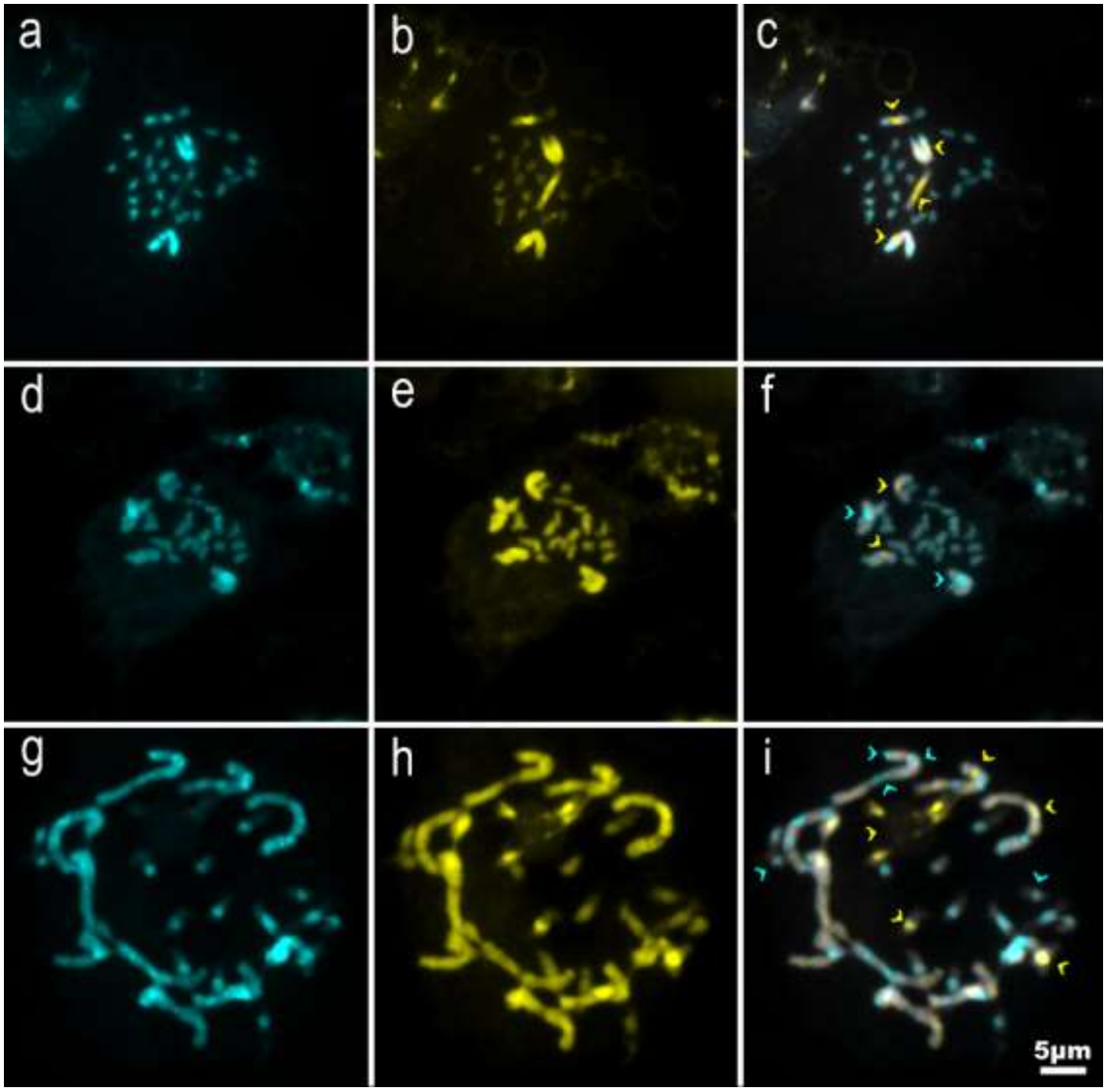


Fig. 2. Metaphases of *C. nitida* (**a**, **b** and **c**), *C. africana* (**d**, **e** and **f**) and *C. angulata* (**g**, **h** and **i**) stained with CMA (yellow) and DAPI (blue). Overlapping in **c**, **f** and **i**. Arrowheads in **c**, **f** and **i** highlight heterochromatic bands in each karyotype

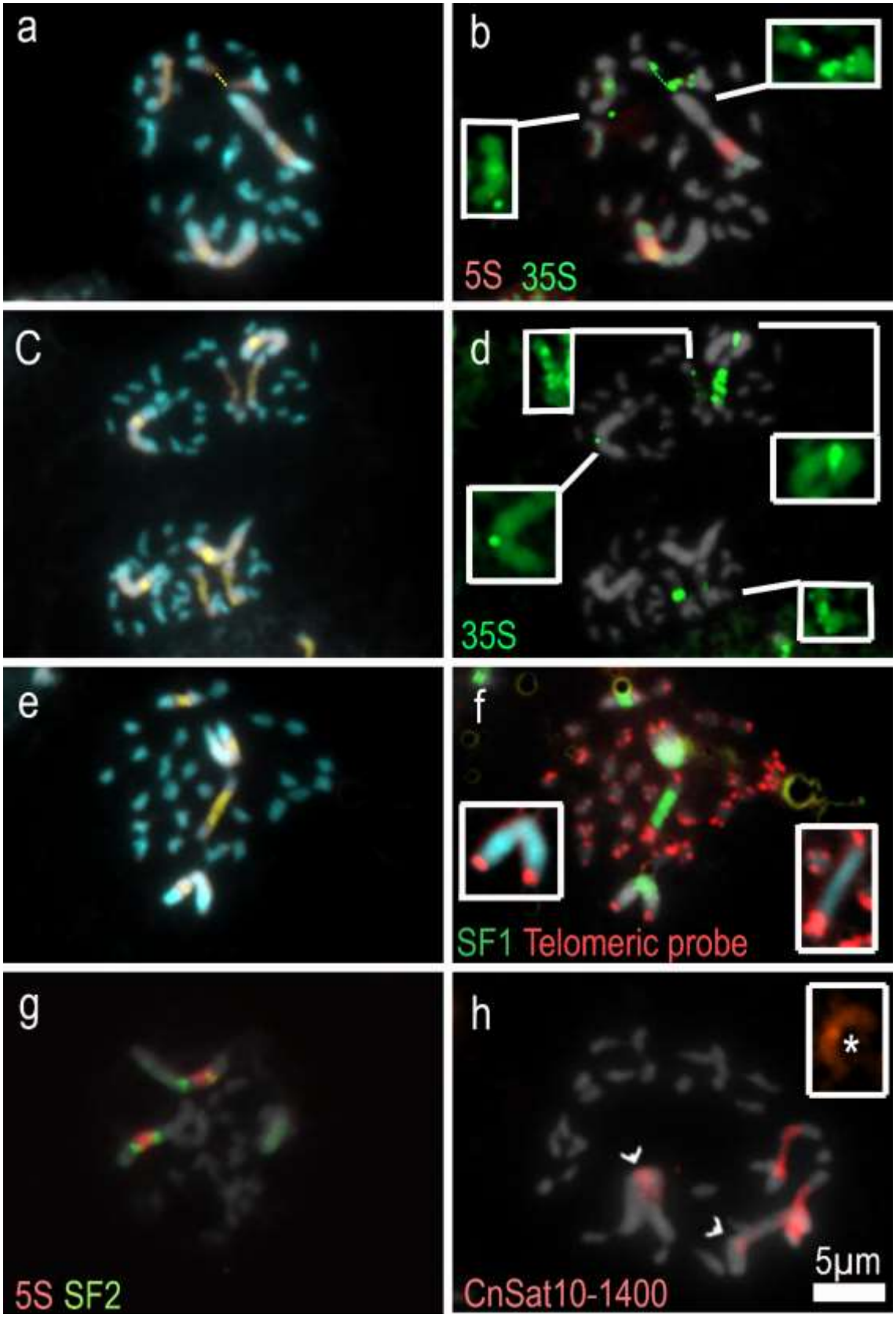


Fig. 3. Metaphases of *C. nitida* showing colocalization of CMA^+ bands (**a**, **c** and **e**) with 5S rDNA in red (**b**) and 35S rDNA in green (**b** and **d**) and SF1 superfamily in green (**f**). Insets in **b** and **d** show smaller 35S sites. In, **f**, the telomeric probe detected only terminal loci. In **h**, the 5S rDNA (in red) is flanked by the SF2 superfamily sites (in green). In **g**, CnSat10-1400 satellite signals in red; inset shows a detail of the satellite

signal on one of the chromosomes of the largest pair. Chromosomes were counterstained with DAPI (blue, **a**, **c** and **e**, or grey)

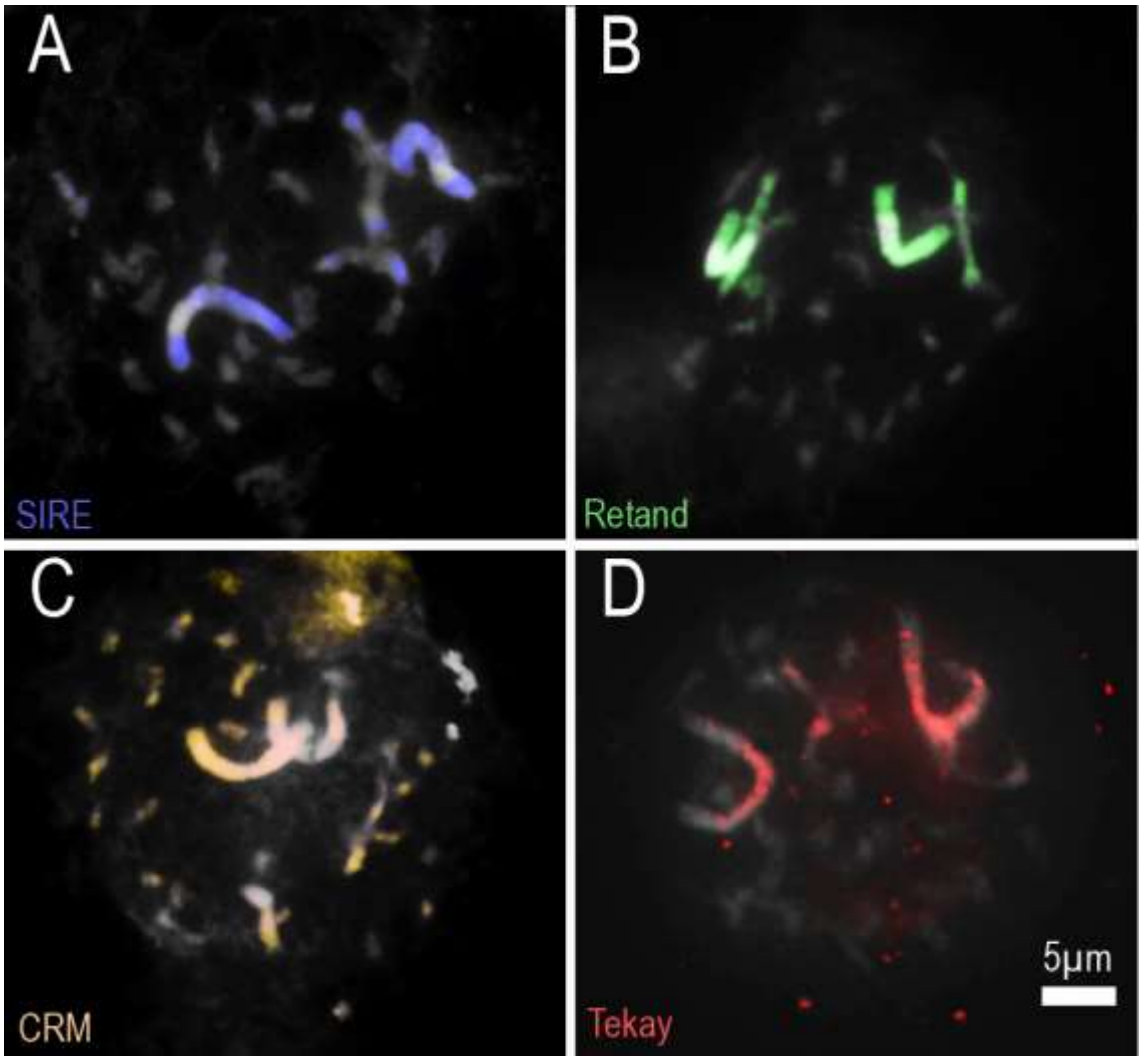


Fig. 4. Metaphases of *C. nitida* showing the distribution of LTR-retrotransposonswith an enrichment in the largest chromosome pairs. In **a**, the element Ty1/Copia SIRE (violet); in **b**, Ty3/Gypsy Retand (green); in **c**, Ty3/Gypsy CRM (yellow); and in **d**, Ty3/Gypsy Tekay (red). Chromosomes were counterstained with DAPI (grey)

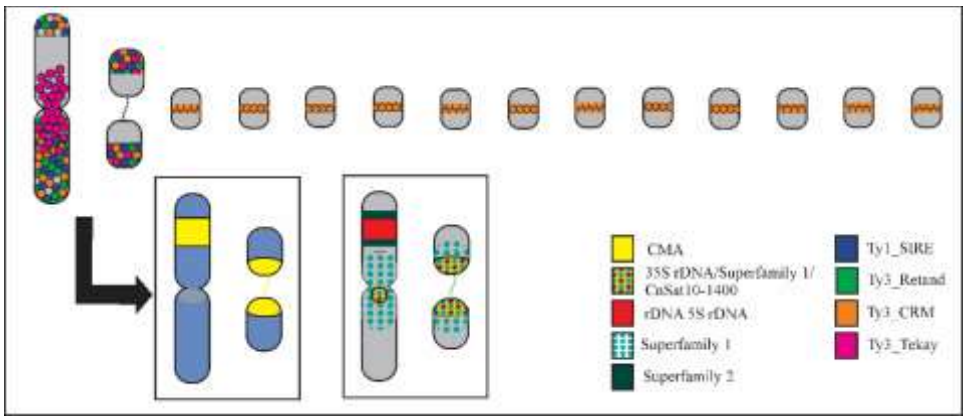


Fig. 5. Schematic representation of the distribution of heterochromatic bands and repetitive sequences in *C. nitida* bimodal karyotype. Ribosomal DNAs (5S e and 35S) and satellite superfamily SF1 colocalized with CMA^+ heterochromatin in the largest chromosome pairs, with SF1 and SF2 also adjacent to it. LTR-

retrotransposons were enriched in the largest pairs, except for the CRM element, also present in the smallest chromosomes

Supplementary data

Table S1. Primer pairs designed for probe amplification for FISH.

Repeats	Annealing temperatures	Forward sequence	Reverse sequence
SF1	65°C	ACGGAAGTCGCTTAGAGGGA	GCATACACCGTTTGTCTTCGT
SF2	55°C	TTGACTCAATGTTGCAAGGAATCC	GGTGAGACCCCGATCGAGTC
CnSat10-1400	55°C	GCAAACAAGCATAGTAGCCCAC	ATCGAGCTCCTGGCATTTGG
CnSat12-1060	55°C	AATGGGAGTAAGAAGTGGGCC	TGACCCATATTTACTCTCCATCTCT
CRM	60°C	AAGGTCCTATTAACTACCTCGGT	GGACTTTGTTTTGGGCTTGC
Retand	60°C	CCGTTCGCTAGATGGGGAAT	GCGATTTGCATTTTCTACTTGGC
SIRE	60°C	CAGTGTCCGATTTCTCCTCTCT	CGCCCATTGGAACTACTACATATG
Tekay	60°C	GAGCATGTCCTCTAAAATCTGGA	GGAAGTGGGAGAACATCACG

Table S2. Consensus sequences of the satellites and integrase domain of the main transposable elements described in *C. nitida*

Sat	Cluster	Consensus

CnSat1-213	CL8	AGGTGGTGGGACACTGAGGTACAGTTACGAAGACAAACGGTGTATGCTTTACGTTTGACCCTACGGAAG TCGCTTAGAGGGAGTTCGGCATGTCCGGTATGCTTGTCAAAGCCGGATAAACGTGACGCGACGGTTCCA TGAGAGAATAATTGACAAGCAATGACCGGGCATGTTGGACGCTGATGGCCGTTATGAATAGTCATAACG GCCATT
CnSat2-295	CL14	TCAAACGTGCTTGGGTCATTGCTTGTCAATTATTMTCTCRTGRAACCGTCGCGTCACGTTTATCGGGGGCT TTGACAAGCCTACMRGACATGCCAAACTCCCTCTAAGCSACTWTCGGAGGGTCGAACKTAAASCATRYWC SGTTCRTATTCRTTAMKGTAAMYCAGYRKCCMACGTGMTKGMCGTTTATGRMTRTTCAWARCGRCCATCA ACGTCCAACATGCCCAGCCATTGCTTGTCAATTATTCTCTCRYGGAACCGTYGCGTCAGGTTTAGCCGGCTTTG ACCAGCATAC
CnSat3-111	CL27	TAGCCAAATTTAAGAATTTTGGAGACTCGATCGGGGTCATACCCCATCTATCTTCCAAATAACGAACGGTTTTC TTCAAATTTGATCCAAGTAAGGATTCCCTCCAACACT
CnSat4-115	CL32	AGAACTTGAAGACTCGATCGGGATCTCATCCCATCTATCTCCAAAACGATGAACGGTAATGGTTCTTCAAATCC ATTCTAAGTAAGGATCACATCCAACACTTAAAACAAATTTA
CnSat5-159	CL62	AAATTTTGGGRAAGTTGGAGGATAAGATCCATATCTTGGCAGGGCAATTCGGGTTGGCCCGATAATTGCAG CCATGGTGGCTCTTATCTCGTCACTACTACCTSCTGGAAAAACAGATGCACCATATTTATTTAGAGGCTRTSAT GATTCCAGGCCACT
CnSat6-289	CL63	TTTTGCAATTTTTATTAAAAATTTACCAGACAACTCGAAAAGTTTTTGAAGATTGTGCTGTATATTTTATTAATT TTATTTTAATTAAATAAATCTATTCAGTTTTTTAATTGTTTTAATTATATTTTATTCATTTAGTTATTTAATTATTT TAAGTTACTATTGAACGTTTCATTTTTAAATTTCAATTATGTTTTGAACTTTTTAAAATTTTAATTATAAAATTTT CGTATATTATCTTTTCTCCAAGTGATTCAAATTTTAATTATTTTTTTCATTCGTTTAATTT

		GCCAGAGTCAGATGGTTGTCGCCGGGTTGTAGTCGCCGGAGTCAGGCAATGGTGGCCGAAGGCTGGTGG
		TTGTCCCCGGAGTCTAGGTCGCCGGAGTCCCGCGGATGTTGCCAGAGTTCGGCGATGATCGCTGGCATCTG
		GCAGTGGTCGCCGGAGTCCCGTGGTGGGCTCCGGAGTCCGACGGTGGTCTCTAGAGTATGGCAGTTTACTC
		TTTTAACATATTAATATTTTAAACCTTCTAAATGTTTCAAGGGTTTATACAATTATTTTAATATTTTAATCATTTA
		ATATTTTTTATTTATTATCCAAAATCATACGGAGTATAATTTGAAAAAATTAAATTAAATGGCTGGTCGTGTTA
		GGCATGGGTAGACAATCTGTCCCGTTTAAATGCCTAGGCCCAGCCCGACCCGACATCGATATTGAATTTTTTT
		AAGTCAGGCCCGAAGCCCGCGAATCCGGCCCATGTTATTTTTTATTAATTGTGATGTGAATTATATAATGTGA
		TTATGTGTTTGAGCCACTTTCATTATTAATATTGGGTCATTTATTTGGAGTATCTAATTCGAAATTATTTGATTA
		TCTAAAGACATAAAGAAGCACTATAATATTTTACTTGCATCACACTTACGGCTAATACTATAAATATATATCCTT
		AATCTCCAATTTTAAATAATTGCTTAATACGCTCTCCAATATGACCATTGGACTTTGTTTATATTTTAAACAATTG
		GTTAATAAGCTCTCCAAATATTACCGTTAGACTCTCTTTTTTAAAGGATAATACTGTTAGACTTGGAGATTGAAG
		AAACCGCTATCCTATAAATATATCGTTCTGCAATTTTAAATAATTGCTTAGCAAGAGTTCTCAAAATATTACCGTT
		AGACTTGGACATTGAAGAAGATGTCGACGATAGGTTTTCAAGTAGCTATGAATCCGCGATTCAGTGTCTCCGTG
		GTTGCCGCCGCCAACGGACAGAGACGCCACAGCAGCTATGTTCGAAAGGACAGAGAGTTCATTAGTGTCTCGC
		GAAAAAGTGTGGATATTTTTCCAGCACCGAAGCGAAGAAGATCGGCGGCTATAACGGTTGCGTGGCGTCTCCA
CnSat7-3702	CL67	GAAAAAGGAGAACGAATTTGTTAAAATTCTGAAACCGTCAGTTTTGCTTACGGAGTAATTCCCGTTAAAAACAG
		AGGAGTTGGTGCCGTTACTGGATATTTTAGCAAACAAAGTGAAGGCCGGAGCTGATGACGACCAAATTCCCTCC
		GGGAACATTCCGGTGAAGGTAAACTGGGCGAATTCACCGTTGTTTTTGATAAAATTATATTTGGGCCCCGCTCAT
		GATTTTTCGTTTGACCTAAAAAGCCATCATTGACTCTCTTCCCCCACCACAAGACTTCCTTTTGAATCTAAACGTGA
		ACCTATGGGACCCAATGCATGACACTGATGAACCTTATATTGCAGGTTCGTGATTCAGAGTTATATACTTAAAGA
		ATGAATCAATTGTCAAGTTCTGGAAGCCAGCAGGTCTCTATGAGAACCCTATTGCGACATTAGATTTGCATCATT
		CTATCATTCTATTATGATGGCCTATAAATAGTGTTATTGCACAGTGATAAAAATGTTTTCATTTTAGTTCCTCAAAG
		TCCAGCAGTTGATCACTTCCAAATATTATTTTACCACCAACTTCCAGTAGAAGCTGGGAAATTTAATGTAGTATCCA
		CAATGCGTAAATAAAAATCCGTCAGGTGGATATATGTCAAACAAAACTTACAGAAGGTTAGTTTTGTATCCATATA
		GGGGGTTTTTAATCATAATAATATAGTTATTTATTTAGTTATATAAATAAACAAATAATTGATGAAAGACTATAGAT
		GATTCCAGGGTGCTTACCTTTCACAATGTCTTCTGGACAAATTGATGTTCCTTAGGATGCCATATAACTAGTGCTATT
		GTACAGTGATAATATTGTTCTCGTTTGAGTCGCCTGATATTATTTCGAGCAGTTGATCATTTTCAAATATGATTGTAC
		AGGTAACTTAAAAAGAAGAAGTCGGATGGGTTTGGGTATAGGCTGTGTTTGGAATTCAGTGAGTTGTTTGCTTTCT
		TTCCACGTTGGTACTTTGGAGTTAGATACTTTATTAAAGAATGAATTGTTTTCAGGTTCTGGAAGCCAGAGAAGGAC
		TCTATGTGAAGCCTGTTGCCATATTAGAGTTTGCATCATTATATCCTTTTATTATGATGGCATATAACTAGTGATATTG
		TACATTGATAATATTGTTCTCGTTTGAATCGCCTGATATTTCCTGTTGATCAGATGATTCACACCGAAACAAGCACG

AGTCAGCAGAGTGCCAATGCTATACCTTCTGGTACTGATTGTTCAATTTTGGTTTCTACTCCGGAAATGTGTCAACA
AGGGAAACCAGCTATGCAGTAGCAACTTGTTGTGCATAGAAATTTAGCAACTGCTGTGGCTCTTCCCTCGAAGCGA
GCGCAACGTCAAATGCCATCGTCCCAGAAAAAGTATGTGAAAATCTTGGAGTGCATGAAGTTAAAACTAAAGTCTG
ATCTGGACAGCTATACAGAGAGCATATTTTGTCCAAGAAGTAAGAACTAAGAAGATCTAATGCTGATGTTTTGAATT
TTTTCTTAAAGATAAGTCCACGACCTACCTAGGTGGATATACTCCATGCATCAAATTACTTATATGAAATAATGACCTT
ACTCGTCGATCAATATTGTAGGTTGTGGAAGCTATCGAGAAAGCATCACTGGTACAGATAAAGAAGTCTCTTCCTAC
TTGGAACATCATCTTATATGAGAGATAAGTACTGAAGATTATTTTTATGATCATCTTTTCTGTTTGAGCCTTAAAAGT
TTCGTCGCTGCATCAATCAAATACTGCATCAATCAGCATTACAACTTTTTTGGAGCGTATTTATTTCAAAATGAATCTA
CGATACTAAACGAATACAAATAGTCTACATTCGTCTGTATCTGCTCTCCTGCTCTTCTTGTCATTCCTTAGAAATGCAT
ATCTCGGCAGTCCACATTATTGTATAATTTTTATTTGAGAAGTGCATTAGTGAAGCTATTGGTTGAGGTGAGAAGCCC
ATGTATTGGCATGCGTATGTAGTGTTCATTTAATGTTGAATTTCTTCATTATAGAATAACACTGAGTATATCCTGAACC
AACAAGAAGGTGGGGCCGTGTGGGAAGTGGGTAGATAATAAAATCATGCTAGTATTTGCTGGTGGCACTCATTTTTA
GTCTGAATTTAGTGTTAAATATCATCAGTGTATGTAGTATAATATTGAGTGATTTGAGTTCATACAATAACGTGACGG
GGAGTGAGTATAGAATAGAGTCTGGTGTCAGTATTCAATAGCTCTCGTGAAGTCATTGCTTGCATAATACTACCTCCAG
AGTCCTGGTTGCTGGAGTCTGACAGATGTCGCCGGAGTCCGACGGTGGTCTCGGAAGTCCGGGGGCTGTGGCAGGAG
TTGTGGATGCCGGAGCGCGGCGGTGGTCGTCCGAGTCCGGCAATGGTCCCCCGGAGTCATGGTTGCTGGTGTCTGGCA
GATTTCGCACGCGTCCGGGTGTTGTAGACGGAGTCCGGCGGTGGTCGCCGCAGTTGTGGTCCCCGGAGTCCGGCGGAT
GTTGCTGGAGTCCGGCAGAGTCTGGTGGTGAATGCCGGAGTCTGGTGGTGGTC

CnSat8-125	CL72	CACGTCCAACACTTAGCCAAATTTGAGAGCAAGAAGACTCGATCGGGGTCCCACCCCGTCTATCTTCCAAATATGGCGA ACGGTTATCGGTGAGCGGACTCTTCAAATTTGACTCAAGTAAGGAC
CnSat9-1185	CL78	ATACAGTAGCAGTGGCTGTGAGTACACGAAACAAAGCTTAACCTAAACAAAAACCAGATTTCTCTTGTAGAGCAACTGC CAAAGACTTGCATTCTTAACCTGTTATTCATTTTTATTTATTCATGACTTATACTATGTGTAAGCAACATTCATATCTCAATG GACCTAATCTTTAAGTCATATTTCCAGATTGTGGAAGAAATGAAGTGCTGATGATGGGATGCTAAAGCCCAACTATTTGT TGAAAACTTCAAGACTTATTGTTTGCAAAAACGTGTTTTCCAACTTAAGCTTTTTATCAGATATTTTATTTATTACTTAAAC TATGTGTATACAATATTTATATCTCAATACATCTAATCTTTTTGCACAGGATTGTTGGATGTGGTCCGGACTCCGAACAAA CCAAGCTTCCGAAGTTATTAGTTCATTGCTGCCTTGTTATAAAGAGTTTTTCAAGTGCTCGATTCCAAGCAAGCTCAATTC TGTGTCCGATTCTATGCATAGCTCTTTCCTTATCGTAAGTTTTTGATTATTGCATGATACATGTGTGGCATGGCTAATTAC GGAGTAACCCTCGGTTGGAAAAAATTAGACTTGATAGCTTTTCTTCATTATCTTATTTTTTCCTAATTTTTAATGGTTTTGA ACATTTTATAATCGATTACAGTTTGTGGTTCATACTAGTACTGCAATCTAAGGCAACTAAATGTGACTTATCAGCAACTAT ACTAATTGTAACTCCCAACACCCTCTACCTATATTCATCATTGATTGAACGAACTATCACCCAAGAGAATTTTAGAACA ATATTTATGTAAGTCAAAATAACAAATGAGTTCCGGCATTAGTAACCCCCAACTAGACTAACCAATCAAATCTGAAAG TATACTTTTTTGTTACACAAGCAGTTTTAGGTTTATAAAAGTCTAGGACACCTAAACACCATTCTACTAGGTTTCAACTC TTGTAGAGTTGCTGCTAGAGGTTTATATATGCCATATAAATCTAATAGCAAGGGGTTCACTATCTAATGATGTTACTGA ACATATGTGTATGCCTTTTAGAACCACAAATTAAAGATTCAACTTATACGAAGTAAATCTCTTATGGATTCAACCCAATC GCAGCGCTGCTACTGCAGTTCATCCACAGACTAGTACATTTCTTTTTCTTCTTGTTTTA

CnSat10-1400	CL92	GACAGAAGAAGGAAGCTCACTATCCATATAATCAAGGACGACGAGGGGAATTGGATAACTGATCAAAGTAAAA TTGCTCAAGAGGCAGTGAATTTCTTCCACTCATTATTCTCTGCAGACAGAAAGAGTACCACCAATCTAGCTATTC TGAACATTGTTCCGAGCAGAGTTACTAAAGAAGATAAAGATGCTCTCACTAAGCTCCCTGACCATGAGGAAAT AAAATCTGTTGTTTGGGCTGTGGATCCAAACAGTGTTCCAGGGCCAGATGGTTTCTCTGGAAAGTTCTACCAAG ACTGTTGGTCCAGTATTAAAATGGATGTTGTTAATATGGTGGATGGCTTCTTTCTAGGGGATTACTTGAGCAAA GGCATTGCTCATTCTGCTATTATTTTGCTCCCAAAGAAGCTTTGCAGACTTCAGACCTATGATATCAATGGCTTG TGGTCCAAATGCCAGGAGCTCGATTAGCCATTTATGATATAGAAATATAACTGAGGGTACAACTTAGCAAACAA GCATAGTAGCCCACAGAGTATAGAAACACCCACAACATAGTTGAACCTGAGGAAGATATGTACTCGGTTTGCCT GACCTGTGAAGAATTCACTCCAAACACTTCTTCTAACAAATTTACATTTTGCATGTAGCGGTCAAACGCCTCATT TTGAACTTTAATGTTGTTGTCCCTGAATTGCTTCAACTTCAAAAATCTCCACTCATTTATGACTTGTGCATCCTGT AAAAGCATGCATGGATGGTGTTAACTGTTACATAAATGTATTACATAACAAATAACTATACGGAATATTATTTTAC TTATCACATTTATAGACCTTTCACCTCTCTGGTTATTGGCTTCCTTGACCTAAGTGGGGTTTGTAAATTACTGAACT CAGCAAAATTTGATGAAAGCTGACGAAATGAAGCAGCTCGGTGTGAAGTCCTGAGTCGGTAAAATACAACGTC ATAAGACACATGCAGTAATGGAAATATGTTACCAAGTAAAAAAGCTAAACCTTTCTGTTTTCTCAATGCCTGGAC AACACATGCTGATTTTAAAATCATTGTTAAATAAGCCTGGGATAAGGAGCATAAAGGATCAGCCATGTATGTTTT TAGTGAGAAACTTAAAGAAACCAGGAAGGCATTACATACATGCAATAAACAGGTGTTTGGGACATCTTTGAGAA GCTAAATGTGCTGGAGAGAAAGGTGAGAGAAGCTGAGGAGGCTGTACAAGAAAATCCAAGTGATTCTAATGTC ATCTTACTAAAAGCTGTGGAAGCACAGTATATTCAGCAACTTTGTGTTGAAGAACTATTTTGGAAGCATTAAGCTC ACCTTAAATGGATCGAGGAAGGTGATAGAAACTTTAAATACTTCCATGCTATAGTCAAA

CnSat11-1357	CL113	CGGATCACTTTGTGTGCGGGTTGGGTCATTTTCGGATCGGGTCATTTTCAGAGTAGGTCAATTTCAGTTTTAACCA TTGTCACCTATGCATATTGATTATTATGACACTTTTTCTTTCTACTTTTTTCTAGCATGATTTACTATTCTTTTCTTCC ATTTGAACACTATAGACATTTAAGACTTATGGAAAACTTTTAGGTCGACACAATTAGACGTGGTTATCTTTCCAAA TGTTCTTAAAACCTGACGGTGCTTTGAGAAGGCGTTTTATTGTTCTTGATATCCGAGGAATGTTTTAATGTTACCA CAAAAAAAATACCAAATCATATGTGAGAAACATGATGTAGAAATGGTAGAAATCCTGAAACTCTAGTTCACTTTG GCTCTTTATATTCTTATTTCCACTGAAGTTTTTCTTCTTTAGGGATTTGTTAGTCAACTCTCTAGCCAGACGAATGG TTCTGAGCTTCCTTTTGGACTCTTGAGTCGCTGGCTTTCTTCCCATCATCATGGTGGGGTACATGATGAGAAGTCT TTTTCTCATCACCTGTAAACGTAATTATAATAATAAAGCTTGATATTGCTTATCACACTGCAAACCTACTAACATCA ACAAAAAGGCTCGTGTTGACGTGTTAGATTAGTGGTTTTGATGAAGGATTATCTCACTAACGAAAACCTACTCGTT GTAGATTAGTTATCATTGTTTTTTTGATATATTTGGTTCGTTCGAGATATGGAAATTAAAGTTTCTCTTTTTGGGTT CAGGCTGAAAGTGCGTTGGATTGAGAAAATTAATGGGGCTATTGCTTCATTGCTGAGTTCCCTTGCTCCTGATAG GGTAATCTTTATTTGAAAGATATATTACAGCATCTCTAAACTCACTTTTGATTCATTGCATTGCTGAGCAAACTGTT TTCGTATTGCTGCATTGGTTTTGCTGATAGTGCAATGGGAAGTGATCATCATCCTTCTTTAAGTCACAATAATTAG CAAGAACTACGGATAAGCAATCCAAGTAAGGGTGACCCAATTTCAACCAAGAAGAAATATGTTCATCGAAAATG CTACCCTTACAGTTTCGCCTCAATGTTTTTCACTCCCCACTTTCTTTTATTTAATGTTCTTTATTATCATCATTAAAGT GATTTCCAACTTTTTGCACTTCTATTAGGCTAACATATTAATCTTTATGGCAGAATATCTTTGATGGAGGCGAGCC TCAAAGATAGACCATTTGGTTTGGGTTCGGTTTGGTTGCCTTCAGGTAGTTCAATGATAAGTTATTTGCTTTCAGT TTGGGTTCTGCTGGTTCAGGTTAATTTGGGTTTGGATTGACTTCGTGTACGTGCTCTTT

CnSat12-1060	CL114	ATTAAATTTTTCATGTCTTATGTCTTTGCTGCTGTGCCATTTATTCTGTGAACTCTCATGTAATATGAGTATATGAC ATGTGGCCACTGCCATATAATTAATGCTCTGAGCAGTTTCACCACCATTAGATATCATATAAATATAAGACTGCAT TACTATTTTTAAATATTAATTGGAAGGCTTTCGTTTTGCAAAAATTTGCAGATTCGATCAGACCACAAAAGATAA AATCATGGTTTTCATTTTAGGCTTTGCTCTAGGTCTTTCTGCCTATGCGAAGGTTGGTCTCTACAGTTTTATCGT TTGATATGCTAAAGCACTGTATTATACTCCCTCCGTCACCCAAAGTTCTTCACACTTTGGGTGAACAAATATTA AGGAATATAATGATTAGGAGTGTAAAAAGACTTAGTTGTCCCTGAAGAATATTTAATGCTATTTTTAAATTAAG AGAGATGGAGAGTAAATATGGGTCAATAGAACAAAAAAAGAGTGATAGGGTAATTAATGGGAGTAAGAAGT GGGCCAGGTGGGAAGGGTAAATTAGGTAGTTTGTATTTTTTTATTTTCTTTTCAAAAAGTGGGAAGATAAATTT GTGACGGTTCAAAATAGAAAGTGTGAAGATAAATTAGAAACGGAGGGAGTACTTGCTTAACCTATATTTAGA ACTGCTTTAAATGACGATGAATTTTTGTGGCAGCTAAGAATTCTATCACTACTTAAAGGCTTGGGAAGCAGGGT AATGCATATTGCTGGTGTTGAGTCACTTTTATTTGAGCTTCTTAAAAGACGCCATGGACATCACCTTGGATATGA CGAGTCATATCCTAAATTGTCAAAGGCTGAAGTTGACATTCTTTGCCTTCTATTAGATGTAAGTCCAAAATTAAC ACAATCTAGTTGGTAACTTGATATAAATAACTTTTAGCTCTTTAATTGAATTTAAGGGAACTATTTATGTCTTTCA GATCTGTACCATCCCTTCTACTTCTGTTACTGGTAAGGAATTGAAGGACTCCATATTGATGGCTTTGCAGGTTTG CTGCCTTTTTCTCT
CnSat13-382	CL117	TGGAGKGGCTCACTTGGGTRACCACYCAYCCTCAAGTMTCCCCTTGTGGGAGGTCTTGAGTTCAAAKCCCACCA AGGKAGGGGAGAAARAAYTGTTTTCTTTCTATGATCMTCCTACCTTCACGGAGTAGYTCAGCGGCTKAGAAGG GAGAGGATGGTAAGAGAGGTAGCGGGTTCGTGTGGTGGGAGGAGCAACCTGTTGATGGGTGATGTAAGTCCT GCTGGATTCGTCCTTTGGAGCCAGTACCAACAAAGTTGACTGAGATCATATCCGGATTTACCTGTCTGTGTGCTT TGGAGGGTAGTTCATAGACTCCAGGAATAGGGGAGCGCAAGCACCGACACCTGGATTAACAAAAAAAAAAAA AATTCAACAACAAAGG

CnSat14-720	CL124	CTGTATGGAGTATGGACTATCTTAATATAGATTGTTTGATGCTGTTATGTTTTCAGATATATGCCACGAAGCTG CTGCTACACTTGGTCAGCTAATAATGAAGGACCTTGTTAGTATCCACATGGAGTTGAGGCAATTCTACGTGG TCAAAGATTCTGCTTTCCCAAAATGGATGAATTATGGAAAAACTTGAGTTGCTTCCAGTTTTCTTCAAGCAAGC AAAGTAATGCCAAAATAATCAGGTTTTATACCTAATTAAAATTTTCTTTTCTTTCTTGCTCACAGTAAAAAAGGC TACCCATTAGTTGGATTGCTCATATTTATGCTTGAGGGTCTAAAATGTATACCTTTTCAGTGCAGCTGAAGTCA CGAGTTGCGTGACTGCGTCCGCTGCTCCTTTTCTCTTCTTCCGTCTATGCATCTGGCTTCTGTCTCTTCTTCCCTT CTTCTAATTGCAAACCGGTTTGGTGGAAGAAAAACAACTAAAATTAGTTGATAATTTTTAAATCCGGGTTTCT GGTTCTGGATAAAACGGATAATCCGGTTTTGTTCTGGTCTCAAATAGGACTTGGGTTTGAACTGGTACGCAG CAGCAGCGGTGGCGCAGCAAAGCAGAGGAAGTAGCAGCAATAGCGGAGGCAGCACCAGAGGTAGCAGGA ACGGCAGCGACATCAGGGGTTCTTAAAAATTAACTTGTAATGTAGTTAATATTTCAGTTAT
CnSat15-990	CL166	TTAAGACACTTGGAAAACTTTCAGGTCRAGACCATTAGACTAGGTTATCTTTCGAAATGTTTTTAAAACTTGA TGGTGATTWKAGAAGGTTTTTTATTTTTCTTGATCGCCGAGGAATGTTTTAATGTTACCACAAACAAAATACC AAATCATATTTGAGTAGCATGCTATACGAATGGWAGAAAACCTGGAAGTCTASTTATATTCTTATTTAGACTG AAAGTTATCTTCTTTAGGGATCTGTCAGTCAACTATCTAGCCAGACAAATGGTTCTGAGCTTCGTTTGGGACTG TTGAGTCGCTGGCTTTCTTCCCATCCTCATGGTGGGGTACATGACGAGGACTCTTTTTCTATGCTGCTGCTAAT ACACTGCTAGTTTTAGATTTGAAGCTCTTACAAGTACTATAAACAGCAAGAACAATGGATAAGCAATCCAAGT AAGGGCGATCTAATTTCAACCCAGGAGAAATATGCTCATCGAAAATGCTGCCCTTTCAGTTTCGGGTCAAGGG TTTTCACTCCCATCGTCTTTTATTTACTGTTCTTTATTATTGTCATTGAATTGATTTCGAATTTTTGCACGTCTATT AGGCTAAAAAGTTGTATTAATCCTTATTGCAGAATATCTTTGGTAGAGGCGAGCCTCAAAGTTAGACCATTGG GTTCCGGTTCAGTTTGGCTGCTTTCAGGTCGGGTCAATGATAAGCTATTTGCTTTAATTTGGGTTCAATTTGGT TCAGCTTAATTTGGGTTAGGGTCGACCTCGGGTATGTGCTCTTTCGGGTCAGTTTGGGTGCGGGTCGAGACA TTTTTTGATCCGTTCATTTTCAGATTAGTTTAAAATCGGATTGGGTCATTTGGGACAAGGTCAACTACAGTTTG AATCATTGTCCGCTATGCATAATGATTTTTATGACTCTTTCTTTCTAGTGTTTTCTAGCATGATTTACTATTCTTT TCTTCCCATAGACATTTAAGACTTGTGGAAA

CnSat16-300	CL171	TAGATAATTAATCTAGCTRTRARTATAATCTAGCGGTGRGAAGATAATCTAGATAATCTAGATTATCTTTAAAG CTACATGAGTTAGCTTTCAGATTTAATCTAGCTTTTCGAAATAATCTAGCTTTTCTACATGTCATCCAAACCAGA CTTATTTAGTYAAACTRGATTATTTGACCAAAAATAATCTAGYTTTAGGAAATAAKCTAGMTTAAATGRCTATC CAAACATAGCCTTAGGGTATGTTTGGATGCAATAGAGGATAATCTAGCGGTAAGTAGATAATCTAGCTGTGG ATAAA
CnSat17-526	CL178	AGCTCTCCAAACCCAAAAYGAAAATGAGTTTTTGACGAAAATCTGGRTTTTGCTYTCAYAAGCTMWWATTTY RTTTGTRGSYGAAAATGTRGTGYTTTTTGGACTGATTTRTACTATTTYTGACCCGAAGAATCGATTTCTACCAT CAGAATTGGCMTAGGAGTGATAGTTTTCGAAATCGTGTGMAAAAKTGACGTYTTTTCAAAAAAATCAAAAA TTTYCAAAAATCGACAGCCTAAACCAAAAGTCTGAGCCAAAAACCGACTCGATATTCGTGATCTACGCCCTC AAATTAGCTAATAAACAATTTTTGAGCCAGATCTGAGGTGCAAAATTTTCAACTTTTTTTTCGAAAAAATCCT AAGGGTTTGACCCCCCAAAAAGTCGATTCAATTTGATCCAGTGTCTTCCTTGGGCAAAAAACTTGTGTGGAT TCAGTCATATCTTATAGATTTTGACCCGTAGATTCGAATTTTGGCATCAGAACTACCCTAAGTTTGATAGTGT TTCCTCCAGAGCCCTCCAA
CnSat18-472	CL284	GAATAGTGGTGTACAATGGAATGTGCGGCCGCTCCTGTCTCTCTCTCTCTGTCTACAAAGAATGCTTTCTCT CAATAATGAATGTCTACTGTACGAACCTCCTGACGAACTCGCGAACTTTTGCAGAAATTAGATTTTGATCA AAAAAACATCACGAAAGTGTTTGATATCAAAAGTAATCGTGCAGAATGTATCTAGAAGCTATTTCTAACACT TCGTATTCAAAGAACGACGCTTGAACACAAAGAAATCAGCATCTATGATGAAGAACAGCTTTTTGCTCAACT CACTCTGAACTTCGAAAATTCACAAAAACTTCGATTTGATTCAATCTGTAACACGATTTGCTCTGATACCAA TTGTAGGATCGCAGTTATCCAATCAATAACATACAAACAGATATAACATCAAGATCAAAGTGGAATCACAA TACAAAATGGAGATTACCACAAAAATGGATTTTCAATGTTAT

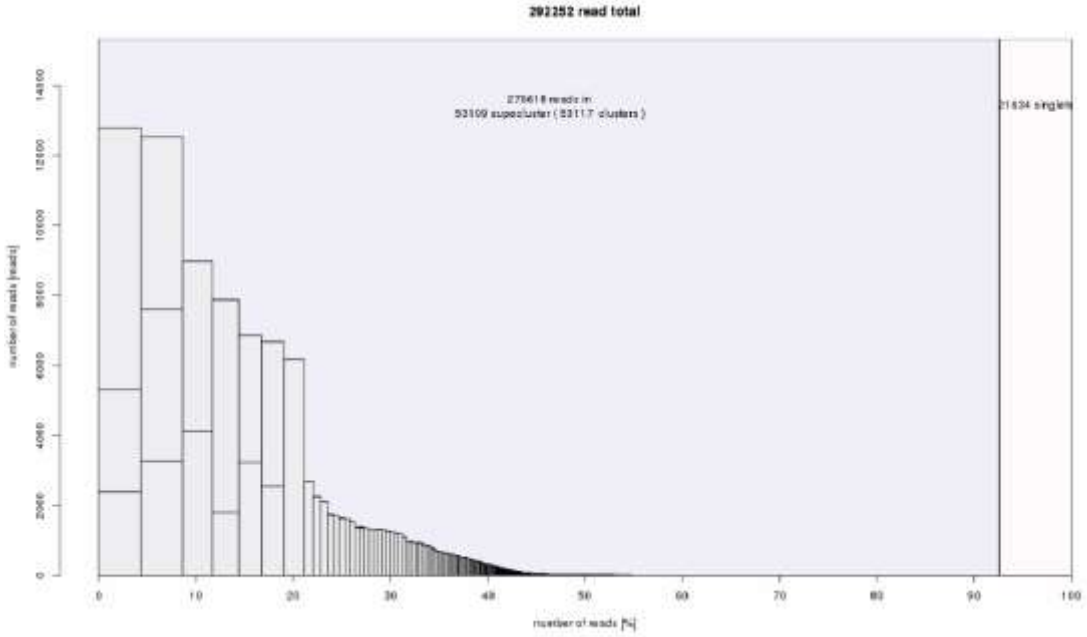
Transposable element	Lenght (bp)	Transposable element

CRM	321	TAAGGTCCTATTAACTACCTCGGTTTGTCCATCCGTTTGGGGATGACACGTAGTAGAAAACAATAACTTAGT ACCTAACTTACCCCACAAAGTCCTCCAAAAGTGACTTAGAAACTTAGTGTCCCTATCACTTACAATGCTCCTA GGAAGTCCATGGAGCCTTACCACCTCCCTAAAGAACAAATCCGCTATGTGGCAAGCGTCGTCCACCTTTTTG CAAGGTATGAAGTGTGCCATCTTTGAGAACCTGTCAACAACCACAAAAACAGAATCTTGTCCATTCTTTGTT CTAGGCAAGCCCAAAACAAAGTCCATGGAAAT
Retand	330	CCGTTCGCTAGATGGGGAATGGACTTGGTCGGAGCCCTACCCAGAGGAACAGGTAACGTACGATATGTGGT AGTGGCAATTGACTACTTTACCAAATGGGTCGAAGCCATACCCTTGGCAAGTATCACTGAAGCCCAAATCCG CAAGTTTGTGAGCAAACACATTCTATGCCGGTATGGCATACCGGAACAGATCATCACGGATAATGGGAGAC AGTTTGACAACGAAAATTTCAGGGAATTTTGTCAAAGCTGGGGCATCAAACATAGCAAGGCATCTGTAGGAT ACCCACAGGCAAATGGCCAAGTAGAAAATGCAAATCGCACAATC
SIRE	342	CAGTGTCCGATTTCTCCTCTCTACTACCCCATTTTGTTGAGGTGTTTTCACGGCTGAGAATTGATGAGTAATCC CTTGTGACTGATAGAAATCACTGAGTACTTGATTTTTGAACTCAGTGCCATTGTCAGTTCTAATCACTTGGATT TTACTCTGCAGTAAGTTCTGTGTGGTCCTCACGAATGTGGTTATCAAGGATGCTGTCTCGTCTTTTGTTCGGAG AAATCTGGTCCATGTGTATCGAGAAAAGTCATCCACAATCACAAGTACATACTTCCGACCATTGATGCTTTGGA TGCGCATTGGACCACAGAGATCCATATGTAGTAGTTCCAATGGGCG
Tekay	354	GAGCATGTCCTCTAAAATCTGGATCACTCGCTCGGACTGCCCATCCGTTTGAGGATGGAACGCCGTACTGAAC TTGAGCTTCGTGCCCATGGCCTCTTGGAACTTCGGCCAAAATCTGGATGTGAACCTCGGATCCCGGTCAGACA CAATGGAGGCTGGTACTCCATGTAGTCGGACGATCTCTTGAATGTAAAGCTCGGCCAGCTTCTCCAACGAATA CGTCATATTGATCGGCAAGAAGTGTGAACTCTTAGTGAGCCGATCAACTATGACCCAAACCGCATCATGAGCC CGGGCTGTCCTCGGCAGCCCGACCACGAAGTCCATCGTGATGTTCTCCCACTTCCATTCTGG

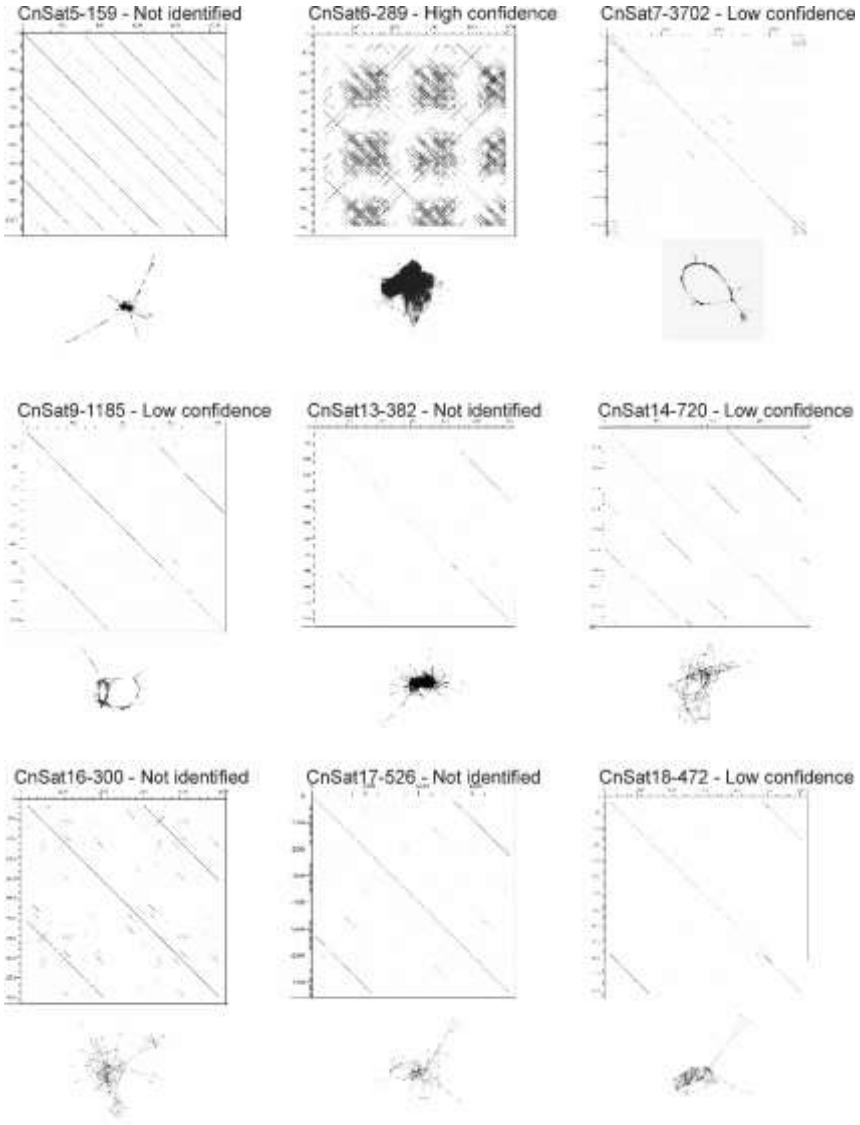
Table S3. Motif and proportion of microsatellites in *C. nitida* genome.

Cluster	Motif	%
34	TTG	0.43
40	TTC	0.34

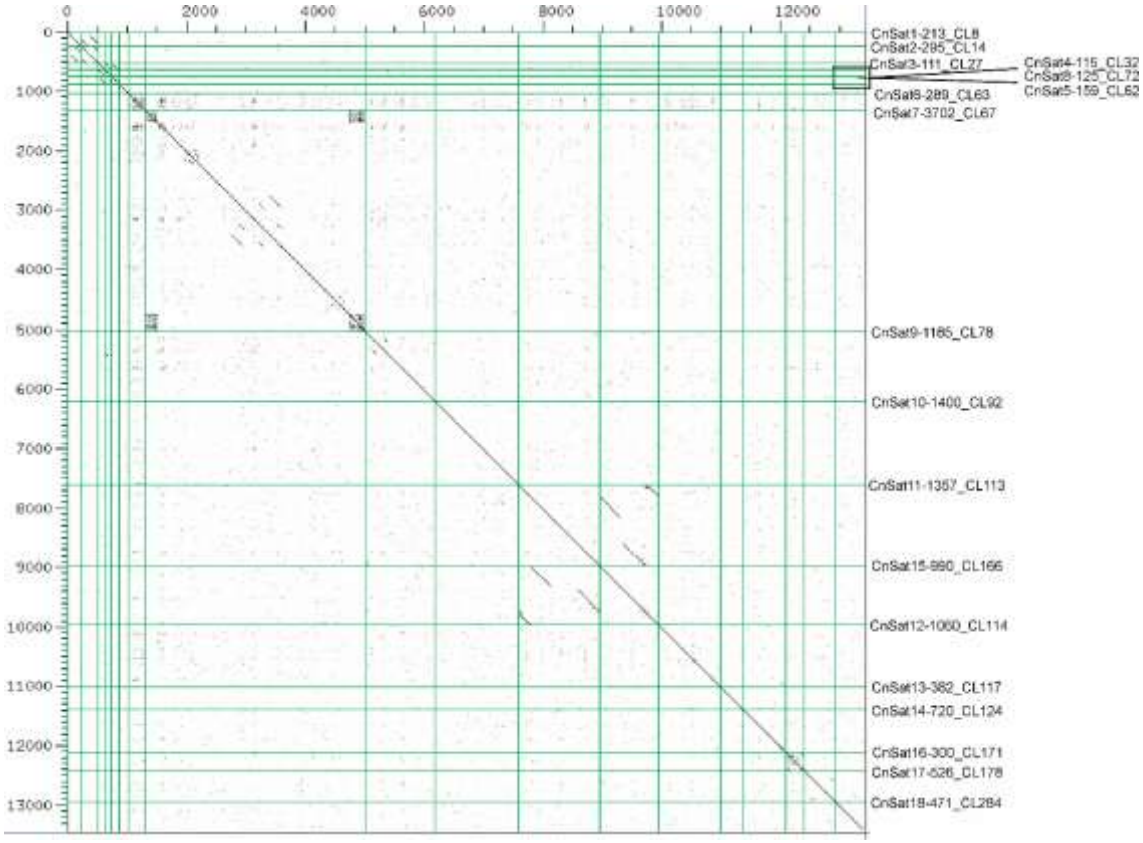
42	ATT	0.31
55	TAAA	0.21
57	TA	0.20
59	AATCTCATCTCATTAACCATTTCTTTTT	0.19
76	CCG	0.12
109	AACCTAGGTT	0.05
126	CGTTCTTCGT	0.04
152	TAA	0.03



FigS1. Histogram showing the distribution of analysed reads in *C. nitida* clusters and supeclusters.


CnSat5-159 - Not identified
CnSat6-289 - High confidence
CnSat7-3702 - Low confidence
CnSat9-1185 - Low confidence
CnSat13-382 - Not identified
CnSat14-720 - Low confidence
CnSat16-300 - Not identified
CnSat17-526 - Not identified
CnSat18-472 - Low confidence

FigS2. Dotplots and cluster graphics of all remaining satellites annotated in *C. nitida* genome. Some satellites were not detected by TAREAN, but the dot plot and cluster graphics characterised them as putative tandem repeats.



FigS3. Comparative dot plot of all satellites described *C. nitida*. This dot plot was used for identifying similarities between different satellites and define satellite superfamilies.

APÊNDICE B – ARTIGO PUBLICADO NA REVISTA FRONTIERS IN THE PLAN SCIENCE

The evolution of cytogenetic diversity in *Cuscuta* (Convolvulaceae)

Amalia Ibiapino[1], Bruno Amorim[2], Mariana Baez[3], Miguel Garcia[4], Mihai Costea[5], Sasa Stefanovic[6] and Andrea Pedrosa-Harand[1*]

[1]Laboratory of Plant Cytogenetics and Evolution, Department of Botany, Federal University of Pernambuco, Recife, PE, Brazil
[2]Postgraduate Program of Biotechnology and Natural Resources of the Amazonia (PPGMBT), State University of Amazonas, Manaus, AM, Brazil
[3] Plant Breeding Department, University of Bonn, Bonn, Germany
[4]Real Jardín Botanico-CSIC, Madrid, Spain
[5]Department of Biology, University of Wilfrid Laurier, Waterloo, Ontario, Canada
[6]Department of Biology, University of Toronto-Mississauga, Ontario, Canada

Corresponding Author: andrea.harand@ufpe.br

To be submitted to Frontiers in Plant Sciences

Abstract

Karyotypes are characterized by different parameters such as chromosome number, which can change through events of whole genome duplication or dysploidy. In *Cuscuta* (Convolvulaceae), chromosome numbers vary more than 18-fold. In addition, species of this genus show the highest diversity in terms of genome sizes among angiosperms, 128-fold, as well as a wide variation in number and distribution of 5S and 35S ribosomal DNA (rDNA) sites. To understand its karyotypic evolution, ancestral character reconstructions were performed for chromosome number, genome size and position of 5S and 35S rDNA sites. For that, previous cytogenetic data were reviewed and complemented with original chromosome counts, genome size estimates and rDNA distribution for one, seven and 10 species, respectively. *Cuscuta* showed $n = x = 15$, with duplications as the main evolutive event. However, the subgenus *Cuscuta* evolved mainly by dysploidy, typical of holocentric karyotypes. The ancestral genome size of *Cuscuta* was approximately 1C = 17 Gbp, with an average genome size of 1C = 2.8 Gbp. This indicates an genome expansion relative to other Convolvulaceae, which may be linked to its parasitic lifestyle. Finally, the positions of rDNA sites varied mostly in species with multiple sites in the same karyotype. This feature may be related to the amplification of rDNA sites in association with other repeats present in the heterochromatin. The data suggest that different mechanisms acted at different subgenera, probably influenced by a shift in centromere organization and by the reduced constrains on genome size due to its lifestyle, generating the wide diversity of karyotypes in *Cuscuta*.

Keywords: Ancestral chromosome number$_1$, character reconstruction$_2$, genome size$_3$, heterochromatin$_4$, karyotype evolution$_5$, ribosomal DNA$_6$

Introduction

Plants, animals and other organisms vary in their chromosome constitution and may be characterized by their karyotypes. Among flowering plants, chromosome number has a wide range of variation from $2n = 4$ to $2n = 640$, reflecting important evolutionary events, such as polyploidy (Heslop-Harrison and Schwarzacher, 2011; Mayrose and Lysak, 2021). The distribution of chromosome numbers in a monophyletic group allows the identification of one or more chromosome numbers that are considered the ancestral haploid number or basic number of each clade, referred as *x* (Guerra, 2008; Mayrose and Lysak, 2021).

In an evolutionary perspective, changes in chromosome number can either occur due to polyploidy, also referred to as whole genome duplications, or due to events of ascending and descending dysploidy, that is, gain and loss of chromosomes due to structural rearrangements. Descending dysploidy is the most frequent mechanism and usually related to diploidization (Guerra, 2008; Schubert and Lysak, 2011). Molecular phylogeny has contributed not only for the definition of the ancestral chromosome number of a clade, but also to the understanding of the direction of chromosome changes. The interpretation of cytogenetic data considering the phylogenetic relationships between species is essential to trace the origin and evolution of karyotype changes involved in species diversification (Vaio et al., 2013a; Sader et al., 2019).

In addition to the chromosome number, the evolution of different karyotype features, such as chromosomal bands, number and distribution of ribosomal DNA (rDNA) sites, genome size, among others, can be understood in the light of the evolution of a group and correlated, for example, to species diversification (Vaio et al., 2013; Costa et al., 2017; Garcia et al., 2017; Sader et al., 2019). Some of these analyses are performed by comparing cytogenetic data with the relationships among species present in a phylogenetic tree. In *Oxalis*, for instance, cytogenetic data supported a new taxonomic circumscription, as proposed for *Oxalis rhombeo-ovata* A. St.-Hil. and *O. psoraleoides* Kunth. Both species share the same number of chromosome arms, number of 5S rDNA sites and positive CMA bands in the long arms and, thus, were proposed to be placed together in the section *Psoraleoideae* (Vaio et al., 2018).

The integration of phylogenetic and cytogenetic data also enables reconstructing the ancestral state of cytogenetic characters and infering trait evolution. Methods based on probabilistic analysis have allowed to test chromosome evolution hypotheses within a phylogeny context, determining ancestral characters such as chromosome number (Revell, 2012; Glick and Mayrose 2014, Maddison and Maddison, 2018; Rice and Mayrose, 2021). The reconstruction of chromosome characters can be done using tools such as ChromEvol, which can estimate the ancestral chromosome number along each branch of a phylogeny, also inferring events such as polyploidy and dysploidy (Glick and Mayrose, 2014). In addition, it is possible to use other methods to reconstruct different characters, such as the R package phytools, which allows the reconstruction of genome sizes (Revell, 2012), or Mesquite, that allows the reconstruction of all sort of characters, including number and position of heterochromatic bands, as well as 5S and 35S rDNA sites (Revell, 2012; Glick and Mayrose, 2014; Maddison and Maddison, 2018). These approaches are particularly relevant when dealing with large samples or highly variable groups.

The genus *Cuscuta* is cytogenetically highly diverse, with chromosome numbers ranging from $2n = 8$ to $2n = 150$ (Pazy and Plitmann, 1995; García and Castroviejo, 2003). Its basic number was proposed to be $x = 15$, however $x = 7$ was also proposed (Fogelberg, 1938; García and Castroviejo, 2003). Most species are diploid with $2n = 30$, but allo- and

autopolyploid species also occur (García et al., 2014, 2018).Including just 200 species, *Cuscuta* presents the highest genome size variation (more than 100-fold) for a single genus among angiosperms, from 0.27 Gbp in *C. australis* Hook. & Arn. to 34.73 Gbp in *C. reflexa* Roxb (Sun et al., 2018; Neumann et al., 2020). This genus is organized into four subgenera, each one with particular cytogenetic features. The subgenus *Cuscuta* Yunck. is characterized by the presence of holocentric chromosomes. *Grammica* (Lour.) Yunck. shows the largest variation in chromosome size and number and in genome size, with confirmed hybridization events and polyploidy. In the subgenus *Monogynella* (Des Moul.) are the species with the largest genomes and chromosomes (Fogelberg, 1938; Pazy and Plitmann, 1994; Guerra and García, 2004; García and Castroviejo, 2003; McNeal et al., 2007; Ibiapino et al., 2019; García et al., 2019; Neumann et al., 2020). The subgenus *Pachystigma* (Engelm.) Baker & C. H. Wright groups species with bimodal karyotypes, suggesting strong phylogenetic signals for cytogenetic characters in the group (García et al., 2019; Ibiapino et al., 2021). Intraspecific chromosome number variation within the genus was also reported. In species such as *C. epithymum* (L.) L. and *C. planiflora* Ten., chromosome number can change among populations. This variation is even more intriguing in *C. epithymum*, which has holocentric chromosomes and show $2n$ = 14, 16, 28, 30, 32 and 34 in different populations (García and Castroviejo, 2003).

Other parameters such as number and position of heterochromatic bands and 5S and 35S rDNA sites are less studied. Nevertheless, the few species investigated revealed enormous variation. *Cuscuta denticulata* Engelm. showed one pair of CMA^+/$DAPI^+$ bands, one pair of 5S rDNA and one pair of 35S, while *C. monogyna* Vahl presented at least 90 CMA^+ bands, 80 $DAPI^+$ bands, 36 5S rDNA sites and 30 35S rDNA sites (Guerra and García, 2004; Ibiapino et al., 2019, 2020). This enormous karyotypic variation combined with a well-resolved phylogeny (García et al., 2014) makes *Cuscuta* an excellent model for studying karyotypic evolution events in plants.

Therefore, the aim of this work was to reconstruct ancestral character states such as chromosome number, genome size and the position of ribosomal DNA sites in the genus *Cuscuta*. For that, we reviewed all available data and expanded the banding and rDNA distribution data for 10 new species from different clades, six new genome size estimates and one new chromosome count, in order to understand how karyotype evolution occurred and the main events involved in these changes between and within each subgenus.

Materials and methods

Sequence sampling and phylogenetic analysis

For the phylogenetic reconstruction, we sampled 58 species from the genus *Cuscuta*, all with available chromosome numbers. The subgenera *Cuscuta*, *Gramica*, *Monogynella* and *Pachystigma* were represented by eight, forty-three, four and three species, respectively. We used 12 terminals as outgroups, including the Convolvulaceae genera *Calystegia* R. Br., *Convolvulus* L., *Cressa* L., *Dichondra* J.R. Forst. & G. Forst., *Dinetus* Buch.-Ham. ex Sweet, *Evolvulus* L., *Jacquemontia* Choisy, *Neuropeltis* Wall., and *Porana* Burm. f., one representant of Solanaceae (*Schizanthus pinnatus* Ruiz & Pav.) and one of Montiniaceae (*Montinia caryophyllacea* Thunb.) to root the tree.

We used a total of 226 sequences of nuclear (nrITS and 26S) and plastid markers (*rbc*L and *trn*L-*trn*F) obtained by Stefanović and Costea (2008) and Garcia et al. (2014) gathered through GenBank database (Benson et al., 2012) (Supplementary Table 1). To align the sequences, the plugin MUSCLE was used in the Geneious v. 7.1.9 software (Kearse et al., 2012).

The phylogenetic relationships were reconstructed using Bayesian Inference (BI) analysis. jModelTest v.2.1.6 (Darriba et al., 2012) was used to select the best model of DNA substitution for each individual marker. We used MrBayes v. 3.2.6. (Ronquist et al., 2012) to perform BI with two independent runs with four Markov Chain Monte Carlo (MCMC), sampling every 1,000 generations in a total of 15,000,000 generations. Both BI runs were evaluated in Tracer v.1.6 (Rambaut et al., 2014) to verify if the estimated sample sizes (ESS) for each parameter was higher than 200. The consensus tree was generated in MrBayes with a burn-in of 25%. The consensus tree with the posterior probability (PP) was visualized and edited in FigTree v. 1.4.2. (Rambaut, 2014). The jModelTest and BI analysis were performed through the CIPRES Science Gateway (Miller et al., 2010).

Slide preparation and FISH

The cytogenetic data of *Cuscuta* were obtained from databases such as Chromosome Count Database (http://ccdb.tau.ac.il/), Plant C-Value Database (https://cvalues.science.kew.org/) and Plant rDNA Database (https://www.plantrdnadatabase.com/). To complement these data, tem species described in Table 2 had young shoot tips or flower buds pre-treated with 2mM 8-hydroxyquinoline at 10ºC for 24 h and fixed in Carnoy (3:1 ethanol/acetic acid) for slide preparation. The meristem or flower buds were washed three times in distilled water for five minutes each and digested in 2% (w/v) cellulase (Onozuka)/20% (v/v) pectinase (Sigma) solution at 37°C for 60 min, squashed in a drop of 45% acetic acid and the coverslip removed in liquid nitrogen.

For double CMA/DAPI staining, slides were aged at room temperature for three days, stained with 8 µl of 0.1 mg/µl chromomycin A3 (CMA) for 60 minutes, mounted in 8 µl of 1µg/ml 4',6-diamidino-2-phenylindole (DAPI) in mounting medium (glycerol:McIlvaine buffer pH 7.0, 1:1, v/v), and aged again for three days at room temperature. The images were captured with a COHU CCD camera attached to a Leica DMLB fluorescence microscope equipped with Leica QFISH software. After image capture, slides were destained for 30 minutes in Carnoy and 1 h in absolute ethanol, and stored for *in situ* hybridization at -20ºC.

The destained slides were subjected to fluorescent *in situ* hybridization (FISH) according to the protocol described in Pedrosa et al. (2002). Two rDNA probes were used: D2 from *Lotus japonicus* (Regel) K. Larsen (5S rDNA; Pedrosa et al. 2002) and p*Ta*71 from wheat (25-28S, 5.8S and 18S rDNA; Gerlach and Bedbrook, 1979). Probes were labelled by nick translation with Cy3-dUTP (5S) and digoxigenin 11-dUTP (35S). The 5S was labelled in a Nick Translation reaction with total volume of 12.5 µL containing 1 µg of amplified DNA, 1× Nick Translation buffer (0.5 M Tris HCl pH 7.5; 50 mM MgCl2), dNTP mix (0.016 mM each of dATP, dCTP, dGTP), 0.08 mM Cy3-dUTP or Alexa-dUTP, 7,5 U of DNA Polymerase I and 0,006 U of DNase I. The mixture was incubated at 15º C for one hour or longer if needed, until most fragments were under 500 bp, and reactions were stopped using 0.5 M EDTA. The 35S was labelled with Nick Translation kit of Invitrogen. The images were obtained as previously described.

Flow cytometry

A total of 12 species had their genome sizes estimated by flow cytometry, six of them for the first time: *C. glomerata* Choisy, *C. partita* Choisy, *C. purpurata* Phil., *C. racemosa* Mart., *C. sandwichiana* Choisy and *C. globosa* Ridil. A suspension of nuclei from shoot tips was prepared using WPB buffer (Loureiro et al., 2007). The nuclei were stained using propidium iodide and the amount of nuclear DNA was estimated using the

CyFlow SL flow cytometer software (Partec, Görlitz, Germany). *Raphanus sativus* L. 'Saxa' (1.11 pg), *Solanum lycopersicum* L. 'Stupické polní rané' (1.95 pg), *Glycine max* Merr. 'Polanka' (2.5 pg) and *Zea mays* L. 'CE-777' (5.43 pg) were used as internal standards (Doležel et al., 2007). The final 1C value was based on three different measurements with 5,000 nuclei in each sample using the equation "Sample peak mean/Standard peak mean 2C DNA content of internal control (pg)" and the software FloMax (Partec) for data processing.

Reconstruction of ancestral chromosome numbers

Haploid chromosome numbers were used to infer basic ancestral numbers for each clade and the genus using the ChromEvol software v. 2.0 (Glick and Mayrose, 2014). To choose the model that best apply to the data set, a first run was made in ChromEvol considering all ten possible models of the program. Then, the model with the smaller AIC value was selected, and this model was submitted to the model adequacy test for adequation of each selected model parameter (Rice and Mayrose, 2021). Due to the numerical chromosome variation reported in *C. epithymum* (*Cuscuta* subgenus, $2n$ = 14, 16, 28, 30, 32 and 34) and *C. planiflora* (*Cuscuta* sugenus, $2n$ = 26 and 28), all the counts found were added to the ChromEvol analysis. First, we perform a standard run, using the parameters given by the model adequacy test mentioned above. Then, we executed two more runs, one removing the holocentric clade (subgenus *Cuscuta*) from the analysis to test for the influence of holocentric chromosomes and intraspecific chromosome number variation. Considering the presence of holocentric and monocentric chromosomes in the genus (Fogelberg, 1938; Pazy and Plitmann, 1994), it is possible that different evolutionary models better applies for different clades (Márquez-Corro et al., 2019). In the second additional run, we fixed n = 15 to the root, since this is the basic number proposed for *Cuscuta* from cytogenetic data (Pazy and Plitmann, 2002), and was supported by the previous analysis. The results were plotted in R using the ChromEvol functions as described in Cusimano et al. (2012). The chromosome number information can be found in Table 1. To run these analyses, we pruned the tree excluding the outgroups using the ape package (Paradis et al., 2004) implemented in R (R Core Team, 2020).

An additional reconstruction of the chromosome number was performed in Mesquite version 2.75 (Maddison and Maddison, 2018) for the purpose of comparison with the results obtained in ChromEvol. However, in Mesquite the haploid chromosome numbers were categorized into nine states: 0 corresponding to n = 4, 1 to n = 5, 2 to n = 7, 3 to n = 10, 4 to n = 13, 5 to n = 14, 6 to n = 15, 7 to n = 16, 8 to n = 19 and 9 representing all polyploids from n = 21 to n = 75. Additionally, we compared the results to the inference of the ancestral state of this character made along the branches using PastML (https://pastml.pasteur.fr/) (Ishikawa et al., 2019). We applied the JOINT method, which reconstructs the states of the scenario with the maximum likelihood. As both models assume only one state per sample, we used for *C. epithymum* the cytotype analysed in the present work, n = 15, and for *C. planiflora*, n = 14.

Reconstruction of genome sizes

The genome size (GS) reconstruction was performed using the 25 species, seven of them represent previously unpublished data (Table 2). This trait was analysed as continuous character in phytools package (Revell, 2012). To run this analysis, we excluded the 34 species that lack GS information from our original tree using the ape package (Paradis et al., 2004). Both phytools and ape packages were implemented in R (R Core Team, 2020). Three species of the subgenus *Cuscuta*, 18 species of the subgenus *Grammica* and four species of the subgenus *Monogynella* were sampled. For this analysis, we considered the

value of 1C in Gbp and for species with different genome sizes published, an average was made between the values. In addition, still for comparative purposes, a reconstruction of the genome size was also performed using Mesquite. As genome size is a continuous character, the maximum parsimony analysis was used.

Reconstruction of rDNA ancestral positions

The reconstruction of ancestral number and positions of the rDNA sites was performed using Mesquite. To run this analysis, 42 species were excluded from the *Cuscuta* tree as previously described and only 18 species with 5S and 35S rDNA information were considered. Both the number and position of sites were transformed into categorical data (discrete characters): centromeric/pericentromeric position, interstitial, terminal/subterminal and 'mix' (when more than one of the previous conditions occurs in the same karyotype) as mentioned in Garcia et al. (2017). For the number of 5S sites the characters were categorized as one site pair, two, three, five or eighteen site pairs. The number of 35S sites were categorized as one, two, five or fifeteen pairs of sites. The "trace character history" function was used and the ancestral state was inferred using maximum likelihood in Mesquite version 2.75 (Maddison and Maddison, 2018), according to Vaio et al. (2013). Due to the low resolution in the rDNA sites number reconstruction using Mesquite, a second reconstruction was performed using the Bayesian Binary MCMC (BBM) tool (Ali et al. 2012) implemented in the software Reconstruct Ancestral State in Phylogenies - RASP 4.2 (Yu et al. 2015; 2019) using the default parameters

Results

Chromosome number, rDNA site and genome size variation in *Cuscuta*

Most *Cuscuta* species with published chromosome number are diploids with up to $2n$ = 38 (47 species). Another 11 species are polyploids, mostly with $2n$ = 60. Only one new count was included in this work, *C. globosa*, with $2n$ = 90. The majority of polyploids belong to the subgenus *Grammica*. Except for some species of the subgenus *Cuscuta*, for example *C. epilinum* Weihe with $2n$ = 42. The smallest number found was $2n$ = 8 in *C. babylonica* Aucher ex Choisy (*Cuscuta*), while the largest number was $2n$ = 150 in *C. sandwichiana* (*Grammica*). *Cuscuta epithymum* (*Cuscuta*) presented intraspecific numerical variation with $2n$ = 14, 16, 28, 30, 32 and 34.

For seven species, 5S and 35S rDNA site number and location were previously published (Table 2; Guerra and Garcia et al., 2004; Ibiapino et al., 2019, 2020). In addition, rDNA data were obtained for 10 additional species, *C. australis* ($2n$ = 30), *C. campestris* Yunck. ($2n$ = 56), *C. epithymum* (with $2n$ = 30), *C. howelliana* P. Rubtzov ($2n$ = 30), *C. partita* (2n = 28), *C. psorothamnensis* Costea & Stefanovic ($2n$ = 60), *C. racemosa* ($2n$ = 60), *C. sandwichiana* ($2n$ = 150), *C. globosa* ($2n$ = 90) and *C. glomerata* ($2n$ = 30). All rDNA sites in *Cuscuta* are colocalized with CMA^+ bands. Most species presented at least one pair of $CMA^+/DAPI^-$ bands colocalized with nucleolus organizer regions (NOR) in proximal regions. Interstitial bands, when present, were weaker and smaller. Only in *C. epithymum* these bands occur in the terminal regions and were present in most chromosomes (Figures 1 and 2). Most species showed only one pair of 5S and one pair of 35S rDNA sites. Usually, 5S sites occurred in interstitial regions, while 35S sites in pericentromeric regions, such as in *C. australis*, *C. howelliana*, *C. partita*, and *C. glomerata* (Figure 1). There was no relationship between the number of sites and the ploidy level of the species. When more than one pair of 5S rDNA was present, these sites were also in interstitial regions, whereas when there was more than one pair of 35S, the extra pairs were interstitial, as in *C. campestris*, *C. racemosa*, *C. psorothamnensis*, *C.*

globosa and *C. sandwichiana* (Figure 2) or terminally located, as in *C. epithymum* (Figure 1). The largest number of rDNA sites in *Cuscuta* species was observed in *C. monogyna*, with approximately 18 pairs sites of 5S and 15 pair of 35S rDNA. Information on the number and distribution of the rDNA in *Cuscuta* can be found in Table 2.

As for the genome size, *Cuscuta* species varied from 0.27 Gbp in *C. australis* (2*n* = 30) to 34.73 Gbp in *C. reflexa* (2*n* = 32), both diploids. Some species showed variation in genome size, such as *C. gronovii*, with five different values reported, ranging from 2.14 Gbp to 6.75 Gbp (Table 2).

Phylogenetic relationships and ancestral character reconstructions

In total, 58 species of *Cuscuta* with DNA sequences and cytogenetic data available were sampled for phylogenetic reconstruction, representing approximately 30% of the 200 known species of the genus. The genus was recovered as monophyletic, as well as its four subgenera, *Monogynella*, *Cuscuta*, *Pachystigma* and *Grammica* (Figure S1). Among the four subgenera, two were better represented in this work: *Cuscuta* with 80% of its total of 10 species, and *Pachystigma* with 60% of its total of five species. The least represented was the subgenus *Grammica*, with 29.33% of the total of 150 species.

The chromosome number reconstruction performed in ChromEvol with the BASE_NUM_DUPL model, which was the best model, indicated *n* = 7 as the basic ancestral chromosome number in *Cuscuta* (Figure S2). This model considers five parameters, the rate of gains and losses of single chromosomes, duplications, in addition to considering a specific chromosome number that characterizes a phylogenetically close group and the number variation rate. Based on this model, the variation in chromosome number in *Cuscuta* is most often related to duplication events (*f* = 11.6), followed by chromosome gains (*f* = 7.7) and losses (*f* = 6.8). The number *n* = 15 was indicated as ancestor of the subgenera *Grammica* and *Monogynella*. The ancestral number of the subgenus *Cuscuta* was also *n* = 7, and *Pachystigma* had *n* = 14, with 50% probability, but *n* = 7 was also very likely, with 40% probability (Figure S2).

The holocentric clade (subgenus *Cuscuta*) include species with intraspecific numerical variation. This clade is the best represented in our sample, so to test if this overrepresentation, as well as its holocentric nature and high variation, influenced the analysis, we ran ChromEvol without the holocentric clade, following the same parameters described above (Figure S3). The basic number in this analysis was *n* = 15, again with chromosome duplication (*f* = 7), followed by chromosome gains (*f* = 5.6) and chromosome losses (*f* = 3.7) as the main evolutive events. The reconstructed ancestral number for the remaining three subgenera was conserved as *n* = 15 (Figure S3). Therefore, we repeated ChromEvol analysis fixing *n* = 15 at the base of the genus (Figure 3). In this scenario, numerical changes were mainly due to chromosome losses (*f* = 14.7), followed by duplications (*f* = 10) and chromosome gains (*f* = 6.6). The reconstructed ancestral numbers for *Grammica* and *Monogynella* were also *n* = 15. For the subgenus *Cuscuta* it was *n* = 8, with 54% probability, but *n* = 7 was also very likely, with 35% probability. The basic number for *Pachystigma* was *n* = 14 (Figure 3).

Other reconstruction tool PastML also indicated *n* = 15 as the basic number in *Cuscuta*, such as Mesquite. The maximum likelihood analysis suggested *n* = 15 for the entire genus, as well as for the subgenera *Monogynella* and *Grammica*. For the subgenus *Pachystigma*, the most likely basic number reconstructed was *n* = 14 (with 50% probability), but *n* = 15 was also very likely (with 43% probability) according to this analysis. Only the *Cuscuta* subgenus had many possible basic numbers. The Mesquite analysis showed *n* = 7 and 14 as the most likely, each with approximately 31% probability (Figure S3). In addition, ancestral chromosome numbers were reconstructed with

PastML. The approach reconstructed *n* = 15 as the ancestral number of both the genus *Cuscuta* and each of its four subgenera. Only subgenus *Pachystigma* has *n* = 14 as probable ancestral number (Figure S5).

Both the reconstructions of the genome size made with phytools and Mesquite demonstrated that the ancestral genome would be of intermediate size, approximately 17 Gbp, considering the smallest genome of 0.39 Gbp and the largest of 34.73 Gbp. Genome sizes expanded in the subgenus *Monogynella*, while they decreased in the other subgenera. Only *C. indecora* showed an expansion of genome size within the subgenus *Grammica* (Figure 4 and Figure S5).

Mesquite has failed to reconstruct the number of ancestral rDNA sites. Both the number of 5S and 35S rDNA sites were inconclusive. The analysis performed using RASP reconstructed 18 site pairs as ancestral for 5S rDNA and 15 site pairs as ancestral for 35S rDNA. The reconstruction of sites number for both Mesquite and RASP can be found in supplementary figures S7 and S8, respectively. The reconstruction of rDNA site positions showed that the interstitial position as ancestor of 5S rDNA. The mesquite failed to reconstruct the ancestral position of the 35S rDNA site. Interstitial 5S rDNA sites were maintained throughout the genus. Only in *C. indecora* Choisy additional sites in terminal positions reappeared. The 35S rDNA were terminal in the subgenus *Cuscuta* and peri/centromeric in *Grammica*. In this subgenus, species with only pair of 35S sites had them always in peri/centromeric position. When more than one pair of 35S rDNA is present, the extra sites were at interstitial positions (Figure 5).

Discussion

The holocentric clade has a differentiated evolutionary path

To estimate the basic ancestral chromosome number of *Cuscuta*, we reconstructed phylogenetic relationships including all species with available chromosome numbers. Our phylogeny was consistent with the phylogenetic hypothesis previously proposed by García et al. (2014), where each of the four subgenera represent very well supported clades, where *Grammica* and *Monogynela* are monophyletic, *Cuscuta* is paraphyletic and the subgenus *Pachystigma*, which was once considered to be just a section of the subgenus *Cuscuta*, as sister to the subgenus *Grammica*. The relationship between the *Pachystigma* and *Grammica* section was well supported, and *Pachystigma* was confirmed as an independent subgenus. The *Monogynella* subgenus was recovery as sister to the other subgenus of *Cuscuta* (Garcia et al., 2014), in agreement to the morphological characteristics, as this subgenus shared pleisiomorphic traits with non-parasitic parents, such as a less reduced vascular system and greater photosynthetic capacity when compared with other *Cuscuta* subgenera (McNeal et al., 2007; García et al., 2014).

Chromosome number variation was almost 19-fold between the smallest (2*n* = 8 in *C. babylonica*) and the largest (2*n* = 150 in *C. sandwichiana*) numbers (Pazy and Plitmann, 2002; García et al., 2019). One new chromosome count was reported for *C. globosa* (2*n* = 90) from *Grammica*. This count is typical for this subgenus, which includes most of the polyploid species and interspecific hybrids of the genus. The best-represented subgenus was *Cuscuta*, with 80% of its total 10 species, showing a variation over 5-fold (from 2*n* = 8 to 2*n* = 42). The variation within this subgenus is also associated to intraspecific numerical variation, found in *C. epithymum* (2*n* = 14, 16, 28, 30 and 32) and *C. planiflora* (2*n* = 26 and 28) (García and Castroviejo, 2003). *Grammica* was the least represented subgenus, only ca. 30% of its 150 species. However, the variation in chromosome number found was similar to that observed for the *Cuscuta* subgenus, just

over 5-fold (2*n* = 28 to 2*n* = 150). In this case, most of this variation is attributed to polyploidy and the underrepresentation of species may suggest that the variation in this subgenus may be underestimated. In the *Monogynella* and *Pachystigma* subgenera, a large variation in chromosome number was not found. The *Monogynella* is represented in this work by 57% of its total seven species and the chromosome numbers found were 2*n* = 28, 30 and 32. On the other hand, *Pachystigma* is represented by 60% of its total five species and the chromosome numbers found were 2*n* = 28 and 30. The limited variation found in these subgenera is due to the fact that they are diploid species and, in the genus *Cuscuta*, chromosomal variation in diploids seems to occur to a larger extend only in the subgenus *Cuscuta*.

We used ChromEvol to reconstruct the basic ancestral chromosome number of the genus *Cuscuta*. The BASE_NUM_DUPL model was the best for our dataset. This model considers the chromosome number n = 15 that happens most frequently in the data sample and the multiples of this number. The parameters considered in this model are the rate of increase of a single chromosome (_gainConstR), the rate of decrease of a single chromosome (_lossConstR), the ate of whole genome duplications (polyploidy) (_duplConstR), rate of transitions per base number (_baseNumberR) and the specified number of chromosomes that characterize a phylogenetic group (_baseNumber), noting that this is not the chromosome number at the root of the phylogeny (Glick and Mayrose, 2014; Rice and Mayrose, 2021). The basic ancestral number indicated for *Cuscuta* was *n* = *x* = 7, as previously suggested by Fogelberg (1938) and Garcia and Castroviejo (2003). But most of the chromosome numbers reported in *Cuscuta* are multiples of 15, since 32 species (55%) are diploids with *2n* = 30 and, among polyploids, 2*n* = 60 is the most frequent number. The highest numbers reported are 2*n* = 90 and 150, which are also multiples of 15. According to the Chromosome Count Data Base (Rice et al. 2015), the most frequent chromosome number in other Convolvulaceae is n = 15, the closest genera to *Cuscuta* with reported chromosome numbers are *Ipomoea* L. and *Dinetus* Buch.-Ham. ex Sweet, with *n* = 15 and *n* = 14 as the most frequent numbers, respectively. Therefore, it would be less likely that, in *Cuscuta*, the number has reduced in its base, followed by independent chromosome duplications and gains in *Monogynella* and *Pachystigma+Grammica*.

One reason for the inferred *n* = *x* = 7 for the genus was probably the presence of lower numbers in the subgenus *Cuscuta*. However, this subgenus is exclusively holocentric and this chromosome type may go through karyotypic changes that are different from the other subgenera. Chromosome fusion and fission events can be favoured in this karyotype type, as they have diffuse kinetochore, which facilitate these types of rearrangements (Mandrioli and Manicardi, 2020). In groups where this different evolutionary dynamics occurs, it is necessary to consider clade specific models, that is, different parts of the phylogeny evolving according to different changes in chromosome numbers (Mayrose and Lysak, 2021). Márquez-Corro et al. (2019) used different methodological approaches to identify diverse patterns of chromosomal evolution in different clades of the Cyperaceae. For this, both a complete tree and subtrees were used, suggesting several evolutionary model transitions in the entire phylogeny of the family. This adequacy in the analysis is particularly relevant in the study of clades containing species with holocentric chromosomes, whose karyotypes can exhibit heterogeneous evolution modes. Therefore, we considered this matter and removed the holocentric clade to evaluate the reconstruction (Figure S3). In this analysis, the base number *n* = *x* = 15 was retrieved for both the *Cuscuta* genus and in each of the subgenera, with duplication events as the most frequent, originating polyploids. This corroborated that the

evolutionary dynamics of holocentrics are indeed different and exerted large influence on the reconstruction of chromosome numbers in *Cuscuta*.

We therefore considered $n = 15$ fixed at the base of the genus *Cuscuta* as the best model to explain the evolution of the chromosome numbers of this genus. In this model, six chromosome losses occurred to originate $n = x = 8$ in the holocentric *Cuscuta* subgenus. However, $n = x = 7$ was also considered another probable ancestor of this clade. The basic numbers $x = 7$ and $x = 8$ had been proposed for *Cuscuta* subgenus (Kaul and Bhan, 1977; García and Castroviejo, 2003). Among the holocentric species, the chromosome numbers found were $2n = 8$, 10, 14, 16, 20, 26, 28, 30, 32, 34 and 42, with both $n = 7$ and $n = 8$ explaining this variation. Some of these numbers may have been generated from the duplication and ascending dysploidy events. In *C. epithymum* ($2n = 14$, 16, 28, 30, 32 and 34), individuals with $2n = 14$ and $2n = 32$ are bimodal, while $2n = 16$ and $2n = 34$ have symmetrical karyotypes (García and Castroviejo 2003; García et al. 2019). This indicates that chromosome fusion or fission events, together with polyploidy, would originate this numerical variation. One of the species of this subgenus, C. *epilinum*, is polyploid with $2n = 42$. The polyploidy of this species may have occurred by duplications in the basic number or by hybridization events, both previously suggested for the *Cuscuta* subgenus (García and Martín, 2007).

Additional analysis of both Mesquite and Past ML corroborated the number $n = x = 15$ at the base of the *Cuscuta* genus. In the Mesquite reconstruction, the nodes of the holocentric clade were inconclusive. It was suggested several likely ancestral numbers for each of them, such as $n = 7$ and $n = 14$, corroborating what was discussed earlier about the holocentric clade having a different evolutionary pathway than monocentric species. PastML analysis was the only one that suggested $n = x = 15$ at the base of the subgenus *Cuscuta*. One possibility for this result may be due to the fact that this analysis does not support more than one number in each branch. In the ChromEvol analysis it is possible to insert more than one number per branch, as was done for *C. epithymum* and *C. palniflora*, where each number of its intraspecific variation was considered. In PastML we inserted only the cytotype $2n = 30$ of *C. epithymum*, since it was the cytotype used in this work for mapping rDNA, and $2n = 28$ in *C. planiflora*, because it is the most frequently cytotype of this species. Nevertheless, considering $n = x = 15$ for the genus and the other subgenus, reduction to $n = x = 7$ or 8 may have occurred within the *Cuscuta* subgenus, after transition from monocentrics to holocentric chromosomes in this lineage.

The data suggest that $n = 15$ is the ancestral haploid number that best explains the numerical evolution of chromosomes in the *Cuscuta* genus. Furthermore, the genus has heterogeneous modes of karyotype evolution, because one of its clades, subgenus *Cuscuta*, has exclusively holocentric chromosomes with frequent chromosome fusions and/or fissions. Although dysploidy is also observed in other clades, in the sugenus *Cuscuta* it is the major evolutionary process. Polyploidy is common in the sugenus *Cuscuta*, but the major mechanism in *Grammica*.

Genome size variation in *Cuscuta* may be linked to its parasitic lifestyle

Genome size data for 25 species of *Cuscuta* were available, varying 128-fold between the smallest (1C = 0.27 Gbp in *C. australis*) and the largest genome (1C = 34.73 Gbp in *C. reflexa*). Intraspecific variation may occur, but cases such as *C. americana* L., with a 1C = 0.01 Gbp of difference between reports, may be due to methodological limitations. On the other hand, large variation as observed for *C. compacta* Juss. ex Choisy, which presented a 1C = 4.43 Gbp difference (between the largest and the smallest genome, 1C = 3.24 Gbp and 1C = 7.67Gbp, respectively), may represent biological differences between samples. McNeal et al. (2007) reported four different accessions of *C. gronovii*

ranging from 1C = 2.14 Gbp to 6.75Gbp. Only one of these accessions had a genome size similar to that reported by Neumann et al. (2020), 1C = 3.58 Gbp. The intraspecific difference in genome size may also be the result of a misidentification of the material used. In contrast to the high variation in cytogenetic features, the reduced presence of diagnostic characters often makes the *Cuscuta* species difficult to identify (Costea et al., 2015).

Genome size variation in *Cuscuta*, however, does not seem to be mainly caused by polyploidy events, despite its high frequency in the genus (Ibiapino et al., 2019; García et al., 2019). Polyploids with higher chromosome numbers, such as *C. globosa* (2*n* = 90) and *C. sandwichiana* (2*n* = 150), both belonging to the *Grammica* subgenus, had small genome sizes, 1C = 1.79 Gbp and 1C = 1.8 Gbp, respectively. The largest genome sizes are found in diploid species such as *C. lupuliformis* Krock. (2*n* = 28) and *C. reflexa* (2*n* = 32), both belonging to *Monogynella* subgenus, with 1C = 21.97 Gbp and 1C = 34.73 Gbp, respectively. Species such as *C. monogyna* (1C = 33.05 Gbp, *Monogynella* subgenus) and *C. indecora* (1C = 24.46 Gbp, *Grammica* subgenus), both diploids, have numerous heterochromatic bands along their chromosomes, indicating that amplification of repetitive sequences in heterochromatin is involved (Ibiapino et al., 2020). However, there is also a relatively large accumulation of heterochromatic bands in the subgenus *Cuscuta*, but there is no drastic increase in genome size, on the contrary. Compared to the sister subgenus *Monogynella*, there was a reduction in genome size in the subgenus *Cuscuta*. This is probably due to the smaller and fewer chromosomes. In *C. europaea* L., for example, the genome size is, on average, 1C = 1.11 Gbp (McNeal et al., 2007; Neumann et al., 2020). This species has satellite DNAs that occupy a large extension of all 14 chromosomes, which varies in sizes from 2.76 to 6.70 µm, approximately (estimated measurements based on mitotic metaphases reported in Oliveira et al., 2020). In *Pachystigma*, the differential accumulation of repeats in just a few chromosomes led to the appearance of bimodal karyotypes. *Cuscuta nitida* (2*n* = 28), for example, has an accumulation of different classes of repetitive DNA in only two chromosome pairs, which are, in average, 12.34 and 8.19 µm long, against 2.67 µm of the smallest pairs which are not enriched with repetitive DNA (Ibiapino et al., 2021). Thus, the accumulation of repetitive DNA can lead to an increase in chromosomes and, genome sizes, especially in the subgenus *Monogynella*.

The ancestral genome of *Cuscuta* was possibly around 1C = 17 Gbp. There was an expansion in the *Monogynella* subgenus, while the other subgenera experienced genome size reductions, except for *C. indecora*, subgenus *Grammica*, which showed an intense expansion in genome size. Neumann et al. (2020) suggested that there is no correlation between genome size and the parasite style of the *Cuscuta* genus. There is a tendency for parasitic plants to have larger genomes, as they escape restrictions imposed by the growth rate of the meristem or the "genomic economy", as they take resources from their hosts. Analyses carried out with Orobanchaceae family showed great dynamics in the genome of these species. The genomes of the autotrophic *Lindenbergia philippensis* and the hemiparasite *Schwalbea americana* are much smaller than those of the holoparasite *Orobanche* and *Phelipanche*, which fits the hypothesis of larger genome sizes in parasitic plants. (Gruner et al., 2010; Piednoël et al., 2012). But, in *Cuscuta*, the subgenus *Monogynella* carries some characteristics of non-parasitic ancestors and has large genomes, and in the other subgenera, where species are more phylogenetically distant from non-parasitic parents, the size of the genomes has decreased. However, according to the Plant DNA C-value Database, other species of the Convolvulaceae family have small genome sizes, with approximately 1C = 0.97 Gbp. An expansion in relation to this value is noted in *Cuscuta*, since the average genome size in *Cuscuta* is 1C

= 2.8 Gbp, 2.8-fold larger than the average of other Convolvulaceae. Besides the increase in *Monogynella*, it is possible to see an increase in genome size in some species of other subgenera. In addition to *C. indecora*, other species showed larger genome sizes compared to the smaller genomes found in the genus, such as *C. campestris*, *C. gronovii* and *C. compacta*, which have 1C = 5.3 Gbp, 6.75 Gbp and 7.67 Gbp, respectively. The data suggest that the evolution of genome size in species of the *Cuscuta* genus may be compatible with the parasitic lifestyle of the species, which might release imposed restrictions leading to the largest variation in genome size observed in a single angiosperm genus. Furthermore, the expansion of genome found in some species of the genus can be caused by polyploidy an the proliferation of repetitive sequences and heterochromatin in these species.

Changes in the position of rDNA sites may indicate the dynamism of tandem repetitive DNA sequences in *Cuscuta*

Most *Cuscuta* species have few rDNA sites: only one pair of 5S and one pair of 35S rDNA sites. Although the number of 5S and 35S rDNA loci has a positive correlation with the ploidy level (Garcia et al., 2017), this does not hold true for this genus. *Cuscuta sandwichiana*, the highest polyploidy reported, has one pair of 5S and two pairs of 35S rDNA sites, while phylogenetically close diploids, such as *C. australis*, have one pair of 5S and one pair of 35S rDNA sites or, also diploid, *C. indecora*, which belongs to the same subgenus, but present five 5S rDNA pairs and two 35S rDNA pairs, more than the number found in *C. sandwichiana*. This may be due to the fact that some *Cuscuta* polyploids are interspecific hybrids, such as *C. sandwichiana* (Stefanović and Costea, 2008). It is often observed in hybrids the occurrence of elimination or gene conversion that results in the presence of rDNA copies from only one of the parents. In allotetraploids from the Dilatata group, genus *Paspalum* L., for instance, the recovered ITS sequences show homogenization towards one paternal genome only (Vaio et al., 2019). In addition, the decrease in the number of expected sites may occur due to the elimination of some sites in terminal regions. The terminal position of the rDNA sites would be selectively favourable compared to the proximal ones, as it would reduce the chances of deleterious chromosomal rearrangements related to unequal recombination and recombination between non-homologous chromosomes (Roa and Guerra, 2012a; Garcia et al., 2017). Besides, the number of parental rDNA sites may be quite conserved in young, artificial allopolyploids. However, in natural allopolyploids, this number is often reduced, especially the 5S rDNA sites (Lee et al., 2011; Volkov, 2017). In *C. veatchii* for example, there is a reduction in the rDNA sites in relation to its parents *C. denticulata* and *C. nevadensis* I.M. Johnst., indicating an old origin of this hybrid (Ibiapino et al., 2019).

It is common for the 5S and 35S rDNA sites to be found at separate locations in the genome, even on different chromosomes. This may be related to the fact that they are transcribed in different cellular compartments, by different enzymes (Garcia and Kovařík, 2013). However, in *Cuscuta*, many species showed rDNA sites located on the same chromosome, in all subgenera. In *C. monogyna*, subgenus *Monogynella* (more ancestral), almost all its 30 chromosomes had 5S and 35S rDNA sites closely positioned, both on the same chromosome arm and on different arms. Co-occurrence of 5S and 35S rDNA sites on the same chromosome is higher in karyotypes with multiple sites and are frequently on the same arm (Roa and Guerra, 2015; Garcia et al., 2017). Even in *Cuscuta* species with few sites, these sites were positioned at the same arm.

In plants, the 5S rDNA usually occupy proximal and less frequently interstitial and terminal regions; while 35S rDNA tends to occupy terminal regions (Roa and Guerra, 2012, 2015). In *Cuscuta*, while 5S rDNA was more frequently found in interstitial

regions, the 35S was frequently found on peri/centromeric regions. In the only three holocentric species of *Cuscuta* with rDNA sites reported in this work, the positions of these sites also diverged from that reported for other groups of plants. Generally, both 5S rDNA and 35S rDNA occupy terminal regions in holocentrics (Roa and Guerra, 2012, 2015). In holocentric *Cuscuta*, however, only the 35S occupied a terminal position, while the 5S rDNA sites were at more interstitial positions.

The ancestral character reconstruction showed the interstitial position as ancestral for the 5S rDNA. This characteristic is present in almost all species, with the exception of *C. monogyna* (*Monogynella*), *C. approximata* (*Cuscuta*) and *C. indecora* (*Grammica*). The 35S was the most variable, and the "mix" condition, when rDNA sites are found in more than one location, was reported several times throughout the phylogeny. Mesquite could not reconstruct the ancestral state of the 35S rDNA position and the number of 5S and 35S sites. The probabilities for each state are very close, therefore inconclusive. The most likely ancestral position for 35S rDNA is the "mix" while the number of sites is one pair. These results are incompatible, as for the "mix" condition to occur it is necessary at least two pairs of 35S rDNA sites. The same happens with the RASP reconstruction, where the most ancestral condition for the 5S rDNA is 18 pairs, which is not compatible with the interstitial ancestral position, since in species with a large number of sites it has various positions (mix). 35S is commonly pericentromeric, however, in species with more sites, these sites are found in interstitial regions and in the subgenus *Cuscuta*, all 35S sites were terminal. These results differ from the position reconstruction of rDNA sites performed by Garcia et al. (2017), where the ancestral position for the 5S and 35S is interstitial and terminal, respectively. This work also shows that the "mix" position occurred more frequently in the 5S.

According to the Plant rDNA Database, within the Convolvulaceae family, only the genus *Ipomoea* has published rDNA data. For the 5S, the peri-centromeric position was the most frequent and the sites varied from one to three pairs. The 35S, on the other hand, occupied predominantly terminal-subterminal positions and the species had between four and nine pairs. Considering these data, in the most ancestral subgenus of *Cuscuta* (*Monogynella*) there is an increase in the diversity of positions in which the rDNA sites were found, suggesting that the "mix" condition is derived and, due to the amplification of these sites, the ribosomal DNAs began to occupy different positions along the chromosome in these species. Recent studies show the possible influence of repetitive DNA amplification on genomic changes in the genus *Cuscuta* (Neumman et al., 2020; Ibiapino et al.,2021) indicating that the increase of these sites in *Monogynella* may be caused by the amplification of rDNA repeats as observed for other tandem repetitive sequences in the genus and possibility not associated to rRNA transcription.

The tandem repetitive DNA in *Cuscuta* appears to be quite complex. Species such as *C. europaea* have species-specific satellite DNA sequences such as the CUS-TR24, colocalized with centromeric proteins, also species-specific (Oliveira et al., 2020). Analysis using long reads showed a complex organization of CUS-TR24. The sequence of this satellite is interspersed with insertions of other repeats, mainly from LINE retrotransposons (Vondrak et al., 2021). In the *Pachystigma* subgenus, there is also evidence of these complex satellites. For example, the CnSat10-1400 found in *C. nitida* in addition to being similar to a LINE type element, co-localizes with 35S signals in the chromosomes of this species. In addition, the most abundant SF1 family of *C. nitida* also co-locates with 35S signals (Ibiapino et al., 2021). This complex organization, with possible insertions of other repetitive DNAs within the rDNA, could influence the diversity of number and position of rDNA sites in *Cuscuta*. In *Allium cepa* L., for example, it was shown that 35S rDNA was able to move from one locus to another in the

genome. Furthermore, it is associated with telomeric DNA and other satellite DNAs, suggesting that the 35S of this species undergoes excision-reintegration mediated by these sequences (Mancia et al., 2015; Fu et al., 2019). Thus, the data suggest that the evolution of rDNA in *Cuscuta* may be influenced by other repeats, therefore contrasting to other angiosperms.

Conclusions

The data support the basic number $n = x = 15$ in *Cuscuta*, with duplications more commonly observed in *Grammica*. As expected, descendent dysploidy occurred predominantly in the holocentric clade (*Cuscuta* subgenus). The genome size evolution in *Cuscuta* may be linked to its parasitic lifestyle. The data showed an expansion of genome size in comparation to the other Convolvulaceae, particularly in some lineages of the genus, and this expansion occurred mostly by repetitive DNA amplification. This amplification of sequences may also have given rise to the great diversity of 5S and 35S ribosomal DNA sites found in the genus, and it seems to contribute to the emergence of "mix"-type karyotypes, in which more than one position is occupied by these rDNA sites. This work analysed data from 58 of the 200 *Cuscuta* species, which represents 29% of *Cuscuta* species, indicating that the karyotypic diversity of the genus may be still underestimated. Nevertheless, *Cuscuta* is one of the exceptionally diverse genera within the angiosperms.

Acknowledgments

We thank the Fundação de Amparo a Ciencia e Tecnologia de Pernambuco (FACEPE) for the financing of the Postgraduate scholarship (process IBPG0993-2.03/17); the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Financial Code 001) for the financial support for the development of the project. BA thanks CAPES for the post-doc fellowship (process #88882.315044/2019-01). Finally, I would like to thanks PhD Pablo E. Rodriguez and MSc Yennifer Mata-Sucre for their help with questions regarding flow cytometry and character reconstruction, respectively.

Author contributions

AI performed the scientific experiments, data collection and writing of the manuscript. BA performed all the phylogenetic analysis and supported the computational analysis. MB contributed by analysing the data and reviewing the manuscript. MG, MC and SS contributed to the collection, identification and molecular data of the plant material and discussions of the data. AP-H supervised and coordinated the project, discussed the data and reviewed the manuscript.

References

Ali, S. S., Yu, Y., Pfosser, M., and Wetschnig, W. (2012). Inferences of biogeographical histories within subfamily Hyacinthoideae using S-DIVA and Bayesian binary MCMC analysis implemented in RASP (Reconstruct Ancestral State in Phylogenies). *Annals of Botany* 109: 95 –107. doi: 10.1093/aob/mcr274.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2012). GenBank. *Nucleic Acids Research* 41, D36–D42. doi:10.1093/nar/gks1195.

Costa, L., Oliveira, Á., Carvalho-Sobrinho, J., and Souza, G. (2017). Comparative cytomolecular analyses reveal karyotype variability related to biogeographic and species richness patterns in Bombacoideae (Malvaceae). *Plant Systematics and Evolution* 303, 1131–1144. doi:10.1007/s00606-017-1427-6.

Costea, M., García, M. A., and Stefanović, S. (2015). A Phylogenetically Based Infrageneric Classification of the Parasitic Plant Genus *Cuscuta* (Dodders, Convolvulaceae). *Systematic Botany* 40, 269–285. doi:10.1600/036364415X686567.

Cusimano, N., Stadler, T., and Renner, S. S. (2012). A new method for handling missing species in diversification analysis applicable to randomly or nonrandomly sampled phylogenies. *Systematic biology* 61, 785–792. doi:10.1093/sysbio/sys031

Darriba, D., Taboada, G. L., Doallo, R., and Posada, D. (2012). jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* 9, 772–772. doi:10.1038/nmeth.2109.

Doležel, J., Greilhuber, J., and Suda, J. (2007). Estimation of nuclear DNA content in plants using flow cytometry. *Nat Protoc* 2, 2233–2244. doi:10.1038/nprot.2007.310.

Fogelberg, S. O. (1938). The Cytology of *Cuscuta*. *Bulletin of the Torrey Botanical Club* 65, 631. doi:10.2307/2481064.

Fu, J., Zhang, H., Guo, F., Ma, L., Wu, J., Yue, M., et al. (2019). Identification and characterization of abundant repetitive sequences in *Allium cepa*. *Sci Rep* 9, 16756. doi:10.1038/s41598-019-52995-9.

García, M. A. (2001). A new western Mediterranean species of *Cuscuta* (Convolvulaceae) confirms the presence of holocentric chromosomes in subgenus *Cuscuta*. *Botanical Journal of the Linnean Society* 135, 169–178. doi:10.1111/j.1095-8339.2001.tb01089.x.

García, M. Á., and Castroviejo, S. (2003). Estudios citotaxonómicos en las especies Ibéricas del género Cuscuta (Convolvulaceae). Anal del Jard Bot de Madrid 60:33–44. Abstract retrieved from Abstracts in Anales del Jardín Botánico de Madrid.

García, M. Á., Costea, M., Guerra, M., García-Ruiz, I., and Stefanović, S. (2019). IAPT chromosome data 31. *TAXON* 68, 1374–1380. doi:10.1002/tax.12176.

García, M. A., Costea, M., Kuzmina, M., and Stefanović, S. (2014). Phylogeny, character evolution, and biogeography of *Cuscuta* (dodders; Convolvulaceae) inferred from coding plastid and nuclear sequences. *American Journal of Botany* 101, 670–690. doi:10.3732/ajb.1300449.

García, M. A., and Martín, M. P. (2007). Phylogeny of *Cuscuta* Subgenus *Cuscuta* (Convolvulaceae) Based on nrDNA ITS and Chloroplast *trnL* Intron Sequences. *Systematic Botany* 32, 899–916. doi:10.1600/036364407783390872.

García, M. A., Stefanović, S., Weiner, C., Olszewski, M., and Costea, M. (2018). Cladogenesis and reticulation in *Cuscuta* sect. Denticulatae (Convolvulaceae). *Organisms Diversity & Evolution* 18, 383–398. doi:10.1007/s13127-018-0383-5.

Garcia, S., Kovařík, A. (2013). Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. *Heredity* 111, 23–33. doi:10.1038/hdy.2013.11 11.

Garcia, S., Kovařík, A., Leitch, A. R., and Garnatje, T. (2017). Cytogenetic features of rRNA genes across land plants: analysis of the Plant rDNA database. *The Plant Journal* 89, 1020–1030. doi:10.1111/tpj.13442.

Gerlach, W. L., and Bedbrook, J. R. (1979). Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucleic Acids Research* 7, 1869–1885. doi: 10.1093/nar/7.7.1869.

Glick, L., and Mayrose, I. (2014). ChromEvol: Assessing the Pattern of Chromosome Number Evolution and the Inference of Polyploidy along a Phylogeny. *Molecular Biology and Evolution* 31, 1914–1922. doi:10.1093/molbev/msu122.

Gruner, A., Hoverter, N., Smith, T., and Knight, C. A. (2010). Genome size is a strong predictor of root meristem growth rate. *Journal of Botany* 2010, 4.doi: 10.1155/2010/390414.

Guerra, M. (2008). Chromosome numbers in plant cytotaxonomy: concepts and implications. *Cytogenetic and Genome Research* 120, 339–350. doi:10.1159/000121083.

Guerra, M., and García, M. A. (2004). Heterochromatin and rDNA sites distribution in the holocentric chromosomes of *Cuscuta approximata* Bab. (Convolvulaceae). *Genome* 47, 134–140. doi:10.1139/g03-098.

Heslop-Harrison, J. S. P., and Schwarzacher, T. (2011). Organisation of the plant genome in chromosomes: Organisation of the plant genome in chromosomes. *The Plant Journal* 66, 18–33. doi:10.1111/j.1365-313X.2011.04544.x.

Ibiapino, A., García, M. Á., Costea, M., Stefanović, S., and Guerra, M. (2020). Intense proliferation of rDNA sites and heterochromatic bands in two distantly related *Cuscuta* species (Convolvulaceae) with very large genomes and symmetric

karyotypes. *Genet. Mol. Biol.* 43, e20190068. doi:10.1590/1678-4685-gmb-2019-0068.

Ibiapino, A., García, M. A., Ferraz, M. E., Costea, M., Stefanović, S., and Guerra, M. (2019). Allopolyploid origin and genome differentiation of the parasitic species *Cuscuta* veatchii (Convolvulaceae) revealed by genomic in situ hybridization. *Genome*, 1–9. doi:10.1139/gen-2018-0184.

Ibiapino, A., Baez, M., García M. A., Costea, M., Stefanović, S., and Pedrosa-Harand, A. (2021). Karyotype asymmetry in *Cuscuta* L. subgenus *Pachystigma* reflects its repeat DNA composition. bioRxiv. doi: https://doi.org/10.1101/2021.08.09.455742.

Ishikawa, S. A., Zhukova, A., Iwasaki, W., and Gascuel, O. (2019). A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Molecular Biology and Evolution* 36, 2069–2085. doi:10.1093/molbev/msz131.

Kaul, M. L. H., and Bhan, A. K. (1977). Cytogenetics of Polyploids: VI. Cytology of tetraploid and hexaploid *Cuscuta reflexa* Roxb. *cytologia* 42, 125–136. doi:10.1508/cytologia.42.125.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., et al. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649. doi: 10.1093/bioinformatics/bts199.

Lee, Y.-I., Chang, F.-C., and Chung, M.-C. (2011). Chromosome pairing affinities in interspecific hybrids reflect phylogenetic distances among lady's slipper orchids (Paphiopedilum). *Annals of Botany* 108, 113–121. doi:10.1093/aob/mcr114.

Loureiro, J., Rodriguez, E., Dolezel, J., and Santos, C. (2007). Two New Nuclear Isolation Buffers for Plant DNA Flow Cytometry: A Test with 37 Species. *Annals of Botany* 100, 875–888. doi:10.1093/aob/mcm152.

Maddison, W. P., and Maddison, D. R. (2018). Mesquite: a modular system for evolutionary analysis. version 0.992. Available from: http://www.mesquiteproject.org/.

Mancia, F. H., Sohn, S.-H., Ahn, Y. K., Kim, D.-S., Kim, J. S., Kwon, Y.-S., et al. (2015). Distribution of various types of repetitive DNAs in *Allium cepa* L. based on dual color FISH. *Hortic. Environ. Biotechnol.* 56, 793–799. doi:10.1007/s13580-015-1100-3.

Mandrioli, M., and Manicardi, G. C. (2020). Holocentric chromosomes. *PLoS Genet* 16, e1008918. doi:10.1371/journal.pgen.1008918.

Márquez-Corro, J. I., Martín-Bravo, S., Spalink, D., Luceño, M., and Escudero, M. (2019). Inferring hypothesis-based transitions in clade-specific models of chromosome number evolution in sedges (Cyperaceae). *Molecular Phylogenetics and Evolution* 135, 203–209. doi:10.1016/j.ympev.2019.03.006.

Mayrose, I., and Lysak, M. A. (2021). The Evolution of Chromosome Numbers: Mechanistic Models and Experimental Approaches. *Genome Biology and Evolution* 13, evaa220. doi:10.1093/gbe/evaa220.

McNeal, J. R., Kuehl, J. V., Boore, J. L., and de Pamphilis, C. W. (2007). Complete plastid genome sequences suggest strong selection for retention of photosynthetic genes in the parasitic plant genus *Cuscuta*. *BMC Plant Biology* 7, 57. doi:10.1186/1471-2229-7-57.

Miller, M.A., Pfeiffer, W., Schwartz, T., (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gatway Computing Environments Worshop (GCE). Available from: http://www.phylo.org/index.php/.

Moraes, A. P., Souza-Chies, T. T., Stiehl-Alves, E. M., Burchardt, P., Eggers, L., Siljak-Yakovlev, S., et al. (2015). Evolutionary trends in Iridaceae: new cytogenetic findings from the New World. *Botanical Journal of the Linnean Society* 177, 27–49. doi:10.1111/boj.12232.

Neumann, P., Oliveira, L., Čížková, J., Jang, T.-S., Klemme, S., Novák, P., et al. (2020). Impact of parasitic lifestyle and different types of centromere organization on chromosome and genome evolution in the plant genus *Cuscuta*. *New Phytol.* 4, 2365–2377. doi:10.1111/nph.17003.

Oliveira, L., Neumann, P., Jang, T.-S., Klemme, S., Schubert, V., Koblížková, A., et al. (2020). Mitotic Spindle Attachment to the Holocentric Chromosomes of *Cuscuta europaea* Does Not Correlate With the Distribution of CENH3 Chromatin. *Front. Plant Sci.* 10, 1799. doi:10.3389/fpls.2019.01799.

Paradis, E., Claude, J., and Strimmer, K. (2004). APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20, 289–290. doi:10.1093/bioinformatics/btg412.

Pazy, B., and Plitmann, U. (1994). Holocentric chromosome behaviour in *Cuscuta* (Cuscutaceae). *Plant Systematics and Evolution* 191, 105–109. doi:10.1007/BF00985345.

Pazy, B., and Plitmann, U. (1995). Chromosome divergence in the genus *Cuscuta* and its systematic implications. *Caryologia* 48, 173–180. doi:10.1080/00087114.1995.10797327.

Pazy, B., and Plitmann, U. (2002). New perspectives on the mechanisms of chromosome evolution in parasitic flowering plants. *Botanical Journal of the Linnean Society* 138, 117–122. doi:10.1046/j.1095-8339.2002.00006.x.

Pedrosa, A., Sandal, N., Stougaard, J., Schweizer, D., and Bachmair, A. (2002). Chromosomal map of the model legume *Lotus japonicus*. *Genetics* 161, 1661–1672. doi:10.1093/genetics/161.4.1661.

Piednoël, M., Aberer, A. J., Schneeweiss, G. M., Macas, J., Novak, P., Gundlach. H. et al. (2012). Next-generation sequencing reveals the impact of repetitive DNA

across phylogenetically closely related genomes of Orobanchaceae. *Molecular Biology and Evolution* 29, 3601–3611. doi: 10.1093/molbev/mss168.

Rambaut A., 2014. FigTree v1.4.2. Available from http://tree.bio.ed.ac.uk/software /Figtree/.

Rambaut A., Suchard, M.A., Xie, W., Drummond, A.J., 2014. TRACER v.1.6, Available from http://tree.bio.ed.ac.uk/software/tracer/.

R Core Team, 2020. R: A Language and Environment for Statistical Computing. Available from http://www.Rproject.org/.

Revell, L. J. (2012). phytools: an R package for phylogenetic comparative biology (and other things): phytools: R package. *Methods in Ecology and Evolution* 3, 217–223. doi:10.1111/j.2041-210X.2011.00169.x.

Rice, A., and Mayrose, I. (2021). Model adequacy tests for probabilistic models of chromosome-number evolution. *New Phytol* 229, 3602–3613. doi:10.1111/nph.17106.

Roa, F., and Guerra, M. (2012). Distribution of 45S rDNA sites in chromosomes of plants: Structural and evolutionary implications. *BMC Evol Biol* 12, 225. doi:10.1186/1471-2148-12-225.

Roa, F., and Guerra, M. (2015). Non-Random Distribution of 5S rDNA Sites and Its Association with 45S rDNA in Plant Chromosomes. *Cytogenetic and Genome Research* 146, 243–249. doi:10.1159/000440930.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D. L., Darling, A., Höhna, S., et al. (2012). MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space. *Systematic Biology* 61, 539–542. doi:10.1093/sysbio/sys029.

Sader, M. A., Amorim, B. S., Costa, L., Souza, G., and Pedrosa-Harand, A. (2019). The role of chromosome changes in the diversification of *Passiflora* L. (Passifloraceae). *Systematics and Biodiversity* 17, 7–21. doi:10.1080/14772000.2018.1546777.

Schubert, I., and Lysak, M. A. (2011). Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends in Genetics* 27, 207–216. doi:10.1016/j.tig.2011.03.004.

Stefanović, S., and Costea, M. (2008). Reticulate evolution in the parasitic genus *Cuscuta* (Convolvulaceae): over and over. *Botany* 86, 791–808. doi:10.1139/B08-033.

Sun, G., Xu, Y., Liu, H., Sun, T., Zhang, J., Hettenhausen, C., et al. (2018). Large-scale gene losses underlie the genome evolution of parasitic plant *Cuscuta australis*. *Nat Commun* 9, 2683. doi:10.1038/s41467-018-04721-8.

Vaio, M., Gardner, A., Emshwiller, E., and Guerra, M. (2013). Molecular phylogeny and chromosome evolution among the creeping herbaceous *Oxalis* species of

sections Corniculatae and Ripariae (Oxalidaceae). *Molecular Phylogenetics and Evolution* 68, 199–211. doi:10.1016/j.ympev.2013.03.019.

Vaio, M., Mazzella, C., Guerra, M., and Speranza, P. (2019). Effects of the diploidisation process upon the 5S and 35S rDNA sequences in the allopolyploid species of the Dilatata group of Paspalum (Poaceae, Paniceae). *Aust. J. Bot.* 67, 521. doi:10.1071/BT18236.

Vaio, M., Nascimento, J., Mendes, S., Ibiapino, A., Felix, L. P., Gardner, A., et al. (2018). Multiple karyotype changes distinguish two closely related species of *Oxalis* (*O. psoraleoides* and *O. rhombeo-ovata*) and suggest an artificial grouping of section Polymorphae (Oxalidaceae). *Botanical Journal of the Linnean Society*. doi:10.1093/botlinnean/boy054.

Volkov, R. A. (2017). Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. *BMC Plant Biol* 17, 21. doi:10.1186/s12870-017-0978-6.

Vondrak, T., Oliveira, L., Novák, P., Koblížková, A., Neumann, P., and Macas, J. (2021). Complex sequence organization of heterochromatin in the holocentric plant *Cuscuta europaea* elucidated by the computational analysis of nanopore reads. *Computational and Structural Biotechnology Journal* 19, 2179–2189. doi:10.1016/j.csbj.2021.04.011.

Yu, Y., Harris, A. J., Blair, C., He, X. (2015). RASP (Reconstruct Ancestral State in Phylogenies): a tool for historical biogeography. *Molecular Phylogenetics and Evolution* 87, 46-49. doi: 10.1016/j.ympev.2015.03.008.

Yu, Y., Blair, C., He, X. (2019). RASP 4: ancestral state reconstruction tool for multiple genes and characters. *Molecular Biology and Evolution* 37(2), 604-606. doi: 10.1093/molbev/msz257.

Table1: Haploid chromosome numbers (*n*) considered for character reconstruction

Espécie	***n***	**References**
Cuscuta approximata	14	Guerra and García (2004)
Cuscuta babylonica	4	Pazy and Plitmann (2002)
Cuscuta capitata	10	Mehra and Vasudevan (1972)
Cuscuta epilinum	21	McNeal et al. (2007)
Cuscuta epithymum	7, 8, 14, 15, 16 and 17	García and Castroviejo (2003)
Cuscuta europaea	7	García and Castroviejo (2003).
Cuscuta pedicellata	5	Pazy and Plitmann (1991)
Cuscuta planiflora	13 and 14	García and Castroviejo (2003)
Cuscuta americana	15	García et al. (2019)
Cuscuta australis	15	García and Castroviejo(2003)
Cuscuta bonafortunae	15	García et al. (2019)
Cuscuta brachycalyx	15	García et al. (2019)
Cuscuta californica	15	Neumann et al. (2020)

Cuscuta campestris	28	Garcia et al (2019)
Cuscuta cephalanthi	30	McNeal et al.(2007)
Cuscuta chapalana	15	Garcia et al. (2019)
Cuscuta chilensis	14	Aryavand (1987)
Cuscuta chinensis	14	Aryavand (1987)
Cuscuta compacta	15	Garcia et al. (2019)
Cuscuta coryli	15	Fogelberg (1938)
Cuscuta corymbosa var. *grandiflora*	15	Garcia et al. (2019)
Cuscuta costaricensis	15	Garcia et al. (2019)
Cuscuta cotijana	30	Garcia et al. (2019)
Cuscuta cuspidata	15	Pazy and Plitmann (1995)
Cuscuta denticulata	15	Raven et al. (1965)
Cuscuta desmouliniana	15	Garcia et al. (2019)
Cuscuta erosa	15	Garcia et al. (2019)
Cuscuta glomerata	15	Garcia et al. (2019)
Cuscuta grandiflora	15	Garcia et al. (2019)
Cuscuta gronovii	30	Love (1982)
Cuscuta howelliana	15	Garcia et al. (2019)
Cuscuta indecora	15	Ibiapino et al. (2020)
Cuscuta indecora var. *neuropetala*	15	Fogelberg (1938)
Cuscuta nevadensis	15	
Cuscuta obtusiflora	15	Garcia et al. (2019)
Cuscuta occidentalis	15	Garcia et al. (2019)
Cuscuta pacifica	15	Garcia et al. (2019)
Cuscuta partita	15	Ibiapino et al (2019)
Cuscuta pentagona	28	Pazy and Plitmann (1995)
Cuscuta psorothamnensis	30	
Cuscuta purpurata	15	Garcia et al. (2019)
Cuscuta racemosa	30	Garcia et al. (2019)
Cuscuta salina	ca. 15	Pazy and Plitmann (1995)
Cuscuta sandwichiana	ca. 75	Garcia et al. (2019)
Cuscuta sidarum	15	Garcia et al. (2019)
Cuscuta subinclusa	15	Garcia et al. (2019)
Cuscuta tinctoria	19	Pazy and Plitmann (1995)
Cuscuta tinctoria var. *floribunda*	15	Garcia et al. (2019)
Cuscuta umbrosa	15	Garcia et al. (2019)
Cuscuta globosa	45	This study
Cuscuta veatchii	30	Ibiapino et al. (2019)
Cuscuta volcanica	15	Garcia et al 2019
Cuscuta japonica	15	Neumann et al. (2020)
Cuscuta lupuliformis	14	McNeal et al. (2007)
Cuscuta monogyna	15	García and Castroviejo (2003)
Cuscuta reflexa	16	Neumann et al. (2020)
Cuscuta africana	14	Ibiapino et al. (2021)

Cuscuta angulata	15	Ibiapino et al. (2021)
Cuscuta nitida	14	Ibiapino et al. (2021)

Table 2: Genome size, 5S and 35S ribosomal DNA sites number and position in species of the genus *Cuscuta*

Species	**1C (Gbp)**	**5S/35S* ****	**References (Genome size/rDNA)**
Cuscuta americana	0.68 and 0.69		Neumann et al., 2020; This study
Cuscuta approximata		2T+4I/2T	Guerra and García, (2004)
Cuscuta australis	0.27, 0.34 and 0.69	2I/2P	Sun et al., 2018; Neumann et al., 2020; This study; This study
Cuscuta californica	0.39		Neumann et al., 2020
Cuscuta campestris	0.45, 0.55 and 0.58	4I/2I+2P	This study; Neumann et al., 2020; Vogel et al., 2018; This study
Cuscuta cephalanthi	3.68 and 3.83		This study; McNeal et al., 2007
Cuscuta chilensis	2.80		McNeal et al., 2007
Cuscuta compacta	3.24 and 7.67		This study; McNeal et al., 2007
Cuscuta denticulata		2I/2P	Ibiapino et al., 2019
Cuscuta epilinum	1.54 and 3.38		Neumann et al., 2020; McNeal et al., 2007
Cuscuta epithymum	0.53	4I/2T	Neumann et al., 2020; This study
Cuscuta europaea	1.051 and 1.17	2I/2T	McNeal et al., 2007; Neumann et al., 2020
Cuscuta glomerata	5.16	2I/2P	This study
Cuscuta gronovii	3.58		Neumann et al., 2020
Cuscuta gronovii (C PA)	6.75		McNeal et al., 2007
Cuscuta gronovii (NJ)	3.70		McNeal et al., 2007
Cuscuta gronovii (OH)	3.51		McNeal et al., 2007
Cuscuta gronovii (SE PA)	2.14		McNeal et al., 2007
Cuscuta howelliana		2I/2P	This study
Cuscuta indecora	22.68, 24.46 and 32.05	6I + 4I/4P	Neumann et al., 2020; Ibiapino et al., 2019; McNeal et al., 2007; Ibiapino et al., 2020
Cuscuta japonica	25.58		Neumann et al., 2020
Cuscuta lupuliformis	21.97		McNeal et al., 2007
Cuscuta monogyna	32.45 and 33.05	36/30	Neumann et al., 2020; Ibiapino et al., 2020; Ibiapino et al., 2020

Cuscuta nevadensis		6I/8I+2P	Ibiapino et al., 2019
Cuscuta nitida		2I/4P	Ibiapino et al., 2021
Cuscuta obtusiflora	0.77		McNeal et al., 2007
Cuscuta partita	1.83	2I/2P	This study; This study
Cuscuta pentagona	0.55 and 0.57		Neumann et al., 2020; McNeal et al., 2007
Cuscuta psorothamnensis		4I/2I+2P	This study
Cuscuta purpurata	2.96		This study
Cuscuta racemosa	1.39	4I/2I+2P	This study; This study
Cuscuta reflexa	34.73		Neumann et al., 2020
Cuscuta sandwichiana	1.80	2I/2I+2P	This study; This study
Cuscuta globosa	1.79	6I/2I+2P	This study; This study
Cuscuta veatchii	2.85	6I/2I+2P	McNeal et al., 2007; Ibiapino et al., 2019

*T = terminal, P= peri/centromeric, I= interstitial

**Ribosomal DNA sites are representented in number not in pair

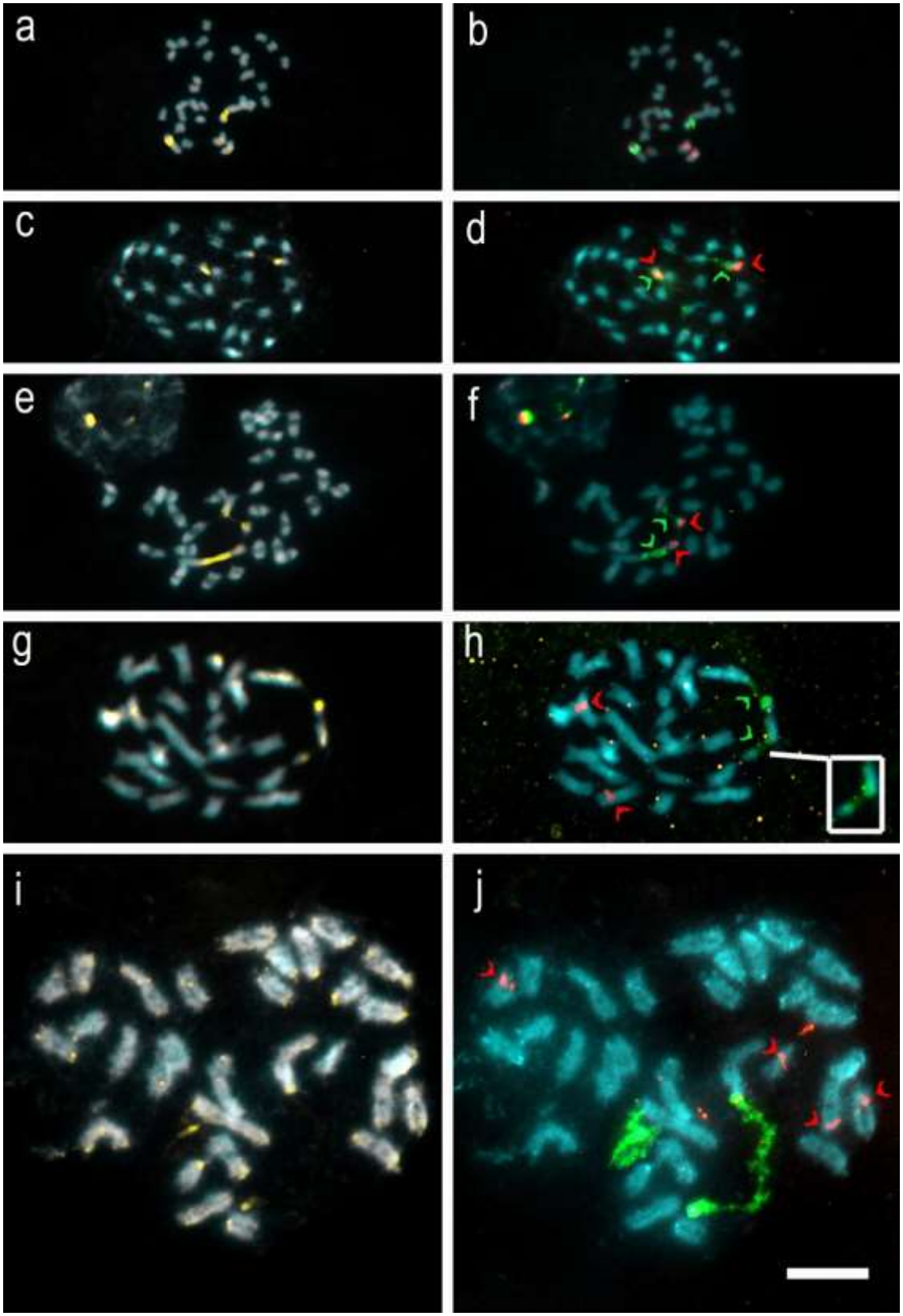


Figure 1: Mitotic metaphases of diploids with karyotypes 2*n* = 30 *C. australis* (**a** and **b**), *C. howelliana* (**c** and **d**), *C. partita* (**e** and **f**), *C. glomerata* (**g** and **h**) and *C. epithymum* (**i** and **j**) stained with CMA (yellow) and DAPI (blue) in **a**, **c**, **e**, **g** and **i**, and with FISH of

rDNA 5S (red) and 35S (green) in **b**, **d**, **f**, **h** and **j**. Red and green arrowheads in **d**, **f**, **h** and **j** showing the 5S and 35S rDNA sites, respectively. Bar in **j** representing 10µm.

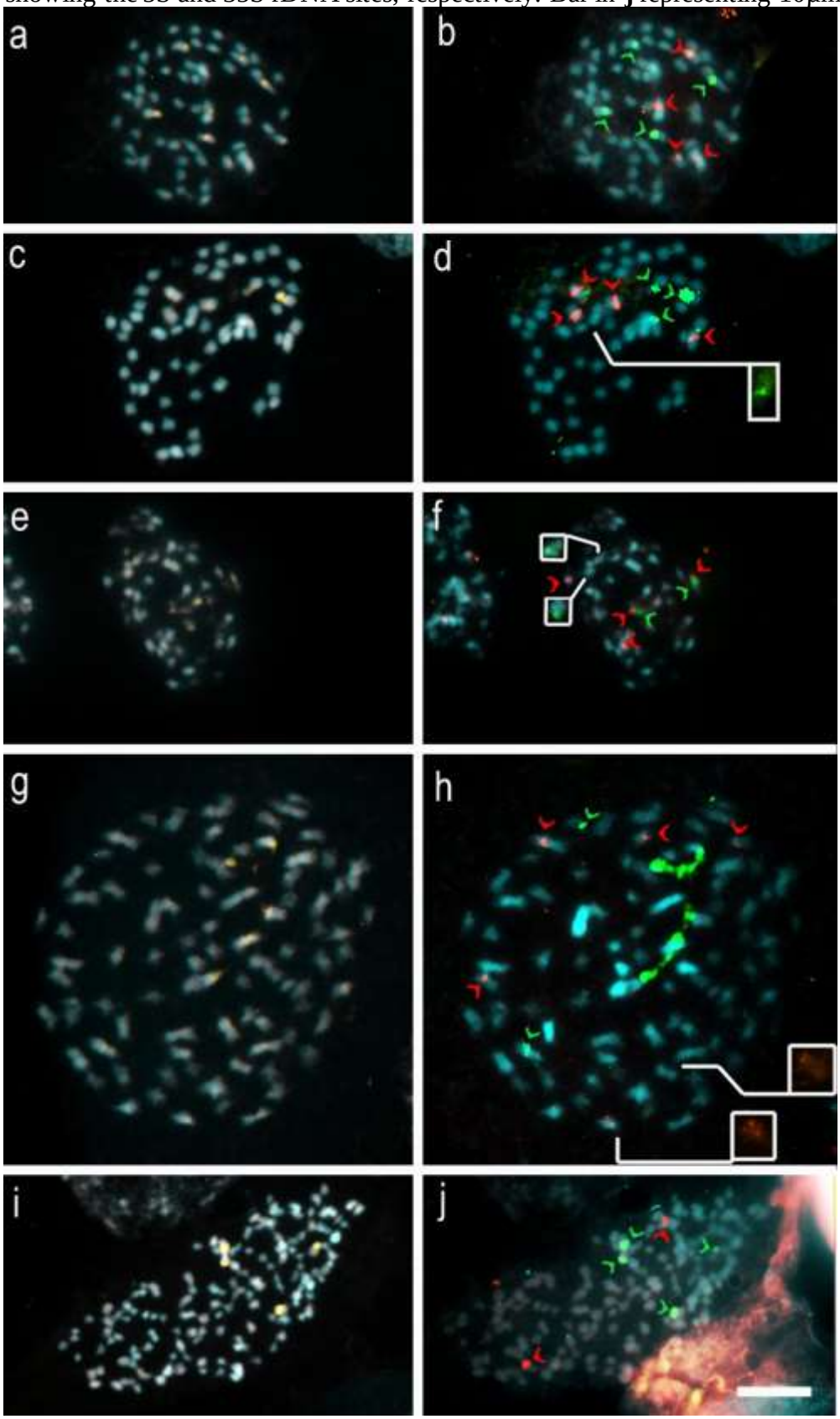


Figure 2: Mitotic metaphases of the polyploids *C. campestris* 2*n* = 56 (**a** and **b**), *C. racemosa* 2*n* = 60 (**c** and **d**), *C. psorothamnensis* 2*n* = 60 (**e** and **f**), *C. globosa* 2*n* = 90 (**g** e **h**) and *C. sandwichiana* 2*n* = 150 (**i** e **j**) stained with CMA (yellow) and DAPI

(blue) in **a**, **c**, **e**, **g** e **i**, and with FISH of rDNA 5S (red) and 35S (green) in **b**, **d**, **f**, **h** and **j**. Red and green arrowheads in **b**, **d**, **f**, **h** and **j** showing the 5S and 35S rDNA sites, respectively Bar in **j** representing 10µm.

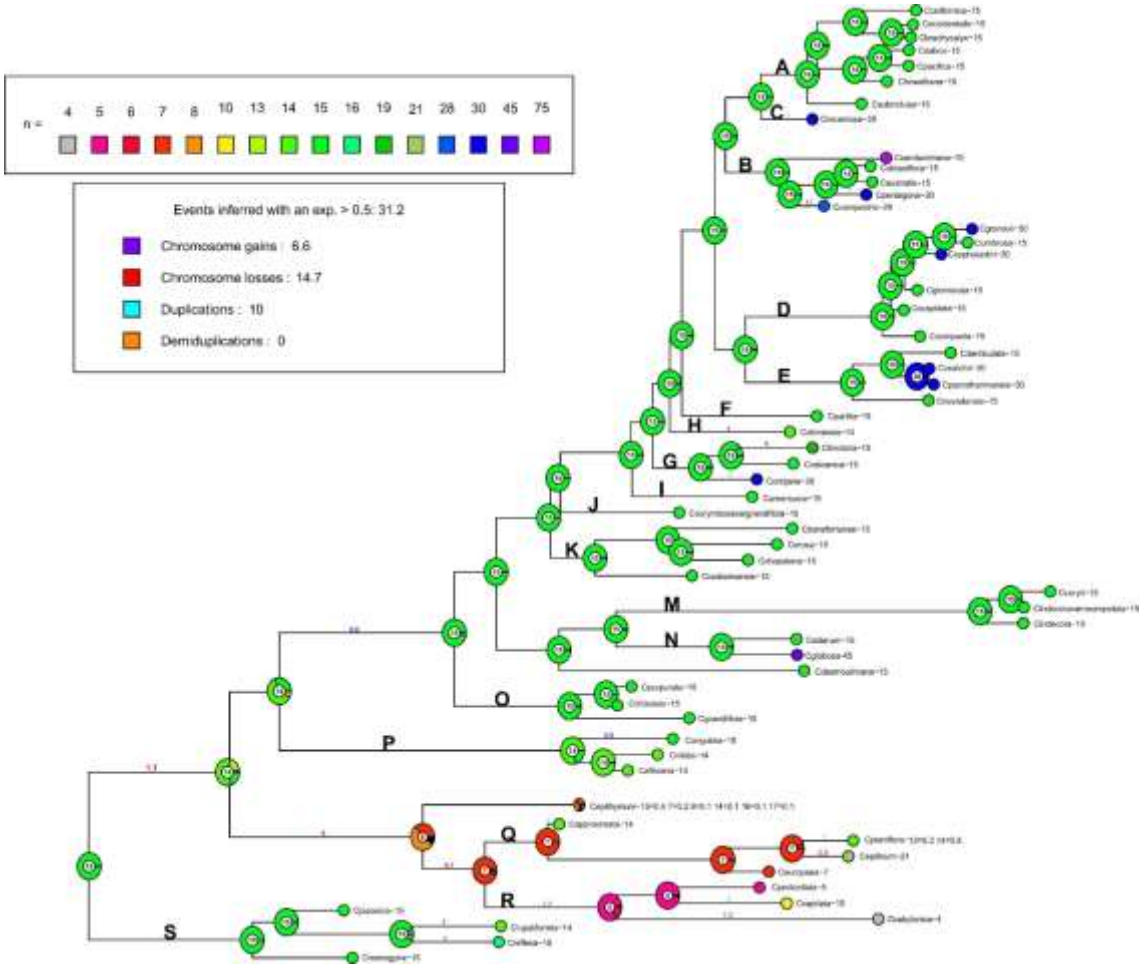


Figure 3: Reconstruction of the chromosome number evolution in *Cuscuta* with the BASE_NUM_DUPL model and the number $n = 15$ fixed in the base. The pie charts at nodes represent the probability of each inferred chromosome number, the numbers along the branches represent the inferred number of the four different types of events (gains, losses, duplications and demiduplications). The bold letters represent the sections described by Garcia et al. (2014), section **S** (subgenus *Monogynella*), sections **R** and **Q** (subgenus *Cuscuta*), section **P** (subgenus *Pachystigma*) and sections **A - O** (subgenus *Grammica*)

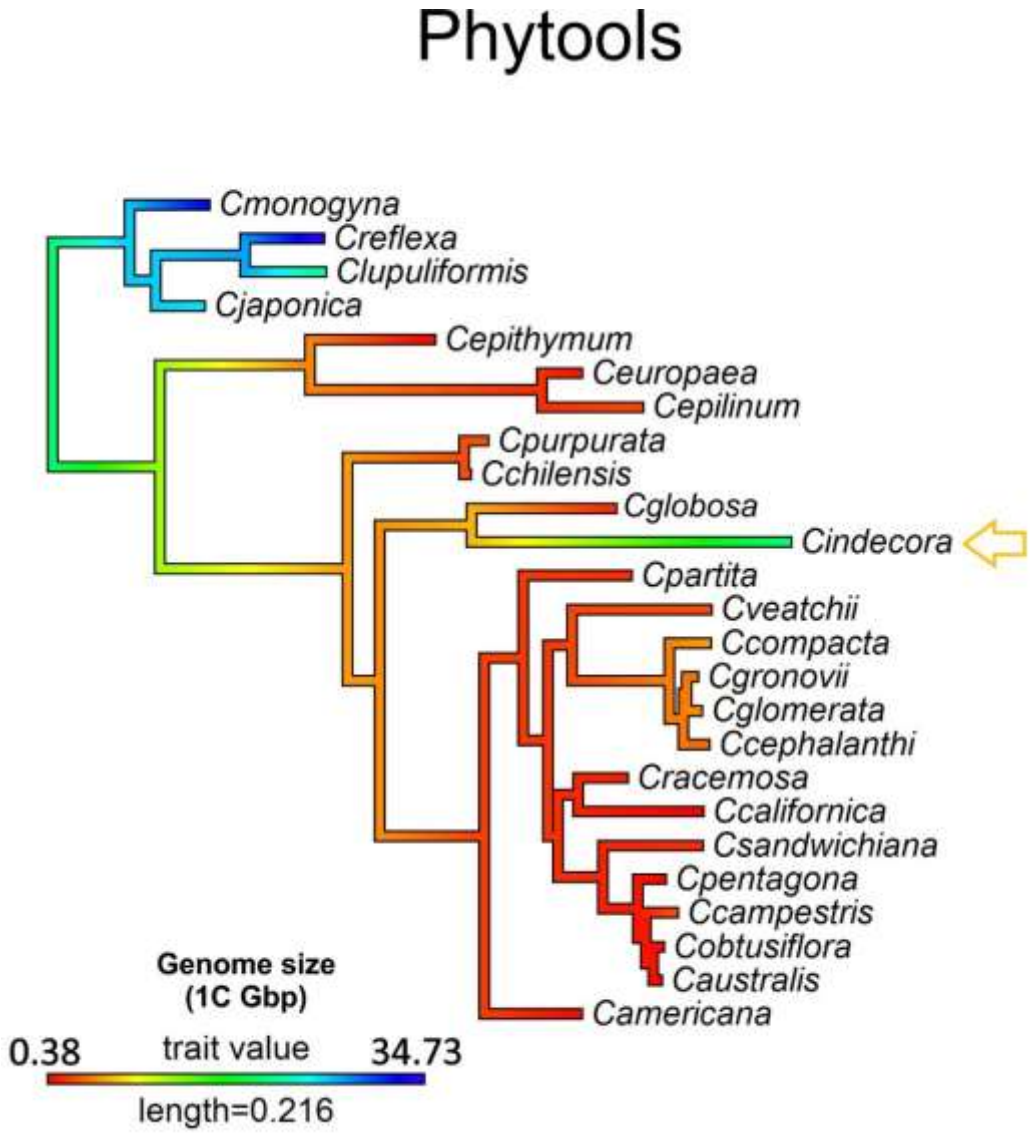


Figure 4: Reconstruction of the genome size made in Phytools R. The variation is shown in a color scale. The largest genomes are showing in shades of blue and in shades of red and yellow are the smaller genomes. Yellow arrow in reconstruction showing *C. indecora*, a species of the subgenus *Grammica* where there was an expansion of the genome.

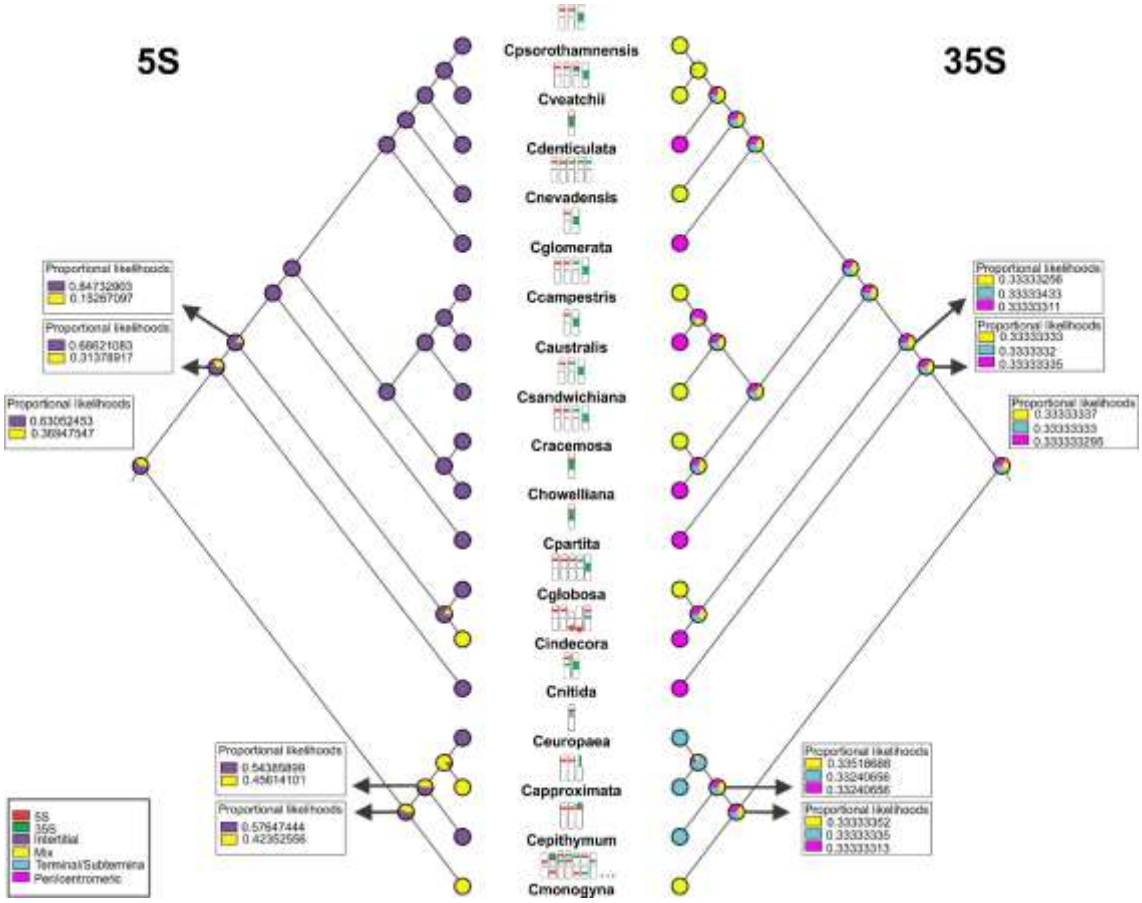


Figure 5: Reconstruction of the position of the rDNA 5S (left) and 35S (right) sites. In the most basal subgenus of *Cuscuta* (*Monogynella*) there is an increase in the diversity of positions in which the rDNA sites were found, suggesting that the "mix" condition is derived and, due to the amplification of these sites, the ribosomal DNAs began to occupy different positions along the chromosome in these species. In *C. monogyna* not all pairs are represented because it is a species that has more than 30 sites of rDNA, in the Figure all the patterns found in this species are outlined.

Supplementary data

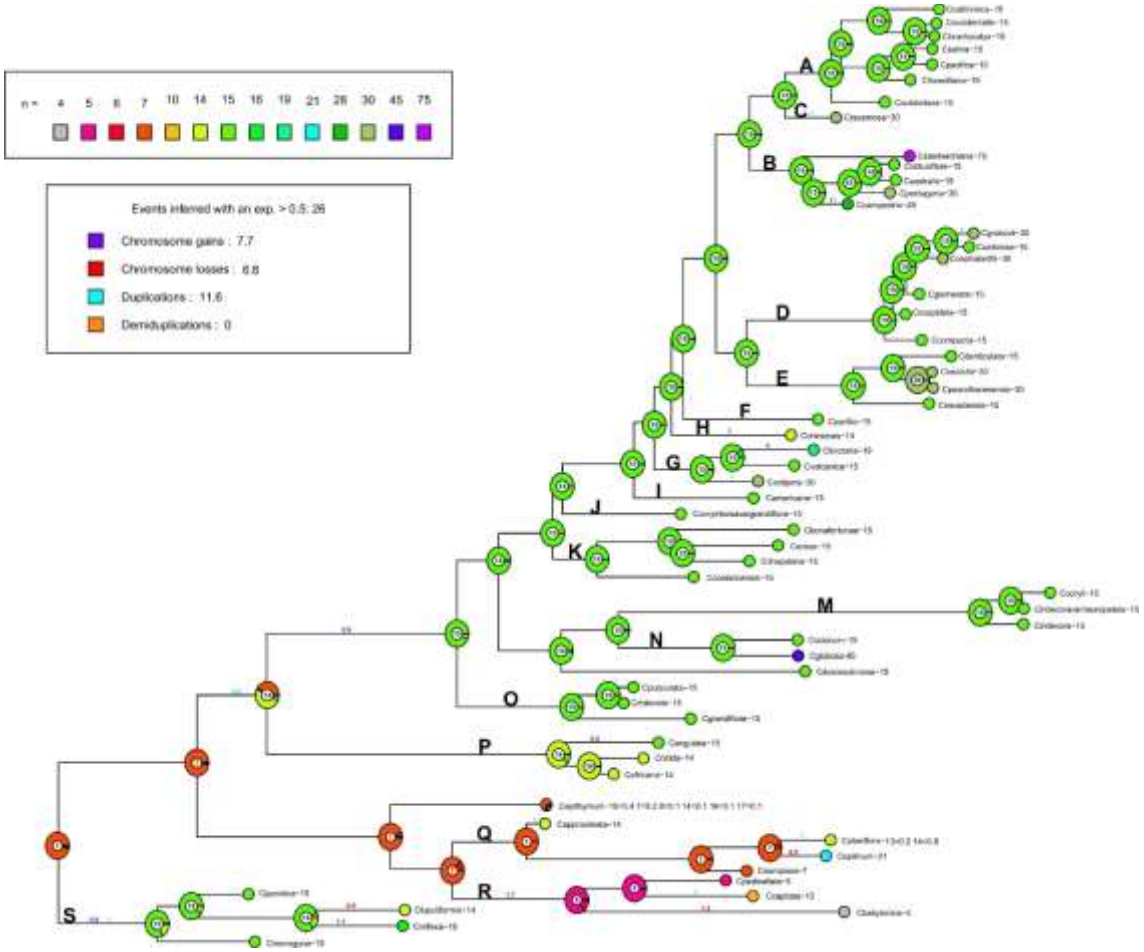
Supplementary Table 1: Sequences accessions used for phylogenetic reconstruction

Subgenera	**Clade**	**Specie**	**rbcL**	**26S**	**trnL-F**	**nrITS**
Cuscuta	Q	*Cuscuta approximata*	KJ436605	KJ400040	EF202557	KY020426
Cuscuta	R	*Cuscuta babylonica*	KJ436611	KJ400047		KU761258
Cuscuta	R	*Cuscuta capitata*	KJ436615	KJ400051		DQ924584
Cuscuta	Q	*Cuscuta epilinum*	KJ436641	KJ400079	AY558849	DQ924610
Cuscuta	Q	*Cuscuta epithymum*	KJ436643	KJ400080	KC569804	AY554400
Cuscuta	Q	*Cuscuta europaea*	AY101060	KJ400082	AY558851	AY554401
Cuscuta	R	*Cuscuta pedicellata*	KJ436700	KJ400151		DQ924582
Cuscuta	Q	*Cuscuta planiflora*	KJ436702	KJ400153	AY558858	AY558822
Grammica	I	*Cuscuta americana*	KJ436602	KJ400038	EF194398	EF194597
Grammica	B	*Cuscuta australis*	KJ436607	KJ400043	EF194457	EF194667
Grammica	K	*Cuscuta bonafortunae*		JN234780	JN234799	JN234809
Grammica	A	*Cuscuta brachycalyx*	EU883441	EU883489		
Grammica	A	*Cuscuta californica*	EU883445	EU883493	EF194479	EF194691
Grammica	B	*Cuscuta campestris*	EU883476	KJ400050	EF194453	EF194661
Grammica	D	*Cuscuta cephalanthi*	KJ436618	KJ400054	EF194413	EF194632
Grammica	K	*Cuscuta chapalana*	KJ436620	JN234783	EF194338	EF194578
Grammica	O	*Cuscuta chilensis*		KJ400056		EF194523
Grammica	H	*Cuscuta chinensis*	KJ436621	KJ400059	EF194368	
Grammica	D	*Cuscuta compacta*	KJ436626	KJ400064	EF194426	EF194640
Grammica	M	*Cuscuta coryli*	KJ436629	KJ400067	EF194288	EF194539
Grammica	J	*Cuscuta corymbosa* var. *grandiflora*	KJ436630	KJ400068	EF194343	EF194584
Grammica	K	*Cuscuta costaricensis*	KJ436633	JN234786	EF194340	EF194580
Grammica	G	*Cuscuta cotijana*	KJ436635	KJ400072	KC485355	KC485379
Grammica	D	*Cuscuta cuspidata*	KJ436637	KJ400075	EF194429	EF194643

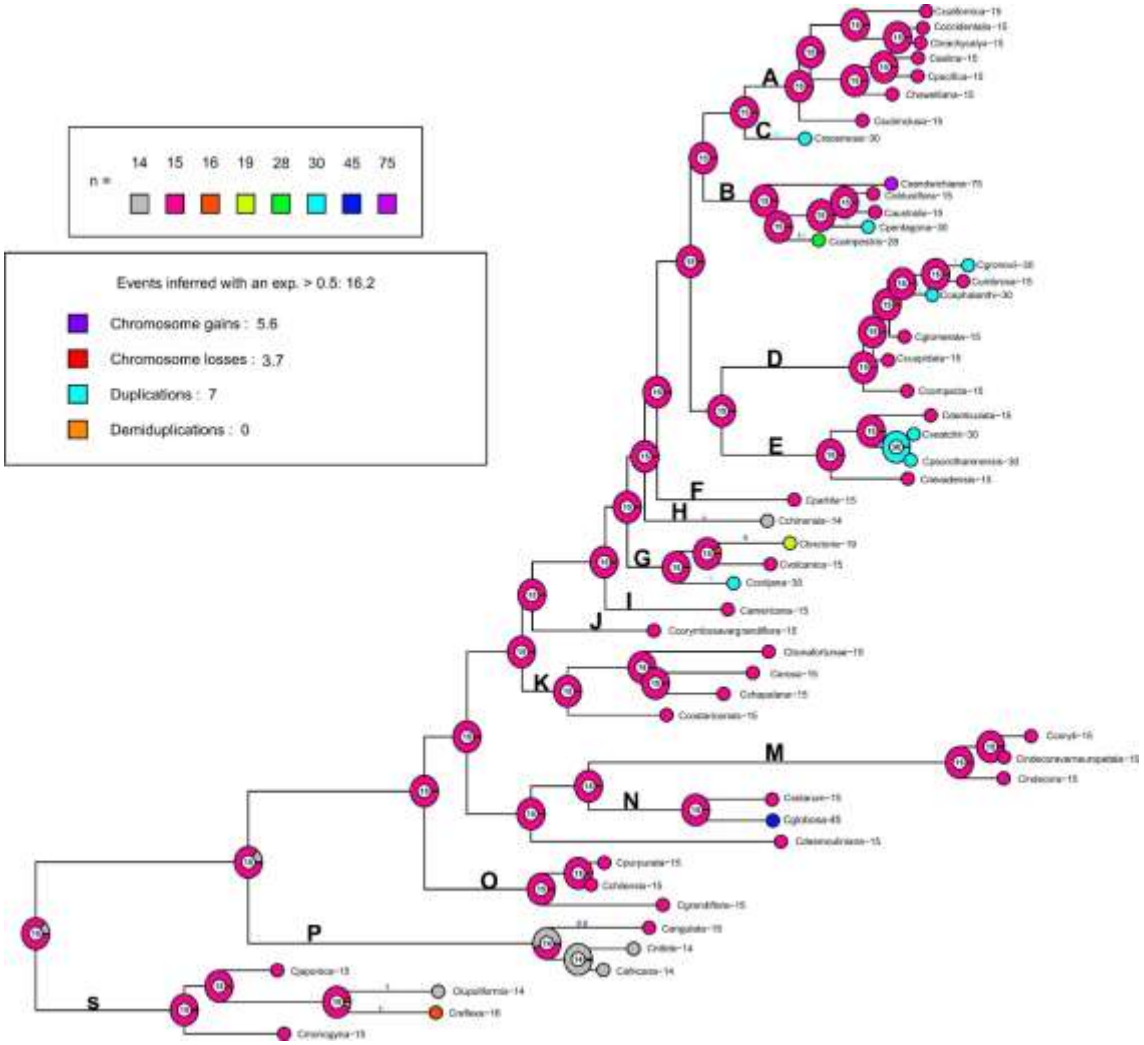
Grammica	E	*Cuscuta denticulata*	KJ436639	KJ400077	EF194411	EF194627
Grammica	L	*Cuscuta desmouliniana*	KJ436640	KJ400078	EU288341	EU288359
Grammica	K	*Cuscuta erosa*		JN234789	JN234804	EF194574
Grammica	D	*Cuscuta glomerata*	KJ436651	KJ400094	EF194431	EF194644
Grammica	O	*Cuscuta grandiflora*		KJ400097		EF194535
Grammica	D	*Cuscuta gronovii*	KJ436654	EU883530	EF194422	EF194639
Grammica	A	*Cuscuta howelliana*	EU883456	EU883505	EF194504	EF194717
Grammica	M	*Cuscuta indecora*	KJ436664	KJ400107	EF194300	EF194549
Grammica	M	*Cuscuta indecora* var. *neuropetala*			EF194301	EF194544
Grammica	E	*Cuscuta nevadensis*	KJ436689	KJ400136	EF194408	EF194630
Grammica	B	*Cuscuta obtusiflora*	KJ436693	KJ400140	EF194463	EF194673
Grammica	A	*Cuscuta occidentalis*	EU883459	EU883509	EF194477	EF194695
Grammica	A	*Cuscuta pacifica*	EU883463	EU883513		
Grammica	F	*Cuscuta partita*	KJ436698	KJ400149	EF194353	EF194591
Grammica	B	*Cuscuta pentagona*	KJ436701	KJ400152	EF194467	EF194679
Grammica	E	*Cuscuta psorothamnensis*			MH920299	MH923173
Grammica	O	*Cuscuta purpurata*		KJ400159		EF194526
Grammica	C	*Cuscuta racemosa*			EF194449	
Grammica	A	*Cuscuta salina*	EU883465	EU883515	GQ254882	GQ254890
Grammica	B H	*Cuscuta sandwichiana*	KJ436712	KJ400165	EU288333	EU288356
Grammica	N	*Cuscuta sidarum*	KJ436713	KJ400167	EF194309	EF194553
Grammica	A	*Cuscuta subinclusa*	EU883470	EU883521	EF194491	EF194703
Grammica	G	*Cuscuta tinctoria*	KJ436721	KJ400175	EF194394	EF194618
Grammica	G	*Cuscuta tinctoria* var. *floribunda*	KJ436723	KJ400177		
Grammica	D	*Cuscuta umbrosa*	KJ436728	KJ400182	EF194435	EF194646
Grammica	N	*Cuscuta globosa*	KJ436729	KJ400183		
Grammica	E	*Cuscuta veatchii*	KJ436730	KJ400184	MH920306	MH923183
Grammica	G	*Cuscuta volcanica*	KJ436734	KJ400188	KC485376	KC485421

Monogynella	S	*Cuscuta japonica*	AY101061	KJ400114	AY101170	KP015829
Monogynella	S	*Cuscuta lupuliformis*	KJ436680	KJ400126	AY558854	KU707914
Monogynella	S	*Cuscuta monogyna*	KJ436686	KJ400133	MH115452	MH109269
Monogynella	S	*Cuscuta reflexa*	KJ436708		AY558859	AY558823
Pachystigma	P	*Cuscuta africana*	KJ436602	KJ400037		DQ924574
Pachystigma	P	*Cuscuta angulata*	KJ436602	KJ400039	EF152065	DQ924575
Pachystigma	P	*Cuscuta nitida*	KJ436690	KJ400137	EF202558	EF202562
		OUTGROUPS				
Convolvulaceae		*Calystegia hederacea Wall.*	LC085875			LC225748
Convolvulaceae		*Convolvulus althaeoides L.*	KC529175		KC786129	KC528819
Convolvulaceae		*Convolvulus arvensis L.*	MH536594	AF479176	AY101102	AY560274
Convolvulaceae		*Convolvulus farinosus L.*	KC529210			KC528841
Convolvulaceae		*Convolvulus glomeratus Choisy*	KC529219		MF621882	KC528887
Convolvulaceae		*Convolvulus lineatus L.*	KC529239		KC786132	KC528963
Convolvulaceae		*Cressa cretica L.*	KY860700			KJ004289
Convolvulaceae		*Dichondra repens J. R. Forst. & G. Forst.*	MN379753			MN380241
Convolvulaceae		*Dinetus racemosus (Roxb.) B. -Ham. ex Sweet*	HQ384920		HQ412975	MG730315
Convolvulaceae		*Evolvulus alsinoides (L.) L.*	MH767511			MH768118
Convolvulaceae		*Jacquemontia tamnifolia (L.) Griseb.*	AY101037	AF148499	AY101146	DQ219865
Convolvulaceae		*Montinia caryophyllacea Thunb.*	L11194	KJ400199	AY206764	
Convolvulaceae		*Neuropeltis acuminata (P. Beauv.) Benth.*	AY101033		AY101142	
Solanaceae Juss.		*Porana paniculata Roxb.*	AY101051		AY101160	
Montiniaceae Nikai		*Schizanthus pinnatus Ruiz & Pav.*	U08619	KJ400202	AY101172	DQ299453

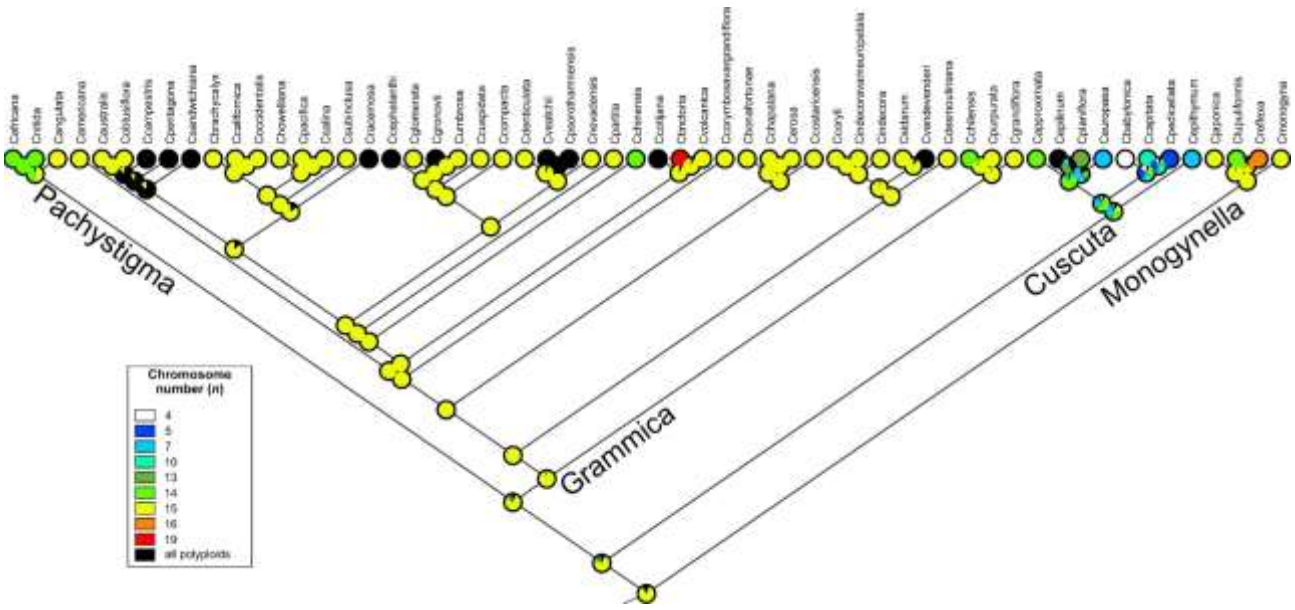
Supplementary figure 1: Concatenated tree made by Bayesian inference from 226 sequences of nuclear (nrITS and 26S) and plastid markers (*rbc*L and *trn*L-*trn*F) gathered through GenBank database



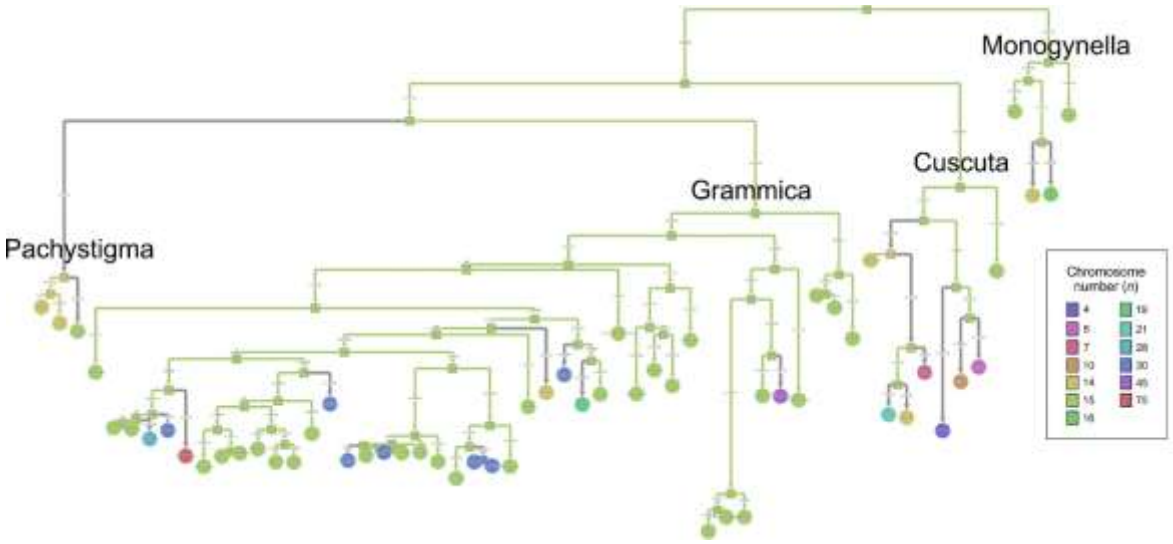
Supplementary figure 2: Reconstruction of the chromosome number evolution in *Cuscuta* with the BASE_NUM_DUPL model. The pie charts on the nodes represent the probability of each inferred chromosome number, the numbers along the branches represent the inferred number of the four different types of events (gains, losses, duplications and demiduplications) The bold letters represent the sections described by Garcia et al. (2014), section **S** (subgenus *Monogynella*), sections **R** and **Q** (subgenus *Cuscuta*), section **P** (subgenus *Pachystigma*) and sections **A** - **O** (subgenus *Grammica*)



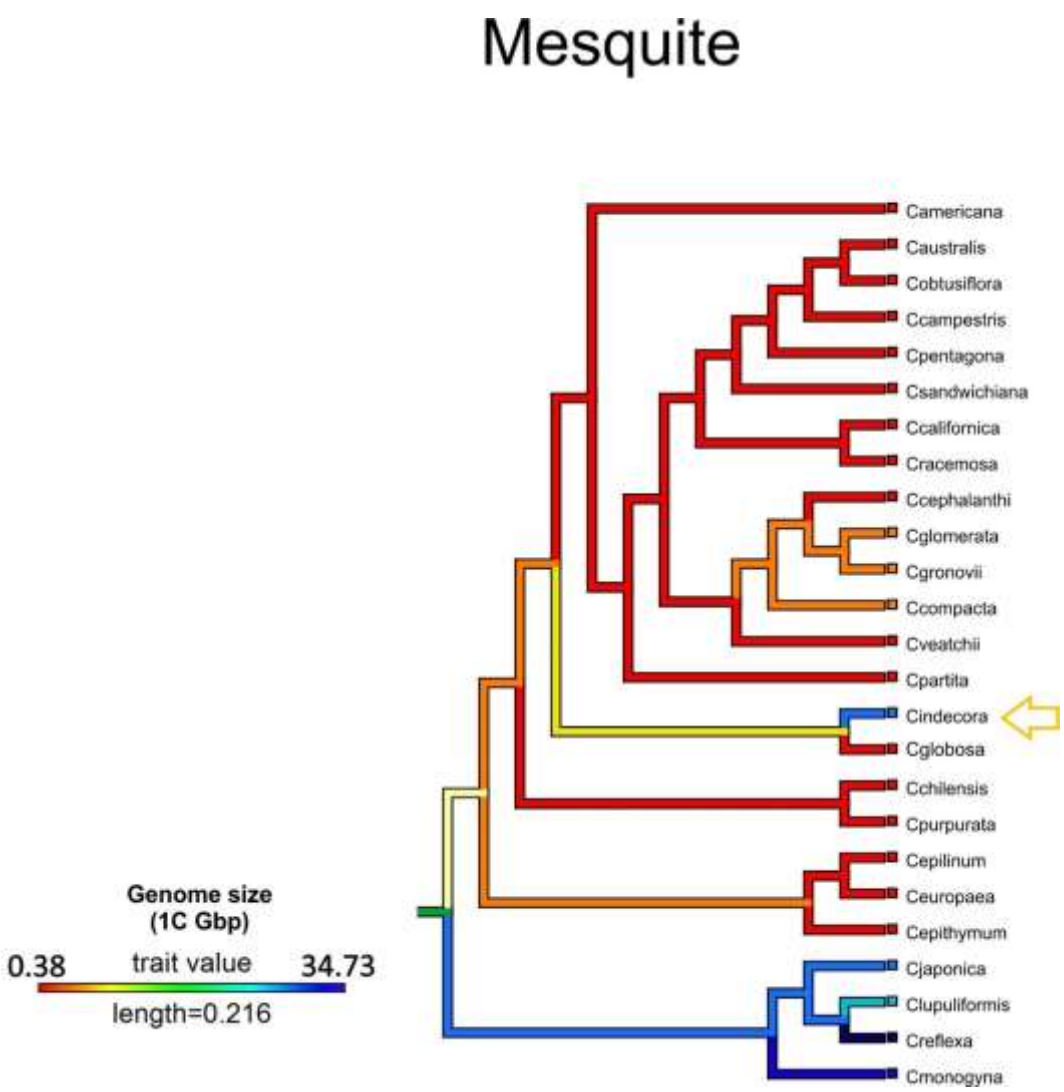
Supplementary figure 3: Reconstruction of the chromosome number evolution in *Cuscuta* with the BASE_NUM_DUPL model, without the subgenus *Cuscuta*. The pie charts on the nodes represent the probability of each inferred chromosome number, the numbers along the branches represent the inferred number of the four different types of events (gains, losses, duplications and de-duplications) The bold letters represent the sections described by Garcia et al. (2014), section **S** (subgenus *Monogynella*), sections **R** and **Q** (subgenus *Cuscuta*), section **P** (subgenus *Pachystigma*) and sections **A** - **O** (subgenus *Grammica*).



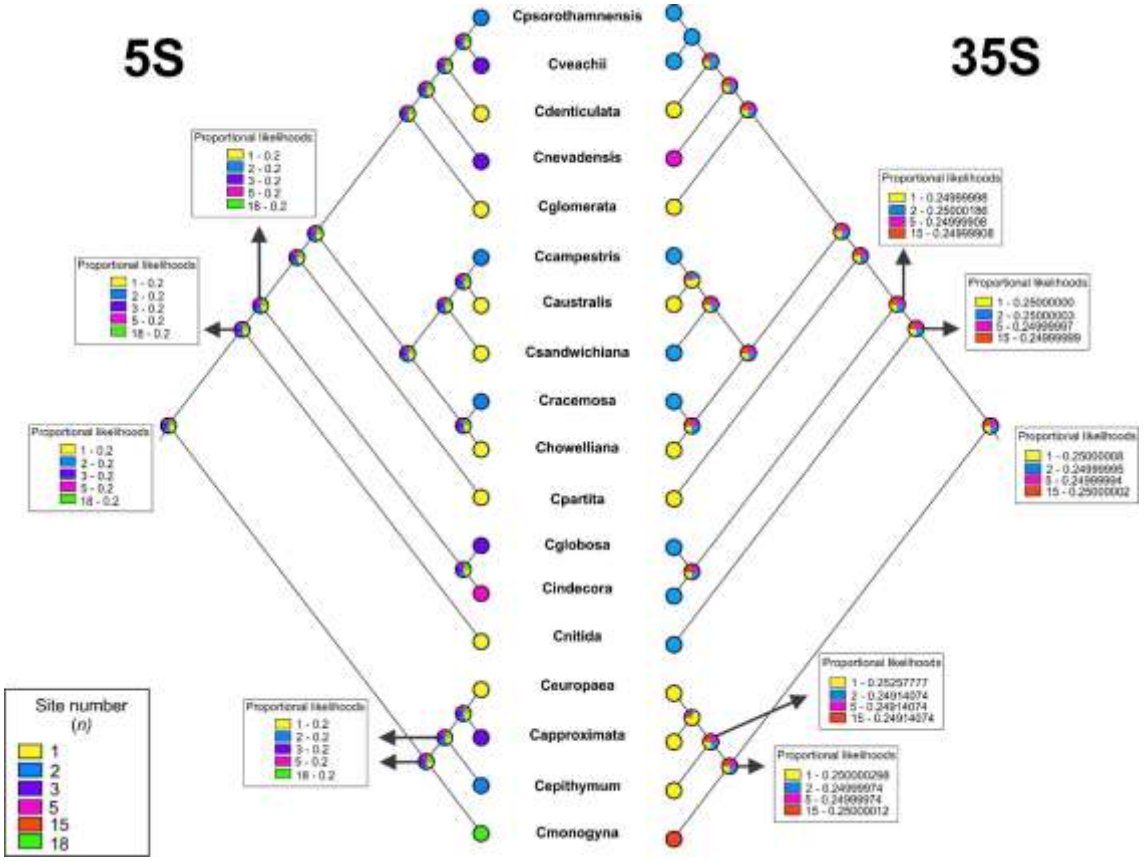
Supplementary figure 4: Reconstruction of ancestral chromosome number in *Cuscuta* performed in Mesquite. The reconstruction model used was of maximum likelihood. Each subgenus is indicated in the tree.



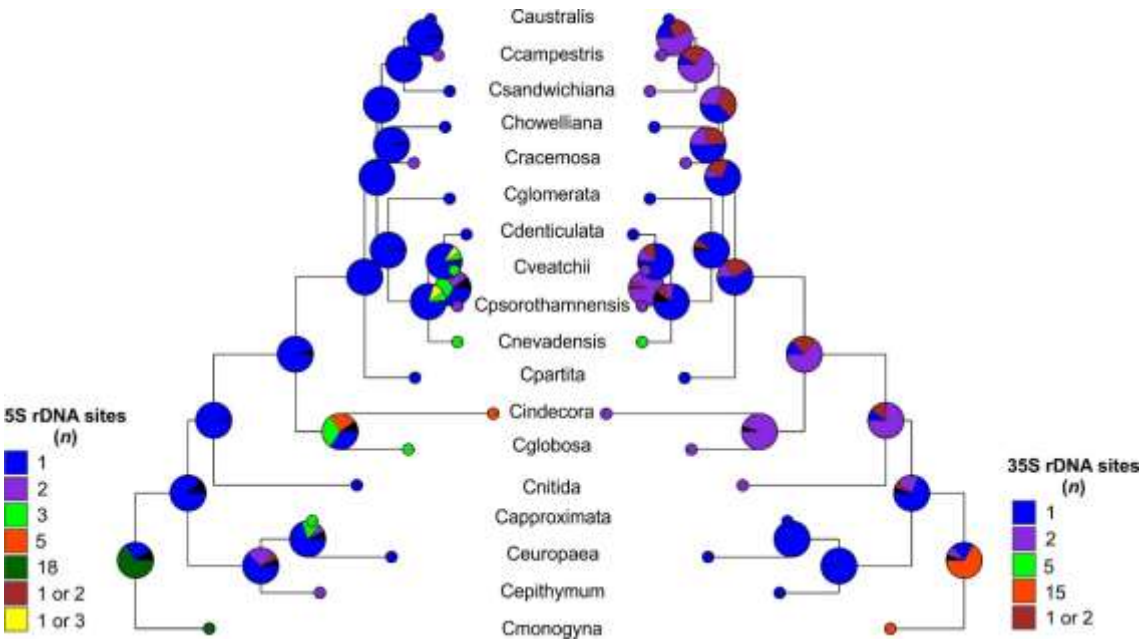
Supplementary figure 5: Reconstruction of ancestral chromosome number with PastML using the JOINT model. The number $n = 15$ was inferred for the entire genus *Cuscuta* and for each of its four subgenera. Each subgenus is indicated in the tree



Supplementary figure 6: Reconstruction of the genome size made in Mesquite. The variation is shown in a color scale. The largest genomes are showing in shades of blue and in shades of red and yellow are the smaller genomes. Yellow arrow in reconstruction showing *C. indecora*, a species of the subgenus *Grammica* where there was an expansion of the genome.



Supplementary figure 7: Reconstruction of the rDNA sites number 5S (left) and 35S (right). Site numbers refer to the haploid number, which can be one pair, two, three, five, 15 or 18 pairs. Mesquite could not resolve the ancestral state of the sites number, placing the same probability for all possible rDNA 5S number of pairs and very close probabilities with each other for rDNA 35S number of pairs. The most likely number of 35S rDNA sites is one site pair.



Supplementary figure 8: Reconstruction of the rDNA sites number 5S (left) and 35S (right) using RASP. It is possible to see the influence of the most ancestral species, *C monogyna* (subgenus *Monogynella*), in the analysis. The numbers of ancestral sites are tending to be the same number reported in this species, 18 pairs of rDNA5S and 15 pairs of rDNA35S.