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

YENNIFER CAROLINA MATA SUCRE

ORIGEM E EVOLUÇÃO DA HOLOCENTRICIDADE NA FAMILIA JUNCACEAE:

uma abordagem filogenética, cito-molecular e genômica

RECIFE
2024

YENNIFER CAROLINA MATA SUCRE

ORIGEM E EVOLUÇÃO DA HOLOCENTRICIDADE NA FAMILIA JUNCACEAE:

uma abordagem filogenética, cito-molecular e genômica

Tese apresentada ao Programa de Pós-Graduação em Biologia vegetal da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Doutora em Biologia Vegetal. Área de concentração: Sistemática e Evolução

Orientador: Dr. Luiz Gustavo Rodrigues Souza

Coorientadora: Dra. Andrea Pedrosa-Harand

Recife

2024

Catalogação na Fonte
Bibliotecário: Marcos Antonio Soares da Silva
CRB4/1381

Sucre, Yennifer Carolina Mata.
Origem e evolução da holocentricidade na família Juncaceae: uma abordagem filogenética, citomolecular e genômica. / Yennifer Carolina Mata Sucre . – 2024.

151 f. : il., fig.; tab.

Orientador: Luiz Gustavo Rodrigues Souza.
Coorientadoras: Andrea Pedrosa-Harand.

Tese (doutorado) – Programa de Pós-Graduação em Biologia Vegetal da Universidade Federal de Pernambuco, 2024.

Inclui referências, apêndice e anexos.

1. Citogenética. 2. Citotaxonomia. 3. Sequencias repetitivas. 4. Juncaceae. 5. Centromero. 6. Holocentricos. 7. Monocentricos. I. Souza, Luiz Gustavo Rodrigues (Orient.). II. Pedrosa-Harand, Andrea (Coorient.). IV. Título.

580 CDD (22.ed.) UFPE/CB – 2024-044

YENNIFER CAROLINA MATA SUCRE

ORIGEM E EVOLUÇÃO DA HOLOCENTRICIDADE NA FAMILIA JUNCACEAE:

uma abordagem filogenética, cito-molecular e genômica

Tese apresentada ao Programa de Pós-Graduação em Biologia vegetal da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Doutora em Biologia Vegetal. Área de concentração: Sistemática e Evolução

Aprovada em: 02/02/24

BANCA EXAMINADORA:

Prof. Dr. Luiz Gustavo Rodrigues Souza (Orientador)

Universidade Federal de Pernambuco

Prof. Drª. Thaís Elias Almeida (Membro interno)

Universidade Federal de Pernambuco

Dr. Lucas Alexandre De Souza Costa (Membro externo)

Universidade Federal de Pernambuco

Prof. Dr. Reginaldo de Carvalho (Membro externo)

Universidade Federal Rural de Pernambuco

Prof. Dr. Cícero Carlos de Souza Almeida (Membro externo)

Universidade Federal de Alagoas

Recife

2024

Dedico esta tesis a quienes nos aventuramos incluso con miedo.

AGRADECIMENTOS

Gostaria de agradecer (em portunhol) a todas essas pessoas que me apoiaram tanto fisica como espiritualmente nesta jornada. E especialmente a:

Minha família, que, embora eles não saibam nem entendam nem um pouco o que eu faço na vida, eles me apoiam em todas as metas e loucuras que eu me proponho. Eternamente agradecida, sem vocês essa jornada teria outro sentido.

Meu orientador Luiz Gustavo, que com sua visão 10/10 vê todo meu potencial, e com seu otimismo semeia o desejo de ir sempre por mais. Obrigada professor, desta vez o Nobel vem. E a minha co-orientadora, chefia e mulher guerreira Andrea Pedrosa-Harand. Embora não tenha aproveitado muito bem o que você podia me oferecer, eu agradeço por cada discussão, ajuda e conselho dado.

Às minhas queridas CitoGatas e C.A., Amália, Mariela, Mariana e meu gatinho Paulo. Obrigada por serem esses exemplos de superação, realização e sucesso. Vocês reforçam todos os dias a ideia de que cercar-se de sucesso leva ao sucesso.

Ao meu povo fanfikeiro Thiago, Amanda Santos e Jessica, agradeço por todos esses momentos compartilhados, os desabafos e as idas na rainha da várzea. Obrigada por manterem minha barriga e meu coração cheios.

A minha ex-pupila Natália, quem ia saber que você não abandonaría o laboratório na primeira semana, hein. Agradeço por terem ficado, me ensinado tantas coisas do mundo dos jovens e me apoiado na doidice dessa próxima etapa. Obrigada mulher, desejo a você muito sucesso na vida. Mentes brilhantes como a sua chegam muito longe.

As pessoas da velha e nova guarda do Citovegetal que me apoiaram nessa jornada: Acalene, Ana, Arthur, Brena, Bruna, Cláudio, Géssica, Gustavinho, Gustavo “alumno”, Julia, Kaiane, Leylson, Lucas, Yhandra e William, obrigada Brasil.

Ao meu grupo alemão no Max Planck. Meu co-orientador André Marques, obrigada pelas oportunidades e conhecimentos brindados. Ao meus colegas de trabalho Gokilavani, Stefan, Meng e Marco que fizeram meus dias na Alemanha mais gratos. Meu grupinho latino obviamente, Rigel y Mijael, gracias eternas por nunca deixar que a festa e a luz se apagasse. "Lo tenemos".

As minhas brasileiras europeias, Leticia e Ludmila, mulheres eu vi vocês umas poucas vezes na vida mas eu já sinto que conhecia vocês de muito tempo atrás, obrigada pelo apoio e pelos conselhos mutuos nos perrengues da vida. Nos veremos novamente.

Agradeço ao Programa de Pós-Graduação em Biologia Vegetal, em especial ao pessoal administrativo, pela ajuda, paciência e constantes esclarecimentos de dúvidas durante a minha permanência no programa.

Agradeço também a todas as pessoas que não mencionei mas que passaram e deixaram marcas na minha vida nesses últimos quatro anos, distantes mas não ausentes.

A todos. Muito obrigada!

RESUMO

Juncaceae é uma família composta por 474 espécies com uma taxonomia complexa e baixa resolução filogenética. Historicamente, era considerada uma família composta exclusivamente por espécies com cromossomos holocêntricos e alta variabilidade cariotípica, mas estudos recentes mostraram que um de seus gêneros mais diversos, *Juncus* (332 spp), apresenta espécies com cromossomos monocêntricos. Essa descoberta desafia o entendimento da origem da holocentricidade na família e cria a necessidade de reinterpretar sua evolução do ponto de vista citogenético, filogenético e genômico. Assim, no Capítulo 1 investigamos a diversidade cariotípica e cito-molecular no gênero *Juncus*, com análises originais de 12 espécies e compilação dos dados disponíveis na literatura. Essa análise realizada num contexto filogenético revelou um número cromossômico ancestral x = 10 a 20 e sugeriu que o gênero seja cariotipicamente mais estável do que reportado na literatura. Os eventos de disploidia e poliploidia foram identificados como fatores significativos na evolução dos cromossomos. Citogeneticamente, duas a quatro bandas grandes de CMA^+ foram observadas na região terminal das espécies, correspondendo a sítios de DNA ribossômico nuclear (DNAr) 35S. No entanto, pequenos sítios CMA^+ pericentroméricos adicionais foram observados em espécies pertencentes a clados diferentes (*J. capitatus*, *J. subsecundus* e *J. compressus*). Além disso, a análise filogenética revelou relações complexas na família Juncaceae e não-monofiletismo do gênero *Juncus* (por exemplo, *J. capitatus* tem relações próximas com os gêneros *Oreojuncus* e *Luzula*) e de suas secções taxonômicamente reconhecidas. No Capítulo 2, foram explorados os repeatomas de representantes de *Juncus*, com foco no significado filogenético dos elementos repetitivos e nos repeats potencialmente associados aos monocentrômeros. Integramos dados de genoma de baixa cobertura de 33 táxons para reavaliar a sistemática do gênero utilizando diferentes abordagens filogenéticas repeat-based. Identificamos tendências clado-específicas na presença/ abundância de repetições para as diferentes linhagens de *Juncus* e a análise filogenética baseada em repeats foi congruente com a topologia do DNAr. Por outro lado, a topologia dos plastomas se apresentou incongruente a esses, sugerindo hibridização antiga na história do gênero. Nesse trabalho, ampliamos a caracterização da monocêntricidade para mais espécies de *Juncus* de diferentes clados, sugerindo que isso seja uma possível sinapomorfia do gênero. Esse estudo, publicado na revista Molecular Phylogenetics, enfatiza o poder da análise dos repeatomas para compreender a sistemática e a evolução das plantas. E, finalmente, no Capítulo 3, foi investifgada a evolução dos repeatomas no clado holocêntrico *Luzula*, irmão de *Juncus*. Foi realizado o sequenciamento do genoma completo e montagem no nível do

cromossomo de *L. sylvatica* visando entender especificamente a organização genômica do seu holocentrômero. Foi identificado um DNA satélite específico, *Lusy1*, predominantemente associada a domínios centroméricos na maioria dos clados de *Luzula*. As análises comparativas sugerem que a redução do número de cromossomos em *Luzula* foi causada por várias fusões de cromossomos inteiros. Sugerimos que a manutenção de repeats centroméricos após fusões cromossômicas pode estar relacionada com a origem da holocentricidade em *Luzula*. Esse estudo propõe uma evolução dinâmica do holocentrômero baseada em repetições, fornecendo evidências para a transição de cromossomos monocêntricos para holocêntricos em plantas.

Palavras-chave: Citogenética, citotaxonomia, filogenomica, sequencias repetitivas, Juncaceae, centromero, holocentricos, monocentricos

ABSTRACT

Juncaceae is a family of 474 species with a complex taxonomy and low phylogenetic resolution. Historically, it was considered a family composed exclusively of species with holocentric chromosomes and high karyotypic variability, but recent studies have shown that one of its most diverse genera, *Juncus* (332 spp), has species with monocentric chromosomes. This discovery challenges the understanding of the origin of holocentricity in the family and creates the need readdress its evolution from a cytogenetic, phylogenetic and genomic point of view. Thus, in Chapter 1 we investigate the karyotypic and cyto-molecular diversity of the genus *Juncus*, with original analyses of 12 species and a compilation of the data available in the literature. This analysis carried out in a phylogenetic context revealed an ancestral chromosome number $x = 10 / 20$ and suggested that the genus is karyotypically more stable than reported in the literature. Dysploidy and polyploidy events were identified as significant factors in chromosome evolution. Cytogenetically, two to four large CMA+ bands were observed in the terminal region of the species, corresponding to 35S nuclear ribosomal DNA (rDNA) sites. However, additional small pericentromeric CMA+ sites were observed in species belonging to different clades (*J. capitatus, J. subsecundus* and *J. compressus*). In addition, the phylogenetic analysis revealed complex relationships within the Juncaceae family and non-monophyly of the genus *Juncus* (for example, *J. capitatus* species has closer relationships with the genera *Oreojuncus* and *Luzula*) and its taxonomically recognized sections. In Chapter 2, we explored the repeatomes of representatives of *Juncus* genus, focusing on the phylogenetic significance of repetitive elements and repeats potentially associated with monocentromeres. We integrated low-coverage genome data from 33 taxa to reassess the systematics of the genus using different repeat-based phylogenetic approaches. We identified clado-specific trends in the presence/abundance of repeats for the different *Juncus* lineages and the repeat-based phylogenetic analysis was congruent with the DNAr topology. On the other hand, the topology of the plastomes was incongruent with these, suggesting ancient hybridization in the history of the genus. In this work, we extended the characterization of monocentricity to six more *Juncus* species from different clades, suggesting that this is a possible synapomorphy of the genus. This study, published in the Molecular Phylogenetics and Evolution journal, emphasizes the power of repeatome analysis to understand plant systematics and evolution. Finally, in Chapter 3 we investigate the evolution of repeatome in the holocentric clade *Luzula*, sister to *Juncus*. We performed whole genome sequencing and assembly at the chromosome level of *L. sylvatica* in order to specifically understand the genomic organization of its holocentromere. We identified

a specific satellite DNA, *Lusy1*, predominantly associated with centromeric domains in most *Luzula* clades. Comparative analyses suggest that the reduction in chromosome number in *Luzula* was caused by several whole chromosome fusions. We suggest that the maintenance of centromeric repeats after chromosome fusions may be related to the origin of holocentricity in *Luzula*. This study proposes a dynamic evolution of the holocentromere based on repeats, providing evidence for the transition from monocentric to holocentric chromosomes in plants.

Keywords: Cytogenetics, cytotaxonomy, phylogenomics, repetitive sequences, Juncaceae, centromere, holocentric, monocentric

SUMÁRIO

1 INTRODUÇÃO

A família Juncaceae inclui oito gêneros e 474 espécies (POWO, 2023). Entre esses gêneros, *Juncus* L. e *Luzula* DC. se destacam como os mais diversos, com 332 e 124 espécies, respectivamente (Drábková et al. 2006). O gênero *Juncus* é dividido em dois subgêneros, *Juncus* e *Agathryon* (Raf.) Záv. Drábk. & Proćków, com seis e quatro seções taxonômicas, respectivamente, propostas com base em características morfológicas. Entretanto, as relações filogenéticas dessas secções e mesmo de gêneros afins (*Oreojuncus* Záv. Drábk. & Kirschner*, Marsippospermum* Desv.*, Rotskovia* Desv.*,* e *Patosia* Buchenau) foram estudadas desde 2005, porém com pouca resolução até agora (Drábková et al. 2006, 2009; Roalson 2005; Brožová et al. 2022).

Citogeneticamente, os organismos podem ser classificados de acordo com o tipo de centrômeros em dois tipos, monocêntricos e holocêntricos. No primeiro caso, os centrômeros estão localizados em locais específicos, enquanto no segundo os centrômeros estão distribuídos ao longo de todo o cromossomo. Nas angiospermas, essa característica aparece aparentemente de forma independente, em famílias como Cyperaceae, Melanthaceae e Juncaceae, levantando questões sobre como e porque essa transição ocorre. Assim, tradicionalmente se considerava que todos os representantes de Juncaceae possuíam cromossomos holocêntricos, associados a uma elevada diversidade cariotípica. Entretanto, estudos recentes revelaram uma interessante transição de tipo centromérico, com cromossomos monocêntricos em espécies do gênero *Juncus* (Guerra et al. 2019; Mata-Sucre et al. 2023; Dias et al. in review).

A investigação de sequências repetitivas de DNA associadas aos centrômeros e a identificação de proteínas centroméricas são etapas cruciais para a compreensão dessa transição de cromossomos monocêntricos para holocêntricos. As sequências repetitivas centrômero-específicas, comuns nos centrômeros monocêntricos, geralmente não são encontradas nos holocêntricos. Entretanto, recentes análises genômicas têm revelado a presença de DNAs satélites centroméricos nos gêneros *Rhynchospora* Vahl (Marques et al. 2015) e *Chionographis* Maxim. (Kuo et al. 2022) levantando questionamentos sobre a participação dessas repetições na origem da holocentricidade.

Os estudos recentes que caracterizaram a monocentricidade em espécies de *Juncus* (Guerra et al. 2019; Hofstatter et al. 2022) abriram uma nova oportunidade para entender como ocorre a transição evolutiva entre esses dois tipos de centrômero nas plantas. Com a oportunidade de investigar o papel das sequências repetitivas e das proteínas centroméricas nesse processo, propomos utilizar a família Juncaceae como modelo para investigar a transição

de diferentes tipos centroméricos. O objetivo deste trabalho foi compreender a evolução cromossômica e os mecanismos de transição de cromossomos monocêntricos para holocêntricos na família Juncaceae. Para isso, os dados do presente trabalho foram organizados em três capítulos que exploram: (i) a diversidade cariotípica do gênero monocêntrico *Juncus* por meio de uma visão citogenética e filogenética; (ii) a filogenômica baseada na fração repetitiva dos genomas para o esclarecimento das relações evolutivas no gênero *Juncus* e, (iii) o estudo dos holocentrômeros baseados em repeats de *Luzula sylvatica* (Huds.) Gaudin (Juncaceae Juss.) para ajudar a esclarecer o processo de transição evolutiva do monocentrismo para o holocentrismo.

2 FUNDAMENTAÇÃO TEÓRICA

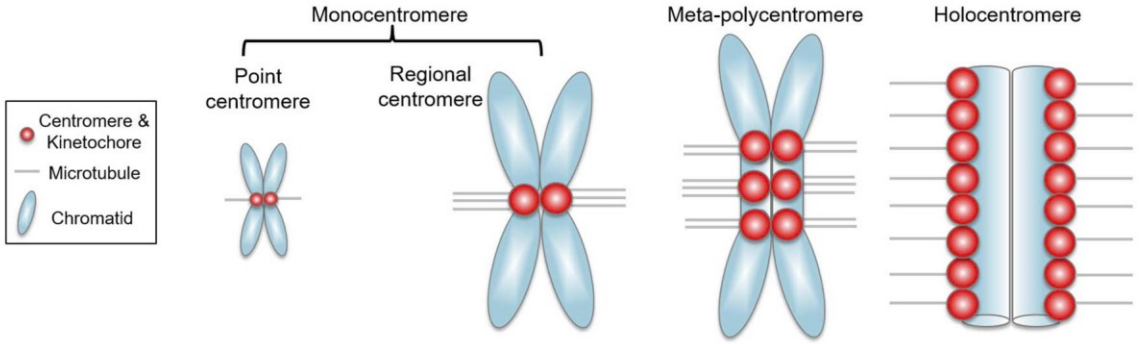
2.1 O CENTRÔMERO E O CINETÓCORO: UMA VISÃO GERAL

O centrômero dos eucariotos é identificado como uma constrição primária na qual cada cromossomo possui uma única conexão de fibra durante a segregação (Flemming 1882). Análises moleculares e citogenéticas sucessivas o definiram, até o momento, como domínios cromossômicos heterocromáticos que direcionam a formação do cinetócoro (Sharp 1934). O cinetócoro é uma estrutura proteica montada no DNA centromérico que o liga aos microtúbulos e conduz a divisão celular (Cuacos et al. 2015; Ramakrishnan et al. 2023). O complexo do cinetócoro é composto por várias proteínas subdivididas em três categorias principais: (i) proteínas que regulam a segregação dos cromossomos, (ii) proteínas do cinetócoro externo associadas aos microtúbulos do fuso e (iii) proteínas do cinetócoro interno associadas ao DNA centromérico (Ishii e Akiyoshi, 2022).

Na maioria dos eucariotos, a identidade do centrômero é determinada pela presença da variante centromérica da histona H3, chamada CENH3 (também chamada CENP-A em humanos, Malik e Henikoff, 2009; Westhorpe e Straight, 2015) e por sequências de DNA centroméricas. A carga da CENH3 depende principalmente das chaperonas CENH3 e dos fatores de licenciamento do centrômero KNL2, M18BP1 e CENP-C (Ramakrishnan et al. 2023). Diferentes funções foram associadas aos centrômeros além da montagem do cinetócoro durante o ciclo celular, incluindo marcação genética/epigenética, ponto de verificação mitótico, coesão e liberação de cromátides irmãs, movimento cromossômico e citocinese (Santaguida e Musacchio, 2009; Perpelescu e Fukagawa, 2011).

Os centrômeros são altamente diversos, variando drasticamente em tamanho e composição entre as espécies, sendo uma característica comum de suas sequências a repetitividade (Schubert et al. 2020). Centrômeros pontuais são encontrados principalmente em leveduras unicelulares de *Saccharomyces cerevisiae* com ~125 bp (Wong et al. 2020; Fig. 1) e centrômeros regionais em *C. albicans* (4-18 kb), levedura de fissão (35-110 kb), organismos multicelulares como milho (300-3000 kb), moscas (420 kb) e humanos (500-1500 kb; Wong et al. 2020).

Figura 1 – Características epigenéticas e genéticas de diferentes organizações de centrômeros. Quatro tipos de arquitetura de centrômero são mostrados esquematicamente. Exemplos de espécies em cada arquitetura de centrômero e seus tamanhos de centrômero.



Example species / size of centromere	*S. cerevisiae*/ 125 bp *C. glabrata*/ 153 bp	*C. albicans*/ 4-18 kb *S. pombe*/ 35-110 kb *Z. mays*/ 0.3-3 Mb *D. melanogaster*/ 420 kb *H. sapiens*/ 0.5-5 Mb	*P. sativum*/ 69-107 Mb	*C. elegans*/ 14-21 Mb *B. mori*/ 8.4-21.5 Mb *L. nivea*/ ~155 Mb **(Size of chromosome)**
CENP-A / CenH3 homologue(s)?	✓ (*S. cerevisiae*)	✓; Exception: ✗ in kinetoplastids *T. brucei*	✓ 3–5 explicit CenH3-containing regions	✓ e.g. *C. elegans*; Exception: ✗ in *Lepidoptera*
Centromere transcription?	✓ (*S. cerevisiae*)	✓ e.g. *H. sapiens* α-satellites	N/A	✗ e.g. *C. elegans*
Pericentric heterochromatin?	✗ (*S. cerevisiae*, but has flanking H2A.Z/Htz1 pericentric regions)	✓ e.g. *S. pombe* outer repeats H3K9me & HP1	N/A	✗ e.g. *C. elegans* and *B. mori* as no pericentric region
Pericentric transcription?	✓ (*S. cerevisiae*, transcription starts in pericentric regions and through centromeric regions)	✓ e.g. *S. pombe* outer repeats *dg* regions	N/A	✗ e.g. *C. elegans* and *B. mori* as no pericentric region
AT-richness?	✓ > 90% (*S. cerevisiae* CDEII)	✓ e.g. ~63% (171-bp monomer of the *H. sapiens* α-satellites)	✓ 67-77%	✓ e.g. *C. elegans* which AT% correlates with CENP-A domain size
DNA sequence repetitiveness?	✗ (*S. cerevisiae*)	✓ e.g. *Z. mays* retrotransposon CRM	✓ satellite DNA and centromeric retrotransposons	✗ e.g. *C. elegans* and *L. elegans*; ✓ e.g. in *Rhynchospora*

Fonte: Won et al. (2020, p. 2).

2.2 O CENTRÔMERO E SUA FRAÇÃO REPETITIVA

Os DNAs presentes no centrômero são uma das sequências de mais rápida evolução nos genomas eucarióticos (Melters et al. 2013; Talbert e Henikoff 2020). Apesar da grande divergência de sequências, uma característica recorrente dos centrômeros regionais em plantas e animais é a presença de elementos transponíveis (ex. transposons, retrotransposons, etc.) e sequencias repetidas em tandem de DNA satélite que se estendem de várias quilobases a megabases e que podem proporcionar condições favoráveis para a formação e/ou função do centrômero (Talbert e Henikoff, 2020).

A maioria das sequencias dos centrômeros de plantas e animais podem diferir mesmo entre espécies irmãs (Melters et al. 2013; Talbert et al. 2018). Por exemplo, os centrômeros de *Oryza sativa* L. possuem repetições centroméricas de DNA satélite de 155 pb que não compartilham nenhuma homologia com a repetição de 154 pb encontrada em *Oryza brachyantha* A.Chev. & Roehr., uma espécie de arroz selvagem próxima (Lee et al. 2005). No entanto, um aspecto que pode ser considerado comum à maioria das espécies estudadas é a alta similaridade de sequências repetitivas de DNA entre os centrômeros do mesmo complemento cromossômico, como em *Arabidopsis thaliana* (L.) Heynh. (Nagaki et al. 2003) ou *O. brachyantha* (Lee et al. 2005). Mas, em *Solanum tuberosum* L. (batata cultivada), por exemplo, cada um dos 12 centrômeros tem uma composição genética diferente das demais: seis deles são compostos por DNA repetitivo, cinco por DNA de cópia única e o último centrômero apresenta os dois tipos de DNA, indicando a rápida evolução de sequências centroméricas (Gong et al. 2012),.

Os centrômeros, além de serem ricos em DNA repetitivo, também podem surgir *de novo* em regiões altamente repetitivas. Esses outros centrômeros foram observados em pacientes humanos (Naughton e Nick Gilbert, 2020) e durante a especiação em plantas e mamíferos (“ENC do inglês evolutionary new centromeres”, Wang et al. 2014; Nergadze et al. 2016; Xue et al. 2022). Os ENC apresentam segregação, frequente correção de erros e normalmente acumulam satélites centroméricos após seu início durante a evolução (Naughton e Nick Gilbert, 2020). Dados sugerem que certas características intrínsecas do DNA centromérico podem facilitar a estabilidade ou a função do centrômero. No entanto, o conhecimento de quais seriam essas características vantajosas permanece elusivo, em parte devido às dificuldades associadas à análise funcional e filogenética de regiões tão grandes, complexas e inacessíveis.

2.3 CARACTERIZAÇÃO DO CENTRÔMERO NO NÍVEL GENÔMICO

As regiões centroméricas têm sido extremamente difíceis de estudar devido ao desafio de montar corretamente longas matrizes de repetições em tandem quase idênticas. Como resultado, a maioria das sequencias de DNA satélites e elementos transponiveis associadas ao centrômero permanecem em grande parte não montados na maioria dos genomas sequenciados (Melter et al. 2020). A montagem linear de repetições em tandem extremamente longas depende de uma cobertura de leitura abrangente para uma única matriz, detecção confiável de variantes

e comprimentos de leitura adequados para garantir que a distância máxima entre duas variantes informativas dentro da matriz possa ser coberta (Miga, 2020).

Os dados de alta qualidade ou alta fidelidade (HiFi) gerados a partir do sequenciamento de consenso circular são uma estratégia promissora para caracterizar repetições e montar sequências centroméricas com *reads* que abrangem centenas de kilobases (Wenger et al. 2019). Foi demonstrado que a combinação de *reads* longas e *reads* HiFi solucionaram os problemas de sequenciamento das regiões do centrômero e do telômero no genoma humano, gerando uma montagem completa do telômero ao telômero (T2T) do cromossomo humano (Chr) X e Chr8 (Miga et al. 2020; Logsdon et al. 2021).

A matriz de satélite centromérico T2T-X forneceu uma montagem de centrômero de referência validada estruturalmente e de alta qualidade. A descoberta mais surpreendente dessa montagem do centrômero foi a alta frequência de heterogeneidade de repetição com variantes menores e de nucleotídeo único, servindo como marcadores exclusivos dentro da região canônica de alta organização ("HOR do inglês high order repeats"; Altemose et al. 2022; Nurk et al. 2022). A matriz de satélites centroméricos humana é definida por 1537 cópias dentro do HOR de ~2 kbp (DXZ1) que são ordenadas em uma matriz em tandem cabeça-cauda em uma única direção (Brouha et al. 2003). A DXZ1 é interrompida uma vez por um único elemento transponível L1HS de comprimento total de 6055 bp, que representa uma subfamília de Elementos Nucleares Longos Intercalados (LINEs) conhecida por saltar de forma autônoma em humanos (Brouha et al. 2003).

Nas plantas, o número de cópias de repetição de satélite centromérico e as variantes de sequência em cada matriz variam significativamente mais do que em humanos. Em duas variedades de arroz (*O. sativa*), Zhenshan 97 e Minghui 63, as regiões centroméricas compartilham motifs de DNA satélites específicos do centrômero conservados com diferentes números de cópias e estruturas (Son et al. 2021). Além disso, a semelhança das repetições do centrômero no mesmo cromossomo é maior do que entre os cromossomos, o que apoia um modelo de expansão e homogeneização local. Os autores oberservaram que ambos os genomas têm mais de 395 genes não TE localizados nas regiões do centrômero, 41% dos quais são ativamente transcritos (Son et al. 2021).

Em *Arabidopsis*, a montagem do centrômero da cultivar "Col" revelou matrizes de satélites repetidos em tandem específicas de cada cromossomo em uma escala de megabases, ocupando a CENH3 e que são densamente metilados no DNA (Naish et al. 2021). Esses

centrômeros são invadidos por retrotransposons da linhagem Athila, que interrompem a organização genética e epigenética. Os autores sugerem que os retrotransposons nessas regiões são mantidos por um processo de homogeneização baseado em recombinação, que mantém a biblioteca de DNA satélite próxima ao consenso ideal para o recrutamento de mais CENH3 (Naish et al. 2021).

Nos centrômeros do genoma da soja (*Glycine max* (L.) Merr.), cinco satélites de centrômero se associam especificamente ao cromossomo 1 (Liu et al. 2023). Esses satélites revelam rearranjos significativos nas estruturas do centrômero do cromossomo em diferentes acessos, afetando, consequentemente, a localização do CENH3. A maioria dos novos centrômeros emergentes se formaram nas proximidades dos centrômeros nativos e alguns centrômeros emergentes foram aparentemente compartilhados em acessos distantemente relacionados, sugerindo que seu surgimento é independente (Liu et al. 2023). Os neocentrômeros se localizam preferencialmente nas regiões ricas em satélites para manter um estado estável, destacando uma função significativa dos satélites na organização dos centrômeros. Em conjunto, esses resultados destacaram a importância dos satélites na restrição das posições do centrômero e no auxílio à função do centrômero.

2.4 DIVERSIDADE CENTROMÉRICA E SUAS VANTAGENS

Apesar da importância funcional dos centrômeros, os citologistas perceberam que uma constrição primária (monocentrômero) não podia ser observada nos cromossomos de todas as espécies e que, em alguns táxons, os microtúbulos do fuso podiam se ligar ao longo de todo o comprimento do cromossomo (White 1977). Seis anos após a descrição de Flemming (1882) da constrição primária, Boveri (1888) descreveu o que hoje é conhecido como cromossomos holocêntricos em *Ascaris megacephala*, cromossomos que não têm uma constrição primária e, em vez disso, têm fibras fusiformes ao longo de seu comprimento, o que poderia ser considerado, retrospectivamente, um indicador da surpreendente diversidade de estruturas centroméricas (Talbert and Henikoff 2020).

Embora seja chamado de holocentrômero, o DNA centromérico não é encontrado em todas as sequências genômicas. Por exemplo, a marca epigenética centromérica, CENP-A, é encontrada em 50% do genoma do *Caenorhabditis elegans* (Gassmann et al. 2022). Citogeneticamente, o holocentrômero aparece nas faces posteriores dos cromossomos mitóticos, nos quais a maior parte do DNA cromossômico não se liga às proteínas centroméricas

(Maddox et al. 2004). Igualmente, nas *Cuscuta* sp. (Convolvulaceae), duas variantes principais de CENH3 são depositadas em uma a três regiões discretas por cromossomo, contendo repetições de DNA satélite, enquanto o restante da cromatina parece ser desprovido de CENH3 (Oliveira et al. 2020). Esse padrão de distribuição da CENH3 contrastava com a distribuição dos microtúbulos do fuso mitótico, que estavam ligados em densidade uniforme ao longo de todo o comprimento do cromossomo. Essa distribuição dos locais de fixação do fuso comprova a natureza holocêntrica dos cromossomos de *C. europaea* e sugeriu que, nessa espécie, a CENH3 perdeu sua função ou atua em paralelo com um mecanismo adicional de posicionamento do cinetocoro sem CENH3 (Oliveira et al. 2020) .

Entre os monocentrômeros e os holocentrômeros, há espécies como a ervilha que contêm metapolicentrômeros, nos quais alguns centrômeros distintos podem ser observados em um cromossomo mitótico individual citologicamente, mas eles estão relativamente próximos uns dos outros em uma escala cromossômica (Fig. 1; Neumann et al. 2012). Acredita-se que as posições proximais dos metapolicentrômeros permitem que eles funcionem juntos e evitem o ciclo de quebra-fusão do cromossomo e a instabilidade do cromossomo, conforme observado nos cromossomos dicêntricos (Wong et al. 2020).

Os cromossomos holocêntricos se separam como hastes rígidas na mitose, sendo esta chamada divisão paralela, ao contrário da separação cromossômica monocêntrica em forma de V, na qual os braços cromossômicos se movem à frente das regiões monocentroméricas (Hofstatter et al. 2021). Nesse cenário, uma potencial vantagem associada à arquitetura cromossômica da holocentricidade está relacionada às quebras de DNA de fita dupla. De fato, as quebras de cadeia dupla nos cromossomos monocêntricos levam a fragmentos cromossômicos que não podem ser herdados adequadamente pelas células, devido à sua incapacidade de se fixar aos microtúbulos do fuso. No entanto, fragmentos de cromossomos holocêntricos podem segregar adequadamente e ser mantidos de maneira estável, uma vez que possuem locais de fixação de microtúbulos por todo o seu comprimento (Maddox et al. 2004; Bureš et al. 2013).

A combinação de cromossomos holocêntricos e a rápida formação de telômeros após as quebras de fita dupla permite a transmissão estável de fragmentos cromossômicos de forma paralela dos cromossomos e, portanto, a rápida evolução do cariótipo (Cuacos et al. 2015; Jankowska et al. 2015). Essa flexibilidade de cariótipo conferida pelos holocentrômeros é refletida em: (i) o número cromossômico extremamente amplo e quase contínuo encontrado

entre as espécies holocêntricas relacionadas, por exemplo, *Carex* L. 2*n* = 12 - 124 (Cyperaceae) , *Eleocharis* R.Br. 2*n* = 6 - 196 ou *Juncus* 2*n* = 18 - 170 (Bureš et al. 2013); (ii) variação do número de cromossomos entre espécies, por exemplo, *Eleocharis kamtschatica* (C.A.Mey.) Kom. com 2*n* = 41 - 47 (Yano e Hoshino 2006) ou (iii) a correlação negativa entre o número de cromossomos e o tamanho dos cromossomos encontrada em *Luzula*, por exemplo (Nordenskiold 1951). Teoricamente, a holocentricidade poderia permitir um incremento na diversidade de números cromossômicos a partir, principalmente, dos eventos de fusão e fissão cromossômica (em vez de duplicação e deleção), também denominados simploidia e agmatoploidia, respetivamente (Roalson et al. 2007; Escudero et al. 2014; Guerra 2016). Esses eventos podem explicar, por exemplo, a rápida taxa de alteração no número cromossômico observada em *Carex* (Roalson 2008; Hipp et al. 2009; Márquez-Corro et al. 2019).

2.5 EVOLUÇÃO DA HOLOCENTRICIDADE

Os cromossomos holocêntricos e seus centrômeros (holocentrômeros) evoluíram independentemente pelo menos 13 vezes em linhagens eucarióticas (Melters et al. 2012; Drinnenberg et al. 2014; Guerra et al. 2019; Seratnate et al. 2022) sendo a monocentricidade a configuração cromossômica provavelmente ancestral (Melters et al. 2012; Schubert et al. 2020). Em plantas com flores, os holocentrômeros são encontrados entre as monocotiledôneas Cyperaceae, Juncaceae (Malheiros et al. 1947; Hakansson 1958) e *Chionographis* (Tanaka e Tanaka 1977; Kuo et al. 2022), bem como em dicotiledôneas como *Cuscuta* subgênero *Cuscuta* (Pazy e Plitmann 1995), ou na árvore da noz-moscada, *Myristica fragrans* Houtt. (Flach 1966).

A origem da holocentricidade ainda é uma pergunta que envolve diferentes hipóteses, porém várias apontam para a participação de proteínas ou sequências de DNA repetitivas no processo. Nagaki et al. (2005), ao analisar *Luzula elegans* Lowe, propuseram que a base da holocentricidade poderia ter ocorrido a partir de uma extensão da área centromérica, como resultado de uma virada de 90° do cinetócoro. Em geral, os cinetócoros são formados ortogonalmente contra os eixos cromossômicos, o que pode explicar a formação de uma constrição nos cromossomos. No entanto, se a direção girar 90° e a formação percorrer os eixos cromossômicos até as regiões teloméricas, isso poderá gerar cromossomos holocêntricos. Nesse sentido, Villasante et al. (2007) propuseram o modelo “telômero para centrômero”, que prediz uma origem de holocentrômeros a partir de uma propagação contínua de sequências teloméricas ao longo dos monocentrômeros durante a evolução. De acordo com isso, foram observados nos

cromossomos holocêntricos de *Caenorhabditis elegans* muitas repetições em tandem da sequência telomérica em vários locais internos, embora sejam mais abundantes nas extremidades (Sedensky et al. 1994).

Por outro lado, uma porção dos domínios centroméricos em *Luzula nivea* (L.) DC. é composta por grupos dispersos de sequencias do tipo satélite (Haizel et al. 2005), sugerindo que o DNA do satélite é um importante determinante do centrômero em pelo menos alguns cromossomos holocêntricos e, no caso de *L. nivea*, um cenário de transição pode ter ocorrido como consequência da disseminação de satélites centrômeros compatíveis (Neumann et al. 2012). E a hipótese do "centromere drive" sugere uma transição dos cromossomos mono- para holocêntricos pela expansão de um satélite ligado à CENH3, fornecendo mais locais de conexão de microtúbulos (Henikoff e Malik 2002). Inclusive, alguns autores sugerem um possível intermediário na transição da monocentricidade para a holocentricidade, encontrado nos cromossomos "meta-policêntricos" da ervilha altamente repetitivos (*Pisum sativum* Lam., Macas et al. 2023). No entanto, até agora não foi descrita nenhuma linhagem que contenha espécies com os dois tipos de centrômeros, metapolicêntricos e holocêntricos (Schubert et al. 2020).

Uma riqueza semelhante de famílias de satélites é encontrada em alguns gêneros holocêntricos. Em *Luzula elegans* o DNA repetitivo compõe 61% do genoma, sendo os retrotransposons do tipo LTR os mais abundantes, compreendendo ~37%. Nessa espécie, existem mais de 30 famílias de DNA satélites que representam 10% do genoma, mas nenhum possui uma distribuição em todo o cromossomo, e um terço reside nas extremidades dos cromossomos. As sequências ligadas à CENH3 permanecem desconhecidas, mas parece improvável que constituam uma única família repetida (Heckmann et al. 2013). Por outro lado, em *Rhynchospora pubera*, a CENH3 se agrupa em linhas holocentroméricas em um sulco ao longo dos cromossomos, e são 58% coincidentes com os sinais FISH de dois satélites relacionados de 172 bp, Tyba-1 e Tyba-2. Esses DNA satélite são altamente enriquecidos em CENH3 e são ambos transcritos (Marques et al. 2015; Hofstatter et al. 2022), sugerindo que ainda compreendemos pouco sobre a organização centromérica em holocêntricos.

A CENH3 é uma proteína presente em centrômeros funcionais na maioria dos organismos relatados, por isso é uma das principais proteínas usada como marcador para identificar os centrômeros funcionalmente ativos (Henikoff e Malik 2002; Panchenko e Black 2009). No entanto, Drinnenberg et al. (2014) relataram que cada uma das quatro transições

independentes para holocentricidade em insetos foi associada à perda da CENH3. Nessa situação, uma nova linhagem de uma proteína centromérica específica evoluiu de alguma forma para conferir a essas linhagens holocêntricas uma capacidade única de realizar mitose e meiose de maneira independente de CENH3, desse modo relaxando a pressão seletiva para manter a CENH3 no genoma (Ross et al. 2013).

Em plantas existem poucas informações referentes à perda/ganho de proteínas centroméricas associada à holocentricidade. Em *Cuscuta europaea* L., a distribuição da proteína CENH3 não apresentou um marcação linear ao longo de todo o holocentrômero, sendo localizada em dois ou três grandes blocos na heterocromatina (Oliveira et al. 2020). Assim, foi proposto que a CENH3, embora presente, não seja uma das proteínas responsáveis pela montagem do cinetócoro nessa espécie (Oliveira et al. 2020). E mais recentemente, foi descrito que os cromossomos holocêntricos de *Cuscuta* se dividem adequadamente durante a mitose, apesar de mudanças drásticas na composição de seus cinetócoros, sugerindo que a determinação do centrômero e a composição do cinetócoro são notavelmente plásticas em organismos holocêntricos (Neumann et al. 2023). Assim, mudanças na arquitetura do centrômero/cinetócoro podem tornar dispensável uma das proteínas mais definidoras associadas à função do centrômero em quase todos os eucariotos.

2.6 A FAMÍLIA JUNCACEAE

Juncaceae é uma família dentro de Poales (APG IV 2016) que inclui oito gêneros e 474 espécies (POWO 2023). No Brasil, são reconhecidas 24 espécies para a familia, ocorrendo nos domínios fitogeográficos Caatinga, Cerrado, Mata Atlântica e Pampa (Groppo e Pirani 2004; Giulietti et al. 2018; Valadares, 2023). Os gêneros mais representativos são *Juncus* L., com cerca de 332 espécies, e *Luzula* DC., com cerca de 124 espécies, ambos com distribuição cosmopolita (Kirschner 2002; POWO 2023). Os demais gêneros incluem entre uma e seis espécies cada. *Oreojuncus* Záv. Drábk. & Kirschner está distribuído em áreas temperadas do hemisfério Norte (Drábková e Kirschner 2013); *Oxychloe*, *Distichia* e *Patosia* ocorrem especialmente nos Andes e outras áreas elevadas do hemisfério Sul, da Colômbia até a Argentina; *Marsippospermum* é endêmico da Patagônia; e *Rostkovia* ocorre disjuntamente nos Andes do Equador, Chile, Bolívia, Argentina e na Nova Zelândia (Balslev 1996; Jones et al. 2007).

O gênero *Juncus* é o mais representado no Brasil, com 23 espécies (Groppo e Pirani 2004; Giulietti et al. 2018; Valadares, 2023). Está dividido em dois subgêneros (subg. *Juncus*

e subg. *Agathryon* Raf.), caracterizados pela presença ou ausência de bractéolas e morfologia da inflorescência (Kirschner 2002), com seis seções no subg. *Juncus* e quatro seções no subg. *Agathryon* (Buchenau 1906) reconhecidas taxonomicamente. Estudos filogenéticos têm demonstrado que *Juncus* é parafilético e aparece dividido em dois clados incluindo os demais gêneros da família, com a maioria de seus táxons agrupados no subg. *Juncus* e subg. *Agathryon*, ambos também parafiléticos (Roalson, 2005; Drábková 2006; Jones et al. 2007; Drábková e Kirschner 2013). *Juncus* subg. *Juncus* forma um clado monofilético fortemente suportado, exceto por *J. trifidus* L., que é colocado próximo ao clado *Luzula* (Roalson 2005). Consequentemente, Drábková e Kirschner (2013) propõem colocar *J. trifidus* no gênero separado *Oreojuncus.* Adicionalmente, filogenias mais recentes apontam para uma resolução mais baixa, sugerindo a delimitação do problemático gênero *Juncus* ao propor seis novos gêneros para recuperar a monofilia nesse grupo (Brožová et al. 2022).

O gênero *Luzula* é principalmente cosmopolita, mas raramente encontrada nos trópicos, com os principais centros de diversidade no sudoeste da Europa, Extremo Oriente, oeste da América do Norte, América do Sul temperada, Austrália e Nova Zelândia (Drábková e Vlček, 2010). No Brasil ocorre apenas uma espécie nos domínios fitogeográficos da Mata Atlântica e Pampa (Valadares, 2023). O gênero é muito variável taxonomicamente e foi recentemente revisado por Kirschner (2002), sendo ele dividido em três subgêneros (*Luzula*, *Marlenia* e *Pterodes*) e sete seções do subgênero *Luzula* (*Alpinae*, *Anthelaea*, *Atlanticae*, *Diprophyllatae*, *Luzula*, *Nodulosae* e *Thyrsnochlamydeae*).

Filogeneticamente *Luzula* é monofilético, mas as relações entre espécies dentro do gênero são difíceis de determinar, principalmente devido à morfologia semelhante, mesmo dentro de táxons geograficamente remotos (Köbele e Tillich, 2002; Drábková e Vlček, 2010). Em um estudo usando o intron *trn*L do plastoma, o espaçador intergênico *trn*L-F e as regiões ribossômicas nucleares ITS1-5.8S-ITS2 em 93 espécies de *Luzula*, econtrou-se sinais filogenéticos incongruentes apontando para uma classificação incompleta da linhagem, bem como para uma hibridização recente nesse grupo (Drábková e Vlček, 2010). Nesse sentido, semelhante ao gênero *Juncus*, nas espécies da seção *Luzula*, a hibridização recente e a classificação incompleta de linhagem ancestrais podem representar fontes potenciais de incongruência entre os dados de cloroplasto e nucleares.

2.7 JUNCACEA DO PONTO DE VISTA CITOGENÉTICO

A evolução do número cromossômico (eventos de fissão, fusão e poliploidização) pode ser inferida ao longo de uma filogenia usando métodos comparativos (Mayrose et al. 2010; Glick e Mayrose 2014; Ribeiro et al. 2018; Sader et al. 2019). Essa estratégia permite estimar probabilisticamente o número de alterações cromossômicas ao longo de cada linhagem, possibilitando inferir os números cromossômicos ancestrais nos ancestrais (Mayrose et al. 2010). No caso de Juncaceae, os três melhores números de cromossomos haploides inferidos nos ancestrais comuns mais recentes da família sob otimização bayesiana foram *n* 5, 6 e 7 (Carta et al. 2020). Até o momento, só um estudo faz uma compilação de números cromossômicos no gênero e uma abordagem sobre evolução cromossômica em Juncaceae (Drábková 2013), mas essa pesquisa não inclui análises estatísticas de reconstrução do número cromossômico e nem considera as relações filogenéticas para a interpretação de dados.

No nível do gênero, só frequência dos números cromossômicos tem sido utilizadas para inferir números básicos, por exemplo Love e Love (1944) sugeriram o número básico $x = 5$ para *Juncus*, e Darlingtone Wylie (1956) os números cromossômicos $x = 8$ ou 10. Snogerup (1963) sugeriu uma série de nove números básicos secundários dentro de *Juncus*, variando de $x = 9$ a $x = 25$, aceitando um ou dois números básicos em cada subgênero. Harriman e Redmond (1976) sugeriram mais números básicos, de modo que cada subgênero poderia ter até três números básicos; e Balslev (1996) sugeriu $x = 10$ para *Juncus* e $x = 6$ para *Luzula*.

A diversificação do gênero *Luzula* é baseada principalmente na diferenciação cariológica, fundamentada tanto na agmatoploidia quanto na poliploidia (Malheiros e Gardé 1950; Nordenskiöld 1951; Bačič et al. 2007; Bozek et al. 2012). Devido à agmatoploidia e alopoliploidia, os táxons do gênero *Luzula* têm cromossomos de três tamanhos diferentes (Nordenskiöld 1951; Bozek et al. 2012). E a alteração do número de cromossomos proporciona isolamento genético eficaz e especiação abrupta (Nordenskiöld 1951, 1956; Kirschner 1992a,b; Bozeket al. 2012).

Citogeneticamente, a holocentricidade tem sido confirmada várias vezes no gênero *Luzula*, mas no gênero *Juncus* não se sabia se os cromossomos holocêntricos estavam presentes (Godward 1985). No entanto, o gênero *Juncus* foi descrito como monocêntrico usando imunodetecção das histonas fosforiladas H3-S10ph e H2A-T133ph, as quais revelaram hiperfosforilação das regiões pericentromérica e centromérica, respectivamente, em uma área restrita como observado nos cromossomos monocêntricos (Guerra et al. 2019). Além disso,

estudos baseados na imunodetecção citológica da proteína CENH3 (Mata-Sucre et al. 2023; Dias et al. em revisão) e no mapeamento do genoma por meio de imunoprecipitação seguida de sequenciamento (Chip-seq) também confirmaram esse cenário monocêntrico (Hofstatter et al. 2022; Dias et al. em revisão).

Sendo assim, a presente tese traz os resultados obtidos de uma abordagem citogenética, filogenômica e genômica de espécies monocêntricas e holocêntricas da família Juncaceae, com o objetivo de discutir dois aspectos importantes na evolução da família: 1) Como é a evolução cromossômica entre espécies holocêntricas (*Luzula*) e monocêntricas (*Juncus*) em Juncaceae? e 2) Qual é o papel das sequências repetitivas na transição dos centrômeros monocêntricos para holocêntricos em Juncaceae?

3 ARTIGO 1 - Karyotypic diversity of the monocentric genus *Juncus* L. (Juncaceae): a cytogenetic and phylogenetic overview

Yennifer Mata-Sucre[1], André Marques[2], Andrea Pedrosa-Harand[1], Gustavo Sousa[1]

Para ser submetido a Botanical Journal of the Linnean Society A2

Karyotypic diversity of the monocentric genus *Juncus* L. (Juncaceae): a cytogenetic and phylogenetic overview

Yennifer Mata-Sucre[1], André Marques[2], Andrea Pedrosa-Harand[1], Gustavo Sousa[1]

[1]Laboratório de Citogenética e Evolução Vegetal, Departamento de Botânica, Centro de Biociências, Universidade Federal de Pernambuco, Recife, PE, 50670-90, Brasil

[2]Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, 50829, Germany

Short running title: Chromosome evolution of *Juncus* (Juncaceae)

ABSTRACT

The genus *Juncus*, the most diverse of the Juncaceae family, is composed of 332 species with complex taxonomic and phylogenetic relationships. Historically, it was a genus considered holocentric with high intra- and interspecific karyotypic diversity ranging from $2n = 8$ to $2n = 170$. However, cytogenetic and genomic studies have shown that *Juncus* has monocentric chromosomes. Assuming that monocentric chromosomes are less tolerant of chromosome fission and fusion events than holocentric, it is unclear how the karyotypic variability reported for *Juncus* was associated with its centromeric type. We performed new cytomolecular analysis for 12 *Juncus* taxa in addition to a new phylogenetic analysis for the entire genus and review of all chromosome numbers available in the literature for Juncaceae. The results were interpreted from a phylogenetic perspective, using a topology of nuclear and plastid loci based. Cytogenetic analyses revealed monocentric, small (< 2 μm) and numerous chromosomes ($2n =$ 20 to 46). Terminal CMA^+ bands, corresponding to 35S rDNA occurred in all specie, with additional pericentromeric sites in *J. capitatus*, *J. subsecundus* and *J. compressus*. Phylogenetic relationships revealed *Juncus* as non-monophyletic and incongruences between nuclear and plastid topology, suggesting ancient hybridization. The frequent karyotype was $2n = 40$ (40%), with an ancestral chromosome number of $n = 10$ or 20 for *Juncus*, $n = 15$ for *Oreojuncus* and $n = 6$ for the holocentric *Luzula* genus. Dysploidy and polyploidy were the most important event throughout *Juncus* chromosome evolution. Our critical review suggests that the reported karyotypic diversity in *Juncus* could be overestimated due to counting errors, particularly in older analysis <1980. Technical difficulties associated with the high chromosome number and small chromosome size in *Juncus* species may explain these possible errors. Remarkably, we did not find a trend towards greater karyotypic diversity in *Luzula* (holocentric) compared to *Juncus* (monocentric), suggesting that in Juncaceae the centromeric type does not seem to impact a higher rate of fixation of chromosomal breaks/fusions.

ADDITIONAL KEYWORDS: ancestral reconstruction – basic chromosome number- chromevol – phylogenetic - rushes

INTRODUCTION

The chromosome number characterization of a taxon is considered one of the most informative features for describing genomic organization (Mayrose and Lysak 2021), that allow to understand evolutionary relationships (Guerra 2012; Figueredo et al. 2016) and helps unravel the macroevolutionary trends (Clark and Donoghue 2018; Ibiapino et al. 2022). The extant chromosome-number variation is an external manifestation of the underlying dynamic genomic processes, encompassing structural chromosomal rearrangements (e.g., aneuploidy, fission, fusion, Robertsonian translocation, etc.) and changes in DNA content (Mayrose and Lysak 2021). Hypervariable and complex chromosome numbers have been reported for some plant genus (Lewis, 1951; Lewis et al. 1971; Brighton et al. 1973; De Assis et al. 2013; Escudero et al. 2015). However, erroneous chromosome counts may account for a significant fraction of the intra- and interspecific variability in reported chromosome numbers, besides chromosomal rearrangements (Merxmüller, 1970; Guerra, 1986; Baum and Oginuma, 1994; Marinho et al. 2014; Figueredo et al. 2016). Therefore, chromosome recounts, followed by more detailed cytomolecular analyses represent an important tool for determining the true karyotypic diversity.

Juncaceae are a cosmopolitan family comprising 474 species (POWO, 2023). Of its eight genera, *Juncus* and *Luzula* (Kirschner, 2002) represent the first and second largest in the family, with 332 and 124 species, respectively (Drábková et al. 2006; POWO, 2023). Phylogenetically, the family has had a complicated history due to the paraphyletism of *Juncus* with other genera. Since Kirshner's sectional treatment in 2002 based on morphological data, different works have shown that the genus *Juncus* is paraphyletic, since then the current debate centers on whether this non-monophyletic genus should be treated as one large genus or divided into several genera (Roalson 2005; Drábková 2006; Jones et al. 2007; Drábková and Vlček 2010; Drábková and Kirschner 2013; Brožová et al. 2022; Mata-Sucre et al. 2023). In this case, Brožová et al. (2022), based on nuclear rDNA (ITS) and two plastid regions (the gene *rbc*L and the spacer *trn*L-F), split the sections into six new genera to recover monophyly in this group.

Historically, Juncaceae were considered as holocentric, and the inter- and intraspecific chromosomal variability reported for *Juncus* was associated with their holocentricity condition (Jorgensen et al. 1958; Mehra and Sachdeva, 1976; Beuzenberg and Hair, 1983; Dalgaard et al. 1991; Schönswetter et al. 2007; Aguilera et al. 2014). Holocentric chromosomes tolerate

chromosome breaks due to their diffuse centromere-like structures, allowing chromosome fragments to segregate normally during meiosis and to be inherited in a Mendelian pattern (Mandrioli and Manicardi, 2020; Senaratne et al. 2022). Therefore, holocentric chromosomes theoretically allow rapid evolution of chromosomal rearrangements through fission and fusion (Hipp et al. 2009; Burchardt et al. 2020; Escudero et al. 2023). However, the discovery that the family has monocentric chromosomes within its representatives has puzzled the interpretation of the range and significance of its karyotypic variability (Guerra et al. 2019; Hofstatter et al. 2022; Mata-Sucre et al. 2023; Dias et al. in review).

Cytogenetically, *Juncus* is characterize by having small (<2 μm) and numerous chromosomes from 2*n* = 18 to 170, but this aspect has been poorly investigated. Among all its eight taxonomic recognized sections, only *Ozophyllum* has been the most cytogenetically analyzed with 37 % of the taxa counted (Drábková, 2013; Choi et al. 2022; Mata-Sucre et al. 2023). On the other hand, for the well-characterized holocentric genus *Luzula* the karyotypes are of medium/large size (~3-5 μm), with numbers ranging from 2*n* = 6 to 70 (Madej and Kuta, 2001; Bozek et al. 2012; Drábková, 2013). For the genus *Orejuncus* Záveská Drábková et Kirschner, a single chromosome number has been reported 2*n* = 30 chromosomes (Jorgensen et al. 1958). Moreover, other genera in the family such as *Patosia* and *Distichia* are completely unknown karyologically.

The evolution of chromosome numbers (fission, fusion and polyploidization events) can be inferred over a phylogeny using comparative methods (Mayrose et al. 2010; Glick and Mayrose 2014; Ribeiro et al. 2018; Sader et al. 2019). This strategy allows for probabilistically estimating the number of chromosomal changes along each lineage, making it possible to infer the ancestral chromosomal numbers (Mayrose et al. 2010). Different basic chromosome numbers (*x*) have been suggested for the Juncaceae family, such as *x* = 5, 8, 9, 10, 13, 14, 15, 17, 19, 20 and 25 (Love and Love, 1944; Darlington and Wylie, 1956; Snogerup, 1963; Harriman and Redmond, 1976; Balslev, 1996; Drábková, 2013; Kaur et al. 2014), however, this remains unresolved.

Genomic analysis suggests a possible whole genome duplication (WGD) event for *Juncus* considering the duplication of some genes (McKain et al. 2016). However, a recent study using the assembled genome of *J. effusus* revealed no evidence of WGD in the origin of *Juncus* (Hofstatter et al. 2022), so high basic chromosome number generated by multiple chromosomal fissions is likely. Nonetheless, detailed analyzes, cytogenetic revisions and chromosome number reconstruction using phylogenetic relationship for data interpretation are lacking. Therefore, in the present work we analyzed the karyotypic diversity of the genus

Juncus based on chromosome counts, morphology, double staining with the fluorochromes chromomycin A3 (CMA) and 4',6-diamidino-2-phenylindole (DAPI), and fluorescent *in situ* hybridization (FISH) with 5S and 35S rDNA. The new chromosome counts were compared with those in the literature to check the occurrence of intra- and interspecific variability in the genus. These data were interpreted in an evolutionary context based on a molecular phylogeny of Juncaceae, which highlighted the patterns of karyotype evolution and diversification of the family.

MATERIAL AND METHODS

PLANT MATERIAL

Seed of 12 *Juncus* taxa available were obtained from Millennium seed bank (Kew Gardens), Leibniz Institute of plant Genetics and Crop Plant Research Genebank (IPK-Gatersleben) or through commercially available cultivars (https://www.plant-world-seeds.com/). These plants were grown and cultivated in the greenhouse of the Laboratory of Plant Cytogenetics and Evolution in Brazil and in the greenhouse of the Max Planck Institute for Plant Breading Research in Germany (MPIPZ; 16h daylight, 20 °C, >70% humidity; Table 1).

CYTOGENETIC ANALYSES

Root tips obtained from germinated seeds were pretreated with 0.002 M 8-hydroxyquinoline for 24 h at 10 °C, fixed in Methanol:acetic acid (3:1 v/v) for 2–24 h at room temperature and stored at -20 °C. For preparation of slides, fixed root tips were washed in distilled water and digested in a 2% (w/v) cellulase (Onozuka)/20% (v/v) pectinase (Sigma) solution at 37 °C, for 90 min. The meristem was macerated in a drop of 45% acetic acid, air-dried and dipped in a 60% acetic acid solution following Carvalho and Saraiva (1993). The CMA/DAPI double-staining technique was used for fluorochrome banding following Mata-Sucre et al. (2020). Slides were aged for three days before their analysis using an epifluorescence Leica DMLB microscope.

Fluorescent *in situ* hybridizations (FISH) of rDNA site were performed following Pedrosa et al. (2002). The 5S rDNA oligo-probe (D2) pre-labeled with Cy3-dUTP (GE) was obtained following Waminal et al. (2018). The pTa71 (18S–5.8S–26S rRNA) probe from *Triticum aestivum* was labelled with Alexa-448 by nick translation (Gerlach et al. 1869). The hybridization mixture contained formamide 50% (v/v), dextran sulphate 10% (w/v), 2× SSC and 50 ng/μL of each probe. The slides were denatured at 75 °C for 5 min and incubated for

24h to 37°. Stringent washes were performed, to give a final stringency of approx. 76%. Images were captured with a Cohu CCD video camera using the Leica QFISH software. Finally, images were edited in Adobe Photoshop CS3 version 10.0 for brightness and contrast.

TAXA SAMPLING AND PHYLOGENETIC ANALYSES

A comprehensive phylogeny of Juncaceae were created by combining an available dataset of 217 species (46%) with representatives of all related genera (*Distichia*, *Juncus*, *Luzula*, *Marsippospermum*, *Oreojuncus*, *Oxychloe*, *Patosia* and *Rostkovia*). *Carex capillifolia* (PRJNA927338) was included as outgroup. Five plastid regions (*mat*K, *psb*A-*trn*H, *rbc*L, *rpo*C1, and *trn*L-*trn*F) and the internal transcribed spacer (ITS) of the ribosomal nuclear region were used. All sequences and alignments were obtained from NCBI GenBank (Benson et al. 2013, https://www.ncbi.nlm.nih.gov/) using the OneTwoTree server (Drori et al. 2018; http://onetwotree.tau.ac.il/). Subgeneric and sectional names are according to Kirschner et al. (2002). The list of the total 217 accessions of Juncaceae family sampled are present in Table S1.

Alignment of plastid and nuclear regions were individually concatenated using the MAFFT software (Katoh and Standley, 2013). jModelTest v.2.1.6 (Darriba et al. 2012) was used to select the best model of DNA substitution. Both matrices were used to construct the final phylogenetic trees through Bayesian Inference under the GTR-I-G nucleotide substitution model using MrBayes v.3.2.6. (Ronquist et al. 2012). The run was conducted with Markov Chain Monte Carlo (MCMC), sampling every 1000 generations for 40,000,000 generations. To verify effective sampling of all parameters and measure the convergence of the independent chains, we examined their posterior distributions in Tracer v.1.6 until MCMC sampling reached an ESS >200. The consensus tree was generated in TreeAnnotator v.1.6.1 (Rambaut and Drummond, 2010) with a burn-in of 25%. The consensus tree and posterior probability (PP) were visualized in FigTree v.1.4.2. (Rambaut, 2014). Support values were considered high when PP was ≥0.95 (Erixon et al. 2003; Pirie, 2015). Since, the individual trees resulted in poorly resolved topologies (see Results), all analyses presented here are based on a comparative analysis of the nuclear locus alignment (ITS) and the concatenated alignment of plastome loci (Plastid).

Divergence time was estimated in BEAST v.1.10.4 (Drummond and Rambaut, 2007) by fixing the topology of the Bayesian trees. An uncorrelated relaxed lognormal clock (Drummond et al. 2006) and a Birth Death Process speciation model (Gernhard, 2008) were applied. Two independent runs of 30,000,000 generations each were performed sampling every

10,000 generations. We verify the effective sampling in Tracer as mentioned before. After removing 25% of the samples as burn-in, a maximum cladistic credibility (MCC) tree was constructed for ITS and plastid regions using TreeAnnotator v.1.8.2. (Drummond et al. 2012). Calibrations were performed using secondary calibrations based on *Juncus*/*Carex* crown age divergence (ca. 84.7 Mya, Ramírez-Barahona et al. 2020) and *Juncus*/*Luzula* crown age divergence (ca. 64.4 Mya; Escudero and Hipp, 2013).

CHROMOSOME NUMBER COMPILATION

The haploid chromosome numbers published to date were collected from the Chromosome Counts Database (Rice et al. 2015, https://ccdb.tau.ac.il/), from the Index to Plant Chromosome Number database (http://www.tropicos.org), from the International Association for Plant Taxonomy (https://www.iaptglobal.org/chromosome-data) and from original counts obtained here (see Table S2). Chromosome number data were filtered based on different categories: a) counts with a photographic record with adequate clarity, resolution and count, b) counts with a drawing record, c) counts without a photographic record but that were confirmed by more than three authors from different sources, and d) counts without a photographic record and without confirmation from at least three authors. Chromosome counts of the holocentric genus *Luzula* were included as outgroups. For each category (a, b, c, d and outgroup), the mean, median and standard deviation were calculated. The dispersion of the data was visualized by violin plots comparatively and to assess the accuracy of chromosome number diversity, the coefficient of variation (CV) of the raw and filtered data was analyzed.

ANCESTRAL RECONSTRUCTION

For phylogenetic reconstruction of ancestral chromosome numbers, due to lack of accuracy of chromosome numbers (see results), the terminals of the final trees (ITS and plastome, see above) with chromosome counts of categories a, and b plus the counts generated in this study were maintained. The terminals were cut using the 'drop.tip' function of the R ape v5.7.1 packages (Paradis and Schliep, 2019), thus 18 and 27 taxa were sampled for the nuclear and plastome matrix, respectively. The sampled subgenera included *Juncus*, *Marsippospermum*, *Oreojuncus*, *Oxychloe*, and *Rostkovia* (Table S3). Additionally, species of the genus *Luzula* with available chromosome numbers were included. For taxa with numbers suggesting different ploidy levels, we used both haploid chromosome number available.

To investigate the haploid ancestral states of chromosome number and evolution we used ChromEvol 2.0 software inference (Glick and Mayrose 2014). The ChromEvol method assesses the fit of several models of chromosome number change across a phylogenetic tree and employs different combinations of (1) duplication, (2) demi-duplication, (3) gain or (4) loss of chromosome pair, (5) gain or (6) loss of a single chromosome pair starting from the current chromosome number or (7) a specified chromosome number characterizing a phylogenetic group and (8) rate for transitions by basic chromosome number (Mayrose et al. 2010; Glick and Mayrose, 2014). The model that best fits the data set was chosen after a first run considering all possible models in the program in both trees. Then, the best-fitted model was selected using Akaike's information criterion (AIC), and submitted to the model adequacy test on-line through the ChromEvol webserver (http://chromevol.tau.ac.il/) for modification of any deviations of the model (Rice and Mayrose, 2021).

We conducted three different runs: (i) first, we performed a standard run, using the haploid and monoploid (ii) chromosome number with the parameters given by the model adequacy test in both trees; and (iii) then, we made a run for independent trees removing the holocentric clade (genus *Luzula*) from the analysis to test the influence of holocentric chromosomes and intraspecific variation in chromosome number following Márquez-Corro et al. (2019). Furthermore, ancestral haploid chromosome numbers were estimated through Bayesian (RevBayes, Freyman and Höhna, 2018) and maximum likelihood reconstructions (PastML, Ishikawa et al. 2019) to compare the results with those originating from ChromEvol. PastML prediction method use a JOINT+F81 evolution model. The same input as for the ChromEvol analysis was used for both reconstructions (Table S3). Since monophyly of genera within Juncaceae has only been demonstrated for *Luzula*, the interpretation of chromosome evolution was based on the distribution of character states at the intragroup level, similar to other character evolution studies conducted in the genus (Drábková, 2013).

RESULTS

NEW CYTOMOLECULAR CHARACTERIZATIONS IN *JUNCUS*

Cytogenetic analyses in 12 taxa of *Juncus* revealed chromosome numbers ranging from $2n = 20$ to 46 (Table 1). The most frequent chromosome count was $2n = 40$ in *J. krausii* (Fig. 1B), *J. subsecundus* (Fig. 1C), *J. compressus* (Fig. 1D), *Juncus* cultiv. "Carman Grey" (Fig. 1E), *J. usitatus* (Fig. 1F), *J. micranthus* (Fig. 1G) and *J. xiphioides* (Fig. 1H). The lowest count was observed in *J. capitatus* with $2n = 20$ (Fig. 1A). Chromosome numbers such as $2n = 44$ were counted in *Juncus* cultiv. "Pincei" (Fig. 1I) and *Juncus* cultiv. "Mallorcan Giant" (Fig.

1J). And the highest chromosome counts of 2*n* = 46 were found in *J. acutus* (Fig. 1K) and *J. cooperi* (Fig. 1L). A primary constriction was evident in most of the chromosomes, as such in *J. micranthus* (Fig. 1G).

In general, all species presented karyotypes with chromosomes of small size (~2 µm) and morphology (Fig. 1). Two to four large $CMA^+/DAPI^-$ bands were evident in the terminal region of most of the species, with the exception of additional pericentromeric sites present in *J. capitatus*, *J. subsecundus* and *J. compressus* species (Fig. 1; Table 1). In addition, a frequency of two 5S rDNA sites at the terminal position were observed, except for *J. capitatus* (Fig. 2A) and *Juncus* cultiv. "Mallorcan Giant" (Fig. 2H), which showed a more interstitial pattern (Fig. 2; Table 1). Likewise, two to four 35S rDNA sites at the terminal position were observed, co-localized with the CMA^+ band varying in size and intensity (Figs. 1 and 2).

CHROMOSOME NUMBER DIVERSITY AND EVOLUTION

Although only about 23% of Juncaceae taxa have at least one chromosome count, raw reports on chromosome number show a large karyotype diversity in *Juncus* with a low coefficient of variation (CV=43.6%; Fig. 3A; Table S4). The basic chromosome number *n* = 20 (40%) was the most frequent for *Juncus*, followed by *n* = 40 (17%; Fig. 3A and C). Filtered data show a significant reduction of the variation in *Juncus*, which is mainly originated by counts in drawn and without photographic record (Fig. 3B; Table S4). Instead, chromosome numbers with photographic record had the lower CV (31.6%; Fig. 3B). The holocentric *Luzula* genus showed a low chromosome number diversity with a high CV (61.5%; Fig. 3A; Table S4), whit *n* = 6 (47%), 12 (24%) and 18 (12%; Fig. 3A and C) as common chromosome numbers. Chromosome numbers multiple of *x* = 20 and 40 were found over 62% of the karyotypes, while multiples of *x* = 15 and *x* = 21 were registered in a small number of accessions, representing each one ~8% of the sample (Fig. 3C). The lowest chromosome count described in *Juncus* is 2*n* = 18 in *J. capitatus*, which differs from our count of 2*n* = 20 (Fig. 1-2). Chromosome numbers compilation was summarized in Table S2.

INCONGRUENCE BETWEEN NUCLEAR AND PLASTID PHYLOGENETIC RELATIONSHIPS

Phylogenetic relationships among species are consistent with previous phylogenetic studies (Drábková et al. 2009; Mata-Sucre et al. 2023; Brožová et al. 2022), revealing a complex phylogenetic relationship among *Juncus* and related genera (Fig. 4). Relationships within Juncaceae showed monophyletism for the genera *Oreojuncus* (~1.8-2.6 Mya) and *Luzula* (~21-

28 Mya) in both ITS and plastid topologies (pp = 1.0; Fig. 4; Fig. S1). However, the relationship between these two genera with the rest of the Juncaceae species was not resolved. In the nuclear topology they appear as sister genera (pp = 0.96) and in the plastid tree *Oreojuncus* appears as sister clade of the other *Juncus* species with low support (pp = 0.69).

Two different accessions of *J. capitatus* grouped as 'sisters' taxa but showed different relationships with the remaining *Juncus* species (hereafter called clade Capitatus; Fig. 4), appearing as the first diverging lineage low-supported related to the *Juncus* "sensu stricto" clade (ITS tree, pp=0.6), or as a high-supported sister group to the genera *Luzula* and *Oreojuncus* (plastid tree, pp=1). *Juncus* emerges as non-monophyletic by including the genera *Distichia, Marsippospermum, Oxychloe, Patosia* and *Rostkovia* among its species and none of the sections described by Kirschner et al. (2002) were monophyletic (Fig. 4).

To facilitate the discussion of the results of the ancestral reconstructions we have designated five main clades in *Juncus* "sensu stricto": Clade I, described above as Clade Capitatus. Clade II contains most of the species of the subgenus *Agathyron* and some species of the sections *Juncus* and *Ozophyllum* (pp = 1.0). Clade III includes sections *Juncus*, *Graminifolii*, *Caespitosa* and the southern hemisphere genera *Marsippospermum*, *Oxychloe* and *Rostkovia* (pp = 0.98-0.57). Clade IV contains the sections *Stygiopsis* and *Juncotypus* (pp = 1). And finally, clade V contains most of the species of section *Ozophyllum* and the only species from *Iridifolii* (pp = 1; Fig. 4).

ANCESTRAL CHROMOSOME NUMBER RECONSTRUCTION

The best model for chromosome number reconstruction in ChromEvol was 'Dys_Dup_Dem', with $n = 7$ indicated as the basic ancestral number for Juncaceae in both ITS and plastid trees (Fig. S2). This model considers up to three parameters that is a constant rate of chromosome gain/loss and similar rate of polyploidy and demiploidy events (Mayrose et al. 2010). According to this model, chromosomal number variation in the family is most frequently related to chromosomal loss events [with a frequency probability of f = 10.5 (ITS) – 2.5 (plastid)], followed by chromosomal gains (f = 0 - 26) and chromosomal duplications (f = 8.2 – 13.9). The number $n = 6$ and 15 were indicated as the ancestral chromosome number of the genus *Luzula* and *Oreojuncus*, respectively in both trees. The ancestral number of the clade Capitatus was $n = 7$ and $n = 10$ were inferred for the rest of the *Juncus* "sensu stricto' species (Fig. S2).

To test whether the holocentric nature of the *Luzula* genus and the high variation in chromosomal number influenced the analysis, we ran ChromEvol without the holocentric clade,

following the same parameters as described above (Fig. S3A). The basic number in this scenario was $n = 10$ for both phylogenies, in this scenario chromosomal gains and duplication were the main chromosome evolution events (Fig. S3A). The reconstructed ancestral number for the genus *Oreojuncus* and *J. capitatus* was $n = 15$ and 10, respectively (Fig. S3A). However, nuclear and plastid trees showed different internal ancestral numbers for the rest of the clades, going from $n = 23$ to $n = 30$ in clades IV and V (Fig. S3A). On the other hand, more stable basic number was observed during monoploid chromosome number reconstruction, where ancestral chromosome was maintained at $n = 10$ or 20, at internal and core nodes (Fig. S3B).

Additionally, Bayesian and PastML reconstruction recovered $n = 20$ as ancestral chromosome number with a highest probability, with some variations in $n = 15$ for *Oreojuncus* and $n = 10$ for *J.* capitatus, similarly multiples events of chromosome losses and duplication were inferred through the phylogeny (Fig. 5B). In this scenario, a stable chromosome number of $n = 20$ was maintained throughout the phylogeny, and numerical changes were mainly due to chromosomal losses, followed by chromosomal duplications and gains. Once again, the reconstructed basic numbers for *Luzula* (except for *L. elegans*) and *Oreojuncus* were $n = 6$ and 15, respectively. For the *Juncus* "sensu stricto" group, different base numbers were observed in both reconstructed trees, where more stable ancestral numbers equal to $n = 15$, 20 or 30 were observe for core node (Fig. 3B).

DISCUSSION

JUNCUS HAS NUMEROUS, DIVERSE AND SMALL MONOCENTRIC CHROMOSOMES

New chromosome counts of different *Juncus* species belonging to the subgenera *Juncus* and *Agathyron* reveals numerous metacentric to submetacentric chromosomes with an evident primary constriction. Centromere characterization in small chromosomes is very difficult and there is no universal method due to limited optical resolution, availability of specific antibodies or the necessary equipment (Baez et al. 2020; Schubert et al. 2020). In this case, the monocentricity observed here in 12 taxa together with the description of monocentricity by centromeric proteins (CENH3) and genomic results in seven other *Juncus* species belonging to phylogenetically distant clades, makes it unlikely to find species with holocentric chromosomes within the genus (Guerra et al. 2019; Hofstatter et al. 2022; Mata-Sucre et al. 2023; Dias et al. in prep.).

Since *Juncus* is monocentric, so far only the genus *Luzula* is holocentric within the family Juncaceae (Heckmann et al. 2011, 2013; Zedek and Bureš, 2016; Schubert et al. 2020; Senaratne et al. 2022). This is not the first case of finding monocentric chromosomes in groups previously considered holocentric, *Prionium serratum* (Thunb.) Drège (Thurniceae) the first diverging lineage of the Cyperid clade (Thurniceae-Juncaceae-Cyperaceae) was described as monocentric (Báez et al. 2020). This finding proposes that monocentricity is the ancestral condition for the Cyperid clade, while holocentricity arose independently at least twice after the differentiation of the three families, once in Cyperaceae and the other in Juncaceae, specifically, in the genus *Luzula* (Heckmann et al. 2011; Bozek et al. 2012).

Karyotypes with $n = 6$, 15 and 20 are the most recurrent among the genera *Luzula*, *Oreojuncus* and *Juncus*, respectively, the latter being confirmed here with 58% of the analyzed species. The wide chromosomal variation reported for *Juncus* was traditionally attributed to the presence of holocentric chromosomes, however, its holocentric sister genus (*Luzula*) shows a similar coefficient of variation than *Juncus*, suggesting that this variation may be may be attributed to other biological processes (Guerra, 2008; Kaur et al. 2014; Mayrose and Lysak, 2021). Some species have been reported as members of species complexes involving cryptic species, such as *J. biglumis* and *J. bufonius* (Schönswetter et al. 2017). Likewise, in *Agathryron* subgenus some species complexes have been reported to be involved in hybridization events between its species (Clifford, 1958; Cope, 1985; O'Mahony, 2002; Smith, 2006; Wilcox, 2010, 2011; Stace, 2020).

Intraspecific and interspecific variations in chromosome counts can also be attributed to technical difficulties, high chromosome numbers, and/or the small size of the chromosomes. The inaccurate counts of numerous small chromosomes such as those of *Juncus* could be influencing this variation, as has been observed in counts performed for *J. bufonius* and *J. minutulus* by Mičieta (1983), where more than 10 different chromosome numbers were reported. Additionally, secondary chromosome constrictions, when positively stained using conventional stains, can resemble extra chromosomes and confuse the observer, leading to imprecise determinations of chromosome numbers (Guerra et al. 1997). However, the low number of populations analyzed so far could be hiding chromosome variability in the evolutionary history of *Juncus*.

The 5S and 35S rDNA sites were not significantly polymorphic among the species analyzed, thus they do not represent a synapomorphy for the genus. However, the observed

variation in the size and intensity of the 35S rDNA bands may indicate preferential amplification of this sequence in the genome of the species. A phylogenomic characterization of repetitive sequences in the genus showed a large abundance of LTR-retroelement, satellite DNA and 35S rDNA in the genome of 33 different *Juncus* species which could explain the variation in intensity observed in this study (Mata-Sucre et al. 2023). Terminal heterochromatin GC-rich bands in terminal position corresponded to 35S rDNA, yet in three species additional pericentromeric bands were observed, suggesting an enrichment of other repetitive sequences in these regions. It has been observed that GC-rich heterochromatin can be a hotspot for the accumulation of repetitive sequences such as transposable elements and satellites DNA (Van-Lume et al. 2019; Mata-Sucre et al. 2020; Ibiapino et al. 2020), and although satellite DNA and LTR-retrotransposon are sequences commonly found in (peri)centromeric regions for *Juncus* (Hofstatter et al. 2022; Dias et al. in review), future cytogenomic analyses could corroborate this scenario in *Juncus*.

NUCLEAR AND PLASTID TOPOLOGY TELLS DIFFERENT HISTORY

Incongruences between ITS and plastid phylogenies have been recurrent in studies on Juncaceae. Many of these incongruences were not detected in some previous studies because they did not analyze these regions separately in a comparative approach (e.g., Drábkova et al. 2003, 2006, 2009, 2010; Roalson, 2005; Jones et al. 2007), allowing some studies to concatenate plastid and nuclear sequence data, in the context of considerably larger data sets (e.g., Drábková, 2010; Brožová et al. 2022). However, we found that relationships between and among sections vary depending on whether plastid or nuclear sequences are considered, as was observed by Mata-Sucre et al. (2023). Even a phylogenetic reconstruction based on the repeatome of *Juncus* species, despite limited sampling, suggested the same topology of ITS sequences (Mata-Sucre et al. 2023). Thus, it appears that, regardless of the amount of data and the concerted evolution of the nuclear or repeated region, the nuclear evolution of the group is accurately reflected.

Our results are matching to those of Drábková et al. (2006), Jones et al. (2007) and Mata-Sucre et al. (2023), positioning the genus *Luzula*, the first genus to diverge, as sister to *J. capitatus* and *Oreojuncus* when using plastid data, and sister to the entire genus *Juncus* with ITS markers. Furthermore, in *J. capitatus*, two contrasting topologies were found, including a split that deeply divides the three genera into two monophyletic clades in the ITS phylogeny, that is missing from the plastid phylogeny. the position of. In the plastid tree, we recovered a

species grouping similar to that of Drábková (2006) and Jones (2007), and although the phylogenies of Brožová et al. (2022) and Mata-Sucre et al. (2023) differ slightly from ours, comparisons are difficult due to different sampling.

Cyto-nuclear incongruities associated with hybridization events have persisted in plant genomes for millions of years (Stull et al. 2023 and references therein). Both homoploid hybridization and polyploidization are common phenomena among the tips of the Poaceae phylogeny (Mason-Gamer, 2004; Marques et al. 2016; Winterfeld et al. 2016). In Juncaceae possible ancient hybridization events have been suggested (Cope and Stace, 1985; O'Mahony, 2002; Smith, 2006; Wilcox, 2011; Mata-Sucre et al. 2023), so it is reasonable to assume that both processes have been important throughout the evolution of grasses.

DYSPLOIDY PLAYS AN IMPORTANT ROLE IN THE CHROMOSOME EVOLUTION OF JUNCACEAE

Juncaceae is a family with holocentric (*Luzula*) and monocentric (*Juncus* and *Oreojuncus*) chromosomes, and a complex transition scenario of chromosome number evolution was shown to best explain the results in different clades as observed in Cyperaceae (Márquez-Corro et al. 2019) and Convolvulaceae (Bozek et al. 2012; Ibiapino et al. 2022). The ChromEvol indicated the Dys_Dup_Dem model as the best for our data set. Through this model, the basic ancestral number for Juncaceae was $x = 7$ as previously suggested (Love and Love, 1944; Darlington and Wylie, 1956; Snogerup, 1963; Harriman and Redmond, 1976; Balslev, 1996; Carta et al. 2020), being $x = 6$ for the genus *Luzula* and $x = 15$ for *Oreojuncus* as suggested by Kaur et al. (2014) and Drábková (2013). Ancestral chromosome number $x = 10$ or $x = 20$ was observed for *Juncus* "sensu stricto", the last one being suggested by Balslev (1996), Kaur et al. (2014) and Drábková (2013).

The inferred ancestral haploid chromosomal number $x = 10$ for *Juncus* "sensu stricto" suggests polyploidy as the most relevant event during evolution, rather than fission/fusion, although other mechanisms cannot be ruled out. In this scenario, a WGD at the base of the tree would explain the maintenance of the diploid $x = 20$ state with a small genome size (Elliot et al. 2022). This is consistent with a possible WGD events in *Juncus* inferred from gene duplication after the divergence from the Cyperids (McKain et al. 2016; Leebens-Mack et al. 2019). Though, most chromosome numbers recorded in *Juncus* are multiples of $x = 20$, as 40% of the species are diploids with $2n = 40$ and, among polyploids, $2n = 60$ and 80 are the most frequent numbers.

On the other hand, ancestral $x = 20$ chromosome number were supported by BayesTraits and PastML analyses as suggested by Drábková (2013) in the lasted review of the chromosome number of the family. In this scenario, dysploidy would be the most dominant mechanism. Since phylogenetically more basal species present $n = 10$ as a base chromosome number, this could suggest a possible paleopolyploid. However, a genome assembly at the chromosome level of *J. effusus* ($2n = 42$) did not reveal any recent duplication event, reinforcing the idea of high ancestral chromosomes numbers (Hofstatter et al. 2022). Therefore, we consider $x = 20$ more likely for the genus *Juncus* "sensu stricto".

For Juncaceae family two mechanism has been involved in the chromosome number evolution, polyploidisation and agmatoploidy, the latter being related to chromosomal fission of *Luzula* holocentric chromosomes (see Bozek et al. 2012; Burés et al. 2013; Drábková, 2013). However, agmatoploidy or agmatopolyploidy in *Juncus* have so far to be confirmed, and the presence of holocentric chromosomes in this genus is becoming less and less likely (Guerra et al. 2019; Hofstatter et al. 2022; Mata-Sucre et al. 2023; Dias et al. in review). In both scenarios $x = 10$ or 20, suggest that the $x = 6$ karyotype for *Luzula* genus has evolved by fusion events, whereas for *Juncus* a combination of polyploidisation with dysploidy could have been originated the different karyotypes, therefore, karyotypes of species with $2n = 44$ (as *Juncus* cultiv. "Mallorcan Giant") and $2n = 46$ (as *J. acutus*) have been arisen by ascending dysploidy. Although more chromosome counts would be useful to establish broader patterns of karyotype evolution in the genus.

CONCLUSIONS

Newly cytomolecular data support a scenario of monocentric chromosomes and relative karyotypic stability in *Juncus* represented by chromosome number, CMA/DAPI pattern and the presence of one or two pairs of 35S rDNA sites and one of 5S rDNA. High intra-inter chromosome number variation in the group may be lower than reported in the literature, where counting errors, especially in early publications, may have overestimated karyotype variability. Our analyses revealed an ancestral haploid chromosome number for the *Juncus* group of $x = 10$ or 20, with both dysploidy and polyploidy playing crucial roles in chromosome number evolution. While descending/ascending dysploidy is equally distributed early and late across the phylogeny, polyploidy is detected mainly near to the tips.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Yennifer Mata-Sucre: Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Andrea Pedrosa-Harand**: Resources, Writing – review & editing, Supervision. **André Marques**: Resources, Writing – review & editing, Supervision, Funding acquisition. **Gustavo Souza**: Conceptualization, Resources, Writing – review & editing, Supervision, Visualization, Funding acquisition.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

ACKNOWLEDGEMENTS

We are thankful to Prof. Dr. Marcelo Guerra (UFPE), Prof. Dr. Leonardo Pessoa Felix (UFPB) and Dr. Erton Mendonça de Almeida for providing the plant material and for the previous study that originated this work. This work was conducted during a scholarship to YMS supported by the International Cooperation Program PROBRAL at the Federal University of Pernambuco. Financed by Capes – Brazilian Federal Agency for Support and Evaluation of Graduate Education within the Ministry of Education of Brazil. This work was supported by a grant awarded to A.P.H (PROBRAL CAPES/DAAD project number 88881.144086/2017-01). G.S. receive productivity fellowship from CNPq (process numbers PQ-312852/2021-5).

Table 1. Table 1. List of *Juncus* taxa investigated cytogenetically, with collection ID, chromosome number (2*n*), previous chromosome number reports, number of CMA bands and ribossomal DNA sites (5S + 35). T: terminal bands. P: proximal bands. I: intersticial bands. CCA: crop commercially available

Species	Collection	2*n*	Previous chromosome reports (2*n*)	References	CMA$^+$	5S-35S
Juncus capitatus Weigel	KEW305163	20	18	Snogerup 1963; Mičieta 1983	4T	2I-4T
Juncus compressus Jacq.	CCA	40	ca.40	Snogerup 1963;	-	2T-4T
			44	Snogerup 1963; Mičieta 1980b; Druskovic 1995		
Juncus krausii Hochst.	KEW262763	40	-		2T	4T-2T
Juncus micranthus Desv.	CCA	40	-		2T	-
Juncus subsecundus N.A.Wakef.	KEW638948	40	-		2T	2T-2T
Juncus usitatus L.A.S.Johnson	CCA	40	-		-	2T-
Juncus xiphioides E.Mey.	CCA	40	40	Snogerup 1963	2T	2T-2T
Juncus cultiv. “Carman Grey”	CCA	40	-		2T	2T-2T
Juncus cultiv. "Pincei"	CCA	44	-		4T	2T-4T
Juncus cultiv. “Mallorcan Giant”	CCA	44	-		4T	2I-4T
Juncus acutus L.	IPK150359	46	ca. 42	Jones and Richards, 1954	4T	2T-4T
			46	Snogerup 1958; Labadie 1976b; Borgen 1969; Dalgaard 1986a		
			48	Snogerup 1963; Dalgaard 1991; Druskovic 1995; Snogerup 1993		
			60	Mesquita 1953		
Juncus cooperi Engelm.	KEW170033	46	-		2T	2T-2T

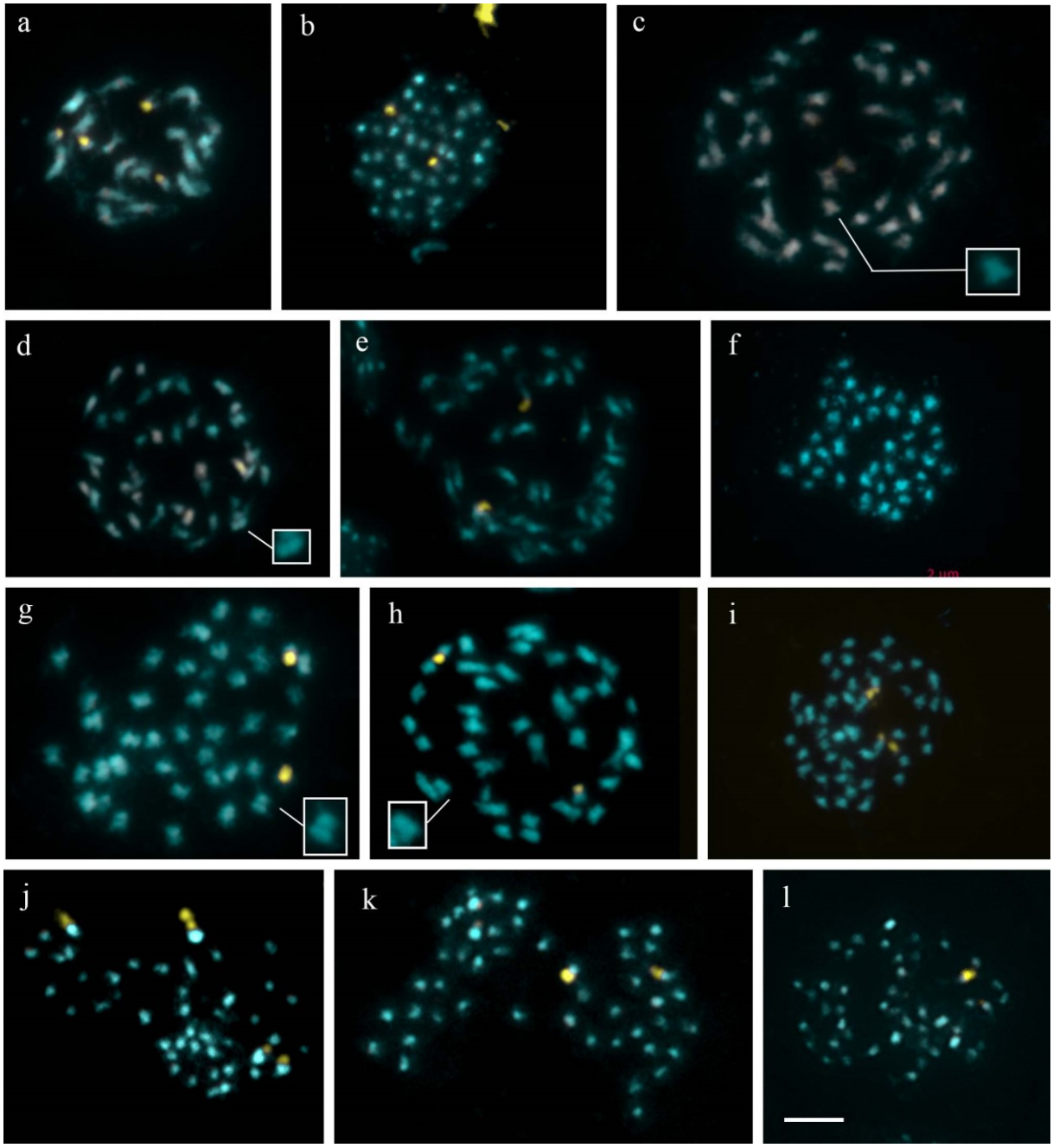


Fig. 1 CMA/DAPI-stained chromosomes of different *Juncus* species. (a) *J. capitatus* (2*n* = 20), (b) *J. kraussii* (2*n* = 40), (c) *J. subsecundus* (2*n* = 40), (d) *J. compressus* (2*n* = 40), (e) *Juncus* cultiv. "Carman Grey" (2*n* = 40), (f) *J. usitatus* (2*n* = 40), (g) *J. micranthus* (2*n* = 40), (h) *J. xiphioides* (2*n* = 40), (i) *J. pincei* (2*n* = 44), (j) *Juncus* cultiv. "Mallorcan Giant" (2*n* = 44), (k) *J. acutus* (2*n* = 46) and (l) *J. cooperi* (2*n* = 46). Yellow sites represent CMA^+ bands. Scale bar = 5µm

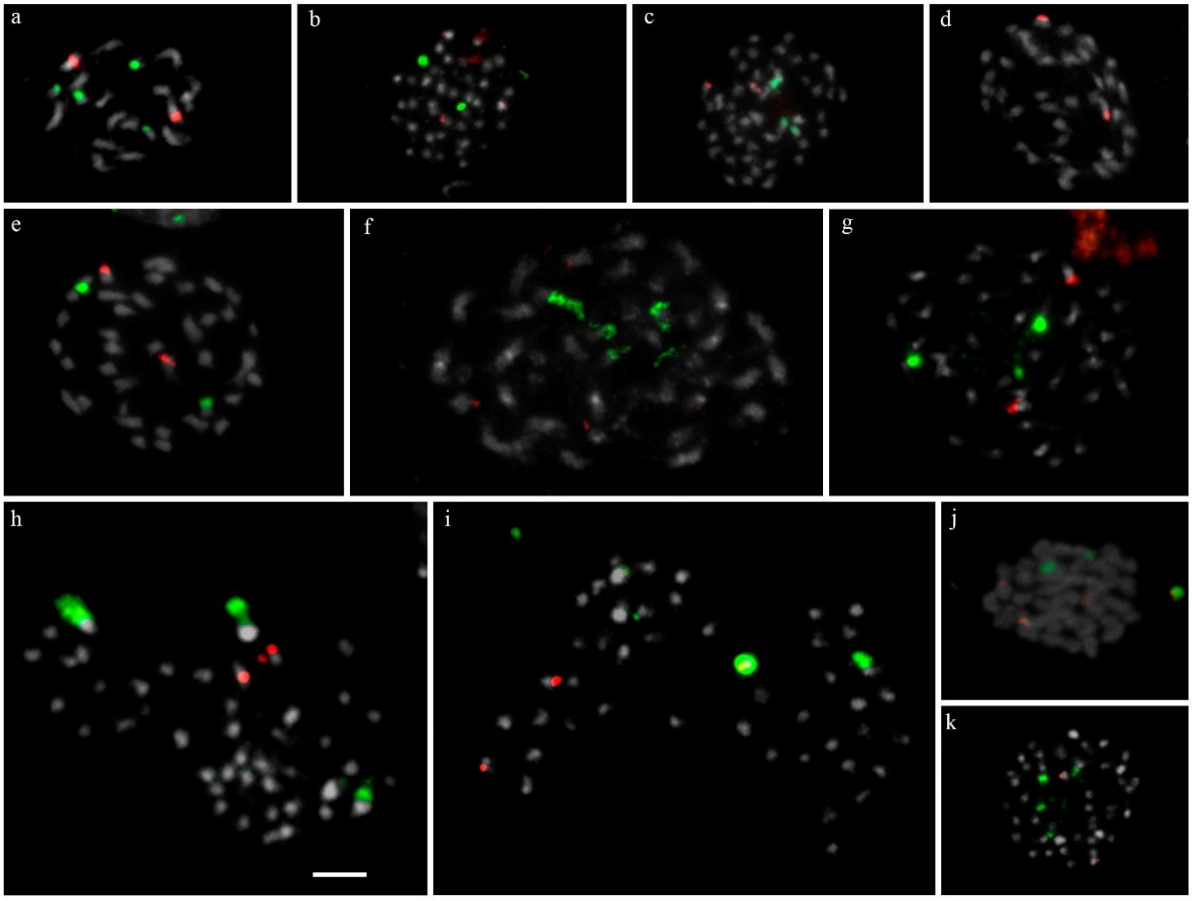


Fig. 2 Hybridization *in situ* fluorescent of *Juncus* species chromosomes with 5S (red) and 35S rDNA sites (green). (a) *J. capitatus*, (b) *J. krausii,* (c) *J. pincei*, (d) *J. usitatus*, (e) *J. xiphioides*, (f) *J. compresus,* (g) *Juncus* cultiv. "Carman Gray", (h) *Juncus* cultiv. "Mallorcan Giant", (i) *J. acutus*, (j) *J. subsecundus*, (k) *J. cooperi.* Scale bar = 5µm

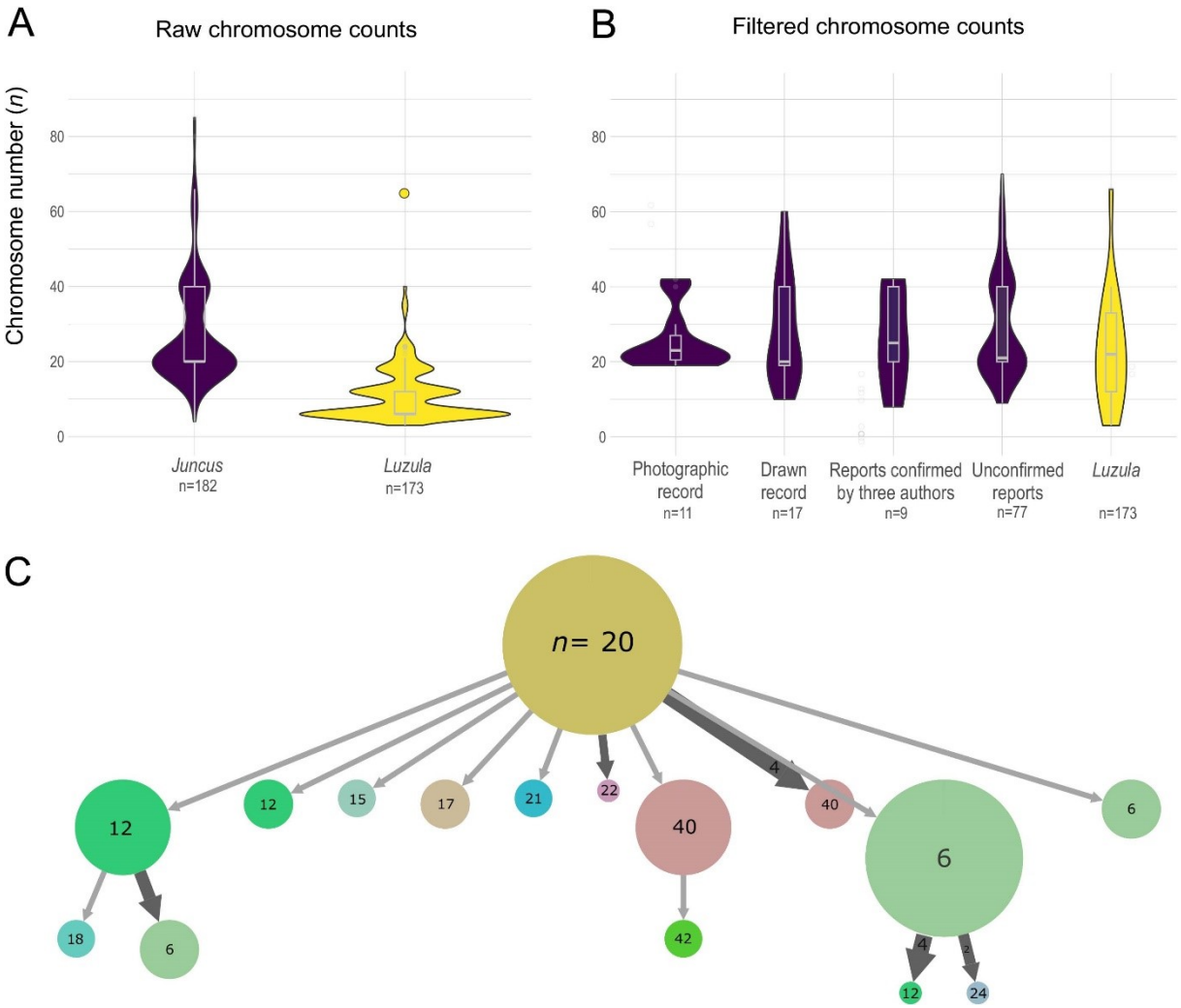


Fig. 3 Compilation of chromosome number variation in Juncaceae. Violin plots showing haploid chromosome number diversity in *Juncus* and *Luzula* species, using raw data (a) and filtered data (b). In the violin plots, the dots indicate the mean values of each group. Core density plots are presented as colored lines delimiting the box plot. For interpretation of the applied filters see materials and methods. (c) Ancestral haploid chromosome frequency produced by PastML using MPPA with a maximum likelihood JOINT+F81 model. Different colors correspond to different chromosome ploidy.

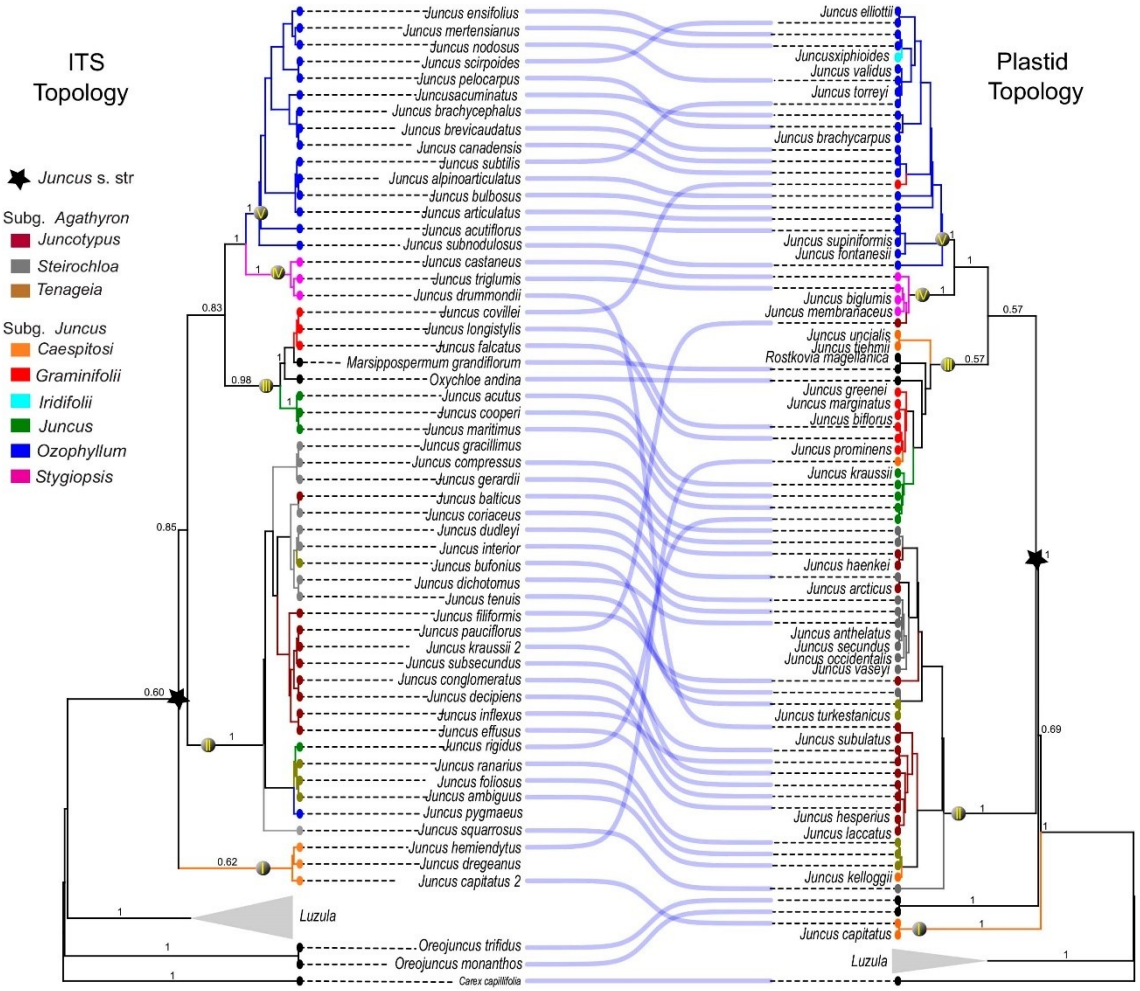


Fig. 4 Comparison of topologies obtained from ITS (left) and plastids (right) data. The number above the branches represents the support of the nodes. Clusters of branches by color represent sections within the genus. The black star designates the clade *Juncus* "sensu stricto". Roman numerals indicate the five clades designated in this study. The blue lines in the center show the correspondence between species.

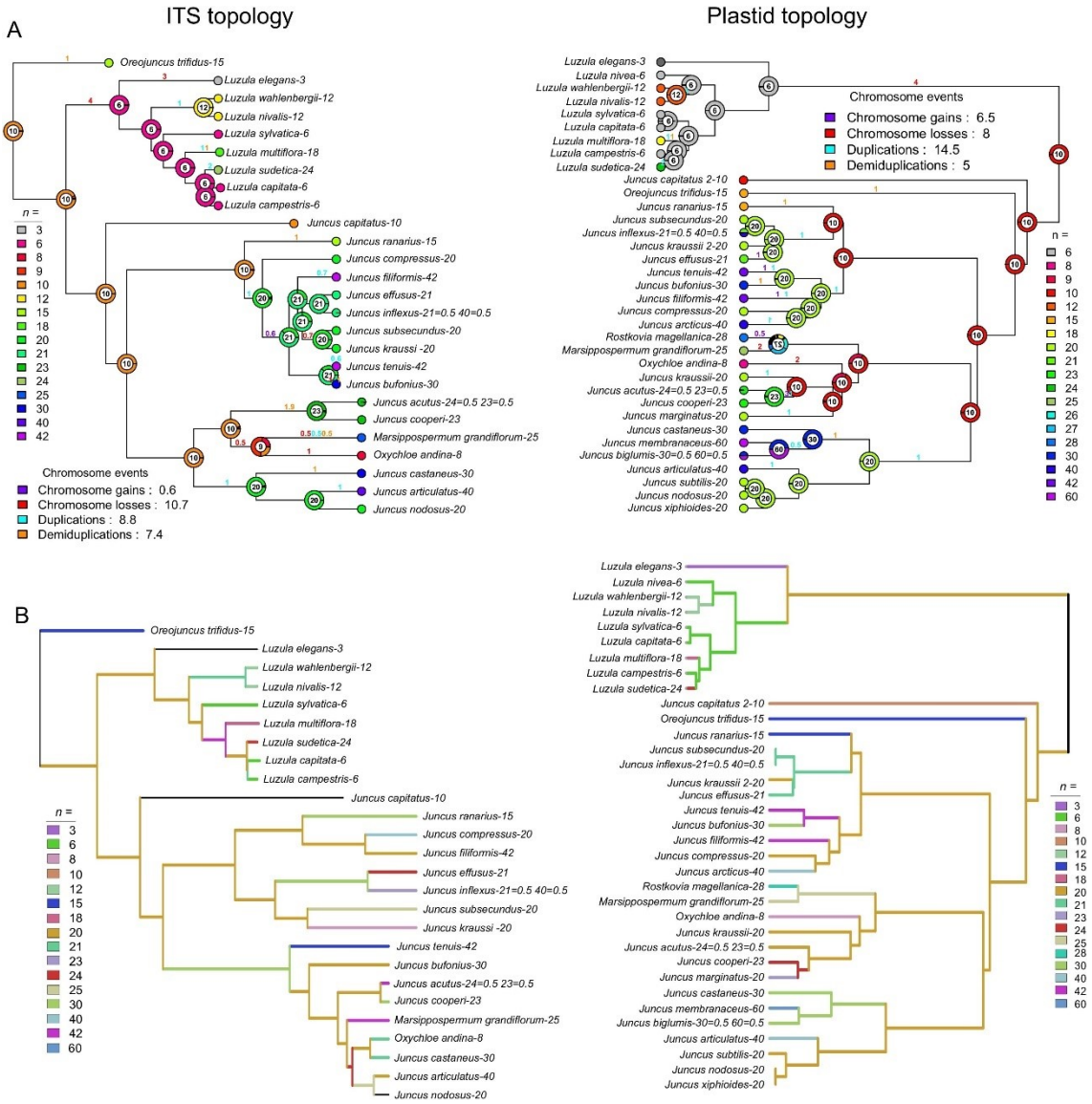


Fig. 5 Ancestral haploid chromosome number reconstruction of Juncaceae using ChromEvol (A) and setting n = 10 in the base along the drop tips Bayesian consensus tree from the ITS (internal transcribed spacer, left) and plastid (atp1, mat*K*, psb*A*-trn*H*, rbc*L*, rpo*B*, rpoC1, and trn*L*-trn*F*, right) regions. Pie charts represent the probabilities of the inferred chromosome numbers, with the ones inside charts having the highest probability. Color coded values above branches represent the inferred frequency of the possible events (fission, fusion or polyploidy) that had a posterior probability higher than 0.5. (B) Ancestral haploid chromosome number reconstruction produced by PastML using MPPA with a maximum likelihood JOINT+F81 model. Different colors correspond to different chromosome ploidy as shown in the tree. The color coding of chromosome numbers and event types is explained in the boxes.

SUPPORTING INFORMATION

Table S1 List and GenBank accession numbers for the Juncaceae family species used in the phylogenetic analysis. Blank spaces represent missing data

Table S2 Karyotype information of *Juncus* species organized in sections according Kirschner (2002) classification. In bold, the species analysed in this study

Table S3 Karyotype information of Juncus taxa used for ancestral chromosome number reconstruction. Each individual was classified according Kirschner (2002) and Brozová et al. (In bold, the species analysed in this study)

Table S4 Statistic of chromosome number variation in Juncaceae. SD = standard deviation. CV = coefficient of variation.

Data	**Media**	**SD**	**CV**
Raw			
Juncus	20	25.9	43.6
Luzula	6	6.77	61.5
Filtered			
Counts with photographic record (a)	20	8.13	31.6
Counts with drawing record (b)	20	14.1	49.5
Counts without photographic record but confirmed (c)	25	12.8	49.3
Counts without photographic record and confirmation (d)	21	13.6	46.6
Luzula	6	6.77	61.5

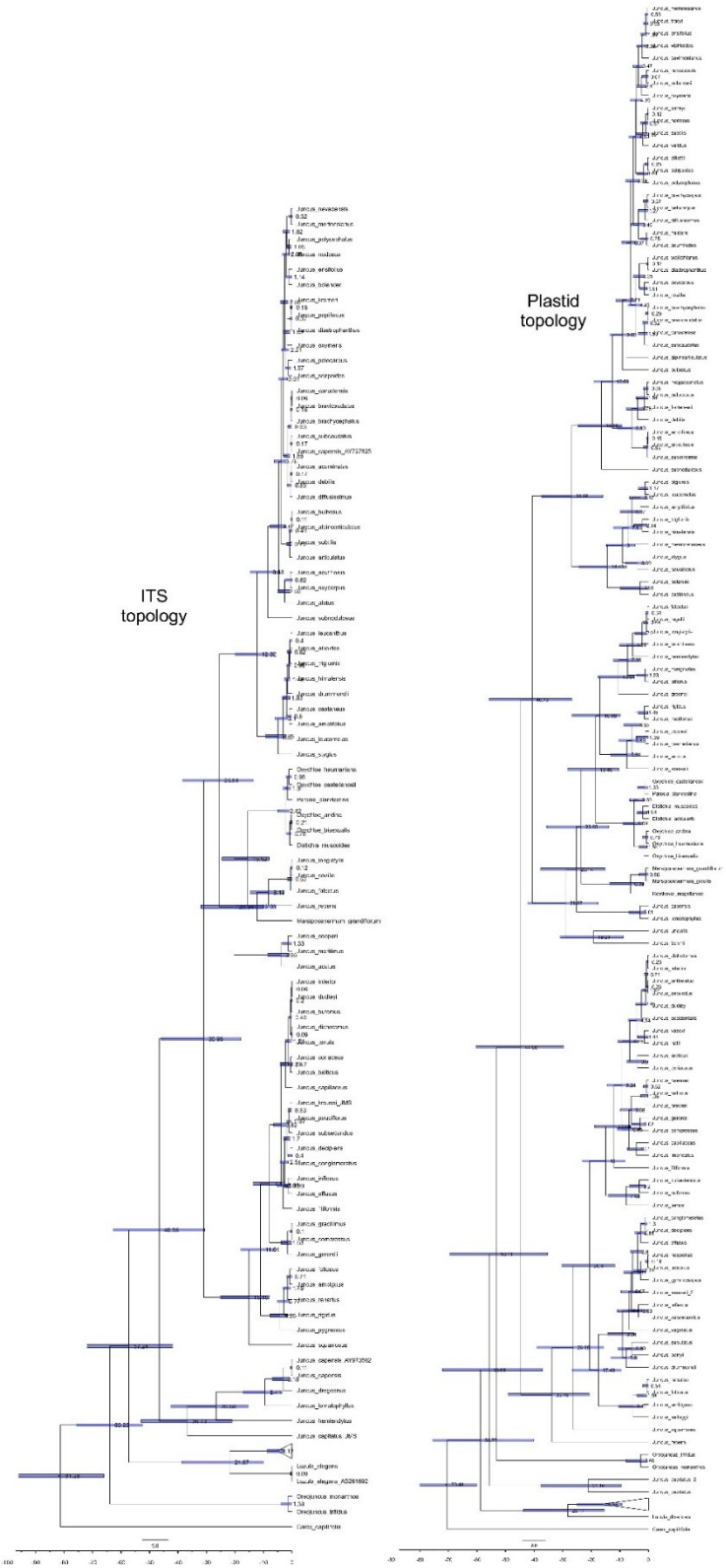


Figure S1 Divergence time of the ITS (internal transcribed spacer, left) and plastid (atp1, mat*K*, psb*A*-trn*H*, rbc*L*, rpo*B*, rpoC1, and trn*L*-trn*F*, right) topologies. The number above the branches represents the age range for each clade

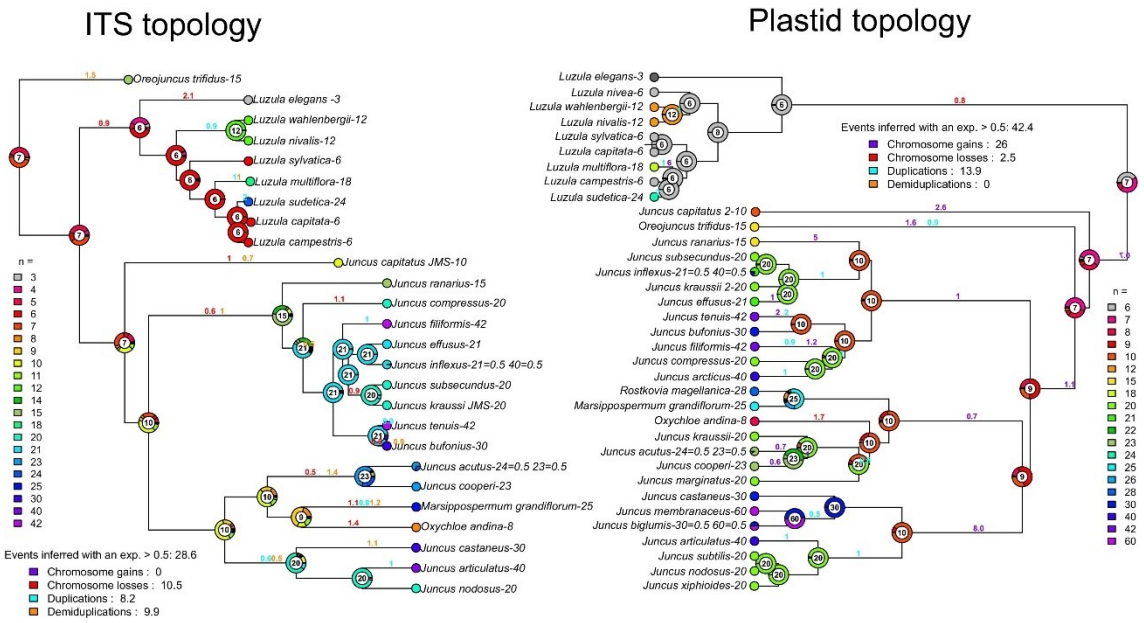


Figure S2 Ancestral haploid chromosome number reconstruction of Juncaceae using ChromEvol, along the Bayesian consensus tree from the ITS (left) and plastid (right) regions. Pie charts represent the probabilities of the inferred chromosome numbers, with the ones inside charts having the highest probability. Color coded values above branches represent the inferred frequency of the possible events (fission, fusion or polyploidy) that had a posterior probability higher than 0.5. The color coding of chromosome numbers and event types is explained in the boxes

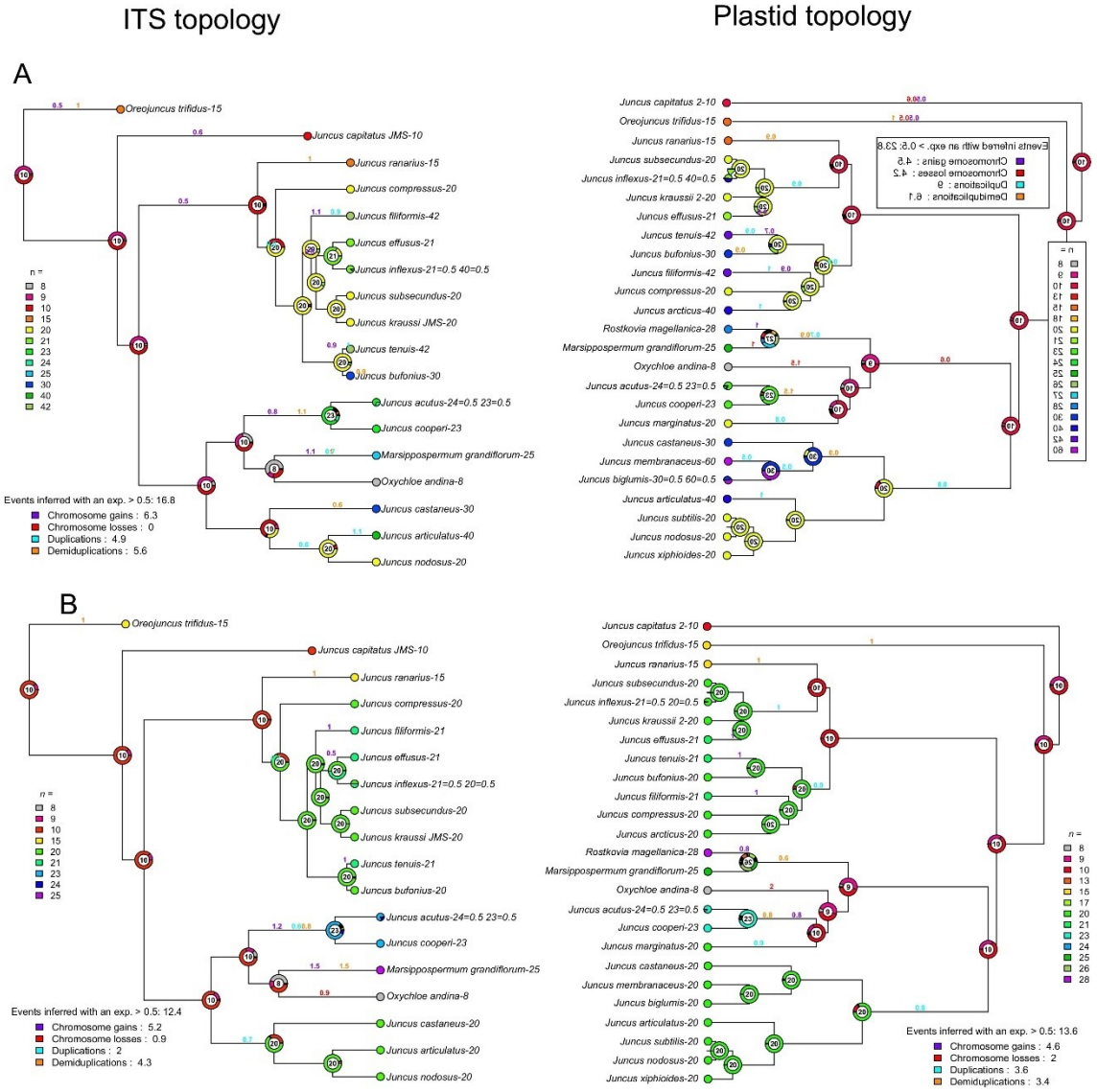


Figure S3 Reconstruction of the ancestral haploid chromosome number in Juncaceae species without holocentric chromosomes (*Luzula* genus) using ChromEvol along the Bayesian consensus tree from the ITS (left) and plastid (right) regions. Dataset including haploid (A) and monoploid (B) chromosome numbers were used. Pie charts represent the probabilities of the inferred chromosome numbers, with the ones inside charts having the highest probability. Color coded values above branches represent the inferred frequency of the possible events (fission, fusion or polyploidy) that had a posterior probability higher than 0.5. The color coding of chromosome numbers and event types is explained in the boxes.

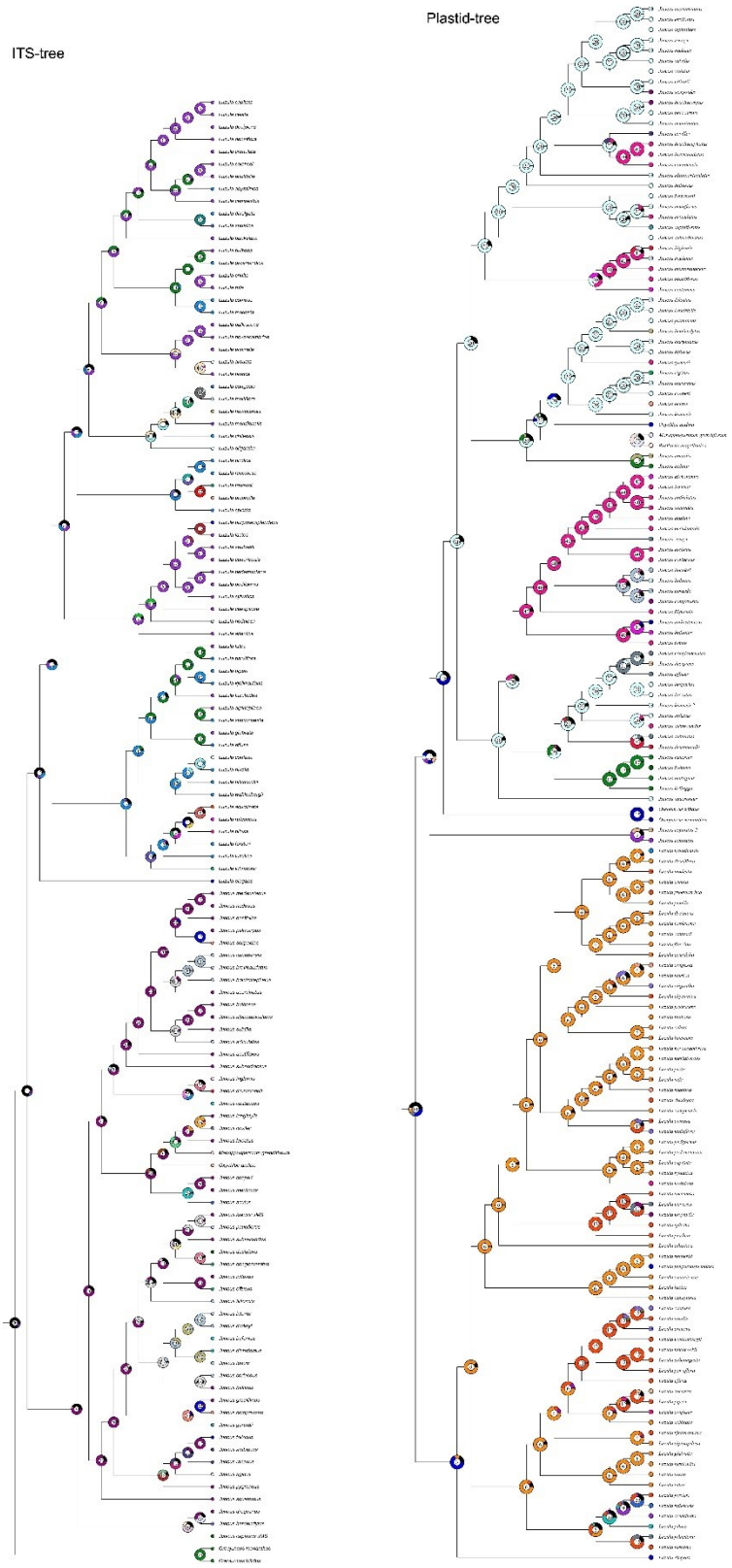


Figure S4 Bayes modelling of the ancestral haploid chromosome number along the ITS (left) and plastid (right) Bayesian phylogeny for Juncaceae. Chromosome number estimation is show inside the pie.

REFERENCES

Baez, M., Kuo, Y. T., Dias, Y., Souza, T., Boudichevskaia, A., Fuchs, J., ... and Houben, A. (2020). Analysis of the small chromosomal Prionium serratum (Cyperid) demonstrates the importance of reliable methods to differentiate between mono-and holocentricity. Chromosoma, 129(3), 285-297.

Balslev, H. (1996). Juncaceae. Flora Neotropica, 1-167.

Baum, D. A., & Oginuma, K. (1994). A review of chromosome numbers in Bombacaceae with new counts for Adansonia. Taxon, 43(1), 11-20.

Beuzenberg, E. J., & Hair, J. B. (1983). Contributions to a chromosome atlas of the New Zealand flora—25 Miscellaneous species. New Zealand journal of botany, 21(1), 13-20.

Bozek, M., Leitch, A. R., Leitch, I. J., Záveská Drábková, L., and Kuta, E. (2012). Chromosome and genome size variation in Luzula (Juncaceae), a genus with holocentric chromosomes. Botanical Journal of the Linnean Society, 170(4), 529-541.

Brighton, C. A., Mathew, B., & Marchant, C. J. (1973). Chromosome counts in the genus Crocus (Iridaceae). Kew Bulletin, 451-464.

Brožová, V., Proćków, J., & Drábková, L. Z. (2022). Toward finally unraveling the phylogenetic relationships of Juncaceae with respect to another cyperid family, Cyperaceae. Molecular Phylogenetics and Evolution, 177, 107588.

Burchardt, P., Buddenhagen, C. E., Gaeta, M. L., Souza, M. D., Marques, A., and Vanzela, A. L. (2020). Holocentric karyotype evolution in rhynchospora is marked by intense numerical, structural, and genome size changes. Frontiers in plant science, 11, 1390.

Bureš, P., Zedek, F., and Marková, M. (2013). Holocentric chromosomes. In Plant genome diversity volume 2 (pp. 187-208). Springer, Vienna.

Carta, A., Bedini, G., and Peruzzi, L. (2020). A deep dive into the ancestral chromosome number and genome size of flowering plants. New Phytologist, 228(3), 1097-1106.

Carvalho, C. R., & Saraiva, L. S. (1993). An air drying technique for maize chromosomes without enzymatic maceration. Biotechnic & histochemistry, 68(3), 142-145.

Choi, Y. M., Choi, B., and Jang, T. S. (2022). New chromosome counts in *Juncus* (Juncaceae) taxa from Korea. Cytologia, 87(3), 221-225.

Christenhusz, M. J., & Byng, J. W. (2016). The number of known plants species in the world and its annual increase. Phytotaxa, 261(3), 201-217.

Clark, J. W., & Donoghue, P. C. (2018). Whole-genome duplication and plant macroevolution. Trends in plant science, 23(10), 933-945.

Clifford, H. T. (1958). On Putative Hybrids between Juncus inflexus L. and Juncus effusus L. Kew Bulletin, 13(3), 392-395.

Cusimano, N. Cope, T. Stadler, S. Renner: A., & Stace, C. A. (1985). Cytology and hybridization in the Juncus bufonius L. aggregate in western Europe. Watsonia.

Dalgaard, V. (1991). Chromosome studies in flowering plants from Macaronesia II. Willdenowia, 139-152.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: more models, new method heuristics and parallel computing. Nature methods, 9(8), 772-772.

Darlington, C. D., & Wylie, A. P. (1956). Chromosome atlas of flowering plants. Chromosome atlas of flowering plants., (2nd Ed).

De Assis, F. N. M., Souza, B. C. Q., Medeiros-Neto, E., Pinheiro, F., Silva, A. E. B., & Felix, L. P. (2013). Karyology of the genus Epidendrum (Orchidaceae: Laeliinae) with emphasis on subgenus Amphiglottium and chromosome number variability in Epidendrum secundum. Botanical Journal of the Linnean Society, 172(3), 329-344.

Drábková, L. (2013). A survey of karyological phenomena in the Juncaceae with emphasis on chromosome number variation and evolution. The Botanical Review, 79, 401-446.

Drábková, L., & Kirschner, J. (2013). Oreojuncus, a new genus in the Juncaceae. Preslia, 85, 483-503.

Drabkova, L., Kirschner, J., & Vlček, Č. (2006). Phylogenetic relationships within Luzula DC. and Juncus L.(Juncaceae): A comparison of phylogenetic signals of trnL-trnF intergenic spacer, trnL intron and rbcL plastome sequence data. Cladistics, 22(2), 132-143.

Drori, M., Rice, A., Einhorn, M., Chay, O., Glick, L., & Mayrose, I. (2018). OneTwoTree: An online tool for handling missing species phylogeny reconstruction. Molecular Ecology Resources, 18(6), 1492-1499.

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., & Rambaut, A. (2006). Relaxed phylogenetics and dating with confidence. PLoS biology, 4(5), e88.

Drummond, A. J., & Rambaut, A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. BMC evolutionary biology, 7(1), 1-8.

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. Molecular biology and evolution, 29(8), 1969-1973.

Erixon, P., Svennblad, B., Britton, T., & Oxelman, B. (2003). Reliability of Bayesian posterior probabilities and bootstrap frequencies in phylogenetics. Systematic biology, 52(5), 665-673.

Escudero, M., & Hipp, A. (2013). Shifts in diversification analysis applicable to randomly or non-randomly sampled phylogeniesrates and clade ages explain species richness in higher-level sedge taxa (Cyperaceae). American Journal of Botany, 100(12), 2403-2411.. Syst. Biol. (2012) 61: 785 – 792

Gerlach WL, Bedbrook JR. Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res. 1979 Dec 11;7(7):1869-85. doi: 10.1093/nar/7.7.1869.

Gernhard, T. (2008). New analytic results for speciation times in neutral models. Bulletin of mathematical biology, 70, 1082-1097.

Glick, L., & Mayrose, I. (2014). ChromEvol: assessing the pattern of chromosome number evolution and the inference of polyploidy along a phylogeny. Molecular biology and evolution, 31(7), 1914-1922.

Guerra, M. S. (1986). Reviewing the chromosome nomenclature of Levan et al. Brazilian Journal of genetics, 9, 741-743.

Guerra, M. (2008). Chromosome numbers in plant cytotaxonomy: concepts and implications. Cytogenetic and Genome research, 120(3-4), 339-350.

Guerra, M., Pedrosa, A., Cornélio, M. T. M., Santos, K., & Soares Filho, W. D. S. (1997). Chromosome number and secondary constriction variation in 51 accessions of a Citrus germplasm bank. Brazilian Journal of Genetics, 20, 489-496.

Guerra, M., Ribeiro, T., & Felix, L. P. (2019). Monocentric chromosomes in Juncus (Juncaceae) and implications for the chromosome evolution of the family. Botanical Journal of the Linnean Society, 191(4), 475-483.

Hofstatter, P. G., Thangavel, G., Lux, T., Neumann, P., Vondrak, T., Novak, P., ... & Marques, A. (2022). Repeat-based holocentromeres influence genome architecture and karyotype evolution. Cell, 185(17), 3153-3168.

Ishikawa SA, Zhukova A, Iwasaki W, Gascuel O (2019) A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. Molecular Biology and Evolution, msz131.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Molecular biology and evolution, 30(4), 772-780.

Kaur, H., Mubarik, N., Kumari, S., and Gupta, R. C. (2014). Chromosome numbers and basic chromosome numbers in monocotyledonous genera of the Western Himalayas (India). Acta Biologica Cracoviensia. Series Botanica, 56(2).

Kirschner J. (ed). 2002. Juncaceae 2: *Juncus* subgenus *Juncus*. – In: Orchard A. E. (ed.), Species plantarum: Flora of the World. Part 7: 1–336. – ABRS, Canberra, Australia. Doi: 10.5962/bhl.title.14705

Kolodin, P., Cempírková, H., Bureš, P., Horová, L., Veleba, A., Francová, J., ... and Zedek, F. (2018). Holocentric chromosomes may be an apomorphy of Droseraceae. Plant Systematics and Evolution, 304(10), 1289-1296.

Harriman, N. A., & Redmond, D. (1976). Somatic chromosome numbers for some North American species of Juncus L. Rhodora, 78(816), 727-738.

Heckmann, S., Schroeder-Reiter, E., Kumke, K., Ma, L., Nagaki, K., Murata, M., ... and Houben, A. (2011). Holocentric chromosomes of Luzula elegans are characterized by a longitudinal centromere groove, chromosome bending, and a terminal nucleolus organizer region. Cytogenetic and genome research, 134(3), 220-228.

Heckmann, S., Macas, J., Kumke, K., Fuchs, J., Schubert, V., Ma, L., ... & Houben, A. (2013). The holocentric species L uzula elegans shows interplay between centromere and large-scale genome organization. The Plant Journal, 73(4), 555-565.

Hipp, A. L., Rothrock, P. E., & Roalson, E. H. (2009). The evolution of chromosome arrangements in Carex (Cyperaceae). The Botanical Review, 75, 96-109.

Ibiapino, A., Báez, M., García, M. A., Costea, M., Stefanović, S., & Pedrosa-Harand, A. (2022). Karyotype asymmetry in Cuscuta L. subgenus Pachystigma reflects its repeat DNA composition. Chromosome Research, 30(1), 91-107.

Ibiapino, A., García, M. Á., Costea, M., Stefanović, S., & 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. Genetics and Molecular Biology, 43.

Ishikawa SA, Zhukova A, Iwasaki W, Gascuel O (2019) A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. Molecular Biology and Evolution, msz131.

Jones, E., Simpson, D. A., Hodkinson, T. R., Chase, M. W., & Parnell, J. A. (2007). The Juncaceae-Cyperaceae interface: a combined plastid sequence analysis. Aliso: A Journal of Systematic and Floristic Botany, 23(1), 55-61.

Jorgensen, C. A. (1958). The flowering plants of Greenland. A taxonomical and cytological survey. K. Dan. Vidensk. Selsk. Skrift., 9, 1-172.

Lewis, H. (1951). The origin of supernumerary chromosomes in natural populations of Clarkia elegans. Evolution, 142-157.

Love, A. & Love, D. (1944). Cytotaxonomical studies on boreal plants. II. Some notes on the chromosome numbers of Juncaceae. Ark. Bot. 31(1): 1-6.

Madej, A., & Kuta, E. (2001). Holokinetic chromosomes of Luzula luzuloides (Juncaceae) in callus culture. Acta Biologica Cracoviensia. Series Botanica, 43.

Mandrioli, M., and Manicardi, G. C. (2020). Holocentric chromosomes. PLoS genetics, 16(7), e1008918.

Marinho, R. C., Mendes-Rodrigues, C., Balao, F., Ortiz, P. L., Yamagishi-Costa, J., Bonetti, A. M., & Oliveira, P. E. (2014). Do chromosome numbers reflect phylogeny? New counts for Bombacoideae and a review of Malvaceae sl. American journal of botany, 101(9), 1456-1465.

Márquez-Corro, J. I., Martín-Bravo, S., Spalink, D., Luceño, M., & 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.

Mata-Sucre, Y., Costa, L., Gagnon, E., Lewis, G. P., Leitch, I. J., & Souza, G. (2020). Revisiting the cytomolecular evolution of the Caesalpinia group (Leguminosae): a broad sampling reveals new correlations between cytogenetic and environmental variables. Plant Systematics and Evolution, 306, 1-13.

Mata-Sucre, Y., Matzenauer, W., Castro, N. M. S., Huettel, B., Pedrosa-Harand, A., Marques, A., & Souza, G. Repeat-Based Phylogenomics Resolves Section-Level Classification within the Monocentric Genus Juncus L.(Juncaceae). biorivix

Mayrose, I., Barker, M. S., & Otto, S. P. (2010). Probabilistic models of chromosome number evolution and the inference of polyploidy. Systematic biology, 59(2), 132-144.

Mayrose, I., and Lysak, M. A. (2021). The evolution of chromosome numbers: mechanistic models and experimental approaches. Genome Biology and Evolution, 13(2), evaa220.

Merxmüller, H. (1970). Provocation of biosystematics. Taxon, 19(2), 140-145.

McKain, M. R., Tang, H., McNeal, J. R., Ayyampalayam, S., Davis, J. I., Depamphilis, C. W., ... & Leebens-Mack, J. H. (2016). A phylogenomic assessment of ancient polyploidy and genome evolution across the Poales. Genome biology and evolution, 8(4), 1150-1164.

Mehra, P. N., & Sachdeva, S. K. (1976). Cytological observations on some W. Himalayan monocots. II. Smilacaceae, Liliaceae and Trilliaceae. Cytologia, 41, 5-22.

Melters, D. P., Paliulis, L. V., Korf, I. F., and Chan, S. W. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. Chromosome Research, 20(5), 579-593.

Merxmüller, H. (1970). Provocation of biosystematics. Taxon, 19(2), 140-145.

Mičieta, K. (1983). Contribution to the chromosome numbers of some species of the genus Juncus L. in Slovakia. Folia geobotanica & phytotaxonomica, 18, 195-198.

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular biology and evolution, 32(1), 268-274.

O'Mahony, T. (2002). The comparative morphology of Juncus conglomeratus L. Compact Rush), J. effusus L. (Soft-rush) and their interspecific hybrid, IRISH BOTANICAL 9, 5-14.

Pagel M., Meade, A., and Barker, D. (2004). Bayesian estimation of ancestral character states on phylogenies. Syst. Biol. 53, 673–684.

Paradis, E., Blomberg, S., Bolker, B., Brown, J., Claude, J., Cuong, H. S., and Desper, R. (2019). Package 'ape'. Analyses of phylogenetics and evolution, version, 2(4), 47.

Pirie, M. D. (2015). Phylogenies from concatenated data: Is the end nigh?. Taxon, 64(3), 421-423.

POWO (2023). "Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; http://www.plantsoftheworldonline.org/Retrieved 13 September 2023."

Rambaut, A., and Drummond, A. J. (2010). TreeAnnotator. version 1.6. 1 http://beast. bio. ed. ac. uk.

Rambaut, A., Drummond, A. J., Xie, D., Baele, G., & Suchard, M. A. (2018). Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic biology, 67(5), 901-904.

Ramirez-Barahona, S., Sauquet, H., & Magallon, S. (2020). The delayed and geographically heterogeneous diversification of flowering plant families. Nature Ecology & Evolution, 4(9), 1232-1238.

Ribeiro, T., Buddenhagen, C. E., Thomas, W. W., Souza, G., & Pedrosa-Harand, A. (2018). Are holocentrics doomed to change? Limited chromosome number variation in Rhynchospora Vahl (Cyperaceae). Protoplasma, 255, 263-272.

Rice, A., Glick, L., Abadi, S., Einhorn, M., Kopelman, N. M., Salman-Minkov, A., ... & Mayrose, I. (2015). The Chromosome Counts Database (CCDB)–a community resource of plant chromosome numbers. New Phytologist, 206(1), 19-26.

Rice, A., & Mayrose, I. (2021). Model adequacy tests for probabilistic models of chromosome-number evolution. New Phytologist, 229(6), 3602-3613.

Roalson, E. H. (2005). Phylogenetic relationships in the Juncaceae inferred from nuclear ribosomal DNA internal transcribed spacer sequence data. International journal of plant sciences, 166(3), 397-413.

Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Höhna, S., ... & Huelsenbeck, J. P. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic biology, 61(3), 539-542.

Sader, M. A., Amorim, B. S., Costa, L., Souza, G., & Pedrosa-Harand, A. (2019). The role of chromosome changes in the diversification of Passiflora L.(Passifloraceae). Systematics and Biodiversity, 17(1), 7-21.

Schönswetter, P., Suda, J., Popp, M., Weiss-Schneeweiss, H., and Brochmann, C. (2007). Circumpolar phylogeography of Juncus biglumis (Juncaceae) inferred from AFLP fingerprints, cpDNA sequences, nuclear DNA content and chromosome numbers. Molecular Phylogenetics and Evolution, 42(1), 92-103.

Schubert, V., Neumann, P., Marques, A., Heckmann, S., Macas, J., Pedrosa-Harand, A., ... and Houben, A. (2020). Super-resolution microscopy reveals diversity of plant centromere architecture. International journal of molecular sciences, 21(10), 3488.

Senaratne, A. P., Cortes-Silva, N., & Drinnenberg, I. A. (2022, July). Evolution of holocentric chromosomes: drivers, diversity, and deterrents. In Seminars in Cell & Developmental Biology (Vol. 127, pp. 90-99). Academic Press.

Smith, P. H. (2006). Revisiting Juncus balticus Willd. in England. WATSONIA-KINGS LYNN-BOTANICAL SOCIETY OF THE BRITISH ISLES-, 26(1), 57.

Snogerup, S. (1963). Studies in the genus Juncus III. Observations on the diversity of chromosome numbers. Bot. Not., 116, 142-156.

Stace, C. A. (2020). Hybrids in Juncus section Juncotypus, with a description of J. x lancastriensis (Juncaceae). British and Irish Botany, 2(4), 266-284.

Van-Lume, B., Mata-Sucre, Y., Báez, M., Ribeiro, T., Huettel, B., Gagnon, E., ... & Souza, G. (2019). Evolutionary convergence or homology? Comparative cytogenomics of Caesalpinia group species (Leguminosae) reveals diversification in the pericentromeric heterochromatic composition. Planta, 250, 2173-2186.

Waminal NE, Pellerin RJ, Kim NS, Jayakodi M, Park JY, Yang TJ, Kim HH. Rapid and Efficient FISH using Pre-Labeled Oligomer Probes. Sci Rep. 2018 May 29;8(1):8224. doi: 10.1038/s41598-018-26667-z.

Wilcox, M. (2010). A novel approach to the determination and identification of Juncus× diffusus Hoppe and J.× kern-reichgeltii Jansen and Wacht. ex Reichg. Watsonia, 28(1), 43.

Wilcox, M. (2011). Hybrid rushes in the UK–sterility and fertility. Deadline for contribution to News 117... 91, 21.

Yu Y, Blair C, He XJ. 2020. RASP 4: Ancestral State Reconstruction Tool for Multiple Genes and Characters. Molecular Biology and Evolution. 37(2):604-606

Zedek, F., & Bureš, P. (2016). Absence of positive selection on CenH3 in Luzula suggests that holokinetic chromosomes may suppress centromere drive. Annals of Botany, 118(7), 1347-1352.

4 ARTIGO 2- Repeat-based holocentromers of *Luzula sylvatica* (Juncaceae) reveal new insights into the evolutionary mono- to holocentric transition

Yennifer Mata-Sucre[1,2#], Marie Kratka[3,4#], Ludmila Oliveira[5], Pavel Neumann[5], Jiri Macas[5], Veit Schubert[6], Andreas Houben[6], Andrea Pedrosa-Harand[2], Gustavo Souza[2], André Marques[1*]

Para ser submetido a Nature Comunications Qualis A1

Repeat-based holocentromers of *Luzula sylvatica* (Juncaceae) reveal new insights into the evolutionary mono- to holocentric transition

Yennifer Mata-Sucre[1,2#], Marie Kratka[3,4#], Ludmila Oliveira[5], Pavel Neumann[5], Jiri Macas[5], Veit Schubert[6], Andreas Houben[6], Andrea Pedrosa-Harand[2], Gustavo Souza[2], André Marques[1*]

[1]Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany

[2]Laboratório de Citogenética e Evolução Vegetal, Departamento de Botânica, Centro de Biociências, Universidade Federal de Pernambuco, Recife PE 50670-901 Brazil

[3]Department of Plant Developmental Genetics, Institute of Biophysics of the Czech Academy of Sciences, Kralovopolska 135, 61200, Brno, Czech Republic

[4]National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Kamenice 5, 625 00 Brno, Czech Republic

[5]Biology Centre, Czech Academy of Sciences, Institute of Plant Molecular Biology, České Budějovice, Czech Republic

[6]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Seeland, Germany

Abstract

Although the centromere is restricted to a single region of the chromosome in most studied eukaryotes, members of the plant family Juncaceae harbor either monocentric (*Juncus)* or holocentric (*Luzula*) chromosomes. This provides an opportunity to study the evolutionary mechanisms involved in the transition to holocentricity. Here, combining genome assembling at chromosome-scale, ChIP-seq for CENH3 and histone modifications, immuno-FISH, and super-resolution microscopy, we report the occurrence of repeat-based holocentromeres in *L.*

sylvatica. An irregular distribution of genes, centromeric units, and most repeats was found along the chromosomes. Centromere function was predominantly associated with the 124 bp-long monomers of the satellite DNA repeat *Lusy1*, with 0.76 arrays/Mb, while other satellite repeats were mostly CENH3-negative. Comparative repeatome analysis of 13 species from the genus revealed that *Lusy1* is abundant in all *Luzula* species, except in the first diverging lineage *L. elegans*, which has repeat-free holocentromeres. Synteny between *L. sylvatica* (*n* = 6) and *J. effusus* (*n* = 21) genomes revealed a likely chromosome number reduction in *Luzula* derived from multiple chromosome fusions of ancestral *J. effusus*-like chromosomes. We propose that multiple fusions of small monocentric chromosomes and subsequent maintenance/ expansion of centromeric domains were followed by colonization with centromeric repeats possibly contributing to the transition to holocentricity.

Key-words: CENH3, ChIP-seq, chromosome fusion, genome assembly, holocentricity, karyotype evolution, satellite DNA

Introduction

Centromeres are specialized chromosomal regions that recruit proteins from the kinetochore and mediate spindle microtubule binding to ensure correct chromosome segregation during mitosis and meiosis (Schubert et al., 2020; Talbert & Henikoff 2020). Most taxonomic groups have chromosomes with a size-restricted centromeric domain confined to the primary constriction, i.e., monocentric (Wong et al., 2020). However, holocentric chromosomes lack a primary constriction and exhibit molecular and epigenetic features that allow centromeric proteins and microtubules to bind extensively along the chromosomes (Heckmann et al., 2013; Schubert et al., 2020; Hofstatter et al., 2022). Remarkably, holocentric chromosomes have evolved repeatedly in animals and plants (Melters et al., 2012; Mandrioli & Manicardi 2020; Senaratne et al., 2021). Consequently, large-scale rearrangements by chromosome fusions and fissions are more likely to be tolerated, as the rearranged chromosomes can maintain kinetochore activity without suffering from segregation problems during cell divisions (Burchardt et al., 2020; Lucek et al., 2022; Escudero et al., 2023).

The lack of conclusive evidence pointing to reversions to monocentricity in any eukaryotic lineage and the sporadic distribution of holocentric versus monocentric organisms

support the unidirectional transition to holocentricity (Melters et al., 2012; Escudero et al., 2016; Senaratne et al., 2021). Numerous evolutionary models and drivers have been proposed to explain the emergence of holocentricity from monocentric ancestors, which include alterations/loss/emergence of protein, genes or repetitive sequences during the process (Drinnenberg et al., 2014; Senaratne et al., 2021; Neuman et al., 2023). The causes of the transitions, however, remain unclear, mainly because only few holocentric species have been studied and because most holocentric groups evolved a long time ago, making the factors involved in the transition difficult to determine (Neumann et al., 2023).

In plants, centromeres are epigenetically specified by the centromeric histone H3 variant (CENH3), the hallmark of a functional centromere. CENH3 domains in monocentrism are typically associated with extended arrays of tandemly repeated sequences (satellite DNA), which are usually highly divergent and fast evolving (Allshire & Karpen 2008; Plohl et al., 2014; Šatović-Vukšić & Plohl 2023). Although the role of these repeats in centromere function has not yet been fully elucidated, several possible advantages of centromeric repeats have been proposed (Naughton & Gilbert 2020). Satellites might have favorable monomer length stabilizing CENH3 nucleosome positioning or contain specific sequences, such as short dyad symmetries, forming non-B-DNA structures and thus aiding CENH3 nucleosome loading (Kasinathan & Henikoff, 2018; Talbert & Henikoff 2020).

Though repeat sequences function in centromeres, only three holocentric genera with centromeric repeats were characterized so far. *Rhynchospora* Vahl. (Cyperaceae) holocentromeres are mainly composed of a 172-bp satellite called *Tyba*, evenly distributed along the chromosomes in ~20 kb domains and specifically colocalizing with the centromere-determining protein CENH3 (Marques et al., 2015; Hofstatter et al., 2022). In *Chionographis japonica* (Willd.) Maxim. (Melanthiaceaenthiaceae), few large (~2Mb) CENH3-positive domains are associated with satellite arrays of 23- and 28-bp-long monomers (Kuo et al., 2023). Similarly, the holocentric *Meloidogyne* root-knot nematodes show conservation of a 19-bp DNA satellite due to its role in centromeric function (Despot-Slade et al., 2021). However, in Juncaceae Juss., the sister family of Cyperaceae, none of the several families of tandem repeats from the holocentric *Luzula elegans* Lowe showed specific association to its holocentromeres (Heckmann et al., 2013). This would suggest that the presence of a specific centromeric satellite may not be necessary for the establishment of holocentricity in this group, although in-depth genomic study may provide further insight into the role of satellites in the centromere transition.

Juncaceae is a cosmopolitan family comprising ~474 species (POWO 2023). *Juncus* L. and *Luzula* DC. represent the largest genera in the family with 332 and 124 species, respectively (Drábková 2006; POWO 2023). An interesting feature of this family is its variation in centromeric organization and chromosomal structure, making it an ideal model to address hypotheses about evolutionary patterns during centromere transition. Although historically the entire Juncaceae family was thought to be holocentric, cytogenetic and genomic studies revealed that the *Juncus* genus is monocentric (Guerra et at., 2019; Hofstatter et al., 2022; Mata-Sucre et al., 2023; Dias et al., in revision). Chromosome-scale genome assembly revealed that *J. effusus* has repeat-based, CENH3-associated monocentromers, consisting mainly of two tandem repeat families underlying one or up to three spaced cores of CENH3-enriched genomic regions (Hofstatter et al., 2022; Dias et al., in revision.). On the other hand, *Luzula* genomes have been up till now characterized as holocentric without specific centromeric repeats (Kuta et al., 2004; Bozek et al., 2012; Heckmann et al., 2011,2013; Schubert et al., 2020), since the lack of a reference genome has hindered detailed studies on *Luzula* centromeres.

Here we performed a comprehensive (epi)genomic characterization of the chromosome-scale genome of *L. sylvatica*, focusing on its holocentromere organization, repetitive fraction and genome evolution. Comparative genomic repeat profiles of 13 *Luzula* species revealed the conservation of repeat-based holocentromeres in the genus, except for *L. elegans*. We show that *L. sylvatica* has a unique repeat-based centromere organization distinct from previously described holocentromere models. Further comparative genomic analyses between *J. effusus* and *L. sylvatica* revealed footprints of extensive chromosomal fusions that potentially played an important role in the transition to holocentricity in this lineage. Our findings suggest an interaction between the centromere and large-scale genome organization, with implications for genome evolution and centromere biology.

Results

Chromosome-scale genome assembly reveals repeat-based holocentromeres of *L. sylvatica*

We estimated a genome size of 1C = 476 Mb for *L. sylvatica* based on k-mer frequencies (**Fig. S1**), and assembled a chromosome-scale reference genome sequence integrating PacBio HiFi reads and a Hi-C chromatin interaction dataset available in www.darwintreeoflife.org (**Fig. 1, Fig. S1**). The *de novo* genome assembly of *L. sylvatica* generated 1,010 contigs totaling 516.08 Mb with a GC content of 33.01 %, N50 of 7.6 Mb, and BUSCO completeness of 93.01% (**Fig.**

1a-b, Table S1). Six pseudomolecules were obtained by Hi-C scaffolding, with a total of 468.44 Mb and an N50 of 78.95 Mb (**Fig. 1c, Table S1**). The assembly was annotated with respect to major genomic sequence types, including genes, tandem repeats and transposable elements (**Fig 1d, Fig. S2**). A dispersed but structurally heterogeneous distribution of sequences along the pseudomolecules were observed, with interstitial regions highly enriched by tandem repeats and lacking genes and transposable elements (**Fig 1d, Fig. S2**).

To identify and characterize the centromeres, as well as eu- and heterochromatin regions on *L. sylvatica* genome, we performed CENH3, and H3K4me3 and H3K9me2 ChIP-seq, respectively (**see Online Methods, Fig. 1d-e**). We detected 358 CENH3-binding regions (hereafter CENH3 domains) distributed across the entire length of all chromosomes (**Fig. 1d-e, S2**). Considering that one CENH3 domain is equivalent to one centromeric unit, we observed an average of 0.76 units/Mb (range 0.64-0.90 units/Mb) per chromosome with average unit length of 183 kb (range 174-197 kb; **Fig. 1e**). Additionally, histone modification marks H3K4me3 and H3K9me2 were observed intermixed along the chromosomes even near the subtelomeric and central chromosomal regions (**Fig. 1d, Fig. S2**).

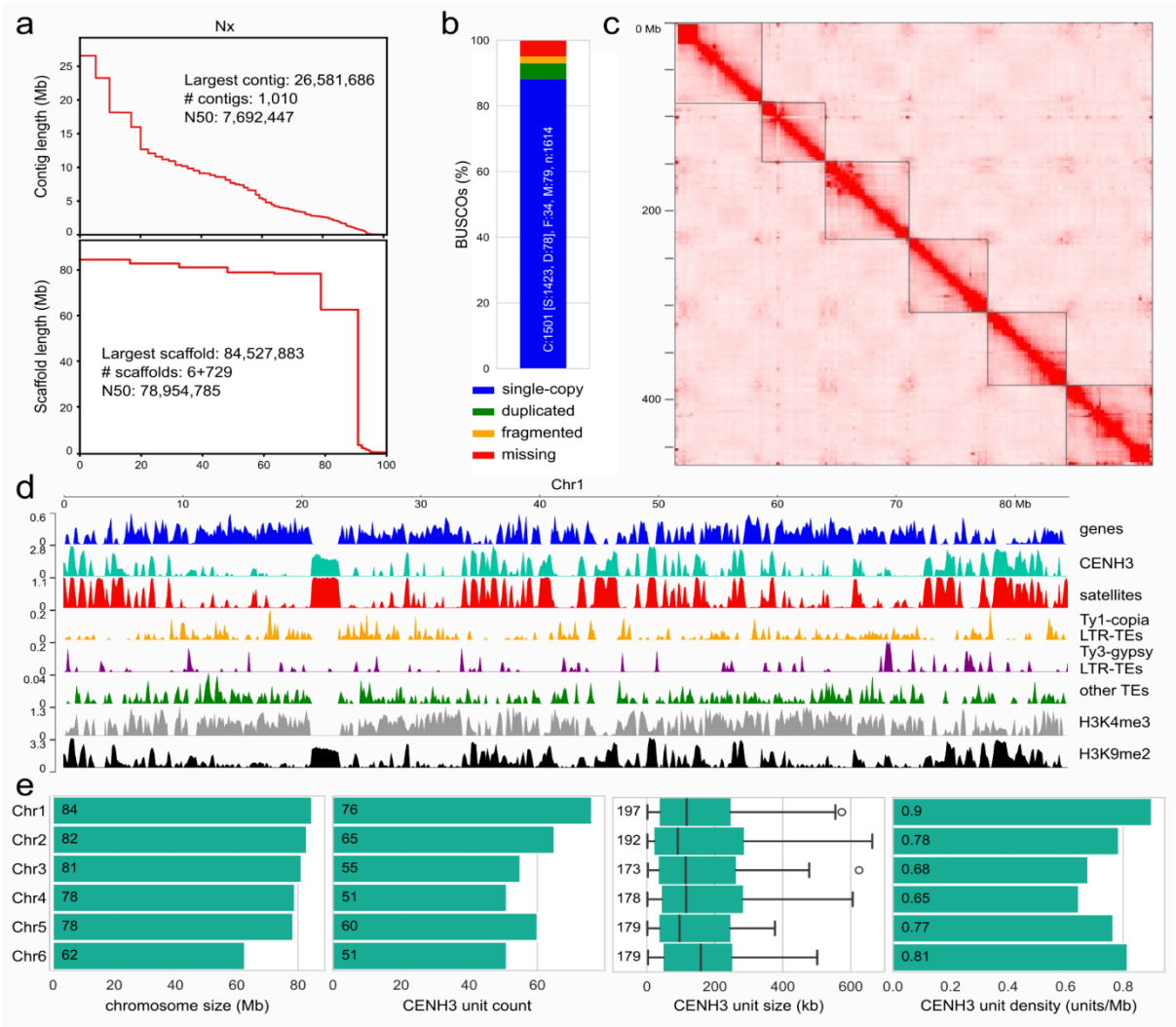


Fig. 1: Genome assembly and annotation of *L. sylvatica*. **(a)** Statistics of the *L. sylvatica* genome assembly (top) and the final scaffolding (bottom). **(b)** BUSCO assessment for completeness of genic space with the viridiplantae_odb10 dataset, using the entire genome assembly. **(c)** Intra-chromosomal contact matrices. Dark squares mark chromosome boundaries. **(d)** *L. sylvatica* Chr1 detailed view showing dispersed distribution of main genomic features: CENH3, gene, tandem repeat, dispersed repeat and eu-heterochromatin marks densities, as typical for holocentric chromosomes. Bin sizes of 100 kb. Distribution of features on all chromosomes is reported in Fig. S2. **(e)** Size of chromosomes, CENH3 units (only values <700 kb shown), number of CENH3 units per chromosome and their densities.

Repeat annotation was based on 175 clustered elements obtained by RepeatExplorer2 representing ~48% of the genome (**Table 1**). Most of these corresponded to satellite DNA sequences with seven families representing 35.31% of the genome, whereas the CL1 and CL2 clusters correspond to the most abundant satellite DNAs with 25.10 % and 7.06 %, respectively (**Table 1 and Fig. S3**). CL1 is a 124-bp satellite, named hereafter as *Lusy1* (**Table S2 and Fig.**

S3b). CL2 is a satellite consisting of two variants of 174 and 175 bp with a 62% similarity, and with 30% similarity to *Lusy1* (hereafter referred to as *Lusy2*; **Table S2 and Fig. S3**). The other five satellite DNAs have monomers with 31 to 197 bp, amounting to less than ~2 % in the genome each (**Fig. 2a and Table S2**). Retrotransposon elements were less abundant than satellites, making up 7.80% of the genome (**Table 1)**. LTR retrotransposons of the Ty1-copia superfamily was the most represented with the Angela lineage being the most abundant (5.87%; **Table 1, Fig. 1d**).

Table 1 Genome proportion of the repetitive sequences in the genome of *Luzula sylvatica*

Element		Proportion (%)
DNA satellites		35.31
	Lusy1_124	25.10
	Lusy2_174	3.61
	Lusy2_175	3.45
	LsylSat3_182	2.00
	LsylSat4_31	01.05
	LsylSat5_119	0.06
	LsylSat6_197	0.03
	LsylSat7_34	0.02
LTR-RE		7.80
	Class_I/Ty1-copia/Angela	5.87
	Class_I/LTR/Ty1-copia/Bianca	0.03
	Class_I/Ty3-gypsy/non-chromovirus/OTA/Athila	0.27
	Class_I/Ty1-copia/SIRE	1.37
	Class_I/Ty1-copia/Tork	0.13
	Class_I/Ty1-copia/Ivana	0.13
35S_rDNA		1.34
5S_rDNA		0.09
Unclassified		3.67
Total		**48.21**

Satellite family distribution varied across the genome (**Fig. 2a**). *Lusy1* was spread throughout all pseudochromosomes, with higher densities in interstitial regions (**Fig. 2a**). In contrast, the other satellite DNAs were found preferentially near telomeres or irregularly distributed on the chromosomes (**Fig. 2a**). We localized *in situ* the two most abundant putative

repeats (*Lusy1* and *Lusy2*) to corroborate the patterns obtained *in silico*. Remarkably, we observed that *Lusy1* shows a line-like distribution across the entire length of each sister-chromatid, in a similar pattern to other holocentromeric repeats (Marques et al. 2015). In contrast, *Lusy2*, which showed a more irregular pattern, sometimes presented internal accumulations compared to *Lusy1* (**Fig. 2b**).

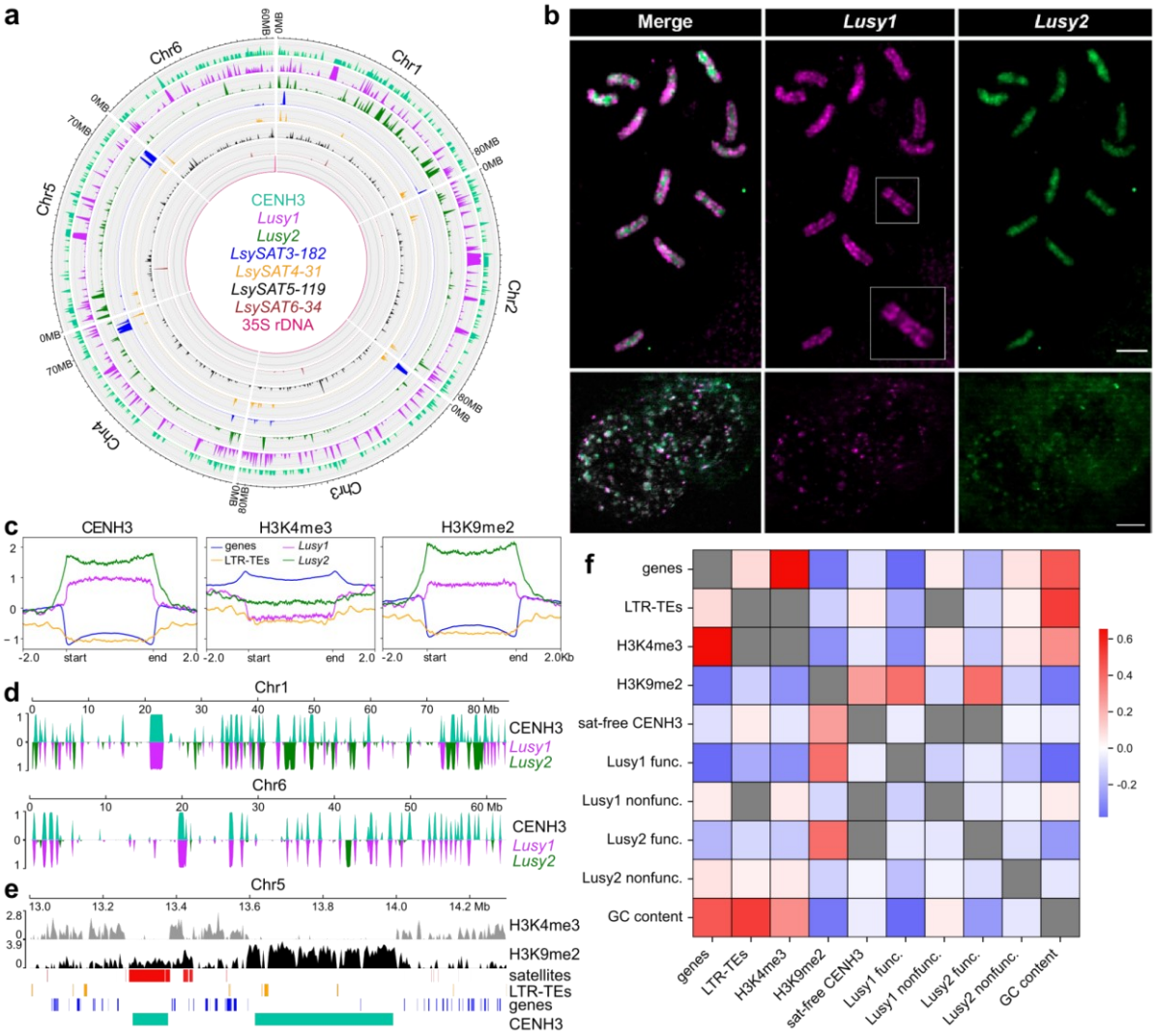


Fig. 2: Holocentromere characterization of *L. sylvatica* chromosomes as satellite repeat based. (a) Circos distribution of the main classes of tandem repeats and CENH3 domains with a 300 kb sliding window. **(b)** FISH showing discontinuous linear-like spreading of the *Lusy1* (magenta) and disperse pattern of *Lusy2* (green) repeats in nucleus and metaphase chromosomes of *L. sylvatica*. **(c)** Metaplots showing the enrichment of CENH3, H3K4me3, and H3K9me2 from the start and end of different types of sequences: genes (blue), LTR transposable elements (yellow), *Lusy1* (magenta) and *Lusy2* repeats

(green). ChIP-seq signals are shown as normalized log2 (RPKM ChIP/input). **(d)** Proportion of CENH3 units (light blue), *Lusy1* (magenta) and *Lusy2* (green) arrays in 100 kb windows. Distribution on all chromosomes is reported in Fig. S4. **(e)** Example of genomic feature distribution in and near satellite-free CENH3 unit. **(f)** Correlogram of genomic features in 100 kb windows. Gray fields indicate values on the diagonal and non-significant values of Spearman coefficient after multiple-testing correction (see Methods).

To investigate whether *L. sylvatica* holocentromeres are repeat-based, we performed comparisons of these satellites with CENH3 ChIP-seq data. CENH3-ChIP-seq showed CENH3 enrichment for *Lusy1* and *Lusy2* repeats and depletion in LTR transposable elements throughout the *L. sylvatica* genome (**Fig. 2c**). By mapping the *Lusy1* and *Lusy2* centromeric repeats together to CENH3 domains, we observed that centromeric units are composed mainly of *Lusy1* (232 out of 358 units) and/or *Lusy2* sequences (96 units), with several cases of satellite-free CENH3 units (33 units; **Fig. 2d and e; Fig. S4**). Although satellite-free CENH3 units were depleted of genes, they often contained transposable elements (16 out of 33 units; Fig. 2e; Fig. S9). Remarkably, Athila elements belonging to Ty3-gypsy family were the most abundant, making up nearly 60 % of LTR retrotransposons inside satellite-free CENH3 units while representing only ~12 % of all genomic LTR retrotransposons (Table S3). Presence of satellite-associated CENH3 units was further confirmed by immuno-FISH analyses, where satellite *Lusy1* signals partially co-localized with CENH3 domains along the chromosome (**Fig. 3a**) and were enriched in heterochromatic regions (**Fig. 3b**). Therefore, *L. sylvatica* represents a new case of repeat-based holocentromeres that are mostly associated, but not exclusively, with *Lusy1* repeats.

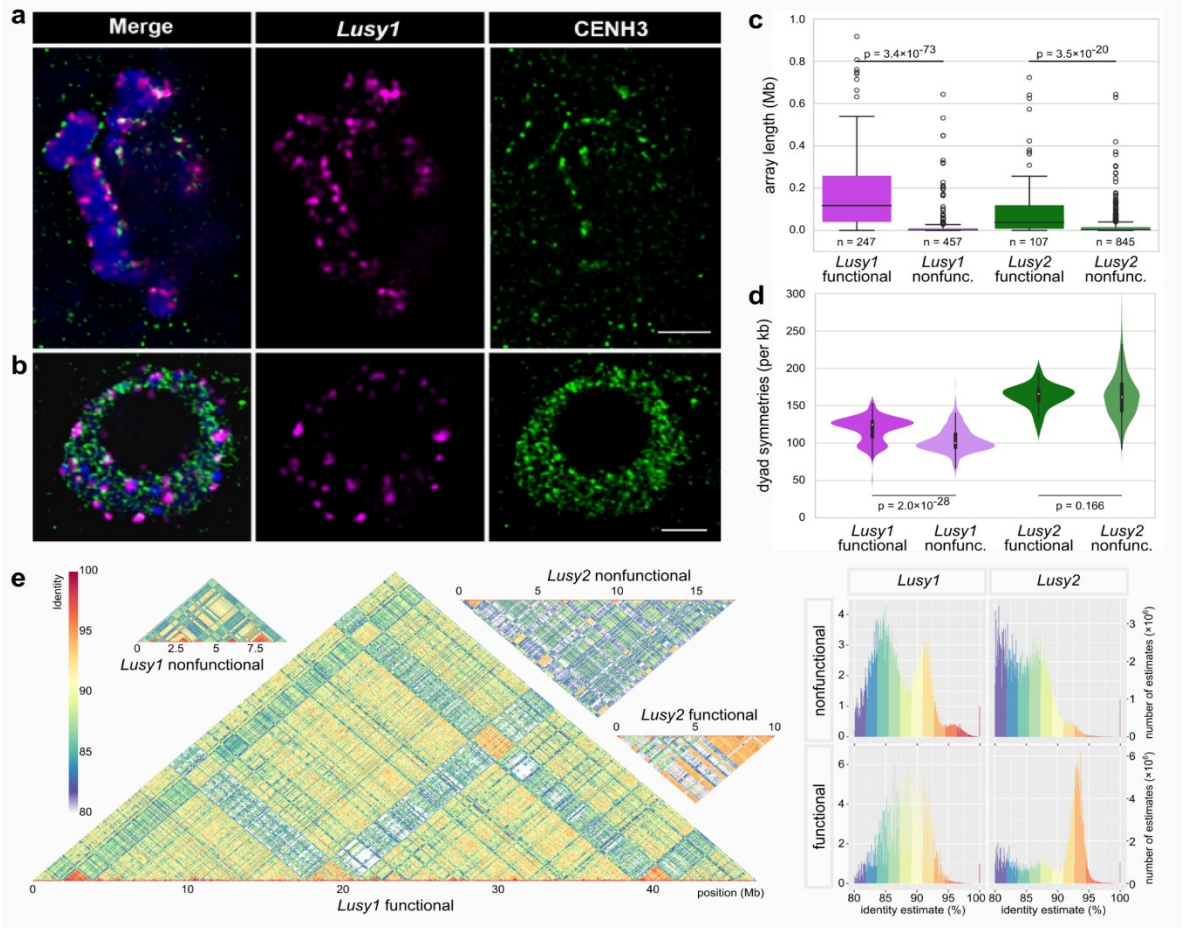


Fig. 3: Association of CENH3 with centromeric satellite arrays. (a) Immuno-FISH showing partial colocalization of CENH3 (green) and *Lusy1* (magenta) repeats in metaphase chromosomes (counterstained with DAPI, blue). **(b)** Immuno-FISH showing colocalization of CENH3 (green) and *Lusy1* (magenta) repeats in interphase nucleus (counterstained with DAPI, blue). **(c)** Box plot of sizes of centromeric satellites *Lusy1* and *Lusy2* arrays associated with CENH3 (functional) or non-centromeric (nonfunctional). Statistical significance tested using two-tailed Mann-Whitney U test. **(d)** Abundance of dyad symmetries in functional and nonfunctional arrays of centromeric satellites. Statistical significance tested using one-tailed Mann-Whitney U test. Array counts are identical to subfigure **c**. **(e)** Homogeneity of functional array fragments (only regions overlapping CENH3 units) and whole nonfunctional arrays of centromeric satellites *Lusy1* and *Lusy2*. Dot plots showing similarity between concatenated arrays, histograms showing frequency distribution of similarity values. Scale bars = 5 µm.

Using *in silico* mapping data, we identified arrays of centromeric satellites *Lusy1* and *Lusy2* which are not associated with CENH3 (nonfunctional). For *Lusy1*, 247 out of 704 arrays

(35 %) overlapped with CENH3 domains (functional). The length of these overlapping regions was 47 Mb out of 77 Mb in total (60 %). For *Lusy2*, 107 out of 952 arrays (11 %) contained CENH3 domains, making up 10 Mb out of 43 Mb (24 %) of total length. Nonfunctional arrays tended to be smaller than the functional arrays of the same satellite family (**Fig. 3c**), with average length of functional / nonfunctional arrays 189 kb/ 20 kb and 94 kb/ 20 kb for *Lusy1* and *Lusy2*, respectively. Functional *Lusy1* arrays contained a higher abundance of dyad symmetries, in *Lusy2,* this difference was not significant (**Fig. 3d**). Functional and nonfunctional arrays also differed in their inter-array sequence similarity. Functional arrays of both *Lusy1* and *Lusy2* satellites had higher average similarity across discrete arrays compared to nonfunctional arrays (88.0 vs 87.2 % and 89.6 vs 85.1 % for *Lusy1* and *Lusy2*, respectively; **Fig. 3e**). Interestingly, nonfunctional *Lusy1* arrays showed a clear bimodal distribution, with one of the groups having a higher similarity than the corresponding functional arrays (**Fig. 3e**). Epigenetic status of the functional array chromatin also showed a striking contrast, since functional centromeric regions (i.e. *Lusy1* and *Lusy2* functional arrays, satellite-free centromeric units) were enriched with heterochromatin mark H3K9me2 and depleted of euchromatin mark H3K4me3, while the nonfunctional arrays were the opposite (**Fig. 2f and S9**).

Kinetochore proteins KNL1 and NDC80 were conserved during the transition to holocentricity in *Luzula*

Transition to holocentricity in *Cuscuta* was associated with massive changes in the centromeric localization of CENH3 and kinetochore proteins KNL1, MIS12 and NDC80, which represent the three complexes of the KMN network (Neumann et al., 2023). However, *Luzula*, unlike *Cuscuta*, still possesses centromeric activity associated with CENH3 (**Fig. 4a and S5a,** Heckmann et al., 2011). To test whether the kinetochore assembles along the poleward chromosome surface, as expected for holocentric chromosomes, we examined by immunodetection the localization of KNL1 and NDC80 in the holocentric *L. sylvatica* and *Luzula nivea* (a sister species of *L. sylvatica*), as well as in the monocentric *J. effusus*. Antibodies against KNL1 showed a similar pattern in both *L. sylvatica* and *L. nivea*, with signals detected as multiple clusters, not as continuous lines, along the poleward surface of chromosomes, where microtubules attached (**Fig. 4b**). However, immunodetected signals for the NDC80 protein were observed only in *L. nivea* (**Fig. S5a)**. In addition, immuno-FISH

experiment showed that the centromeric repeat *Lusy1* exhibited signals in the most inner part of the kinetochore protein KNL1 in a cluster pattern, not as continuous lines as KNL1 where the microtubules attach (**Fig. 4c)**. In *J. effusus*, these antibodies showed a specific dot-like pattern in the primary constriction region of the chromosome, also associated to microtubule attachment sites (**Fig. S5b**). Both proteins provided spindle-binding sites detected with antibodies against α-tubulin, indicating that both proteins have a conserved kinetochore function in *Luzula* and *Juncus*.

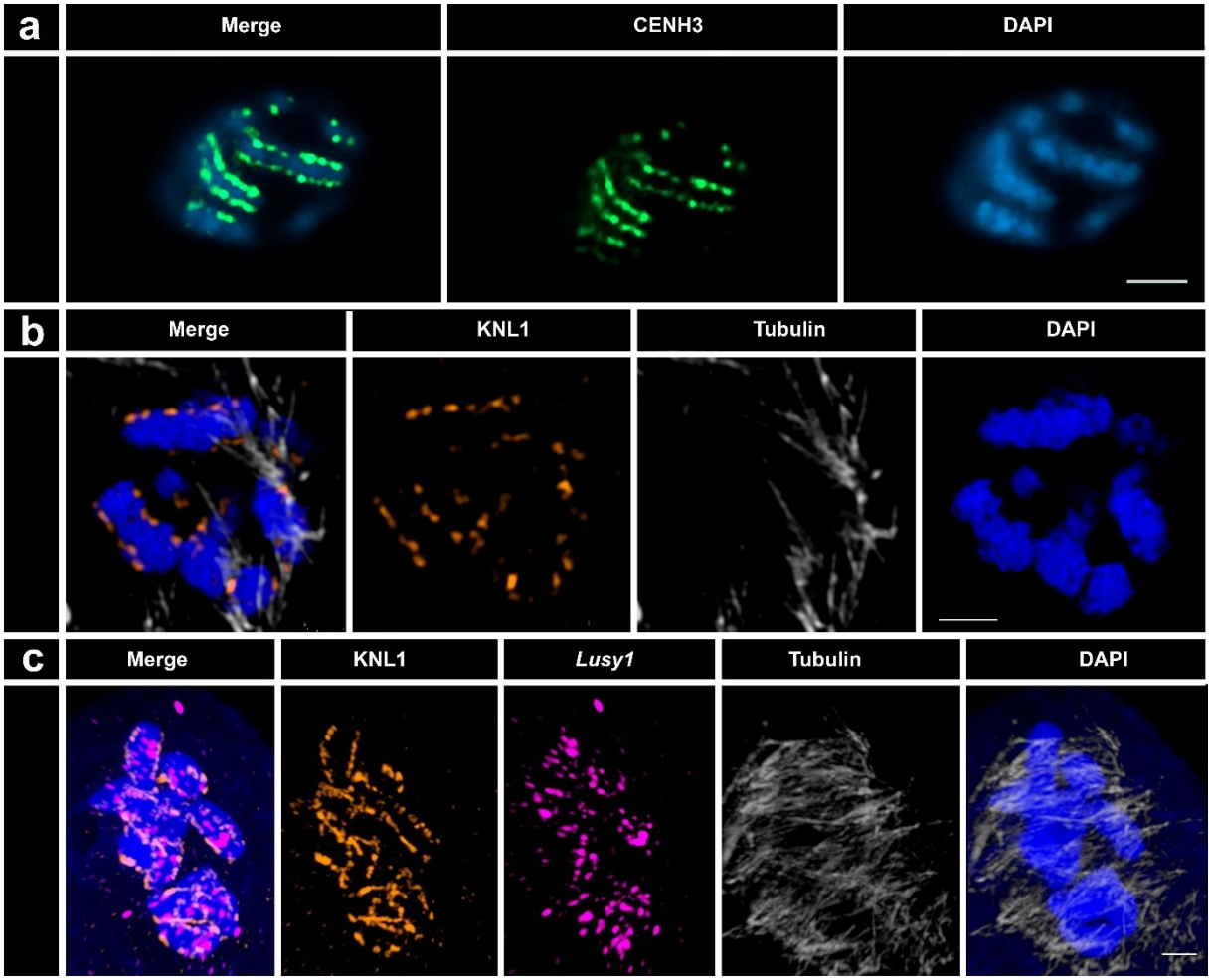


Fig. 4: Detection of α-tubulin with centromere and kinetochore proteins in chromosomes of *L. sylvatica*. The CENH3 (green) centromere **(a)** and KNL1 **(**orange, **b)** proteins localize specifically to the centromere surface, where microtubules (gray) bind. **(c)** Detection of α-tubulin (gray) with KNL1 (orange) and *Lusy1* (violet) repeat in chromosomes of *L. sylvatica*. Microtubules attach to the kinetochore KNL1 proteins regularly, while some functional *Lusy1* sequences are found in the most inner part of some kinetochore regions.

Evolutionary dynamics of the centromeric satellites in *Luzula* species

To determine whether repeat-holocentromeres are conserved across other species of the genus *Luzula*, both an individual and comparative analysis of the repeatome using RepeatExplorer were performed in 13 species, including *L. sylvatica* (**Table S4-S6**). The global genomic proportion of repetitive DNA varied from 35.24% (*L. pilosa*) to 66.29% (*L. wahlenbergii*) (**Table S5**). In general, satellite DNA was the most abundant class of repeats, comprising up to 49% of the *L. sudetica* genome. The centromeric satellite *Lusy1* was one of the most abundant among all satellites, representing up to 47.41% of *L. sudetica* genome but only 3.33% of the *L. nivea* genome (**Table S5**). *Lusy2* also showed variation in abundance among species, ranging from 0.34% (*L. multiflora* subsp. *frigida*) to 31.03% (*L. luzuloides*), being also found in the *L. elegans* genome (0.72%; **Table S5**), a species with previously undetected holocentromeric repeat (Heckmann et al. 2013). LTR retrotransposons revealed variable abundances among species, with the Ty1/copia superfamily being the most represented (1.21% in *L. pilosa* to 41.35% in *L. elegans*; **Table S5**).

Comparative repeat analysis resulted in 166 shared clusters (**Fig. 5a, Table S6**). Because a similar pattern of repeat abundance was observed in the known genome size species sample (**Fig. S6**), we have decided to discuss our analysis with the full species matrix. Variants of *Lusy1*, the most abundant satDNA in *Luzula*, were found in all analyzed species, except in *L. elegans*, where this satellite was not detected even in an additional fine search of the raw sequencing reads (**Fig. 5a**). Different variants of the *Ty1-copia Angela* lineage were found in high abundance among the species, being more dominant in the genomes of *L. arcuata* and *L. elegans* (**Fig. 5a; Table S6**). *Ivana* and *SIRE Ty1-copia* lineages were also shared among all species, although they exhibited lower abundance than *Angela* (**Table S5-6**).

Because *Lusy1* and *Lusy2* were the most abundant in the comparative analysis, consistent with the observation from the *L. sylvatica* genome, we performed fluorescence *in situ* hybridization to confirm their distribution also in *L. nivea*, the species with the lowest abundance of *Lusy1*. Similar to what was described for *L. sylvatica*, the FISH signals of *Lusy1* in *L. nivea* showed a non-uniform distribution along the chromosomes, exhibiting regions both enriched and depleted of this satDNA. Furthermore, *Lusy2* showed clustered signals enriched at interstitial and terminal regions (**Fig. 5b**).

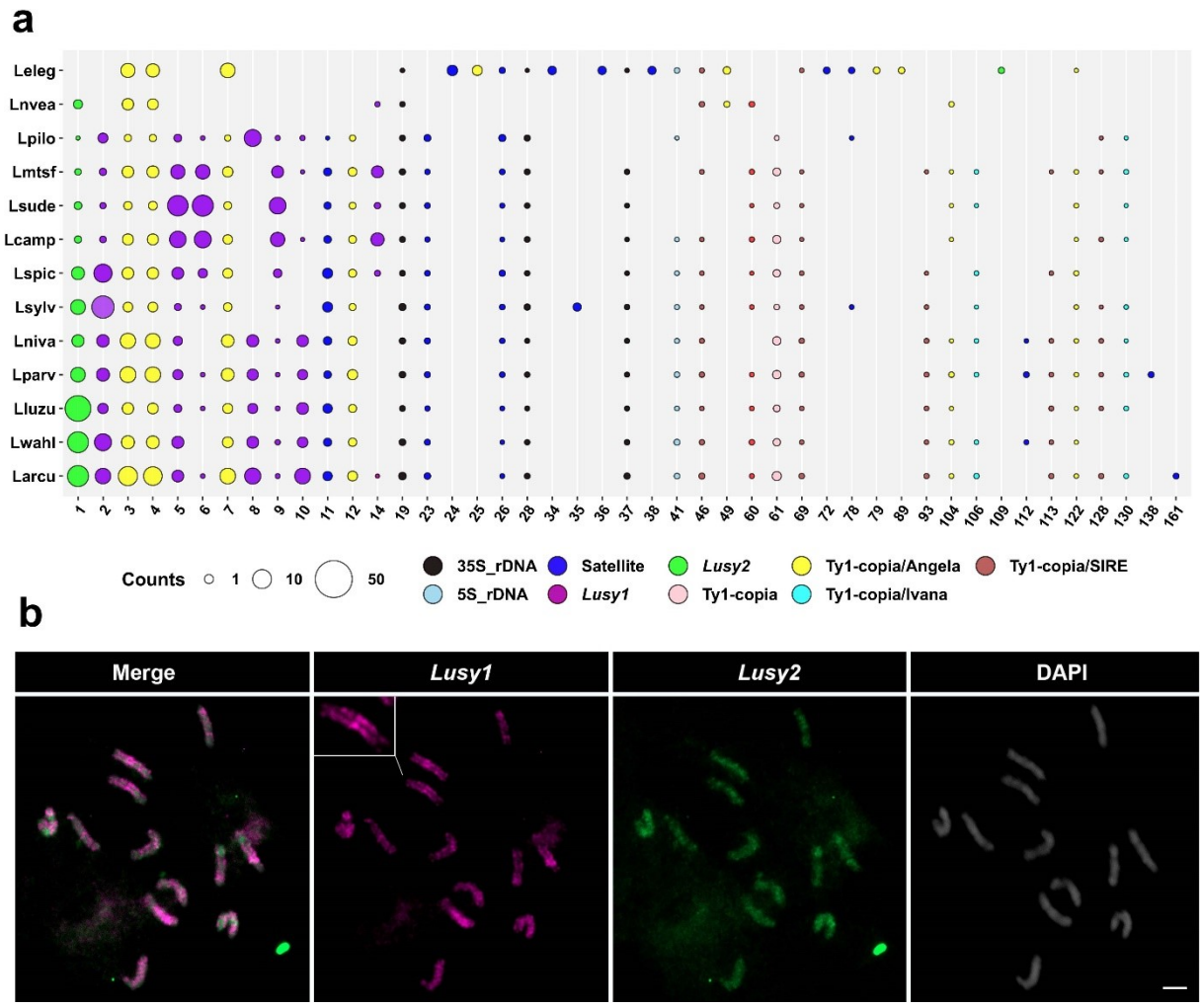


Fig. 5: Comparative analyses of the main types of repetitive sequences in *Luzula* species. (a) Comparative abundance of the main types of repetitive sequences in *Luzula* species. The size of the ball is proportional to the genome abundance of that cluster for each species. The colors of the balls correspond to different repetitive sequence types (see Table S4 for details). **(b)** FISH showing wide-spreading of DNA satellite family *Lusy1* (red) and dispersed pattern of *Lusy2* (green) repeats in metaphase chromosomes of *L. nivea*. Chromosomes were counterstained with DAPI. Scale bars = 2 µm.

Chromosome fusions drive karyotypic evolution in *Luzula*

Chromosomes from some grasses and several holocentric species have undergone extensive karyotypic rearrangements through fusions (Wang et al., 2021). To investigate the possible association between holocentricity and chromosome fusions, we analyzed synteny between the genomes of the holocentric *L. sylvatica* and the monocentric *J. effusus* (**Fig. 6, Fig. S7**).

Considering *n* = 20 as the putative ancestral karyotype for the family (Drábková et al. 2013; Mata-Sucre et al. in prep.), synteny analysis between the two genomes revealed that the chromosomes of *L. sylvatica* consist of fused blocks originating from *J. effusus* chromosomes, resulting in a descending dysploidy to *n* = 6, supported by the presence of interstitial telomeric signals by FISH (**Fig. 6a-b**). Despite their high chromosome number and centromere-type differences, large syntenic blocks were identified between both genomes, indicating well-conserved genomic structures in this family, with a total of 86.4% (23,016 gene pairs) of the *J. effusus* genome being syntenic with *L. sylvatica* (**Fig. 6c, Table S7, Fig. S7-8**).

We have recently shown that the holocentromeric repeat *Tyba* can be involved in facilitating end-to-end chromosome fusions in *Rhynchospora* (Hofstatter et al., 2022). To assess the possible role of *Lusy1* and *Lusy2* in the fusion points observed in *Luzula* genomes, we looked for specific enrichment of these repeats at the fusion points. Size of the fusion point, defined as the space between collinear blocks, ranged from 10 kb to 8 Mb in size. At the block boundaries, we observed a higher gene density (**Fig. 6e, S9**). Larger fusion regions (>100 kb) also often contained regions enriched with satellites (especially *Lusy2* and less frequent *Lusy1*) and TEs (**Fig. 6d**). However, this repeat enrichment pattern is not universal enough to be recognizable at the scale of genome-wide colocalization between features (**Fig. 6e**).

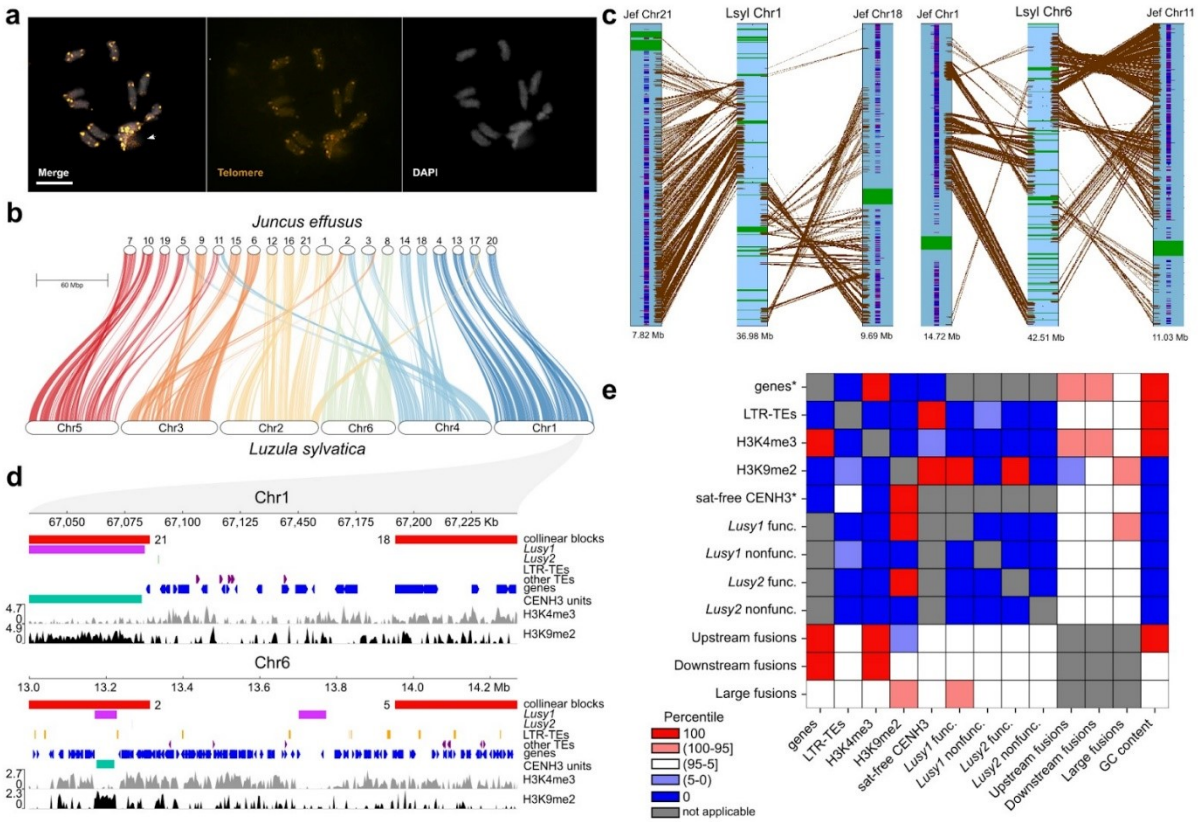


Fig. 6: Genome synteny comparison of *Luzula sylvatica* and *Juncus effusus*. (a) FISH of telomeric DNA showing chromosomes with remnant interstitial telomeric sites, suggesting ancestral fusion events. (**b**) Genome synteny patterns showing macro-conserved blocks of *J. effusus* that were fused into *Luzula* chromosomes. (**c**) *L. sylvatica* (Lsyl) chromosome 1 and 6 showing the fusion of syntenic blocks of *J. effusus* (Jef) 1, 11, 18 and 21. The synteny stops near the centromeric CENH3 (green) sites. Genes, CENH3 and telomere domains are annotated as blue-purple, green and black stripes, respectively. (**d**) Examples of fusion sites on Chr1 and Chr6 in *L. sylvatica*. (**e**) Colocalization between major genomic and epigenomic features and fusion points (columns) based on comparison of overlap with simulated regions (rows). Heatmap values show the rank of real overlap value in a distribution of overlaps with simulated regions. * Signifies exclusion of satellite array positions from permitted simulated region locations, for the rest of the features, the whole genome was used for simulated regions.

Discussion

Here we provide novel findings about the dynamics and functionality of holocentromeres in the genus *Luzula*. We show that repeat-holocentromeres in *L. sylvatica* are organized in kilobase-scale, non-uniformly spaced CENH3-positive centromeric units, mainly made up of *Lusy1*, a 125 bp-long satellite repeat. This arrangement, unlike monocentric chromosomes, intermingles with coding and non-coding regions, and associates with heterochromatic domains across the genome. The diffused holokinetic centromere structure in *L. sylvatica* organized into multiple CENH3-positive subunits is similar to what has been reported in other holocentric *Luzula* (Nagaki et al., 2005) and in *Rhynchospora* species (Hofstatter et al., 2022).

Among the *Luzula* species analyzed, *L. sylvatica* genome is enriched in satellite *Lusy1*, comprising 23.14% of its genome, and ranking second after *L. sudetica* with approx. 36%. Interestingly, while *L. elegans* possesses a repeat-rich genome, constituting 61%, its repeats exhibit a non-centromeric pattern (Heckmann et al., 2013; Jankowska et al., 2015; Nóvak et al., 2017), distinguishing it from the repeat-based holocentromere organization observed in this study. The holocentromeres analyzed here are similar to other holocentric species harboring repeat-based centromeres, e.g., *Rhynchospora pubera* (Marques et al., 2015; Hofstatter et al., 2022), *Chionographis japonica* (Kuo et al., 2022), and in the nematode *Meloidogyne incognita* (Despot-Slade et al., 2021).

Previous studies have identified a possible 178 bp potential centromeric tandem repeat in *L. nivea* and other *Luzula* species with similarity to the centromeric satellite RCS2 from rice (Haizel et al., 2005). Indeed, the 178-bp satellite shares 87% similarity with the *Lusy2* satellite,

which we found in all species, including *L. elegans*. However, unlike *Lusy1*, which largely encompasses the entire holocentromere, *Lusy2* shows an enrichment in centromeric regions but does not cover the entirety of the chromosomes. Unlike typical centromeres, which are usually associated with HOR, neocentromeres can arise in heterogeneous genomic regions that are subjected to rapid cycles of invasion and purification of repetitive sequences through satellite homogenization (Wlodzimierz et al., 2023). This distinction of *Lusy* satellites implies that both variants may be competing for centromere dominance through homogenization process, and that *Lusy1* probably serves as the primary centromeric satellite in most *Luzula* species.

We evaluated the genomic landscape of species with different centromere types: the monocentric *J. effusus* and the holocentric *L. sylvatica*. Our results support dysploidy as the main driver of karyotype evolution in holocentric organisms (Bozek et al., 2012; Guerra 2016; Senaratne et al., 2022), since a descending dysploidy through fusion resulted in *L. sylvatica* genome with $n = 6$. Although phylogenetic relationships are poorly resolved in the genus, descending dysploidy has been observed in species from different *Luzula* clades (*L. elegans* in the Marlenia clade and *L. purpureo-splendens* in the Nodulosae clade), indicating that this process of chromosomal evolution has occurred independently at least twice during the evolution of the genus (Bozek et al., 2012).

Fission and fusion events have been described in sedges, leading to dysploid karyotypes in the holocentric genera *Rhynchospora* (Hofstater et al., 2022) and *Carex* (Ning et al., 2023), as well in some holocentric butterflies (Cicconardi et al., 2021). However, the fusion of ancestral chromosomes that resembles chromosomes from the sister genus *Juncus* resulting in the dysploid *L. sylvatica* is intriguing, since it involves a switch of centromere organization in parallel. Recently, the repeat-based centromeres of *J. effusus* were described as an atypical monocentromere, with up to three centromere cores and different types of centromeric organization, resembling a wide-spread metapolycentric organization (Dias et al., in review.). In *L. sylvatica*, these 21 putative ancestral-like chromosomes merged into six chromosomes, with different rearrangements and genomic reshuffling. Furthermore, we observed an abundance of well-conserved genomic regions across both genomes, except in the regions enriched by centromeric satellites, possibly resulting from repetitive DNA turnover in the past 56 million years of divergence (Elliott & Davies 2019; this study). However, in contrast to repeat-mediated chromosomal fusions observed in *Rhynchospora* (Hoffstater et al., 2022), many of the fusion sites in *L. sylvatica* lack centromeric repeats, suggesting another mechanism of chromosome fusions in karyotype evolution in *Luzula*. Alternatively, the footprints of

genomic features mediating ancestral fusions were masked by additional chromosomal rearrangements and dynamic repeat turnover.

Holocentromeres have evolved from a monocentric ancestor multiple times during the evolution of eukaryotes, but these few events involved different changes and preconditions that were necessary for their formation in each group (Marques-Corroz et al., 2019; Lucek et al., 2022; Senaratne et al., 2022; Kuo et al., 2023). In *C. europea*, the transition to holocentricity was associated with CENH3-independent kinetochore assembly and extensive changes in kinetochore structural and regulatory protein genes (Oliveira et al., 2020; Neumann et al., 2023). In insects, four independent transitions to holocentricity were associated with losses of CENH3 domains and relative preservation of the inner kinetochore complex (Drinnenberg et al., 2014). Here, we detected multiple clusters of signals for KNL1 and NDC80 proteins along *Luzula* chromosomes, without the typical line-like distribution of CENH3 observed for *L. elegans* and *L. nivea* (Nagaki et al., 2005; Heckmann et al., 2011), as well as a conserved monocentric-like distribution in *J. effusus* species. In contrast to *Cuscuta* species, *R. pubera* species also exhibit a holocentromere-like distribution of NDC80 and KNL1 (Neumann et al., 2023), suggesting that the independent transitions to holocentricity in Cyperaceae and Juncaceae might be associated with differences in kinetochore assembly in these groups (more or less continuous along chromosomes, respectively), although maintaining the same main players in kinetochore formation.

Numerous evolutionary models and drivers have been proposed to explain the emergence of holocentricity from monocentric ancestors, one of them is based on satellite propagation (Neumann et al., 2012). Although the *Luzula* species analyzed have a centromeric satellite DNA (*Lusy1*) present, *L. elegans*, one of the first lineages to diverge in the genus (Drábková & Vlček 2010; present work), lacks centromeric satellite repeats. Similarly, to *L. elegans*, early lineages diverging from the genus *Rhynchospora* lack the holocentromeric Tyba repeats and instead have diverse satellite arrays arranged in block-like patterns (Ribeiro et al., 2017; Costa et al., 2022), indicating that the colonization of centromeric satellites might occur later and independently of the transition to holocentricity.

A possible explanation for this finding is that new evolutionary centromeres, initially devoid of satellite DNA, may undergo a "maturation" process during their evolution through the acquisition of heterochromatin epigenetic modifications, accumulation of heterochromatin-specific transposable elements, and invasion of satellite repeats (Piras et al., 2010; Capelleti et

al., 2022). Indeed, the transition to holocentricity in Cyperaceae is believed to have occurred much earlier than the emergence of Tyba in *Rhynchospora* (Costa et al., 2022). From this point of view, the presence of repeat-free centromeres in *L. sylvatica* and *L. nivea*, as well as the uneven distribution of its centromeric satellite repeats and the cluster-like distribution of outer kinetochore proteins, could be interpreted as evidences of intermediate stages of holocentromere "maturation". We hypothesize that, after multiple fusions of interspersed monocentric chromosomes (*Juncus*-like), the centromeric domains were initially conserved in the larger chromosomes (*Luzula*-like), forming polycentric chromosomes. Later, satellites could evolve by a combination of satellite DNA library diversification and concerted evolution, as observed for *Tyba* repeats (Costa et al., 2023). Although neither *Tyba* nor *Lusy1* were probably required for holocentromere formation, it is possible that their sequence confers advantages in CENH3 loading (Kasinathan & Henikoff, 2018) or nucleosome formation and positioning, which resulted in their propagation and conservation as a centromere stabilization factor (Talbert & Henikoff 2020; Hofstatter et al., 2022; Costa et al., 2022). Indeed, we propose that the evolution of satellites in *L. sylvatica* is prominently affected by their centromeric function, since arrays associated with CENH3 are dispersed along the chromosome, have larger size, and contain a higher frequency of dyad symmetries, suggesting selection pressure acting upon them (Fig. 7). Yet, the factor(s) triggering this process and the molecular mechanisms involved remain to be elucidated.

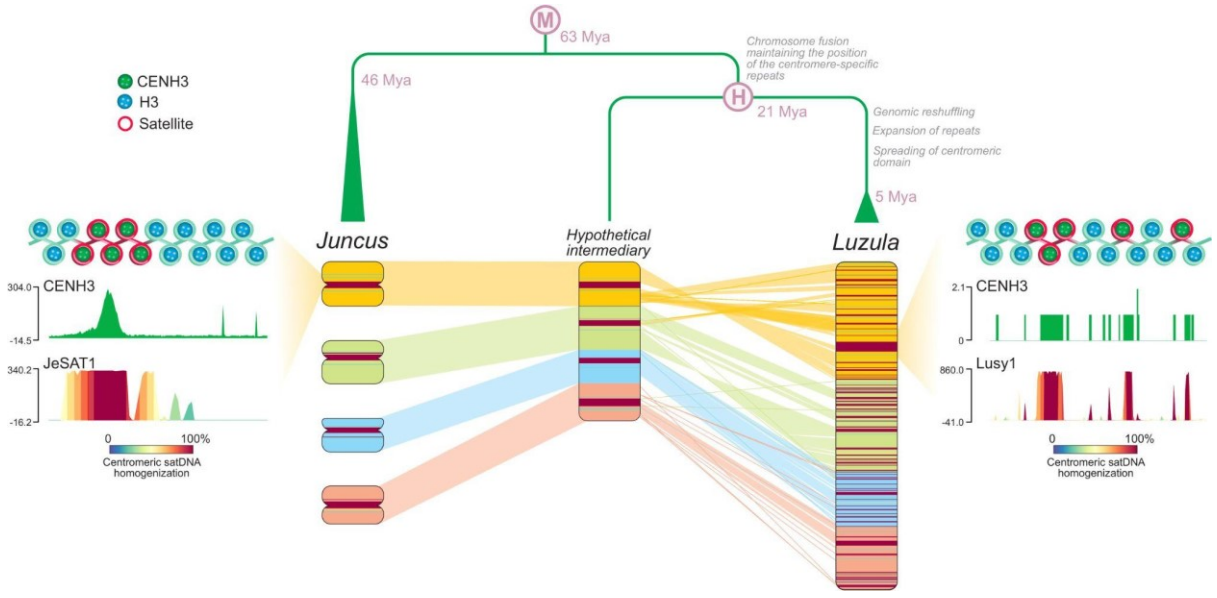


Fig. 7: Model for the origin of holocentricity in Juncaceae. After several fusions of whole monocentric chromosomes (*Juncus*-type), centromeric domains were initially conserved on the larger

chromosomes (intermediate state), forming polycentric chromosomes. Subsequently, expansion of these repeats may have extended the length of the centromeric domain, and genomic rearrangement gave rise to the holocentric condition.

Declaration of interests

The authors declare no competing interests.

Acknowledgements

We acknowledge the excellent technical assistance of Christina Philipp and Ursula Pfordt. We thank Dra. Magdalena Vaio for her comments, which led to an improvement in the quality of the work. We thanks to the Darwin Tree of Life Project at the Wellcome Sanger Institute by make the data available and can be obtained from https: //www.darwintreeoflife.org/project-resources. We also thank the ELIXIR-CZ Research Infrastructure Project (LM2015047) for providing computational resources for RepeatExplorer analysis and e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic for providing computational resources for the analysis of ChIP-seq data.This work was funded by the Max Planck Society and Deutsche Forschungsgemeinschaft (DFG grant number MA 9363/3-1 granted to AM) and supported by a scholarship to Y.M.S. from the International Cooperation Program PROBRAL (CAPES/DAAD project number 88881.144086/2017-01). Work of M.K. is supported by grant 21–00580S from Czech Science Foundation. A.P.H. and G.S. received productivity fellowships from CNPq (process numbers PQ- and PQ-312852/2021-5, respectively).

Methods

Plant material

Plants from natural populations of *L. sylvatica* were collected in Germany, and further cultivated under controlled greenhouse conditions (16h daylight, 26 °C, >70% humidity). Ornamental plant *L. nivea* was commercially obtained and cultivated under controlled greenhouse conditions (16h daylight, 20°C).

Genome assembly and Hi-C scaffolding

HiFi and Hi-C reads obtained through the Darwin Tree of Life database (www.darwintreeoflife.org) were assembled using Hifiasm (Cheng et al., 2021), available at https://github.com/chhylp123/hifiasm, following the command: "*hifiasm -o output.asm -t 40 reads.fq.gz*". Preliminary assemblies were evaluated for contiguity and completeness with BUSCO (Seppey et al., 2019) and QUAST (Gurevich et al., 2013).

Hi-C reads were first mapped to the primary contigs file obtained from the Hifiasm assembler using BWA (Li & Durbin, 2009) following the hic-pipeline (https://github.com/esrice/hic-pipeline). Hi-C scaffolding was performed using SALSA2 (https://github.com/marbl/SALSA, Ghurye et al., 2019) with default parameters using '*GATC, GAATC, GATTC, GAGTC, GACTC*' as restriction sites. After testing several minimum mapping quality values of bam alignments, final scaffolding was performed with MAPQ10. Several rounds of assembly correction guided by Hi-C contact maps and manual curation of scaffolds were performed to obtain the six pseudomolecules.

Genome size estimate was obtained from HiFi reads using findGSE (Sun *et al*., 2018). First, histogram of k-mers was created using jellyfish (Marçais & Kingsford, 2011), and then the findGSE R package was used for model fitting according to package documentation (https://github.com/schneebergerlab/findGSE).

Chromatin immunoprecipitation (ChIPseq) sequencing and analysis

ChIP experiments were performed following Hofstatter et al. (2022). *L. sylvatica* leaves were harvested and frozen in liquid nitrogen until sufficient material was obtained. The samples were fixed in 4% formaldehyde for 30 min and the chromatin was sonicated until 300 bp fragments. Then, 40 ng of sonicated chromatin was incubated with 2 ng of antibody overnight. Immunoprecipitation experiments were carried out for the rabbit anti-LeCENH3 (Ma et al., 2016), rabbit anti-H3K4me3 (abcam, ab8580), and mouse anti-H3K9me2 (abcam, ab1220). Recombinant rabbit IgG (abcam, ab172730) and no-antibody inputs were used as controls. Two experimental replications were also maintained for all the combinations. ChIP DNA was quality-controlled using the NGS-assay on a FEMTO-pulse (Agilent); then, an Illumina-compatible library was prepared for all immunoprecipitants with the Ovation Ultralow V2 DNA-Seq library preparation kit (Tecan Genomics) and single-end 1× 150-bp reads were sequenced on a HiSeq 3000 (Illumina) device. For each library, an average of 20 million reads were obtained.

The raw sequencing reads were trimmed by Cutadapt (Martin, 2011) to remove low-quality nucleotides (with quality score less than 20) and adapters. Trimmed ChIPed 150-bp single-end reads were mapped to the respective reference genome with bowtie2 (Langmead & Salzberg 2012), where all read duplicates were removed and only the single best matching read was kept on the final alignment BAM file. ChIP vs input signal was calculated as the log2 ratio of read coverages normalized by reads per kilobase per million mapped reads (RPKM) using the bamCompare tool from deepTools package (Ramírez et al., 2016). Averaged signal from both replicates was visualized using pyGenomeTracks (Lopez-Delisle et al., 2021). Metaplots obtained by the plotProfile function from deepTools were used to compare the distribution with other genomic features (Ramírez et al., 2016). To concretize enriched domains, we performed peak-calling by MACS3 (Zhang *et al*., 2008) and epic2 (Stovner & Sætrom, 2019) and filtered only the peaks identified by both tools in both replicates. Based on analysis of CENH3 peak clustering, peaks closer than 150 kb were merged to obtain uninterrupted centromeric units (code available on GitHub). To further analyze colocalization of genomic features and epigenetic marks, 1000 rounds of random distribution or random distribution excluding satellite array locations was simulated for genomic and epigenomic features. Then, overlap with all other studied features was calculated for simulated and real feature distributions as a proportion of overlapping bases to all bases covered by the feature. Percentile of real overlap proportion in the distribution of simulated values was reported (code available on Github).

Synteny analysis

The synteny analysis between *L. sylvatica* and *J. effusus* was performed with CoGe SynMap platform (https://genomevolution.org/coge/SynMap.pl, Lyons et al. 2008) and SyMAP v. 5.0.6 (Soderlund et al., 2006, 2011). For this analysis, CDS sequences, centromeric and telomeric repeats of both species were used. Synteny plots were obtained with GENESPACE (https://github.com/jtlovell/GENESPACE, Lovell et al., 2022). Orthologs were identified following the steps and parameters described by Montenegro et al. (2022): (1) using the BlastZ tool; (2) synteny analysis was performed using DAGChainer, with 25 genes as the maximum distance between two matches (-D) and 20 genes as the minimum number of aligned pairs (-A); (3) Quota Align Merge was used to merge syntenic blocks, with 50 genes as the maximum distance between them; and (4) orthologous and paralogous blocks were differentiated according to the synonymous substitution rate (Ks) using CodeML (where 2 was the maximum value of log10), and represented with different colors in the dot plot (Fig. S7a).

For the characterization of the regions involved in fusion points, we followed Hofstatter et al. (2022). The synteny alignment between *L. sylvatica* and *J. effusus* genomes obtained in SyMAP allowed us to pin the putative regions around the borders of the fusion events. In order, to identify the underlying sequences at the fusion regions, we loaded annotation features for genes, TEs, and tandem repeats on SyMAP alignments. This allowed us to detect the sequence types in the putative fused regions. Further inspection and characterization of such regions were done by checking the genome coordinates and annotation features with Geneious (Kearse et al., 2012).

Repeat characterization

Illumina reads were filtered by quality with 95% of bases equal to or above the quality cut-off value of 10 using RepeatExplorer pipeline (https://repeatexplorer-elixir.cerit-sc.cz/; Novák et al., 2013). The clustering was performed using the default settings of 90% similarity over 55% of the read length. For the comparative analyses, we performed an all-to-all similarity comparison across all species following the same approach. Because the genome size is unknown for some analyzed species, each set of reads was down sampled to 1,000,000 for each species. Additionally, a subsample of eight species with known genome size were analyzed to compare the results. Samples from each species were identified with the four-letter prefixes shown in Table 1, and concatenated to produce datasets as input for RepeatExplorer (Novák et al., 2020) graph-based clustering.

The automatic annotation of repeat clusters was manually inspected and revised, and followed by a recalculation of the genome proportion of each repeat type where applicable. DANTE and DANTE-LTR annotation pipelines were used to identify full length LTR retrotransposons containing a set of protein domains from REXdb (Novák et al., 2020). Overall repeat composition was calculated excluding clusters of organelle DNA (chloroplast and mitochondrial DNA). All contigs with tandem repetitions identified by TAREAN (Novák et al., 2017), were compared in order to verify their homology with DOTTER (Sonnhammer & Durbin 1995). All repeat features were individually mapped to the genome by BLAST using Geneious (Kearse et al., 2012), converted to BED and used as input track for genome-wide overview with ShinyCircos (Yu *et al*., 2018).

Characterization of centromeric units

Centromeric units from ChIP-seq analysis were grouped by chromosome and their size, 0 and density were calculated. Next, centromeric units were overlapped with locations of satellites to

obtain locations and extract sequences of functional array fragments (precise regions where centromeric satellites *Lusy1*/*Lusy2* and CENH3 units overlap), nonfunctional arrays (whole *Lusy1*/*Lusy2* arrays not overlapping CENH3 units), and satellite-free units (CENH3 units not overlapping any satellites). Sequences of discrete arrays of each type were concatenated and their homogeneity assessed using ModDotPlot (Sweeten, 2023). Dyad symmetries were identified in each array using the EMBOSS palindrome tool as described in Kasinathan & Henikoff (2018) with -nummismatches parameter set to 0. Statistical significance of the increase of dyad symmetry abundance for functional arrays was tested using one-tailed Mann-Whitney U test from scipy package (Virtanen et al., 2020). Proportions of functional and nonfunctional arrays as well as other genetic and epigenetic features in 100-kb windows were correlated using Spearman's rank correlation from scipy package and resulting correlation coefficients were plotted in a heatmap (code available on Github). Additional packages were used for data handling and visualization (Dale et al., 2011; Harris et al., 2020; Hunter, 2007; Waskom, 2021; McKinney, 2011).

Cytogenetic and immunostaining of CENH3 protein

Mitotic preparations were made from root meristems fixed in 4% paraformaldehyde and Tris buffer (10 mM Tris, 10 mM EDTA, 100 mM NaCl, 0.1% Triton, pH 7.5) for 30 min on ice in vacuum and for another 20 min only on ice. After washing twice in 1× PBS for 10 min, the roots were digested in a cellulase-pectinase (2% w/v /20% v/v solution) containing PBS buffer and squashed in PBS. The coverslips were removed in liquid nitrogen and the slides were air dried and stained in 2 µg/mL DAPI/Vectashield mounting medium for slide selection under the epifluorescence microscope. The slides with the highest number of cells in division were incubated in 3% (w/v) bovine serum albumin (BSA) containing 0.1% Triton X-100 in PBS. Immunostaining was performed using the primary antibodies rabbit anti-LeCENH3 (dilution 1:100; Ma et al., 2016), rabbit anti-KNL1 (dilution 1:1000; Neumann et al., 2023), rabbit anti-NDC80 (dilution 1:1000; Neumann et al., 2023) and mouse anti αtubulin (dilution 1:100, Sigma-Aldrich, St. Louis, MO; catalog number T6199). As the secondary antibody, goat anti-Rabbit IgG antibody (Invitrogen) or goat anti-mouse Alexa Fluor 488 (ImmunoResearch; catalog number: 115-545-166) were used in a 1:500 dilution. Slides were incubated overnight at 4 °C, washed 3 times in 1× PBS and then the secondary antibody was applied, incubated at room temperature for 3 h and washed 3 times in 1× PBS. The slides were counterstained with 2 µg/mL DAPI in Vectashield mounting buffer. Microscopic images were recorded using a

Zeiss Axiovert 200M microscope equipped with a Zeiss AxioCam CCD. Images were analyzed using the ZEN software (Carl Zeiss GmbH).

Oligo probes from the most abundant tandem repeats (*Lusy1* and *Lusy2*) and the *Arabidopsis* telomeric sequence (TTTAGGG) were used for fluorescent *in situ* hybridization (FISH). Mitotic chromosomes from roots pretreated with 2mM 8-hydroxyquinoline for 24 h at 4°C and fixed with ethanol:acetic acid (3:1 v/v) for 2 h, were prepared using the air-drying method (Ribeiro et al., 2017). FISH was performed as described by Pedrosa et al. (2002). The slides were counterstained with 2 µg/mL DAPI in Vectashield mounting medium. The images were captured as described above. Immuno-FISH was performed following Houben et al. (2007), the immunostained slides were washed with PBS for 15 min, postfixed in 4% paraformaldehyde in PBS for 5 min, and then probed with the satellite *Lusy1*. We applied super-resolution spatial to analyze the chromatin ultrastructure by structured illumination microscopy (3D-SIM) using a 63x/1.40 Oil Plan-Apochromat objective of an Elyra PS.1 microscope system (Carl Zeiss GmbH, Germany). Maximum intensity projections from image stacks were calculated via the Zeiss ZENBlack software. Zoom-in sections were presented as single slices to indicate the subnuclear chromatin structures at the super-resolution level.

CRediT author statement

Y.M.S: Investigation, Validation, Formal analysis, Data Curation, Writing- Original draft preparation. **M.K.:** Formal analysis, Data Curation, Writing- Reviewing and Editing. **L.O.:** Investigation, Resources. **P.N. and J.M.:** Resources, Writing- Reviewing and Editing. **V.S. and A.H.:** Resources, Writing- Reviewing and Editing. **A.P.H and G.S.:** Supervision, Resources, Writing- Reviewing and Editing. **A.M.:** Conceptualization, Supervision, Resources, Funding acquisition, Writing- Reviewing and Editing.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and/or its supplementary information files.

Code availability

Referenced code used for analysis is available in a GitHub repository (https://github.com/437364/Repeat-based-holocentromeres-of-Luzula-sylvatica).

Supplementary files

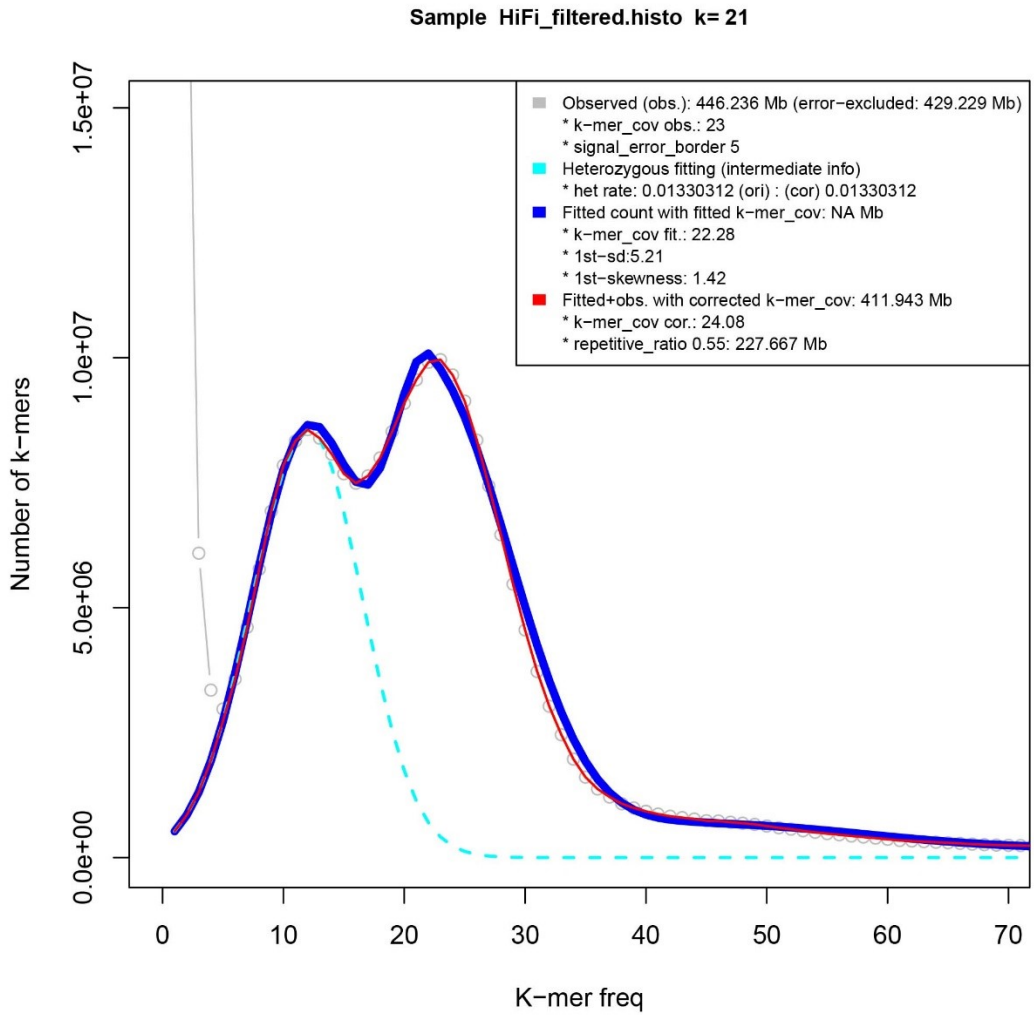


Fig. S1 K-mer-based genome size estimation in *Luzula sylvatica*.

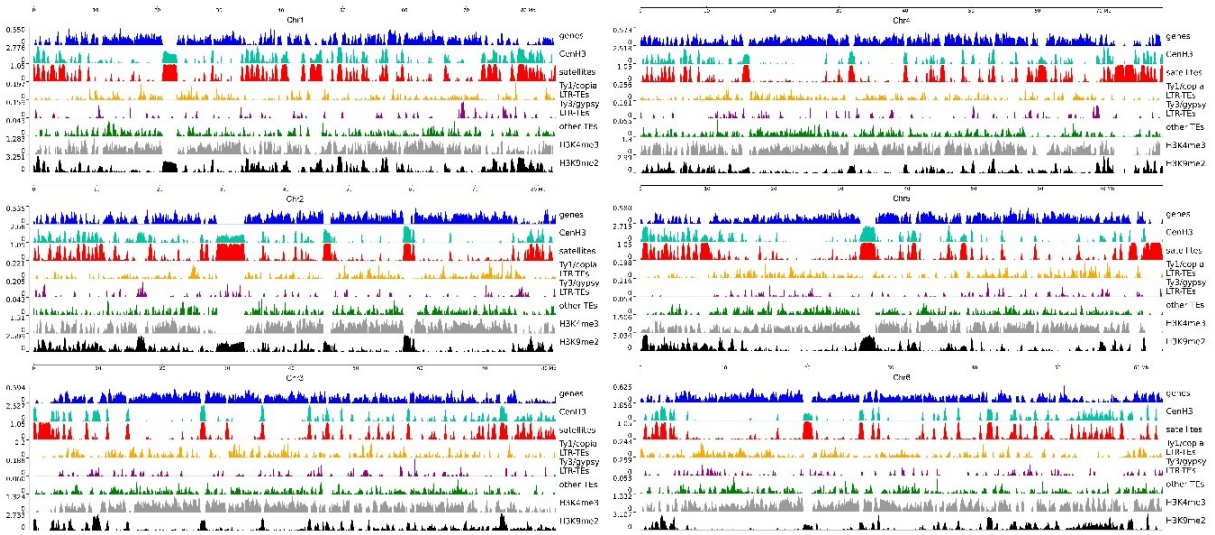


Fig. S2 Detailed view of *Luzula sylvatica* chromosomes showing the dispersed distribution of the main genomic features: CENH3, gene, tandem repeat, dispersed repeat and eu-heterochromatin mark densities, typical of holocentric chromosomes. Window sizes of 100kb.

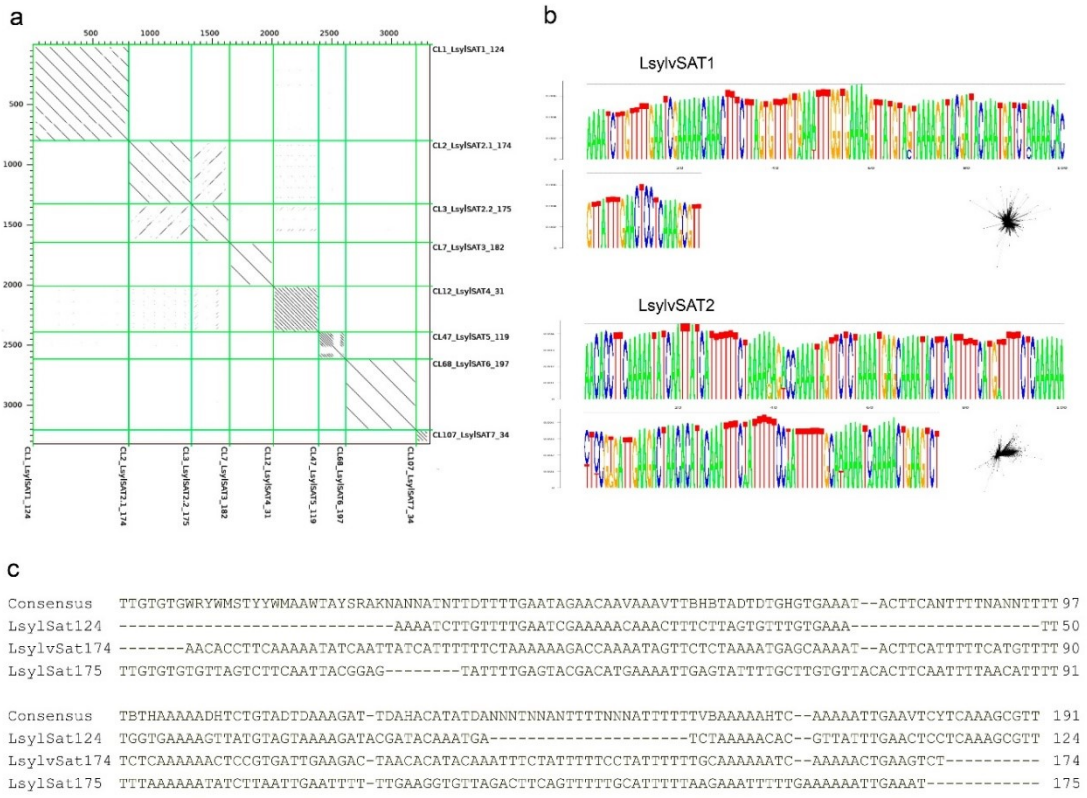


Fig. S3 *Luzula sylvatica* satellite. **a)** Dot plot showing similarities between groups of tandem repeats. **(b)** Sequence logo the most abundant satellite clusters LsylSatCL1 (Lusy1) and LsylSatCL2 (Lusy2). **(c)** Alignment between the two variants of Lusy2 (LsylvSat174 and LsylSat175) and Lusy1.

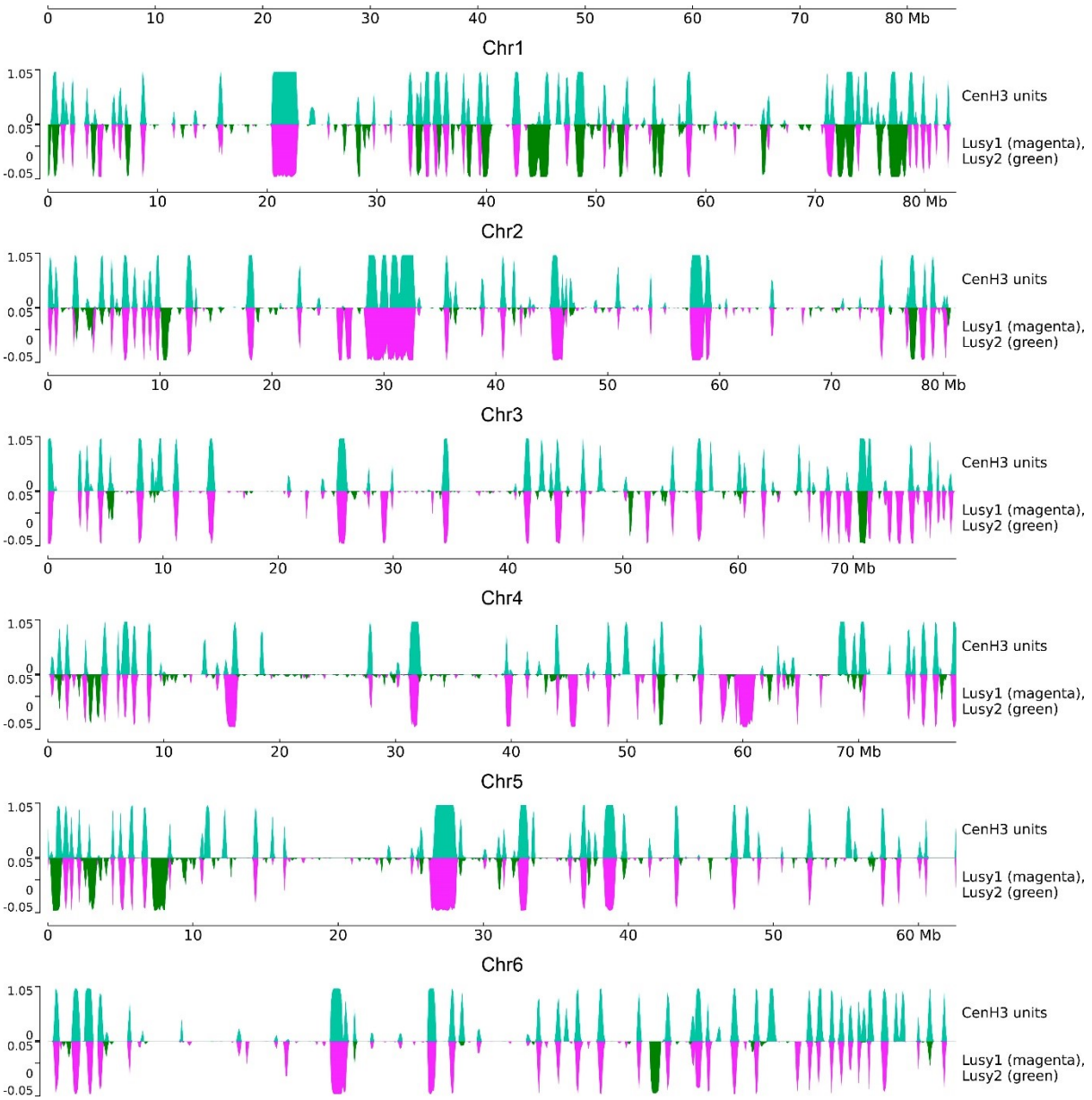


Fig. S4 Proportion of CENH3 units (blue), *Lusy1* (magenta) and *Lusy2* (green) arrays in 100kb windows.

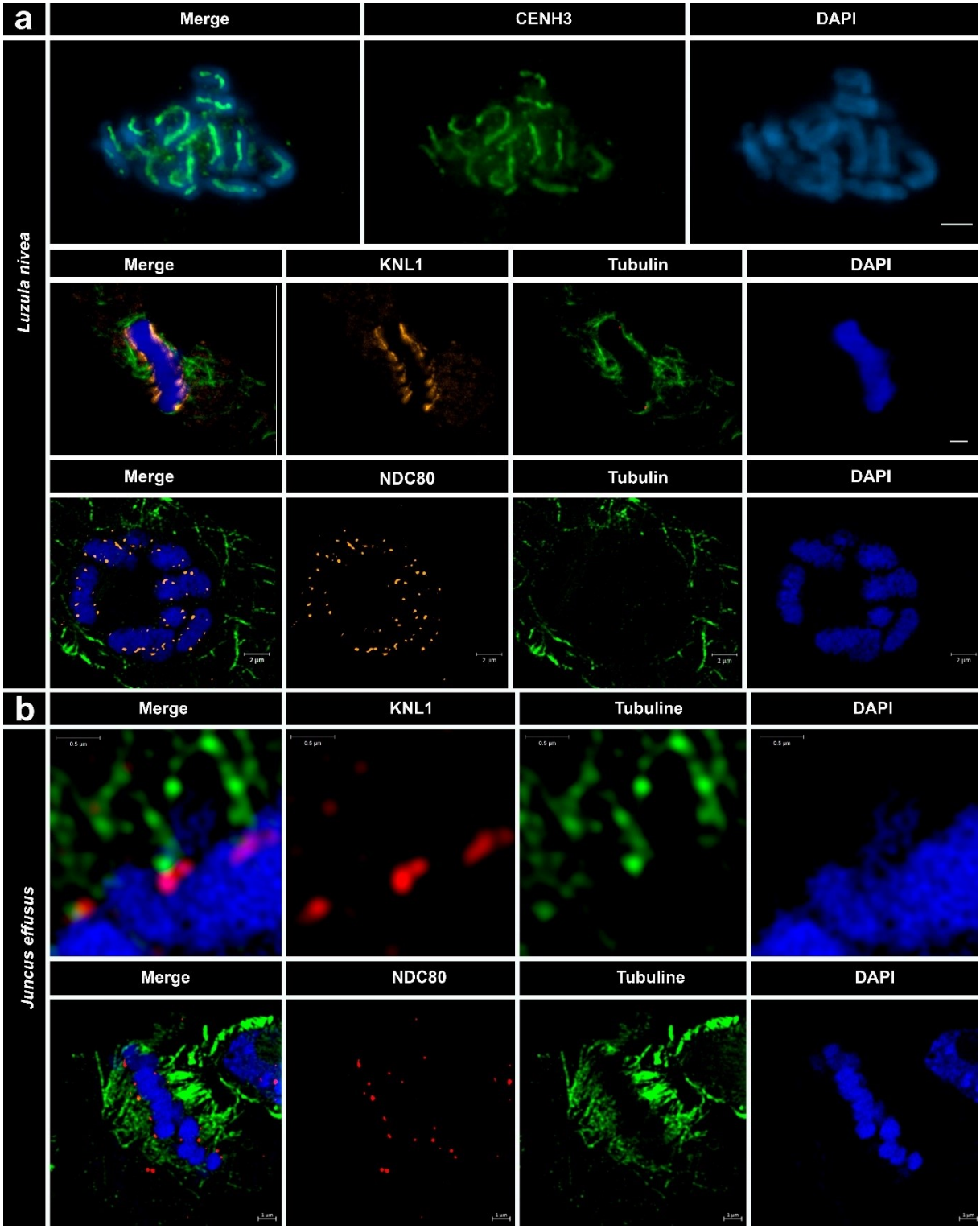


Fig. S5 Immunostaining of CENH3, KNL1 and NDC80 proteins in *L. nivea* metaphase **(a)**. **(b)** Immunostaining of KNL1 and NDC80 proteins in *Juncus effusus* chromosomes. The two proteins localize specifically to the surface of the centromere, where microtubules bind.

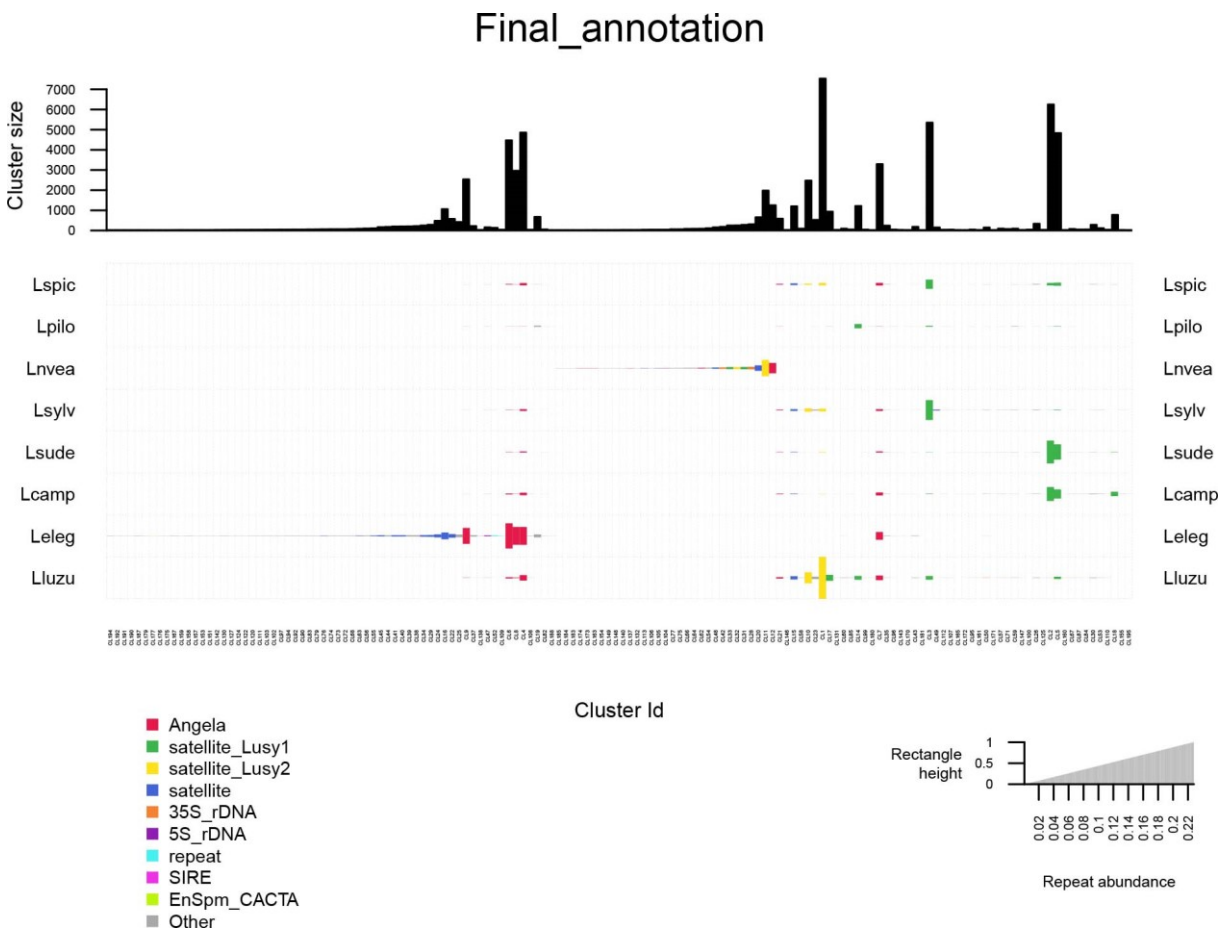


Fig. S6 Comparative abundance of the main types of repetitive sequences in *Luzula* species. The size of the rectangle is proportional to the genome abundance of that cluster for each species. The colors of the balls correspond to different repetitive sequence types. The proportion of each cluster was adjusted according to genome size.

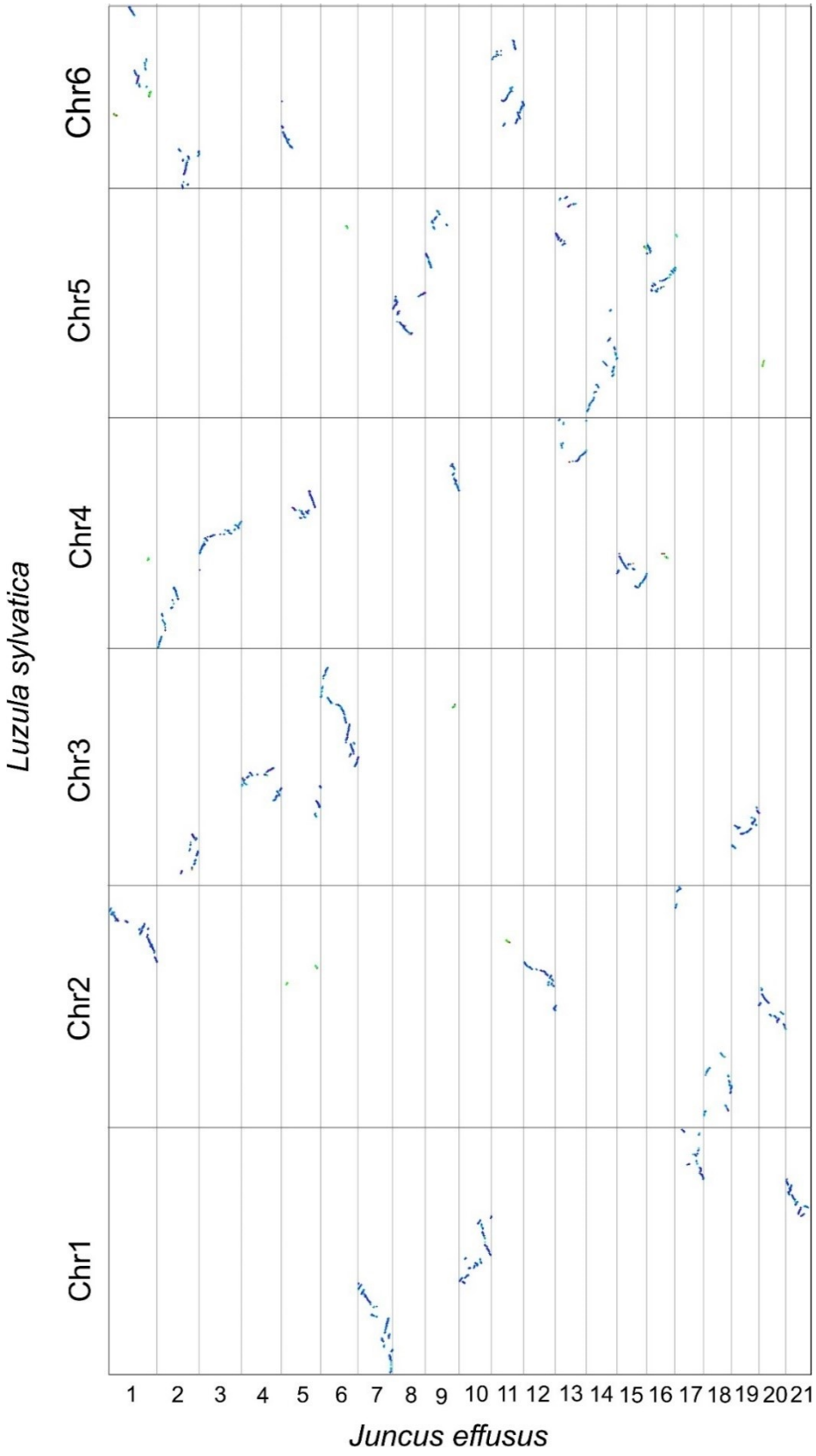


Fig. S7 Genome synteny patterns showing macroconserved blocks of *J. effusus* that are part of the *Luzula sylvatica* chromosomes.

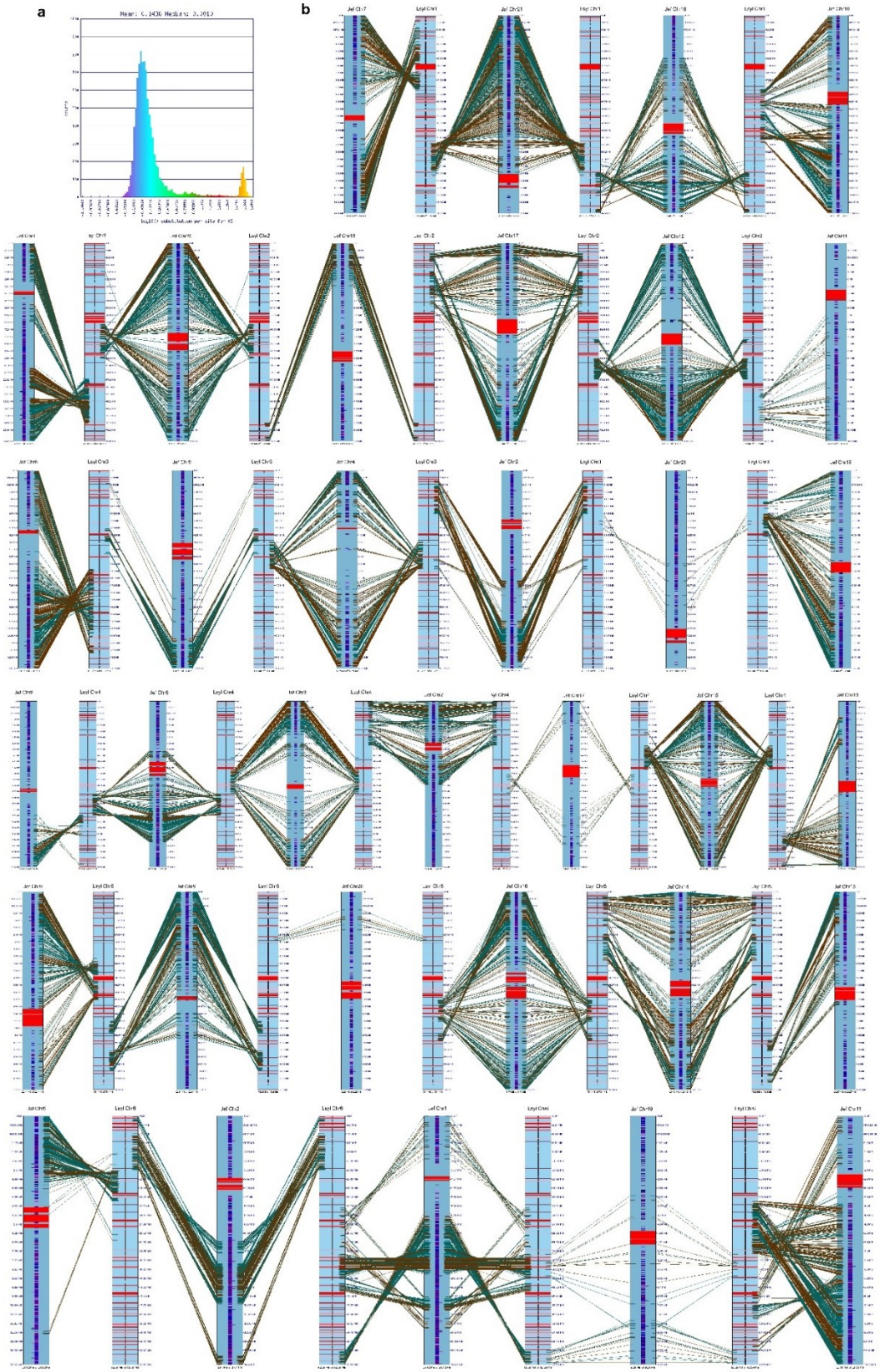


Fig. S8 *L. sylvatica* chromosomes showing the fusion of syntenic blocks of *J. effusus* (Jef). The synteny stops near the centromeric CENH3 (green) sites. Genes, CENH3 and telomere domains are annotated as blue-purple, green and black stripes, respectively.

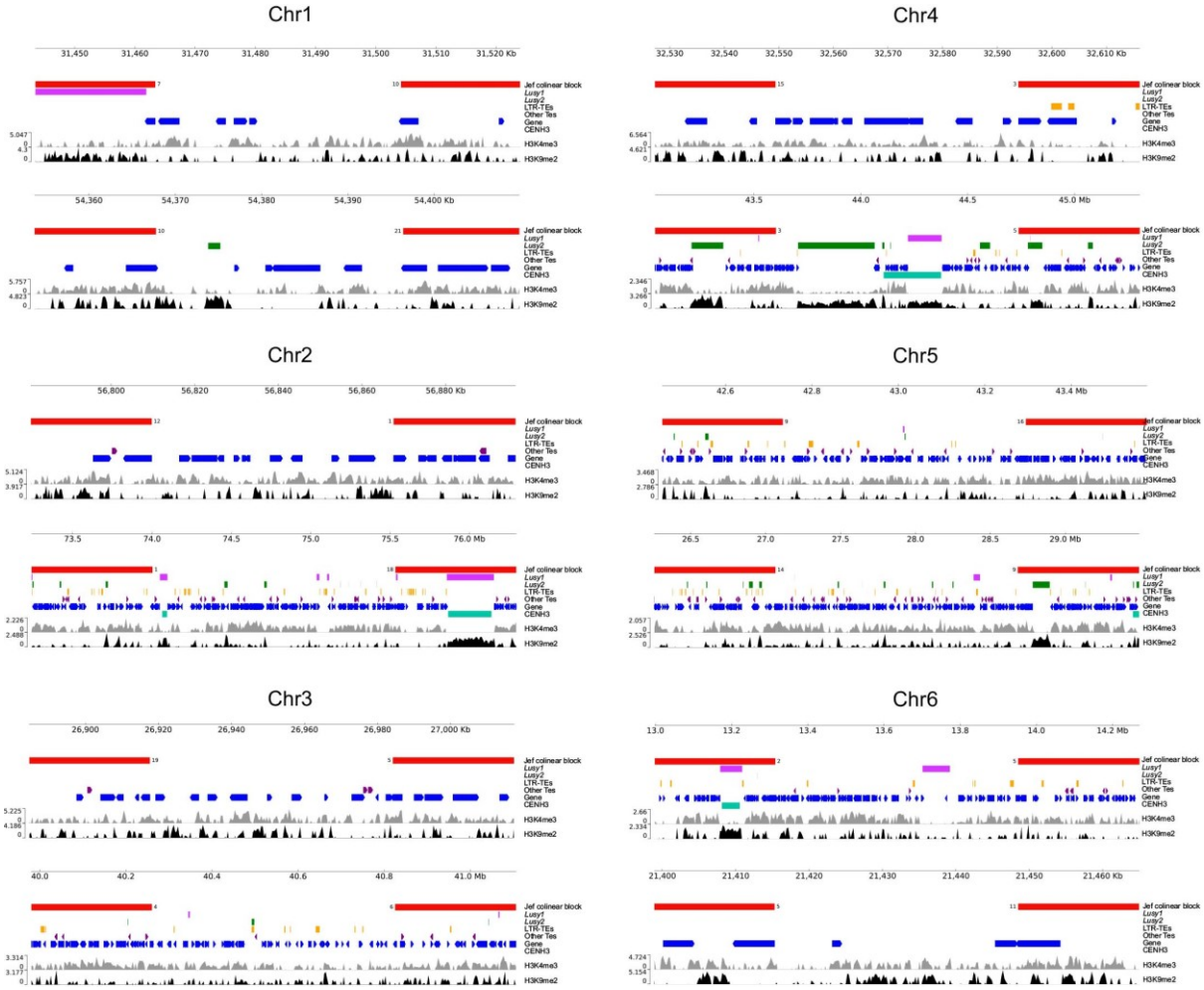


Fig. S9 Characterization of fusion sites in *L. sylvatica* genome. Similar fusion signatures are shared between some chromosomes, where regions enriched with gene, TE or *Lusy2* repeats are located either up or downstream of the fusion point.

Table S1 Characteristics of the *Luzula sylvatica* genome assembly

	Assembly	**Scaffold**
Total length of contigs (bp)	516,083,138	468,439,054
Largest contig	26,581,686	84,527,883
N50 (bp)	7,692,447	78,954,785
L50	21	4
L90	90	6
Number of unanchored scaffolds		729
GC content (%)		33.01
AT (%)		66.99
BUSCO		
Complete BUSCOs	1,501 (93.01%)	
Complete and single-copy BUSCOs	1,423 (88.2%)	
Complete and duplicated BUSCOs	78 (4.8%)	
Fragmented BUSCOs	34 (2%)	
Missing BUSCOs	79 (5%)	
Total BUSCO groups searched	1.614	
Average gene length (bp)	1.614	
Number of annotated genes	32.644	

Table S2 Monomer sequence of the satellite DNA of *Luzula sylvatica* genome

Name	**Abundance**	**Monomer size**	
Lusy1	25.1	124	AAAATCTTGTTTTGAATCGAAAAACAAACTTTCTTAGTGTTTGTGAAATT
Lusy2_1.1	3.61	175	TTGTGTGTGTTAGTCTTCAATTACGGAGTATTTTGAGTACGACATGAAAA
Lusy2_1.2	3.45	174	AACACCTTCAAAAATATCAATTATCATTTTTTCTAAAAAAGACCAAAATA
LsylSAT3	2.00	182	TATTTTTTGTTAATTCGGGTTCAAAATACATAATTTCCCAAATTCGTTCGG
LsylSAT4	1.05	31	TTTTTTGAAGAAAACCACAAACTAATTTTCT
LsylSAT5	0.06	119	TGATCGTATTCATCGGATAATTAACTATATTTTAGTAAATTCAATCGTAT
LsylSAT6	0.03	197	ATTGATTATTCATGAACTTCATTTGTCACCAATAAAATAGATACCAACAA
LsylSAT7	0.02	34	TGCAAATCCCGCCAAAAACACAATAGGCGGGATT

Table S3 Proportions of transposons identified in *Luzula sylvatica* genome assembly

LTR retrotransposon type	# in genome	% in genome	# in sat-free CENH3 units	% in sat-free CENH3 units
Class_I|LTR|Ty1/copia|Angela	1720	40.73	9	28.13
Class_I|LTR|Ty1/copia|Ivana	537	12.72	0	0.00
Class_I|LTR|Ty3/gypsy|non-chromovirus|OTA|Athila	516	12.22	19	59.38
Class_I|LTR|Ty1/copia|Ale	323	7.65	0	0.00
Class_I|LTR|Ty1/copia|SIRE	318	7.53	1	3.13
Class_I|LTR|Ty3/gypsy|chromovirus|Reina	293	6.94	2	6.25
Class_I|LTR|Ty1/copia|Tork	258	6.11	1	3.13
Class_I|LTR|Ty1/copia|Bianca	92	2.18	0	0.00
Class_I|LTR|Ty3/gypsy|chromovirus|Tekay	50	1.18	0	0.00
Class_I|LTR|Ty1/copia|Alesia	46	1.09	0	0.00
Class_I|LTR|Ty1/copia|TAR	44	1.04	0	0.00
Class_I|LTR|Ty1/copia|Ikeros	15	0.36	0	0.00
Class_I|LTR|Ty3/gypsy|chromovirus|CRM	11	0.26	0	0.00

Table S4 *Luzula* species, ENA codes and available genome size with their references used for comparative analyses

Species	ENA code	Name	1C (Mbp)	Reference
Luzula arcuata (Wahlenb.) Sw.	ERR5554955	Larcu	-	-
Luzula campestris (L.) DC.	ERR5529684	Lcamp	449.88	Bačič et al. 2016; 2019
Luzula elegans Lowe	ERX125774	Leleg	1506.12	Mukerjee and Sharma,1993
Luzula luzuloides (Lam.) Dandy & Wilmott	ERR5555243	Lluzu	880.20	Zonneveld 2019
Luzula multiflora subsp. f*rigida* (Buchenau) V.I.Krecz.	ERR5555376	Lmtsf	-	-
Luzula nivalis (Laest.) Spreng.	ERR5529777	Lniva	-	-
Luzula nivea (L.) DC.		Lnvea	880.20	-
Luzula parviflora (Ehrh.) Desv.	ERR5554954	Lparv	-	-
Luzula pilosa (L.) Willd.	ERR5554983	Lpilo	224.94	Smarda et al, 2019; Zonneveld 2019
Luzula spicata (L.) DC.	ERR5554743	Lspic	391.20	Smarda et al, 2019
Luzula sudetica (Willd.) Schult.	ERR5554826	Lsude	420.54	Bačič et al,,2019
Luzula sylvatica (Huds.) Gaudin	ERR5554782	Lsylv	469.44	Smarda et al, 2019
Luzula wahlenbergii Rupr.	ERR5554894	Lwahl	-	-

Table S5 Individual genome proportion in porcentage of the repetitive sequences in *Luzula* species

	Larcu	Lwahl	Lluzu	Lparv	Lniva	Lsylv	Lspic	Lcamp	Lsude	Lmtsf	Lpilo	Lnvea	Leleg
Tandem													
35S_rDNA	0.72	0.74	0.38	0.93	0.89	1.34	0.53	0.79	0.53	0.85	1.02	2.14	0.04
5S_rDNA	0.08	0.21	0.06	0.12	0.07	0.09	0.09	0.02	-	0.02	0.06	0.16	0.21
Satellite_Lusy1	15.89	22.10	9.44	12.82	13.66	25.1	20.02	33.68	47.41	23.65	13.35	3.33	-
Satellite_Lusy2	10.12	20.28	31.03	7.95	4.44	7.08	5.41	0.43	0.76	0.34	-	15.54	0.72
Other Satellites	1.69	1.67	4.07	1.87	1.45	3.15	2.56	0.78	0.78	1.18	18.72	1.21	9.69

LTR													
Ty1_copia/Alesia	0.13	0.05	-	0.11	0.16	-	-	-	-	-	-	0.03	0.02
Ty1_copia/Angela	21.49	16.04	8.84	25.45	22.5	5.87	10.49	10.85	3.69	12.49	1.21	12.36	37.61
Ty1_copia/Bianca	-	-	-	0.2	0.31	0.03	0.02	0.15	0.03	0.27	-	0.12	-
Ty1_copia/Ivana	0.26	0.14	0.02	0.34	0.38	0.13	0.06	0.02	-	0.04	-	0.19	2.26
Ty1_copia/SIRE	3.25	2.02	0.07	4.18	4.52	1.37	2.1	0.15	-	1.64	-	2.14	1.46
Ty1_copia/Tork	0.03	-	-	0.07	0.17	0.13	-	-	-	-	-	0.2	-
Ty3_gypsy/chromovirus/Tekay	-	0.02	0.05	0.38	0.29	-	-	-	-	-	-	1.1	0.94
Ty3_gypsy/non-chromovirus/Athila	1.43	-	-	0.16	1.78	0.27	0.07	0.11	-	0.17	-	0.04	-
No-LTR													
LINE	-	-	-	-	-	-	-	-	-	-	-	-	0.3
Pararetrovirus	-	-	-	-	-	-	-	-	-	-	-	0.01	-
TIR/EnSpm_CACTA	2.34	0.03	0.03	1.72	2.02	-	-	-	-	-	-	2.19	2.16
TIR/hAT	0.29	-	-	-	-	-	-	-	-	-	-	-	0.03
TIR/MuDR_Mutator	-	-	-	0.04	0.03	-	-	-	-	-	-	-	0.22
Unclassified	2.78	2.99	4.89	4.44	3.34	3.67	2.68	2.43	1.71	3.72	0.88	1.91	7.79
Total	60.49	66.29	58.88	60.78	56.01	48.23	44.03	49.41	54.91	44.37	35.24	42.67	63.45

Table S6 Comparative genome proportion of the repetitive sequences in *Luzula* species

Elem	cluster	Larcu	Lwahl	Lluzu	Lparv	Lniva	Lsylv	Lspic	Lcamp	Lsude	Lmtsf	Lpilo	Lnvea	Leleg
Lusy2	1	14.85	14.59	24.20	5.56	3.23	5.78	3.88	0.45	0.63	0.28	0.01	1.22	0.00
Lusy1	2	6.62	8.68	2.00	4.00	3.53	17.10	10.31	0.32	0.23	0.48	1.71	0.00	0.00

Ty1-copia/Angela	3	11.74	4.07	2.75	7.08	6.34	1.57	2.94	2.45	0.89	2.90	0.48	2.78	5.06
Ty1-copia/Angela	4	10.88	3.80	2.21	6.51	5.88	1.59	2.68	2.43	1.08	3.06	0.49	2.30	4.29
Lusy1	5	2.98	3.18	0.74	1.82	1.31	0.42	3.01	8.04	14.03	5.22	0.67	0.00	0.00
Lusy1	6	0.02	0.00	0.02	0.02	0.00	0.02	1.38	8.57	14.84	5.45	0.02	0.00	0.00
Ty1-copia/Angela	7	6.74	2.20	1.71	3.72	3.55	0.96	1.60	1.42	0.70	1.96	0.23	0.00	5.57
Lusy1	8	7.51	3.06	1.79	2.55	3.27	0.00	0.00	0.00	0.00	0.00	8.55	0.00	0.00
Lusy1	9	0.04	0.08	0.04	0.02	0.02	0.01	0.99	5.30	8.31	3.15	0.06	0.00	0.00
Lusy1	10	6.93	2.08	2.97	1.82	2.81	0.00	0.00	0.01	0.00	0.01	0.08	0.00	0.00
All/repeat/satellite	11	1.34	0.82	1.40	0.62	0.91	1.76	1.80	0.61	0.48	0.76	0.01	0.00	0.00
Ty1-copia/Angela	12	1.64	0.88	0.75	1.83	1.21	0.42	0.76	0.65	0.37	1.03	0.22	0.00	0.00
Lusy1	14	0.01	0.00	0.00	0.00	0.00	0.00	0.19	4.19	0.27	3.12	0.00	0.07	0.00
35S_rDNA	19	0.62	0.33	0.15	0.31	0.24	0.47	0.15	0.17	0.18	0.22	0.17	0.09	0.02
All/repeat/satellite	23	0.30	0.19	0.08	0.20	0.21	0.18	0.14	0.08	0.15	0.10	0.40	0.00	0.00
All/repeat/satellite	24	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.02
Ty1-copia/Angela	25	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.74
All/repeat/satellite	26	0.04	0.02	0.04	0.12	0.11	0.13	0.24	0.04	0.06	0.13	0.46	0.00	0.22
35S_rDNA	28	0.14	0.09	0.05	0.09	0.06	0.24	0.09	0.14	0.10	0.16	0.22	0.00	0.01
All/repeat/satellite	34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.94
All/repeat/satellite	35	0.00	0.00	0.00	0.00	0.00	0.85	0.00	0.00	0.00	0.00	0.00	0.00	0.00
All/repeat/satellite	36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.86
35S_rDNA	37	0.24	0.11	0.06	0.06	0.06	0.09	0.03	0.02	0.04	0.09	0.00	0.00	0.02
All/repeat/satellite	38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.81
5S_rDNA	41	0.13	0.18	0.03	0.11	0.05	0.05	0.03	0.02	0.00	0.00	0.01	0.00	0.10
Ty1-copia/SIRE	46	0.20	0.09	0.02	0.09	0.06	0.02	0.02	0.02	0.00	0.03	0.00	0.11	0.06
Ty1-copia/Angela	49	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.18	0.58
Lusy1	60	0.04	0.11	0.03	0.03	0.00	0.02	0.01	0.07	0.01	0.07	0.00	0.13	0.00
Ty1-copia	61	1.25	0.54	0.39	0.96	0.89	0.12	0.56	0.83	0.17	0.67	0.05	0.00	0.00
Ty1-copia/SIRE	69	0.11	0.02	0.02	0.07	0.09	0.02	0.02	0.01	0.01	0.01	0.00	0.00	0.02
All/repeat/satellite	72	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.37

All/repeat/satellite	78	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.02	0.00	0.27
Ty1-copia/Angela	79	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.33
Ty1-copia/Angela	89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.28
Ty1-copia/SIRE	93	0.07	0.02	0.03	0.05	0.05	0.02	0.01	0.00	0.00	0.01	0.00	0.00	0.00
Ty1-copia/Angela	104	0.03	0.05	0.01	0.10	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.08	0.00
Ty1-copia/Ivana	106	0.09	0.02	0.00	0.03	0.02	0.01	0.03	0.00	0.01	0.02	0.00	0.00	0.00
Lusy2	109	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.24
All/repeat/satellite	112	0.00	0.05	0.00	0.16	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Ty1-copia/SIRE	113	0.07	0.02	0.05	0.03	0.03	0.00	0.02	0.00	0.00	0.01	0.00	0.00	0.00
Ty1-copia/Angela	122	0.02	0.01	0.01	0.02	0.02	0.02	0.05	0.01	0.03	0.02	0.00	0.00	0.01
Ty1-copia/SIRE	128	0.07	0.00	0.02	0.02	0.03	0.01	0.00	0.03	0.00	0.01	0.01	0.00	0.00
Ty1-copia/Ivana	130	0.04	0.00	0.03	0.03	0.01	0.02	0.00	0.01	0.01	0.04	0.01	0.00	0.00
All/repeat/satellite	138	0.00	0.00	0.00	0.19	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
All/repeat/satellite	161	0.16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

Table S7 Conserved collinear blocks between *Luzula sylvatica* and *Juncus effusus* genome

Juncus effusus **Chr**	**Start**	**End**	***Luzula sylvatica*** **Chr**	**Start**	**End**
1	51433	3183432	3	70363021	75329258
1	10100862	16389583	3	56270627	70062240
1	5428881	6590660	3	69713915	70969951
1	4955851	5219128	3	70246702	70357775
1	6419584	8660074	4	59222634	62351670
1	8666460	14017739	4	30087466	40523834
1	359973	2826449	4	24704582	29206939
1	12094647	12676081	4	42831408	43911396
1	12293132	12500531	4	40588465	41503016
1	12233790	12280498	4	42165724	42373860
2	7999654	9110949	2	4224848	5625944
2	10929431	11237284	2	12038264	12900212
2	11260392	11512518	2	14000663	14325542
2	11958403	13338874	2	15733364	17813551
2	11552150	13910537	2	6017075	12033933
2	8834905	9131821	4	8110353	9886682
2	8787489	8837599	4	9544971	9732269
2	7268584	7966362	4	12953057	13570334
2	8976282	10913835	4	4292581	11192455
2	8512797	8767572	4	198622	1314837
2	8911312	9008149	4	2211188	2322502
2	13913258	14022952	4	11201127	11528901
2	10312728	10432051	4	1319560	1398290
2	14184276	14285240	4	12030284	12201931
2	14339213	14385948	4	12803310	12869632
2	14115374	14169337	4	12528395	12735463
2	14490	7008521	5	235042	21651067
2	7023641	7239531	5	14304402	17229375
3	79222	5052348	5	32591660	39048943
3	6299618	14303370	5	38930858	43639650
3	27144	1274802	5	26910056	29561462
4	11000514	13635366	2	29210091	33438038
4	7337778	11376941	2	37991307	40885670
4	5426131	5545754	2	37931095	38309427
4	3577950	3741224	2	37803680	37926686
4	319462	1744287	2	34642808	36986752
4	1914940	3263793	2	37072566	38852497
4	213107	317540	2	34459587	34625887
4	142996	190660	2	34192929	34306967
5	12929586	13254788	2	33459312	34138975

5	11775608	13003276	2	26940284	29189849
5	11306660	11768069	2	23753302	24718805
5	332362	1784009	3	49142220	51724044
5	90464	4134568	4	12803630	22397121
5	37612	230146	4	29807983	31297385
5	11528601	11683773	4	13581124	13886368
5	9162613	11291951	5	48426257	54179416
5	5750861	6975518	5	44411639	46434257
5	8906078	9151684	5	47015822	47261551
5	3959727	4508345	5	47406591	48404276
5	8289681	8888534	5	46350917	46813061
5	7055974	7986974	5	44143621	46286297
5	6736347	6830307	5	46955866	47011727
5	215266	2526380	6	62266276	65532391

References

Allshire, R. C., & Karpen, G. H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks?. Nature Reviews Genetics, 9(12), 923-937.

Bozek, M., Leitch, A. R., Leitch, I. J., Záveská Drábková, L., & Kuta, E. (2012). Chromosome and genome size variation in Luzula (Juncaceae), a genus with holocentric chromosomes. Botanical Journal of the Linnean Society, *170*(4), 529-541.

Burchardt, P., Buddenhagen, C. E., Gaeta, M. L., Souza, M. D., Marques, A., & Vanzela, A. L. (2020). Holocentric karyotype evolution in Rhynchospora is marked by intense numerical, structural, and genome size changes. Frontiers in Plant Science, 11, 536507.

Cappelletti, E., Piras, F. M., Sola, L., Santagostino, M., Abdelgadir, W. A., Raimondi, E., ... & Giulotto, E. (2022). Robertsonian fusion and centromere repositioning contributed to the formation of satellite-free centromeres during the evolution of zebras. *Molecular Biology and Evolution*, *39*(8), msac162.

Christenhusz, M. J., & Byng, J. W. (2016). The number of known plants species in the world and its annual increase. *Phytotaxa*, 261(3), 201-217.

Cicconardi, F., Lewis, J. J., Martin, S. H., Reed, R. D., Danko, C. G., & Montgomery, S. H. (2021). Chromosome fusion affects genetic diversity and evolutionary turnover of

functional loci but consistently depends on chromosome size. *Molecular Biology and Evolution*, *38*(10), 4449-4462.

Costa, L., Marques, A., Buddenhagen, C. E., Pedrosa-Harand, A., & Souza, G. (2023). Investigating the diversification of holocentromeric satellite DNA Tyba in Rhynchospora (Cyperaceae). *Annals of Botany*, 131(5), 813-825.

Dale, R. K., Pedersen, B. S., & Quinlan, A. R. (2011). Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics, 27(24), 3423-3424.

Drábková, L. Z., & Vlček, Č. (2010). Molecular phylogeny of the genus Luzula DC.(Juncaceae, Monocotyledones) based on plastome and nuclear ribosomal regions: a case of incongruence, incomplete lineage sorting and hybridisation. Molecular Phylogenetics and Evolution, 57(2), 536-551.

Drinnenberg, I. A., deYoung, D., Henikoff, S., & Malik, H. S. (2014). Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. *Elife*, *3*, e03676.

Elliott, T. L., & Davies, T. J. (2019). Phylogenetic attributes, conservation status and geographical origin of species gained and lost over 50 years in a UNESCO Biosphere Reserve. *Biodiversity and Conservation*, *28*, 711-728.

Escudero, M., Márquez-Corro, J. I., & Hipp, A. L. (2016). The phylogenetic origins and evolutionary history of holocentric chromosomes. Systematic Botany, 41(3), 580-585.

Guerra, M. (2016). Agmatoploidy and symploidy: a critical review. Genetics and Molecular Biology, 39, 492-496.

Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072-1075.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357-362.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(03), 90-95.

Heckmann, S., Macas, J., Kumke, K., Fuchs, J., Schubert, V., & Ma, L. (2013). Nová k P, Neumann P, Taudien S, Platzer M, et al. The holocentric species Luzula elegans shows

interplay between centromere and largescale genome organization. Plant J, 73(4), 555-565.

Heckmann, S., Schroeder-Reiter, E., Kumke, K., Ma, L., Nagaki, K., Murata, M., ... & Houben, A. (2011). Holocentric chromosomes of Luzula elegans are characterized by a longitudinal centromere groove, chromosome bending, and a terminal nucleolus organizer region. Cytogenetic and genome research, 134(3), 220-228.

Hofstatter, P. G., Thangavel, G., Lux, T., Neumann, P., Vondrak, T., Novak, P., ... & Marques, A. (2022). Repeat-based holocentromeres influence genome architecture and karyotype evolution. Cell, 185(17), 3153-3168.

Jankowska, M., Fuchs, J., Klocke, E., Fojtová, M., Polanská, P., Fajkus, J., ... & Houben, A. (2015). Holokinetic centromeres and efficient telomere healing enable rapid karyotype evolution. Chromosoma, 124, 519-528.

Kasinathan, S., & Henikoff, S. (2018). Non-B-form DNA is enriched at centromeres. Molecular biology and evolution, 35(4), 949-962.

Kirschner, J., Balslev, H., Clemants, S. E., Ertter, B., Alvarez, M. C. F. C., Hämet-Ahti, L., ... & Wilson, K. L. (2002). Juncaceae 2: Juncus subg. Juncus, Species Plantarum: Flora of the World, 7, 1-336.

Kuo, Y. T., Câmara, A. S., Schubert, V., Neumann, P., Macas, J., Melzer, M., ... & Houben, A. (2023). Holocentromeres can consist of merely a few megabase-sized satellite arrays. Nature Communications, 14(1), 3502.

Kuta, E., Bohanec, B., Dubas, E., Vizintin, L., & Przywara, L. (2004). Chromosome and nuclear DNA study on Luzula-a genus with holokinetic chromosomes. Genome, 47(2), 246-256.

Lucek, K., Augustijnen, H., & Escudero, M. (2022). A holocentric twist to chromosomal speciation?. Trends in Ecology & Evolution, 37: 655-662

Ma, W., Schubert, V., Martis, M. M., Hause, G., Liu, Z., Shen, Y., ... & Houben, A. (2016). The distribution of α-kleisin during meiosis in the holocentromeric plant Luzula elegans. *Chromosome Research*, *24*, 393-405.

Macas, J., Ávila Robledillo, L., Kreplak, J., Novák, P., Koblížková, A., Vrbová, I., ... & Neumann, P. (2023). Assembly of the 81.6 Mb centromere of pea chromosome 6

elucidates the structure and evolution of metapolycentric chromosomes. PLoS Genetics, 19(2), e1010633.

Mandrioli, M., & Manicardi, G. C. (2020). Holocentric chromosomes. PLoS Genetics, 16(7), e1008918.

Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics, 27(6), 764-770.

Marques, A., Ribeiro, T., Neumann, P., Macas, J., Novák, P., Schubert, V., ... & Houben, A. (2015). Holocentromeres in Rhynchospora are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. Proceedings of the National Academy of Sciences, 112(44), 13633-13638.

Márquez-Corro, J. I., Martín-Bravo, S., Pedrosa-Harand, A., Hipp, A. L., Luceño, M., & Escudero, M. (2019). Karyotype evolution in holocentric organisms. eLS, 1-7.

McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. Python for high performance and scientific computing, 14(9), 1-9.

Melters, D. P., Paliulis, L. V., Korf, I. F., & Chan, S. W. (2012). Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. Chromosome Research, 20, 579-593.

Nagaki, K., Kashihara, K., & Murata, M. (2005). Visualization of diffuse centromeres with centromere-specific histone H3 in the holocentric plant Luzula nivea. The Plant Cell, 17(7), 1886-1893.

Naughton, C., & Gilbert, N. (2020). Centromere chromatin structure–Lessons from neocentromeres. *Experimental Cell Research*, *389*(2), 111899.

Neumann, P., Navratilova, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerova, V., Chocholova, E., ... & Macas, J. (2012). Stretching the rules: monocentric chromosomes with multiple centromere domains. PLoS genetics, 8(6), e1002777.

Neumann, P., Oliveira, L., Jang, T. S., Novák, P., Koblížková, A., Schubert, V., ... & Macas, J. (2023). Disruption of the standard kinetochore in holocentric *Cuscuta* species. *Proceedings of the National Academy of Sciences*, *120*(21), e2300877120.

Ning, Y., Li, Y., Dong, S. B., Yang, H. G., Li, C. Y., Xiong, B., ... & Xia, X. F. (2023). The chromosome-scale genome of Kobresia myosuroides sheds light on karyotype evolution and recent diversification of a dominant herb group on the Qinghai-Tibet Plateau. DNA Research, 30(1), dsac049.

Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P., & Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic acids research*, *45*(12), e111-e111.

Oliveira, L., Neumann, P., Jang, T. S., Klemme, S., Schubert, V., Koblížková, A., ... & Macas, J. (2020). Mitotic spindle attachment to the holocentric chromosomes of Cuscuta europaea does not correlate with the distribution of CENH3 chromatin. Frontiers in Plant Science, 10, 1799.

Plohl, M., Meštrović, N., & Mravinac, B. (2014). Centromere identity from the DNA point of view. Chromosoma, 123, 313-325.

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Research. 2016 Apr 13:gkw257.

Ribeiro, T., Marques, A., Novák, P., Schubert, V., Vanzela, A. L., Macas, J., ... & Pedrosa-Harand, A. (2017). Centromeric and non-centromeric satellite DNA organisation differs in holocentric Rhynchospora species. Chromosoma, 126, 325-335.

Senaratne, A. P., Cortes-Silva, N., & Drinnenberg, I. A. (2022, January). Evolution of holocentric chromosomes: drivers, diversity, and deterrents. In Seminars in Cell & Developmental Biology. Academic Press.

Šatović-Vukšić, E., & Plohl, M. (2023). Satellite DNAs—From Localized to Highly Dispersed Genome Components. *Genes*, *14*(3), 742.

Schubert, V., Neumann, P., Marques, A., Heckmann, S., Macas, J., Pedrosa-Harand, A., ... & Houben, A. (2020). Super-resolution microscopy reveals diversity of plant centromere architecture. *International journal of molecular sciences*, *21*(10), 3488.

Stovner, E. B., & Sætrom, P. (2019). epic2 efficiently finds diffuse domains in ChIP-seq data. Bioinformatics, 35(21), 4392-4393.

Sun, H., Ding, J., Piednoël, M., & Schneeberger, K. (2018). findGSE: estimating genome size variation within human and Arabidopsis using k-mer frequencies. Bioinformatics, 34(4), 550-557.

Sweeten, A. (2023). *ModDotPlot*. GitHub. https://github.com/marbl/ModDotPlot

Talbert, P. B., & Henikoff, S. (2020). What makes a centromere?. *Experimental cell research*, *389*(2), 111895.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., ... & Van Mulbregt, P. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. Nature methods, 17(3), 261-272.

Wang, J., Zi, H., Wang, R., Liu, J., Wang, H., Chen, R., ... & Zong, J. (2021). A high-quality chromosome-scale assembly of the centipedegrass [Eremochloa ophiuroides (Munro) Hack.] genome provides insights into chromosomal structural evolution and prostrate growth habit. *Horticulture Research*, *8*.

Waskom, M. L. (2021). Seaborn: statistical data visualization. Journal of Open Source Software, 6(60), 3021.

Wong, C. Y. Y., Ling, Y. H., Mak, J. K. H., Zhu, J., & Yuen, K. W. Y. (2020). Lessons from the extremes: Epigenetic and genetic regulation in point monocentromere and holocentromere establishment on artificial chromosomes. Experimental Cell Research, 390(2), 111974.

Zedek, F., & Bureš, P. (2016). Absence of positive selection on CenH3 in Luzula suggests that holokinetic chromosomes may suppress centromere drive. *Annals of Botany*, *118*(7), 1347-1352.

5 CONSIDERAÇÕES FINAIS

- O gênero *Juncus* apresenta cariótipos diversos, menos variáveis do que reportado na literatura, com números cromossômicos predominantemente $2n = 40$ e um provável cariótipo ancestral $x = 20$.
- Bandas terminais CMA^+ observadas nas espécies estudadas correspondentes aos sítios 35 rDNA, e bandas pericentroméricas adicionais observadas em *J. capitatus*, *J. subsecundus* e *J. compressus*, sugerem um enriquecimento de outros tipos de sequências repetidas.
- *Juncus* apresenta uma clara incongruência entre topologias nucleares (DNAr) e plastidiais, o que parece ser reflexo de eventos antigos de hibridação. Isso pôde ser detectado tanto em abordagens filogenômicas quanto em filogenias tradicionais com sequenciamento Sanger.
- Em *Juncus*, a topologia do DNAr se ajusta melhor com a evolução dos repeatomas (topologia repeat-based) e com as classificações taxonômicas em nível de secção. Assim, a fração repetitiva dos genomas (abundância/ similaridade dos repeats e dados brutos de NGS) parecem bons marcadores de relações de parentesco.
- A evolução dos repeatomas das espécies de *Juncus* é influenciado pela expansão de repeats, pincipalmente de elementos do tipo Ty1-copia/Angela, e a dinâmica desses elementos estão impulsionando uma diferenciação no tamanho do genoma e nas relações filogenéticas
- *Luzula sylvatica* é uma especie com holocentrômeros baseados em repeats, como observado para o gênero *Rhynchospora* e *Chionographis*.
- As fusões de cromossomos monocêntricos de espécies ancestrais de *Juncus* ($2n = 40$) parecem estar envolvidas na redução do número cromossômico em espécies holocêntricas de *Luzula*.
- Os dados gerados aqui permitiram propor um modelo para a origem da holocentricidade: Após múltiplas fusões de cromossomos monocêntricos inteiros (*Juncus*-like), os domínios centroméricos foram inicialmente conservados nos cromossomos maiores (*Luzula*-like), formando cromossomas policêntricos.

Subsequentemente, a expansão desses repeats pode ter ampliado a extensão do domínio centromérico levando à condição holocêntrica.

REFERÊNCIAS

BALSLEV, Henrik. Juncaceae. **Flora Neotropica**, p. 1-167, 1996

BUREŠ, Petr; ZEDEK, František; MARKOVÁ, Michaela. Holocentric chromosomes. In: **Plant genome diversity volume 2**. Springer, Vienna, 2013. p. 187-208.

BOVERI, Theodor. Zellen-Studien: Die Befruchtung und Teilung des Eies von Ascaris megalocephala. G. **Fischer**, 1888.

CHRISTENHUSZ, Maarten JM; BYNG, James W. The number of known plants species in the world and its annual increase. **Phytotaxa**, v. 261, n. 3, p. 201-217, 2016.

CUACOS, Maria et al. Atypical centromeres in plants—what they can tell us. **Frontiers in plant science**, v. 6, p. 913, 2015.

CONESA, Ana et al. A survey of best practices for RNA-seq data analysis. **Genome biology**, v. 17, n. 1, p. 13, 2016.

DARLINGTON, Cyril Dean et al. Chromosome atlas of flowering plants. **Chromosome atlas of flowering plants.**, n. 2nd Ed, 1956.

DRÁBKOVÁ, Lenka; KIRSCHNER, Jan; VLČEK, Čestmír. Phylogenetic relationships within *Luzula* DC. and *Juncus* L. (Juncaceae): A comparison of phylogenetic signals of *trnL-trnF* intergenic spacer, *trnL* intron and *rbcL* plastome sequence data. **Cladistics**, v. 22, n. 2, p. 132-143, 2006.

DRINNENBERG, Ines A. et al. Recurrent loss of CENH3 is associated with independent transitions to holocentricity in insects. **Elife**, v. 3, p. e03676, 2014.

ESCUDERO, Marcial et al. Karyotypic changes through dysploidy persist longer over evolutionary time than polyploid changes. **PloS one**, v. 9, n. 1, 2014.

ESCUDERO, Marcial; MÁRQUEZ-CORRO, J. Ignacio; HIPP, Andrew L. The phylogenetic origins and evolutionary history of holocentric chromosomes. **Systematic Botany**, v. 41, n. 3, p. 580-585, 2016.

FLACH, M. Diffuse centromeres in a dicotyledonous plant. **Nature**, v. 209, n. 5030, p. 1369-1370, 1966.

FLEMMING, Walther. Zellsubstanz, kern und zelltheilung. **FCW Vogel**, 1882.

GLICK, Lior; MAYROSE, Itay. 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.

GONG, Zhiyun et al. Repeatless and repeat-based centromeres in potato: implications for centromere evolution. **The Plant Cell**, v. 24, n. 9, p. 3559-3574, 2012.

GROPPO JR, MILTON; PIRANI, José Rubens. Flora da Serra do Cipó, Minas Gerais: Juncaceae. **Boletim de Botânica da Universidade de São Paulo**, p. 25-27, 2004.

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

GUERRA, Marcelo; RIBEIRO, Tiago; FELIX, Leonardo P. Monocentric chromosomes in *Juncus* (Juncaceae) and implications for the chromosome evolution of the family. **Botanical Journal of the Linnean Society**, v. 191, n. 4, p. 475-483, 2019.

HÅKANSSON, ARTUR. Holocentric chromosomes in *Eleocharis*. **Hereditas**, v. 44, n. 4, p. 531-540, 1958.

HARRIMAN, Neil A.; REDMOND, Darrell. Somatic chromosome numbers for some North American species of *Juncus* L. **Rhodora**, v. 78, n. 816, p. 727-738, 1976.

HECKMANN, Stefan et al. The holocentric species *Luzula elegans* shows interplay between centromere and large-scale genome organization. **The Plant Journal**, v. 73, n. 4, p. 555-565, 2013.

HECKMANN, Stefan et al. The E2F transcription factor family regulates CENH3 expression in *Arabidopsis thaliana*. **The Plant Journal**, v. 68, n. 4, p. 646-656, 2011.

HENIKOFF, Steven; MALIK, Harmit S. Centromeres: selfish drivers. **Nature**, v. 417, n. 6886, p. 227-227, 2002.

HIPP, Andrew L.; ROTHROCK, Paul E.; ROALSON, Eric H. The evolution of chromosome arrangements in *Carex* (Cyperaceae). **The Botanical Review**, v. 75, n. 1, p. 96-109, 2009.

JANKOWSKA, Maja et al. Holokinetic centromeres and efficient telomere healing enable rapid karyotype evolution. **Chromosoma**, v. 124, n. 4, p. 519-528, 2015.

JONES, Eleanor et al. The Juncaceae-Cyperaceae interface: a combined plastid sequence analysis. **Aliso: A Journal of Systematic and Evolutionary Botany**, v. 23, n. 1, p. 55-61, 2007.

KANESAKI, Yu et al. Identification of centromere regions in chromosomes of a unicellular red alga, *Cyanidioschyzon merolae*. **FEBS letters**, v. 589, n. 11, p. 1219-1224, 2015.

KIRSCHNER, J. et al. Juncaceae 1: *Rostkovia* to *Luzula*, Species Plantarum. In: Flora World Part 6, pp. 1– 237. ABRS, Canberra, Australia. 2002.

LEE, Hye-Ran et al. Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. **Proceedings of the National Academy of Sciences**, v. 102, n. 33, p. 11793-11798, 2005.

LEE, Jiyoung et al. Comparing time series transcriptome data between plants using a network module finding algorithm. **Plant methods**, v. 15, n. 1, p. 61, 2019.

LÖVE, A.; LÖVE, Doris. Cytotaxonomical studies on boreal plants. II. Some notes on the chromosome numbers of Juncaceae. **Arkiv för Botanik**, v. 31, n. 1, p. 1-6, 1944.

MADDOX, Paul S. et al. " Holo" er than thou: chromosome segregation and kinetochore function in *C. elegans*. **Chromosome Research**, v. 12, n. 6, p. 641-653, 2004.

MALHEIROS, Nydia; DE CASTRO, D.; CAMARA, A. Chromosomas sem centromero localizado. O caso da *Luzula purpurea* Link. **Agronomia lusitana**, v. 9, n. 1, p. 51-71, 1947.

MANDRIOLI, Mauro; CARLO MANICARDI, Gian. Unlocking holocentric chromosomes: new perspectives from comparative and functional genomics?. **Current genomics**, v. 13, n. 5, p. 343-349, 2012.

MARQUES, André et al. Holocentromeres in *Rhynchospora* are associated with genome-wide centromerespecific repeat arrays interspersed among euchromatin. **Proceedings of the National Academy of Sciences**, v. 112, n. 44, p. 13633-13638, 2015.

MÁRQUEZ-CORRO, José Ignacio et al. Inferring hypothesis-based transitions in clade-specific models of chromosome number evolution in sedges (Cyperaceae). **Molecular phylogenetics and evolution**, v. 135, p. 203-209, 2019.

MAYROSE, Itay; BARKER, Michael S.; OTTO, Sarah P. Probabilistic models of chromosome number evolution and the inference of polyploidy. **Systematic biology**, v. 59, n. 2, p. 132-144, 2010.

MELTERS, Daniël P. et al. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. **Genome biology**, v. 14, n. 1, p. R10, 2013.

MELTERS, Daniël P. et al. Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. **Chromosome Research**, v. 20, n. 5, p. 579-593, 2012.

MUTZ, Kai-Oliver et al. Transcriptome analysis using next-generation sequencing. **Current opinion in biotechnology**, v. 24, n. 1, p. 22-30, 2013.

NAGAKI, Kiyotaka et al. Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres. **Genetics**, v. 163, n. 3, p. 1221-1225, 2003.

NEUMANN, Pavel et al. Stretching the rules: monocentric chromosomes with multiple centromere domains. **PLoS genetics**, v. 8, n. 6, 2012. **Japanese Journal of Genetics**, v. 59, n. 5, p. 465-472, 1984.

NORDENSKIÖLD, Hedda. cyto-taxonomical studies in the genus *Luzula*: i. somatic chromosomes and chromosome numbers. **Hereditas**, v. 37, n. 3, p. 325-355, 1951.

OLIVEIRA, Ludmila 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.

PANCHENKO, Tanya; BLACK, Ben E. The epigenetic basis for centromere identity. In: **Centromere**. Springer, Berlin, Heidelberg, 2009. p. 1-32.

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

PEREA-RESA, Carlos; BLOWER, Michael D. Centromere biology: transcription goes on stage. **Molecular and cellular biology**, v. 38, n. 18, p. e00263-18, 2018.

REVELL, Liam J.; REVELL, Maintainer Liam J. Package 'phytools'. 2020.

RIBEIRO, Tiago et al. Are holocentrics doomed to change? Limited chromosome number variation in *Rhynchospora* Vahl (Cyperaceae). **Protoplasma**, v. 255, n. 1, p. 263-272, 2018.

RIBEIRO, Tiago et al. Centromeric and non-centromeric satellite DNA organisation differs in holocentric *Rhynchospora* species. **Chromosoma**, v. 126, n. 2, p. 325-335, 2017.

ROALSON, Eric H. A synopsis of chromosome number variation in the Cyperaceae. **The Botanical Review**, v. 74, n. 2, p. 209-393, 2008.

ROALSON, Eric H. Phylogenetic relationships in the Juncaceae inferred from nuclear ribosomal DNA internal transcribed spacer sequence data. **International journal of plant sciences**, v. 166, n. 3, p. 397-413, 2005.

ROSS, Benjamin D. et al. Stepwise evolution of essential centromere function in a *Drosophila* neogene. **Science**, v. 340, n. 6137, p. 1211-1214, 2013.

SADER, Mariela 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, Veit et al. Super-Resolution Microscopy Reveals Diversity of Plant Centromere Architecture. **International Journal of Molecular Sciences**, v. 21, n. 10, p. 3488, 2020.

SEDENSKY, Margaret M. et al. Identification of a mariner-like repetitive sequence in *C. elegans*. **Nucleic acids research**, v. 22, n. 9, p. 1719-1723, 1994.

SHARP, Lester Whyland et al. Introduction to cytology. **Introduction to cytology.** n. 3rd ed, 1934.

SNOGERUP, S. Studies in the genus Juncus III. Observations on the diversity of chromosome numbers. **Bot. Not.**, v. 116, p. 142-156, 1963.

TALBERT, Paul B.; HENIKOFF, Steven. What makes a centromere?. **Experimental Cell Research**, p. 111895, 2020.

TALBERT, Paul B.; KASINATHAN, Sivakanthan; HENIKOFF, Steven. Simple and complex centromeric satellites in *Drosophila* sibling species. **Genetics**, v. 208, n. 3, p. 977-990, 2018.

TANAKA, Noriyuki; TANAKA, Nobunori. Chromosome Studies in Chionographis (Liliaceae). **Cytologia**, v. 42, n. 3-4, p. 753-763, 1977.

VILLASANTE, Alfredo; ABAD, José P.; MÉNDEZ-LAGO, María. Centromeres were derived from telomeres during the evolution of the eukaryotic chromosome. **Proceedings of the National Academy of Sciences**, v. 104, n. 25, p. 10542-10547, 2007.

WHITE, Michael James Denham. **Animal cytology and evolution**. CUP Archive, 1977.

YANO, Okihito; HOSHINO, Takuji. Cytological studies of aneuploidy in *Eleocharis kamtschatica* (Cyperaceae). **Cytologia**, v. 71, n. 2, p. 141-147, 2006.

ZAVESKA DRABKOVA, Lenka; KIRSCHNER, Jan. *Oreojuncus*, a new genus in the Juncaceae. **Preslia**, v. 85, p. 483-503, 2013.

ZEDEK, František; BUREŠ, Petr. Holocentric chromosomes: from tolerance to fragmentation to colonization of the land. **Annals of botany**, v. 121, n. 1, p. 9-16, 2018.

APÊNDICE A – Repeat-based phylogenomics resolves section-level classification within the monocentric genus *Juncus* L. (Juncaceae)

Yennifer Mata-Sucre[1], William Matzenauer[2], Natália M. Souza Castro[1], Bruno Huettel[3], Andrea Pedrosa-Harand[1], André Marques[4], Gustavo Souza[1*]

Molecular Phylogenetics and Evolution, Volume 189, December 2023, 107930

https://doi.org/10.1016/j.ympev.2023.107930

Original Article

Repeat-based phylogenomics resolves section-level classification within the monocentric genus *Juncus* L. (Juncaceae)

Yennifer Mata-Sucre[1], William Matzenauer[2], Natália M. Souza Castro[1], Bruno Huettel[3], Andrea Pedrosa-Harand[1], André Marques[4], Gustavo Souza[1*]

[1]*Laboratório de Citogenética e Evolução Vegetal, Departamento de Botânica, Centro de Biociências, Universidade Federal de Pernambuco. Recife PE 50670-901 Brasil*

[2]*Laboratório de Morfo-Taxonomia Vegetal, Departamento de Botânica, Centro de Biociências, Universidade Federal de Pernambuco, Recife PE 50670-901 Brasil*

[3]*Max Planck Genome-Centre Cologne, Max Planck Institute for Plant Breeding Research, Cologne, Germany*

[4]*Department of Chromosome Biology, Max Planck Institute for Plant Breeding Research, Cologne, Germany*

*Corresponding author:
Gustavo Souza
Universidade Federal de Pernambuco
Centro de Biociências
Departamento de Botânica
Laboratório de Citogenética e Evolução Vegetal
R. Prof. Moraes Rego, s/n, CDU
50670-901 Recife PE Brazil

Abstract

The repetitive fraction of eukaryotic genomes is diverse and usually fast evolving, being an important tool for clarify plant systematics. The genus *Juncus* L. comprises 332 species, karyotypically recognized by having holocentric chromosomes. However, four species were recently described as monocentric, yet our understanding of its genome evolution is largely masked by unclear phylogenetic relationships. Here, we reassess the current *Juncus* systematics using low-coverage genome skimming data of 33 taxa to construct repeats, nuclear rDNA and plastome-based phylogenetic hypothesis. Furthermore, we characterize the repetitive fraction and chromosomal distribution of *Juncus*-specific centromeric repeats/CENH3 protein to test the monocentricity reach in the genus. Repeat-base phylogenies revealed topologies congruent with the rDNA tree, but not with the plastome tree. The incongruence between nuclear and plastome chloroplast dataset suggest an ancient hybridization in the divergence of *Juncotypus* and *Tenageia* sections 40 Mya ago. The phylogenetic resolution at section level was better fitted with the rDNA/repeat-based approaches, with the recognition of four monophyletic sections (*Juncus*, *Juncotypus*, *Stygiopsis* and *Tenageia*). We found specific repeatome trends for the

main lineages, such as the higher abundances of TEs in the *Caespitosi* and *Iridifolii+Ozophyllum* clades. CENH3 immunostaining and centromeric satDNA *in situ* hybridization confirmed the monocentricity of *Juncus*, which can be a generic synapomorphy. The heterogeneity of the repeatomes, with high phylogenetic informativeness, identified here may be correlated with their ancient origin (56 Mya) and reveals the potential of comparative genomic analyses for understanding plant systematics and evolution.

Keywords: Centromere, Chromosome, High-throughput sequencing, Repeatome, Systematic

1. Introduction

Within the repetitive fraction of the eukaryotic genome, referred as repeatome, transposable elements (TEs) and satellite DNA (satDNA) are the major components, accounting for up to 90% of plant nuclear DNA (Fu et al., 2019). TEs can be considered one of the most influential components of plant genomes due to their ability to move and insert at novel locations, thereby shaping coding and non-coding DNA landscapes (Wicker et al., 2001, 2007; Rostoks et al., 2002; Lisch, 2013). In contrast, satDNA, which consists of long, late replicating and generally non-coding arrays of tandemly arranged monomers is predominantly concentrated in the heterochromatic regions of the chromosomes (Biscotii et al., 2015; Satović et al., 2016; Thakur et al., 2021). While satDNAs are non-mobile, recent studies suggests their involvement in various functions ranging from chromosome organization and pairing to cell metabolism and the modulation of gene functions (see review by Garrido-Ramos et al., 2017). Extensive and intricate relations between TEs and satDNAs exists within eukaryotic genomes, creating a complex network of repetitive sequences that has a crucial effect on genome structure, function and evolution (Meštrović et al., 2015; Chumova et al., 2022).

In plants, the repeatome is highly diverse and fast evolving, making it a useful tool for better understand plant systematics. In the quest to understand its diversity, it was proposed that the genomes have been considered as ecosystems “populated” by repeats, where repeat lineages are analogous to “species” and the number of copies of a given repeat lineage to the number of “individuals”, each one interacting with other repeats, genes and regulatory sequences (Brookfield et al., 2005; Venner et al., 2009; Stitzer et al., 2021; Schley et al., 2022). The similarities between the dynamics of repeats in genomes and the dynamics of communities in an ecosystem allow the use of analogous metrics between them such as the Shannon index (Shannon, 1948). This index helps to describe quantitative aspects of repeat “communities” such as the diversity of repeat lineages and the amount of the genome they occupy (Schley et al., 2022). Thus, this metric can be used as a parameter in comparative genomics.

Another genomic parameter widely used to discuss plant systematics is the genome size (GS). This trait varies greatly across the flowering plants and most of this variation is based on differences in repeat content and polyploidization (Pellicer et al., 2018; Wendel et al., 2018; Wang et al., 2021). GS is generally considered an evolutionary character whose evolution can be correlated with various structural and eco-physiological plant traits (Hloušková et al., 2019; Souza et al., 2019; Trávníček et al., 2019; Mata-Sucre et al., 2020a). It has been shown, for

instance, that genome sizes in palms are influenced by repeat expansion, and that the dynamics of these repeat "communities" are moderated by aridity, through the selective pressure that aridity exerts on repeat amplification and genome size (Schley et al., 2022). Different studies have already highlighted the relationship between repeat dynamics and genome size variation in plants (Novák et al., 2020; Vitales et al., 2020a; Sader et al., 2021; Ibiapino et al., 2022).

Repetitive elements have also been shown to be informative for inferring phylogenetic relationships (Vitales et al., 2020b; Breman et al., 2021; Costa et al., 2021; Herklotz et al., 2021). Bioinformatics pipelines, such as the Repeat Explorer, have made it possible to study the repetitive genomic fractions of plants using low coverage sequencing by Next Generation Sequencing (NGS) (Mata-Sucre et al., 2020b; Sader et al., 2021; Ibiapino et al., 2022; Zuo et al., 2022). The repeatome is not a homogeneous fraction during its evolution it can accumulate genetic changes at different rates. It can vary in abundance, and can either be conserved throughout groups of species, or rapidly evolving, with some of them showing a strong phylogenetic significance (Hartley and O'Neill, 2019; Enriquez-Gasca et al., 2020; Louzada et al., 2020; Breman et al., 2021; Costa et al., 2021; Oliveira et al., 2021). Using repeatome dynamics in a phylogenetic context can shed light on the evolution of the different repeat-classes and can be a powerful tool for understanding relationships in phylogenetically unresolved groups (Dodsworth et al., 2017).

Juncaceae Juss. is a cosmopolitan family comprising eight genera and 524 species (Kirschner, 2002a,b; Christenhusz and Byng, 2016; POWO, 2023). *Juncus* L. represent the largest genus in the family with 332 species (Drábková, 2006; POWO, 2023). It is morphologically divided into two subgenera, subgenus *Juncus* and subgenus *Agathryon* Raf. (Kirschner, 2002), with six and four sections, respectively (Buchenau 1906). Phylogenetic relationships within Juncaceae are still poorly understood, especially within *Juncus*, where its subgenera and sections appeared as non-monophyletic (Roalson, 2005; Drábková, 2006; Jones et al., 2007; Drábková and Vlček, 2010; Drábková and Kirschner, 2013). The most recent phylogeny by Brožová et al. (2022), based on nuclear rDNA (ITS) and two plastid regions (the gene *rbc*L and the spacer *trn*L-F), yielded inconclusive results in the delimitation of the paraphyletic genus. Therefore, six new genera, elevating sections to generic status were proposed to recover monophyly in this group. However, this proposal was criticized in later publications (Elliott et al., 2023) indicating the need for taxonomy and evolutionary studies of other markers (e.g. genomic) before adopting any taxonomic change.

The genus *Juncus* has been the subject of recent cytogenetic and genomic studies, mainly because it was discovered to four species have monocentric chromosomes, either by cytomolecular or genomic approaches, instead of the traditional view that all Juncaceae would have holocentric chromosomes (Guerra et al., 2019; Hofstatter et al., 2022, Dias et al., unpublished). These monocentric chromosomes have been characterized genomically only for *Juncus effusus* L., showing a very abundant satellite DNA in a centromeric position (Hofstatter et al., 2022; Dias et al., unpublished), yet the reach of the monocentricity throughout the whole genus is still unknown. From a cytogenetic point of view, few studies report chromosome counts or genome size estimation for *Juncus*. However, an evolution through polyploidy/dysploidy events has been suggested as crucial drivers of changes in the

karyotype/genome of the species (Jorgensen et al., 1958; Mehra and Sachdeva, 1976; Beuzenberg and Hair, 1983; Dalgaard et al., 1991; Schönswetter et al., 2007; Aguilera et al., 2014; Guerra et al., 2019). Here, we took advantage of genome skimming data of 33 taxa of *Juncus* to characterize their repeatomes, to establish plastome and repeat-base phylogenies, and to contribute for understanding their evolutionary history. Furthermore, we characterized the repeat fraction of these species to understand the evolution of these sequences and the chromosomal localization of *Juncus*-specific centromeric repeats/ CENH3 protein to test whether the entire genus is monocentric. We are specifically interested in the following questions: (1) Repeat abundances support congruent phylogenetic relationships in the genus? (2) Is there any repeatomic synapomorphies supporting the main *Juncus* clades? (3) In a phylogenetically varied sample, are the centromeres of *Juncus* species monocentric and can they be characterized by centromeric repeats?

2. Materials and methods

2.1 Taxa sampled, DNA isolation and Illumina sequencing

We included 33 taxa from the two major *Juncus* subgenera (subgenus *Juncus* and subgenus *Agathyron*) proposed by Kirschner (2002; Table 1). Total genomic DNA for 12 species were extracted from approximately 200 mg of young leaves using the CTAB extraction protocol of Ferreira and Grattapaglia (1995), slightly modified (without preheating the extraction buffer before placed in the sample). Libraries for 12 species were prepared using the TPase-based DNA library (Illumina) and sequenced on an Illumina HiSeq3000 sequencer at The Max Planck Genome Center Cologne, generating 150-bp paired-end reads. For the rest of the taxa, we obtained sequence data available from the European Nucleotide Archive (ENA, https://www.ebi.ac.uk/ena/browser/home, Table 1). We include *Luzula elegans* since the genus *Juncus* retrieved as not monophyletic.

2.2 Estimation of nuclear genome size

Genome size estimates were performed here for the first time for *Juncus acutus* L., *J. capitatus* Weigel, *Juncus* cultv. "Pincei", *J. kraussii* Hochst. In C.Krauss s. str., *J. pallidus* R. Br., *J. usitatus* L. A. S and *Juncus* cultiv. "Mallorcan Giant" Johnson by flow cytometry. Samples were prepared according to Loureiro et al. (2007). Young leaves of each of the studied plant were chopped simultaneously with the reference standard *Raphanus sativus* 'Saxa' (1.11 pg/2C, Dolezel et al., 1992) in a Petri dish (kept on ice) containing 2 mL of Galbraith Plant Buffer (GPB) or WPB isolation buffer (Loureiro et al., 2007). The sample was then filtered through a 30-µm disposable mesh filter (REF: 04-0042-2316, Sysmex, Germany) with subsequent addition of 50 µL propidium iodide (1 mg/mL). Three replicates per species were made on different days (5000 nuclei per analysis). The samples were measured using a CyFlow Space flow cytometer (Sysmex) in Germany and a BD Accuri C6 Plus (BD) equipped with a green laser (532 nm) in Brazil. Histograms of relative fluorescence were obtained using the software Cytexpert. Mean fluorescence and coefficient of variation were assessed at half of the

fluorescence peak. The absolute DNA content (pg per 2C) was calculated by sample peak mean/standard peak mean × 2C DNA content of the standard (pg).

2.3 Repeat identification, classification and dynamics from sequence data

An individual similarity graph-based clustering analysis for each species was performed following the method as described by Novák et al. (2020) and implemented in the RepeatExplorer2 pipeline (https://repeatexplorer-elixir.cerit-sc.cz/; Novák et al., 2013). All sequence reads were filtered by quality with 95% of bases equal to or above the quality cut-off value of 10. We used the pipeline default parameters and included a database of *Juncus effusus* repeats (Hofstatter et al., 2022). The clustering was performed using the default settings of 90 % similarity over 55 % of the read length. This analysis resulted in the clustering of overlapping reads, and these clusters represented different families of repetitive sequences. For comparative analyses, we performed an all-to-all similarity comparison following the same approach as above on two datasets: (i) a 1,000,000 read set for each of the 33 taxa, as the genome size of some species is unknown, and (ii) a sample of 0.1x coverage reads for 21 species with known genome size (Table 1). Samples from each species were identified with the four-letter prefixes mentioned in Table 1, and concatenated to produce datasets as input for RepeatExplorer2 graph-based clustering (Novák et al., 2020).

The repeat classification was performed using a combined approach that involved similarity searches with DNA and protein databases, as implemented in the RepeatExplorer2 pipeline (Novák et al., 2013), and improved by including a custom *Juncus* repeats database. Clusters that remained unclassified were annotated (if possible) by examining the shape of the cluster graph and by similarity searches using BLASTN against NCBI public databases (https://blast.ncbi.nlm.nih.gov/Blast.cgi) and the TIGR Plant Repeat Databases (www.tigr.org/tdb/e2k1/plant. repeats/index.shtml). All these sources were combined and used for final manual annotation and quantification of repeats from clusters that represented at least 0.01 % of the genome. Repeat composition was calculated excluding clusters of organellar DNA (chloroplast and mitochondrial DNA) representing possible contamination.

All contigs with tandem repetitions identified by TAREAN (Novák et al., 2017) were confirmed with DOTTER (Sonnhammer and Durbin, 1995). The consensus sequences of all identified satellites were compared in order to verify similarity among them. The consensus monomers that showed similarity in DOTTER were aligned using MAFFT in Geneious (Kearse et al., 2012). 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). Correlation analysis between the total proportion of repetitive sequences and genome size expressed in base pairs, and between tandem and dispersed sequences with genome size were undertaken with the package stats implemented in the software R v. 4.0.2 (R Core Team, 2019). Plots were constructed with the R package ggplot2 (Wickham, 2016).

2.4 Analysis of repeat diversity in Juncus *genomes*

In order to test for the relationship between repeat diversity and genome size in the analyzed species, we employed the Shannon-Wiener Index (Shannon, 1948). This index is commonly used to measure the diversity of a given community, in this case treating a genome like a community. To calculate the Shannon index of the genomes, the lineages identified by RepeatExplorer were treated as "species" within a "community" (genome). We then used the *diversity*() function of the Vegan (Oksanen et al., 2013) package from R software (R Core Team, 2019) to calculate the Shannon index, as outlined by Schley et al. (2022). The abundance of all lineages was taken into consideration, providing us with a diversity index for each genome.

To examined the relationship between the Shannon's index and genome sizes for 21 species, we used linear regression to determine whether genome size could be correlated with variations in repeat diversity (measured by Shannon-Wiener Index). Statistical analyses and their results were visualized using the ggplot2 package in R (Villanueva & Chen, 2019).

2.5 Phylogenetic analyses and divergence time

We employed a repeat abundance-based phylogenetic inference method (see details in Dodsworth, 2015) to assess if repeat abundance of the elements could be used to resolve phylogenetic relationships. A comparative clustering analysis (simultaneous clustering of all species on the dataset) on RepeatExplorer2 with default settings (Novák et al., 2013) was performed as described above. As the sequences were coded with the species names, we could identify the number of reads that each species contributed to each of the generated clusters, which is proportional to the abundance of each repeat in the genome of each species. Parsimony analysis using repeat abundances as quantitative characters was undertaken as described by Dodsworth (2015). Furthermore, the phylogenetic signal λ (Pagel, 1999) was calculated for the repeat abundances across all species. For this, the *phylosig* function from the phytools package (Revell, 2012), implemented in the Rstudio software (R Core Team, 2022) was used. In addition, to test the phylogenetic signal of repetitive elements based on sequence similarity, we used the AAF approach (Fan et al., 2015) using all reads identified as repeats by RepeatExplorer2. AAF constructs phylogenies directly from unassembled genome sequence data, bypassing both genome assembly and alignment. Thus, it calculates the statistical properties of the pairwise distances between genomes, allowing it to optimize parameter selection and to perform bootstrapping.

In order to compare our repeat abundance-based phylogeny with a traditional phylogenetic approach, we performed a plastome and nuclear rDNA phylogenetic trees. The plastome (Lu et al., 2021) and rDNA (NCBI: OX326994) of *J. effusus* was used as a reference to assemble the plastome and rDNA of the 33 new *Juncus* taxa sampled here. For this, the raw Illumina reads were mapped against the references plastome and rDNA using Geneious v. 9.1.8 (Kearse et al., 2012). The alignments were made using MAFFT (Katoh and Standley, 2013). For simplification, we used the most general model of DNA substitution, GTR+I+G (Abadi et al., 2019). Phylogenetic relationships were inferred using FastTree (Price et al., 2010) as implemented in Geneious. Trees were visualized using FigTree v. 1.3.1. *Luzula elegans* Lowe,

Prionium serratum (L.f.) Drège and *Fimbristilis* sp. were used as outgroups for phylogenomic analysis. To check the incongruences among the different topologies, the Cophylo package in software R was used (Revell, 2012).

Divergence time estimates were performed in BEAST v.1.10.4 (Drummond and Rambaut, 2007) fixing the tree topology of the rDNA FastTree analyses. An uncorrelated relaxed lognormal clock (Drummond et al., 2006) and a Birth Death Process speciation model (Gernhard, 2008) were applied. Two independent runs of 10,000,000 generations each were performed, sampling every 10,000 generations for the full rDNA alignment. In order to verify the effective sampling of all parameters and assess the convergence of independent chains, we examined their posterior distributions in Tracer v.1.6 until MCMC sampling reached ESS >200. After removing 25% of samples as burn-in, the independent runs were combined and a maximum clade credibility (MCC) tree was constructed using TreeAnnotator v.1.8.2. (Drummond et al., 2012). Calibrations were performed using the secondary calibrations of Bouchenak-Khelladi et al. (2014) [*Juncus*/*Prionium* divergence approx. 104.2 Mya] and Escudero et al. (2013) [*Juncus*/*Luzula* divergence approx. 64.4 Mya].

2.6 Cytogenetic mapping and immunostaining of CenH3 protein

The JeCENH3: VRTKHFSSRPAGSGRPRKR-C peptide was used for centromeric protein immunodetection (Dias et al., unpublished). Mitotic preparations were made from root meristems fixed in paraformaldehyde and Tris buffer (10 mM Tris, 10 mM EDTA, 100 mM NaCl, 0.1% Triton, pH 7.5) for 30 min on ice in a vacuum and for another 20 min only on ice. After washing twice in 1× PBS for 10 min, the roots were digested in a cellulase-pectinase solution (2% w/ 20%v) for 60 min containing PBS buffer and then macerated in PBS following Carvalho and Saraiva (1993). The coverslips were removed in liquid nitrogen and the slides were air dried and stained in DAPI:Vectashield for slide selection under the epifluorescence microscope. The slides with the highest number of mitotic cells were incubated in 3% (w/v) bovine serum albumin (BSA) containing 0.1% Triton X-100 in PBS. Immunostaining was performed using the primary antibody rabbit anti-JeCENH3 (diluted 1:300). As secondary antibody, a goat anti-Rabbit IgG antibody (Invitrogen) was used in a 1:500 dilution. Slides were incubated overnight at 4 °C, washed 3 times in 1× PBS and then the secondary antibody was applied, incubated at room temperature for 3 h and washed 3 times in 1× PBS. The slides were counterstained with 2 µg/mL DAPI in Vectashield (Vector) mounting buffer. Microscopic images were captured using a Zeiss Axiovert 200M microscope equipped with a Zeiss AxioCam CCD. Images were analyzed using the ZEN software (Carl Zeiss GmbH).

A Cy5-labeled oligo probe was designed from the centromeric JeSat1-155 satDNA (5'AGGTGCAAACAAAGATTGTGGAGAAATATATTTTAAA; Dias *et al*., unpublished) and used for fluorescence in situ hybridization (FISH). Mitotic chromosomes were prepared using the air-drying method, after enzymatic digestion with 2% cellulase Onozuka and 20% pectinase Sigma (Ribeiro *et al.*, 2017) using pretreated roots with 2 mM 8-hydroxyquinoline for 24 h at 4°C, fixed in ethanol:acetic acid 3:1 (v/v) for 2 h and stored at −20°C. FISH was

performed as described by Pedrosa *et al.* (2002). The slides were counterstained with 2 µg/mL DAPI in Vectashield (Vector) mounting buffer. The images were captured as described above.

3. Results

3.1 Phylogenetic tree and genome size estimations

For phylogenetic analyses, a combined assembled plastome matrix containing 186,091 bp for 33 taxa was generated, where 36,984 nucleotide sites were informative for phylogenetic analyses. This analysis provided a well-supported strict consensus tree with maximum support (PP= 1) for all nodes (Fig. S1). The rDNA alignment (7,031 bp) showed 4,539 informative sites with good resolution and support at most nodes (Fig. 1A). However, this topology was partially different from the plastome tree, mainly in the relationship of the *Tenageia* and *Juncotypus* sections, where the latter appears as monophyletic only in the nuclear rDNA topology (Fig. 1A; S1). In total eight sections within the genus were sampled here, five from subg. *Juncus* and three from subg. *Agathyron*. Of the four sections that form the subgenus *Agathyron*, we could not include *Forskalina* in our analyzes because of the difficulty in obtaining material, since the only species in the section, *J. subulatus* Forssk., is restricted to the Mediterranean region and Central Asia (Kirschner et al., 2002). Of the eight sections sampled, plastome data recovered only two sections as monophyletic: sect. *Stygiopsis* Grand. ex Kuntze and sect. *Tenageia* Dumort. (all of them with PP= 1; Fig. S1). On the other hand, rDNA topology recovered as monophyletic the sections *Juncotypus*, *Stygiopsis* and *Tenageia* (all of them with PP= 1; Fig. 1A). In both plastome and nuclear topologies, the sect. *Ozophyllum* Dumort was not monophyletic due to the inclusion of the only analyzed species of the sect. *Iridifolii*, *J. xiphioides* E. Mey. (PP= 1; Fig. 1A; S1). Furthermore, non-monophyly of the *Juncus* section was caused by the inclusion of the species *J. kraussii* Hochst. within the *Juncotypus* section (PP=1; Fig. S1). Section *Caespitosi* Cout. (*J. capitatus* Weigel) was not tested for monophyly due to the presence of only one species in our sampling; however, this species is positioned sister to the outgroup *Luzula elegans* and out of the other *Juncus* species in both plastid and nuclear topologies (Fig. S1). The sect. *Steirochloa* Griseb. presents a complex relationship among species, being polyphyletic in plastid and rDNA analyses. The two species sampled in this study (*J. compressus* and *J. gerardi* subsp. *atrofuscus*) appear mixed within species of the subgenus *Agathyron* in plastome phylogeny strongly supported, however, they appear mixed between *Juncotypus* and *Juncus* sections with low support in the rDNA tree (Fig. S1). All topology reveals that the subgenus *Agathyron,* as well *Juncus*, are non-monophyletic (Fig. S1). Despite the unclear systematics of *Juncus* cultv. Pincei and *Juncus* cultv. Mallorcan Giant phylogenetic position based on plastome and rDNA trees highly support that they belong to *Juncus* section (PP=1; Fig. S1).

Our data suggest an origin of the genus *Juncus* to 56 Mya (51-80 Mya) in the Paleocene-Eocene boundary (Fig. 1; S2). The crown age of its main clades ranged from 5.5 to 17 Mya, with *Stygiopsis* as the youngest clade and *Ozophyllum*+*Iridifolii* the oldest (Fig. 1A; S2). The stem age of the divergence *Juncotypus*/*Tenageia*, the incongruent clade between rDNA/repeat and plastome-based topologies, was estimated to be 40 Mya (25-60 Mya; Fig. S2) in the Eocene (Fig. 1A). New genome size estimates for seven species revealed 2-fold variation in 2C content,

ranging from 0.25 pg in *Juncus usitatus* to 0.82 pg in *J. xiphioides* (Table 1, Fig.1B). Previously reported nuclear DNA content were generally higher, with 2C ranging from 0.50 in *J. compressus* to 3.58 in *J. articulatus* (Table 1, Fig.1B). In general, species of the sections sampled here showed a small genome size (mean = 0.58 pg), except for species of the sections *Tenageia* and *Ozophyllum*, with sizes up to 2C of 3.58 pg (Table 1, Fig.1B).

3.2 Characterization and diversity of the repetitive fraction of Juncus *genome*

We used the Shannon index to determine the diversity of repeat abundance within genomes by section (Table 1; Fig. 1C). We observed that there is no correlation between diversity and genome size, as clades *Caespitosi*, *Iridifolii*, *Juncotypus*, *Ozophyllum, Steirochloa*, and *Tenageia* (except *J. ranarius*) have a higher diversity of repeats and a more even distribution of their abundances among the species analysed compared to clades *Stygiopsis* and *Juncus* (Table 1; Fig. 1C; S3). In *Stygiopsis* and *Juncus* sections there is a dominance of one type of repeats, mainly satDNAs, with very little diversity and abundance (<5%) of other types of repeats among species (specially LTR elements, Fig. 1C; Table S1).

Individual repetitive analyses revealed fractions varying from 19.61% of the genome in *J. kraussii* to 57.95% in *J. articulatus* (Table S1; Fig. 2A). The abundance of repetitive elements showed a similar proportion within each section, where the abundances of both dispersed and tandem repeats allowed distinguishing the eight sections sampled (Fig. 2A). The repetitive fraction of these genomes was composed mainly of LTR elements (LTR-TEs), except for the *J. roemerianus* genome, which had only 0.13% of LTR-TEs within its genome (Table S1; Fig. 2). Ty1-copia superfamily elements were the most abundant within the LTR-TEs, mostly those belonging to the Angela and SIRE lineages (Table S1; Fig. 2A). Although these elements are highly enriched in all species, they show significant variation in abundance between clades, such as in species from the *Ozophyllum* and *Iridifolii* clades, which have as much as twice the relative content as the other clades (Table S1; Fig. 2A). The Ty3-gypsy elements were less abundant among *Juncus* genomes, with Athila and Tekay lineages more represented in the *Juncotypus* and *Ozophyllum* clades, respectively (Table S1). *Oreojuncus trifidus* and *J. capitatus* differed in diversity and abundance of repetitive elements from the rest of the *Juncus* species (Table S1; Fig. 2A). Unlike the other species, *O. trifidus* shows a high abundance of the elements Ty3-gypsy Retand (2.46%) and Ty1-copia Ivana (2.27 %) lineages, whereas *J. capitatus* presents a high abundance of 35S rDNA (17.30%) and the lowest abundance of satDNA sequences in its genome (0.525%, Table S1).

Some *Juncus* species have an enrichment in tandem repeats (Fig. 2B). Species belonging to the *Ozophyllum* clade presented the lowest abundance and diversity of satDNA sequences, contrary to the rest of the species of the other clades, which showed a higher abundance and diversity of satDNA in their genomes (Fig. 2A-B; Table S1). The highest proportion of satDNA was observed in species of the section *Stygiopsis*, reaching up to 49 % of their genome (Fig. 2B; Table S1). The 35S rDNA showed a large variation in abundance, from 0.01% in *J. inflexus* reaching up to 17.3% in *J. capitatus* (Fig. 2B; Table S1). A significant positive correlation was observed between repeat proportion and genome size ($r = 1$; $p < 0.001$;

Fig. 2C-D), especially between the Ty1-copia elements and the genome size (r = 0.70; p < 0.001; Fig. 2C). Among the Ty1-copia, the Angela (r^2 = 0.89; p < 0.001) and SIRE (r^2 = 0.82; p < 0.001) lineages are the most significant contributors to genome size increase (Fig. S3). Furthermore, an interesting negative correlation was observed between the Ty1-copia and the satDNA proportion (r = -0.71; p < 0.001; Fig. 2C-D).

3.3 Phylogenomic and dynamics of shared repeats within Juncus

Comparative clustering analysis of a combined dataset of 33 *Juncus* genomes show 231 shared clusters with a strong phylogenetic signal (λ = 0.83, p < 0.001), where the abundance of shared repeats elements allows distinguishing seven of the eight sections sampled here, except for *Steirochloa* (Fig. 3; Table S2). Since comparative analyses with a sample size of 1,000,000 reads and coverage of 0.1x the genome size showed similar trends to those observed in our larger dataset, we continue our discussion with the data obtained from the 33 taxa (Fig. S4). Several of the most abundant shared elements belong to the LTR group, such as Angela, SIRE and Tekay lineages. In the case of Angela, a higher proportion was observed among species of the section *Ozophyllum*; however, several variants (e.g., clusters 78, 114) appears to be absence or in low frequency in this clade compared to the other clades (Fig. 3A; Table S2). We also found some lineage-specific tandem repeats, such as CL 1 (SatCL1_154 bp) and CL 55 (124 bp) satellites of the same superfamily, and CL 188 (41 bp) present among species of the clade *Stygiopsis* (Fig. 3A; Fig. S5). Similarly, CL 124 is a 132-bp species-specific satellite, only present in the genome of *J. roemerianus* (Fig. 3A; Fig. S5).

3.4 Repeat-based phylogenomics

We used repeat abundances obtained by comparative clustering of 33 *Juncus* taxa to construct phylogenetic hypotheses. By using the first 120 most abundant clusters of the comparative analysis (repeat-abundance approach), we were able to reconstruct phylogenetic relationships among *Juncus* species with relative high bootstrap support (BS>60) for most of the clades (Fig. 4A). Using raw reads from NGS sequencing, the AAF (repeat-similarity approach) analysis generated congruent topology with repeat-abundance tree for all the clades with high support, however, the relationships between the species were not completely resolved (BS = 100; Fig. 4A). Remarkable, both AAF and repeat-abundance (Fig. 4A) topologies were in general congruent with the rDNA topology, recovering the monophyly of the sections *Juncus*, *Juncotypus*, *Stygiopsis* and *Tenageia* (Fig. S6). On the other hand, the repeat-based topologies showed more incongruences with the plastome topology, involving species from the *Juncotypus*, *Stygiopsis* and *Tenageia* sections (Fig. 4B). In the case of the *Juncus* section, unlike the repeat-based tree, the plastome tree reveals all sections of the *Juncus* subgenus as monophyletic. Furthermore, in the case of the species *J. capitatus*, the repeat-based topology maintains this species distinct from the other *Juncus*, whereas the plastome-based topology includes it as sister to *Luzula* genus (Fig. 4B).

3.5 Centromere identification and composition of Juncus *species*

We have performed *in situ* immunodetection of JeCENH3 in six species of the genus, one representative of each of the clades/sections *Caespitosi, Iridifolii*, *Juncus*, *Juncotypus*, *Ozophyllum* and *Steirochloa* (Fig. 5). This antibody revealed expected patterns of monocentric CENH3 distribution on mitotic (pro)metaphase or metaphase chromosomes for *J. capitatus*, *J. xiphioides*, *Juncus* "Mallocan Giant", *J. subnodulosus*, *J. usitatus* and *J. compressus* with several chromosomes exhibiting dot-like signals at the putative primary constriction (insets in Fig. 5).

Furthermore, to check the localization of JeSat1 we performed FISH in (pro)metaphase mitotic cells with a probe derived from its monomer (Dias et al., unpublished). In some species, signals were observed preferentially located in the middle and/or pericentromeric regions of the chromosomes (Fig. 5, *J. acutus* and *J. subnodulosus*); although signals also formed larger clusters on chromosomes, as observed in a few chromosomes of *J. acutus*, *J. usitatus* and *J. subnodulosus* (Fig. 5). Finally, immunostaining with centromere-specific JeCENH3 in combination with probe FISH for JeSat1 in *J. usitatus* resulted in CENH3 signals around and overlapping with JeSAT1, suggesting that this satellite may have a centromeric distribution but also extend to the pericentromeres of chromosomes (Fig. S7).

4. Discussion

4.1 Repeat-based phylogeny supports a new Juncus *intrageneric classification*

Here, we use different phylogenomic datasets (repetitive sequences, raw NGS reads, full plastome and nuclear rDNA assembly) to stablishing evolutionary relationship in the phylogenetically complex genus *Juncus*. The phylogenetic relationships in Juncaceae have been very difficult to unravel, showing a scenario of very complex and unresolved relationships (Roalson, 2005; Drábková, 2006; Jones et al., 2007; Drábková and Kirschner, 2013). The previous analyzes reveal that the genus is sister to *Luzula* and paraphyletic with unclear delimitation among *Juncus* and other genera of the family (*Distichia*, *Marsippospermum*, *Oxychloe*, *Patosia* and *Rostkovia*). These genera form a clade called SHC (Southern Hemisphere Clade), which is positioned close to members of the sections *Graminifolii* Engelm., *Caespitosi* and *Juncus* (Drábková et al., 2003; Drábková and Kirschner, 2013). Although it is a new approach, the repeat-based phylogenies have revealed relationships within different flowering plant groups, (Oliveira et al., 2021; Zuo et al., 2022; Costa et al., 2021;2023). Our data suggest a high potential of this type of analysis for studying the systematics of the genus *Juncus*. The rDNA and repeat-based analysis fits better to the morphological data, allowing better recognition of section relationships.

We found concordance between rDNA and repeat-based phylogenies, which allowed us to discuss *Juncus* section classifications. A recent phylogenetic analysis based on loci *rbc*L, *trn*L, *trn*L-*trn*F and nuclear (ITS1-5.8S-ITS2) suggested that only sections *Juncus* and *Stygiopsis* would be monophyletic (Brožová et al., 2022). This contrasts with the better resolution identified here, with the recognition of four monophyletic sections (*Juncus*, *Juncotypus*, *Stygiopsis* and *Tenageia*) and a strongly supported clade bringing together

representatives of the sections *Ozophyllum* and *Iridifolii* (despite having only one species in our sample). Thus, our data suggest that sections *Ozophyllum* and *Iridifolii* (subg. *Juncus*) should be unified into a single monophyletic section. This proposal has been suggested in many works and is morphologically supported by the similar anatomy of the leaves, unitubulate, pluritubulate or flattened and perfectly septate, forming the so-called " *Juncus septati* group" (*Ozophyllum*, *Iridifolii*) proposed by Buchenau in 1980 (Drábková et al., 2003; 2006; 2010; Roalson et al., 2005).

We suggested that the low phylogenetic resolution previously reported for the genus *Juncus* may be partially explained by the incongruence between plastome and nuclear data, which may be associated with ancient hybridization events and/or reticulated evolution, since in contrast to nuclear data, plastome data reflect the relationships of genetic material from one parent species (Oliveira et al., 2021). It is possible that there was a hybridization event in the common ancestor that gave rise to sections *Juncotypus* and *Tenageia* 40 Mya (Cope and Stace, 1985; O'Mahony, 2002; Smith, 2006; Wilcox, 2011), generating the incongruence detected here. The monophyletism of the sections *Tenageia* and *Juncotypus*, identified in rDNA and repeat-based approaches, is corroborate by morphological characteristics, such as leafless stems and the lower bract forming an apparent extension of the stem, which gives the inflorescence a pseudolateral aspect (Balslev, 1996; Kirschner, 2002a, b). Genome skimming approach stands out for allowing the assembly of full plastome and rDNA allowing robust verification of ancient hybridization events (Dodsworth et al., 2015; Oliveira et al., 2021). These hybrid genomes may present a greater diversity of repetitive sequences that facilitate genetic exchange promoting recombination and contributing to the generation of a new unique genome diversity that is reflected in the relationships observed here. The high congruence between repeat-based and rDNA trees observed here corroborates the hypothesis that the repeatomic phylogeny reflects a nuclear topology (Oliveira et al., 2021).

The non-monophyletism of the subgenera *Juncus* and *Agathyron* was observed in previous studies using different plastid markers (Drábková, 2006; Drábková and Kirschner, 2013; Brožová et al., 2022). Here it was identified that most of the phylogenetic complexity is given by the species *J. compressus* and *J. gerardii* subsp. *atrofuscus* of section *Steirochloa* (subgenus *Agathyron*), which show genomic differentiation between their sequences that is reflected in their phylogenomic relationships. These morphologically similar species have already appeared phylogenetically close in previous phylogenies based on plastid markers (Drábková, 2006; Drábková and Kirschner, 2013), so an error in the identification of any species cannot be ruled out. However, this apparently polyphyletic section, is composed by 35 species widespread in temperate regions (Kirschner, 2002a, b). They are taxa of very varied morphology or even with absent or rare traits within the genus *Juncus*, with petiolate or only basal leaves, solitary flowers on pedicels, sessile or in distinct racemes, flat or nearly round leaves (Buchenau, 1890, 1906; Novikov, 1990). And more recently two species belonging to this section were taxonomically transferred to the new genus *Oreojuncus* (*Oreojuncus trifidus* and *O. monanthos*, Drábková and Kirschner, 2013). *Steirochloa* shares several characteristics with the other species of *Juncus*, including genomic traces observed here and being basically separated by the presence of lacerated-fimbriate auricles, which are unique within all Juncaceae (Drábková and Kirschner, 2013), so the splitting of this section would not be a surprise.

Another group with an unstable position within *Juncus* is the sect. *Caespitosi* subg. *Juncus*, represented here by *J. capitatus*, which stayed here outside the genus *Juncus* and formed a clade together with *Luzula*. The unexpected phylogenetic placement of *J. capitatus* has been frequently described (Drábková et al., 2006; Roalson, 2005; Drábková and Kirschner, 2013; Drábková and Vlček, 2009; Brožová et al., 2022). Mophologically, *J. capitatus* is a perennial plant with long-flowering chasmogamous and short-flowering cleistogamous flowers and the presumed primitive type of the thin outer coat of its seeds that emphasizes its special position (Buchenau, 1906). In this case, evolution to annuality may have occurred independently in two or three times (Kirschner et al., 2002) and sect. *Caespitosi* represents an inconsistent group of taxa with a distinct genomic composition that is reflected in the phylogenies. The different chromosome number $2n = 18$ (Mičieta, 1983), the distinct genome composition and its phylogenetic distant position from other species of the genus make this species particularly interesting. The presence of monocentric chromosomes observed here corroborates a relationship with species belonging to the genus *Juncus*, however, further studies are needed to understand its complex relationship with the genus *Luzula*.

Our low sampling and the absence of some sections limit the taxonomic treatment discussion of our work. However, based on the data obtained here, there are indications that support the notion of unification of some sections but not of the genus. Our analysis reveals genomic markers or patterns shared among the species sampled, but different among larger groups which could suggest underlying genetic features that extend across sectional boundaries.

4.2 Monocentric Juncus *and their evolutionary dynamics of tandem repeats*

Mapping centromeric sequences (JeCENH3) and satDNA repeats (JeSat1) in mitotic (pre)metaphase or anaphases chromosomes showed a preferential localization in (peri) centromeric regions in species belonging to six of the eight main clades, confirming their monocentricity, as observed for *J. effusus* (Hofstatter et al., 2022; Dias et al., unpublished). Colocalization of JeSat1 with JeCENH3 implies a centromere-specific role of this element at the structural level in *Juncus*. These results together with those published by Guerra et al. (2019) for *J. effusus*, *J. marginatus* Rostk, *J. microcephalus* Kunth and *J. tenuis* Willd confirm that all the species analyzed so far are monocentric. However, due to the fact that only few species have been cytogenetically analyzed and to the presence of holocentric chromosomes in the sister genus *Luzula*, the occurrence of several transitional events to holocentricity in the group, although rare, cannot be ruled out.

Comparative genomic analysis of repetitive sequences obtained by NGS has provided insight into the organization of the *Juncus* genomes revealing a heterogeneity of DNA repeats, so they can be separated into repeat-rich and repeat-poor clades in terms of abundance. An inverse correlation between transposable elements and satDNA sequences suggests that these sequences govern the evolutionary dynamics of the genome. In this case, a contrasting pattern of loss/gain of sequences tends to eliminate/reduce LTRs and increase satDNAs (or vice versa) leading to a predominance of one of these elements in the genome. This trend was also observed in *Aegilops speltoides* (Poaceae; Raskina et al., 2011), *Pteronia* (Asteraceae, Chumova et al.,

2022) and *Rhynchospora* (Cyperaceae; Costa et al., 2023), the latter being a genus with holocentric chromosome, where both LTRs-RT and satDNA traits presented opposing patterns. This can be explained by the mechanism of burst purification cycles, when a certain number of TE copies are "cleared" from the genome mainly through ectopic recombination (Devos et al., 2002; see Le Rouzic and Deceliere, 2005; Chumova et al., 2022). TEs and satDNAs sometimes share a similar sequence and organizational structure, if homologous regions during ectopic recombination events are located near satDNA sequences, the recombination process can lead to TE deletion and the genesis, mobility, and duplication or amplification of satDNA segments, as has been observed in *Pisum sativum* (Macas et al., 2009; Kejnovsky and Jedlicka, 2022). Knowing that TE dynamics can lead to reproductive isolation between populations (Rankasvia, 2011; Belyayev, 2014; Chumova et al., 2022), altering chromosomal segregation and leading to a new repetitive landscape (Ferree and Barbash, 2009), it is likely that this event could have caused the difference and separation of lineages at the section level in *Juncus*, as observed in our phylogenomic analysis.

Evolutionarily, tandem repeats are a feature of most eukaryotic genomes and are one of the most rapidly evolving sequences, which can differ dramatically even between closely related species (Melters et al., 2013; Talbert and Henikoff, 2020; Oliveira et al., 2021). A high abundance and diversity of satDNAs sequences were observed among *Juncus* genomes, although no highly conserved satellite DNA sequences were found throughout the genus. Unique tandem repeats are frequently found in other plant species, and because of their genus or species specificity, they have been widely used in molecular cytogenetics (Mata-Sucre et al., 2020b; Ribeiro et al., 2022; Šimoníková et al., 2022). Here, clusters with large satellite repeat abundance representing clade and/or specie-specific markers were observed, such as CL 1, CL 55 and188, which can therefore be considered a molecular synapomorphy of the section *Stygiopsis*. The present work resulted in the identification of additional putative section- or species-specific putative tandem-arranged repeats that could be used as cytogenetic markers in the genus (Table S3). Although further wide sampling is needed to confirm these trends, these observations expand the number of potential cytogenetic markers for comparison and identification of chromosomes in the karyotype of other *Juncus* species and reveal the phylogenenomic significance of satDNAs.

4.3 LTR retrotransposons drive genome evolution in Juncus

In this study, we performed comparative repeat analyses to analyzed the impact of DNA repeats on genome size in 33 *Juncus* taxa. Using low-coverage sequencing data including five sections of the subgenus *Juncus* (5/6) and three sections of the subgenus *Agathryron* (3/4), we confirmed that the content of dispersed and tandem repeats is an important contributor to genome size variation. The Ty1-copia superfamily was the largest genome constituent (45%), with the LTR-Angela lineage being one of the most abundant elements within the genus and an obvious increase in its abundance was observed in species with larger genome size, as in *Tenageia* and *Ozophyllum* clade. The evolutionary history of *Juncus* seems to be associated with multiple chromosomal fissions, so it is possible that these chromosomal rearrangements also impact repeat dynamics (Hofstatter et al., 2022) since there is a relationship between chromosomal

rearrangements and repetitive DNA amplification (see Li et al., 2017). Essentially, in the absence of evidence for a Juncaceae specific WGD event (not paleopolyploidy, Leebens-Mack et al., 2019; Hofstatter et al., 2022), it becomes evident that the high proportion of LTRs elements may have caused the differences in genome size within the *Juncus* genus, reflecting different clade-specific evolutionary trends.

In Juncaceae, records on genome size vary 7.8-fold, with 2C value ranging from 0.25 pg in *J. usitatus* to 3.58 pg in *J. articulatus* (Šmarda et al., 2019). Ty1-copia/Angela elements are not only associated with larger genome size, as in *Juncus*, but also with stress response genes, as observed different plant groups (Cavrak et al., 2014; Pietzenuk et al. 2016; Galindo-Gonzalez et al., 2017; Schley et al., 2022). Members of the Ty1-copia superfamily have evolved a tendency to insert close or inside the exons, given the adaptive benefit of evading cell methylation and silencing, making them prone to expansion under stress conditions (Lockton and Gaut, 2009; Galindo-González and Deyholos, 2012; Galindo-González et al., 2017). Some hypothesis suggests that stress conditions, such as acute or chronic exposure events to clastogens of varying intensity and duration, possibly explain the repeated origin of holocentric chromosomes in eukaryotes, where their tolerance to fragmentation leads to an advantage in clastogenic conditions and environments (Márquez-Corro et al., 2018; Zedek and Bureš, 2018; Zedek et al., 2021). Similar to the high abundance of Angela elements in *Juncus* genome, in the sister holocentric specie *L. elegans*, over 33% of their genome is composed of Angela, which are uniformly dispersed along the chromosomes (Heckmann et al., 2013). The amplification of Ty1-copia sequence could be, therefore, a significant driving force behind the Juncaceae genomes evolution and could have some role in the monocentric/ holocentric transitions.

5. Conclusions

In this study, we sequenced, assembled the plastid genomes and analyzed the repeat fraction of 33 *Juncus* taxa, providing a valuable genomic resource for this genus. Based on phylogenomic analysis, we have clarified the relations between the different sections of the genus previously proposed, allowing us to support both the union and separation of the sections, but not the merging of the all genus *Juncus*, since the species show shared evolutionary traits (repeatomic synapomorphies). Immunostaining with the CENH3 antibody confirmed the monocentricity of the genus and its combination with a probe for centromeric satDNA indicated a possible role of this sequence in centromere function. Regarding the repetitive fraction, in general, the patterns observed in this study show that genome size within monocentric *Juncus* species is influenced by the expansion/contraction of repeats. In *Juncus* the dynamics of repetitive elements are leading the underlying mechanisms that have driven a differentiation in genome size and phylogenetic relationships. Our results show that as long as LTR repeats were abundant within the genome, the increase in satDNA would be limited. Similarly, we also show that certain retrotransposon lineages (e.g. Ty1-copia/Angela and SIRE elements) have amplified in some clade. The high diversification of repeat fractions in these genomes may be correlated with genus divergence (56 Mya) and reveals the potential of repeat-based comparative genomics for understanding plant systematics and evolution.

Data availability

All sequencing data used in this study were submitted to the NCBI under the Bioproject number PRJNA984183. All other data needed to evaluate the conclusions in the paper are provided in the paper and/or the Supplementary Materials. Additional data related to this study may be requested from the authors.

CRediT authorship contribution statement

Yennifer Mata-Sucre: Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **William Matzenauer:** Resources, Writing – review & editing. **Natália Souza:** Data curation, Writing – review & editing, Visualization. **Bruno Huettel:** Data curation, Writing – review & editing. **Andrea Pedrosa-Harand:** Resources, Writing – review & editing, Supervision. **André Marques**: Resources, Writing – review & editing, Supervision, Funding acquisition. **Gustavo Souza**: Conceptualization, Resources, Writing – review & editing, Supervision, Visualization, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We acknowledge the excellent technical assistance of Christina Philipp in preparing DNA for genome sequencing. We are thankful to Prof. Dr. Marcelo Guerra (UFPE), Prof. Dr. Leonardo Pessoa Felix (UFPB) and Dr. Erton Mendonça de Almeida for providing the plant material and for the previous study that originated this work. This work was conducted during a scholarship to YMS supported by the International Cooperation Program PROBRAL at the Federal University of Pernambuco. Financed by Capes – Brazilian Federal Agency for Support and Evaluation of Graduate Education within the Ministry of Education of Brazil. This work was supported by a grant awarded to A.P.H (PROBRAL CAPES/DAAD project number 88881.144086/2017-01). G.S. receive productivity fellowship from CNPq (process numbers PQ-312852/2021-5). We also thank the ELIXIR-CZ Research Infrastructure Project (LM2015047) for providing computational resources for RepeatExplorer analysis.

References

Abadi, S., Azouri, D., Pupko, T., & Mayrose, I. 2019. Model selection may not be a mandatory step for phylogeny reconstruction. Nat. Commun., 10(1), 934. doi: 10.1038/s41467-019-08798-8

Aguilera, P. M., Debat, H. J., García, Y. S., Martí, D. A., & Grabiele, M. 2014. IAPT/IOPB chromosome data 18. Taxon, 63(6), 1387-1393. doi: 10.12705/636.16

Balslev, H. 1996. Juncaceae. Fl. Neotrop., 1-167.

Balslev, H. 1998. Juncaceae. In Flowering Plants· Monocotyledons (pp. 252-260). Springer, Berlin, Heidelberg.

Belyayev, A. 2014. Bursts of transposable elements as an evolutionary driving force. J. Evol. 27, 2573–2584. doi: 10.1111/jeb.12513

Beuzenberg, E. J., & Hair, J. B. 1983. Contributions to a chromosome atlas of the New Zealand flora-25 Miscellaneous species. N. Z. J. Bot., 21(1), 13-20. doi: 10.1080/0028825X.1983.10427564

Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. 2015. Repetitive DNA in eukaryotic genomes. Chromosome Res., 23(3), 415-420. doi: 10.1007/s10577-015-9499-z

Bouchenak-Khelladi Y, Muasya AM, Linder HP. 2014. A revised evolutionary history of Poales: origins and diversification. Bot. J. Linn. Soc. 1, 175(1), 4-16.

Breman, F. C., Chen, G., Snijder, R. C., Schranz, M. E., & Bakker, F. T. 2021. Repeatome-Based Phylogenetics in *Pelargonium* Section Ciconium (Sweet) Harvey. Genome Biol. Evol., 13(12), evab269. doi: 10.1093/gbe/evab269

Brookfield, J. 2005. The ecology of the genome — mobile DNA elements and their hosts. Nat. Rev. Genet., 6, 128–136. doi: 10.1038/nrg1524

Brožová, V., Proćków, J., & Drábková, L. Z. 2022. Toward finally unraveling the phylogenetic relationships of Juncaceae with respect to another cyperid family, Cyperaceae. Mol. Phylogenet. Evol., 177, 107588. doi: 10.1016/j.ympev.2022.107588

Buchenau, F. 1890. Monographia Juncacearum. Bot. JahrbPflanzengesch. Pflanzengeogr., 12, 1-459.

Buchenau, F. 1906. Juncaceae, pp. 1–285. In H. G. A. Engler [ed.], Das Pflanzenreich, Vol. 4 (36). W. Engelmann, Leipzig, Germany.

Carvalho C.R., Saraiva L.S. 1993. An air drying technique for maize chromosomes without enzymatic maceration. Biotech Histochem 68: 142–145. doi.org/10.3109/10520299309104684

Cavrak V.V., Lettner N., Jamge S., Kosarewicz A., Bayer L.M., Mittelsten Scheid O. 2014. How a retrotransposon exploits the plant's heat stress response for its activation. PLoS Genet. 10 (1): e1004115. DOI: 10.1371/journal.pgen.1004115.

Christenhusz, M. J., & Byng, J. W. 2016. The number of known plant species in the world and its annual increase. Phytotaxa, 261(3), 201-217. doi: 10.11646/phytotaxa.261.3.1

Chumova, Z., Belyayev, A., Mandáková, T., Zeisek, V., Hodkova, E., Šemberová, K., ... & Trávníček, P. 2022. The relationship between transposable elements and ecological

niches in the Greater Cape Floristic Region: A study on the genus *Pteronia* (Asteraceae). Front. Plant Sci. 13, 982852. doi.org/10.3389/fpls.2022.982852

Cope, T. A., & Stace, C. A. 1985. Cytology and hybridization in the Juncus bufonius L. aggregate in western Europe. Watsonia. 15, 309-320.

Costa, L., Marques, A., Buddenhagen, C. E., Pedrosa-Harand, A., & Souza, G. (2023). Investigating the diversification of holocentromeric satellite DNA Tyba in *Rhynchospora* (Cyperaceae). Ann. Bot., 131(5), 813-825. doi: 10.1093/aob/mcag264

Costa, L., Marques, A., Buddenhagen, C., Thomas, W. W., Huettel, B., Schubert, V., ... & Pedrosa-Harand, A. 2021. Aiming off the target: recycling target capture sequencing reads for investigating repetitive DNA. Ann. Bot., 128(7), 835-848. doi: 10.1093/aob/mcab117

Dalgaard, V. 1991. Chromosome studies in flowering plants from Macaronesia II. Willdenowia, 21, 139-152.

Devos, K. M., Brown, J. K. M., Bennetzen, J. L. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res. 12, 1075–1079. doi: 10.1101/gr.132102

Dodsworth, S. 2015. Genome skimming for next-generation biodiversity analysis. Trends Plant. Sci., 20(9), 525-527. https://doi.org/10.1016/j.tplants.2015.05.004

Dodsworth, S., Jang, T. S., Struebig, M., Chase, M. W., Weiss-Schneeweiss, H., & Leitch, A. R. 2017. Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). Plant Syst. Evol., 303(8), 1013-1020. https://doi.org/10.1007/s00606-017-1412-3

Doležel, J., Sgorbati, S., & Lucretti, S. 1992. Comparison of three DNA fluorochromes for flow cytometric estimation of nuclear DNA content in plants. Physiol. Plant, 85(4), 625-631. https://doi.org/10.1111/j.1399-3054.1992.tb04764.x

Drábková, L., & Kirschner, J. 2013. *Oreojuncus*, a new genus in the Juncaceae. Preslia, 85, 483-503.

Drábková, L., Kirschner, J., & Vlček, Č. 2006. Phylogenetic relationships within *Luzula* DC. and *Juncus* L.(Juncaceae): A comparison of phylogenetic signals of trnL-trnF intergenic spacer, trnL intron and rbcL plastome sequence data. Cladistics, 22(2), 132-143. https://doi.org/10.1111/j.1096-0031.2006.00095.x

Drábková, L. Z., & Vlček, Č. 2010. Molecular phylogeny of the genus *Luzula* DC. (Juncaceae, Monocotyledones) based on plastome and nuclear ribosomal regions: a case of incongruence, incomplete lineage sorting and hybridisation. Mol Phylogenet. Evol., 57(2), 536-551. https://doi.org/10.1016/j.ympev.2010.07.015

Drábková L., Kirschner J., Seberg O., Petersen G. & Vlček Č. 2003: Phylogeny of the Juncaceae based on rbcL sequences, with special emphasis on *Luzula* DC. and *Juncus* L. – Plant Syst. Evol., 240: 133-147. doi:10.1007/s00606-003-0018-y

Drummond, A. J., Ho, S. Y. W., Phillips, M. J., and Rambaut, A. 2006. Relaxed Phylogenetics and Dating with Confidence. Plos Biol. 4, e88. doi:10.1371/ journal.pbio.0040088

Drummond, A. J., and Rambaut, A. 2007. BEAST: Bayesian Evolutionary Analysis by Sampling Trees. BMC Evol. Biol. 7, 214. doi:10.1186/1471-2148- 7-214

Drummond, A. J., Suchard, M. A., Xie, D., and Rambaut, A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. doi:10.1093/molbev/mss075

Elliott, T. L., Larridon, I., Barrett, R. L., Bruhl, J. J., Costa, S. M., Escudero, M., ... & Muasya, A. M. 2023. Addressing inconsistencies in Cyperaceae and Juncaceae taxonomy: Comment on Brožová et al. (2022). Mol. Phylogenet. Evol., 179, 107665. https://doi.org/10.1016/j.ympev.2022.107665

Enriquez-Gasca, R., Gould, P. A., & Rowe, H. M. 2020. Host gene regulation by transposable elements: the new, the old and the ugly. Viruses, 12(10), 1089. https://doi.org/10.3390/v12101089

Escudero, M., & Hipp, A. 2013. Shifts in diversification rates and clade ages explain species richness in higher-level sedge taxa (Cyperaceae). Amer. J. Bot. 100(12), 2403-2411.

Fan, H., Ives, A. R., Surget-Groba, Y., & Cannon, C. H. 2015. An assembly and alignment-free method of phylogeny reconstruction from next-generation sequencing data. BMC Genomics, 16(1), 1-18. https://doi.org/10.1186/s12864-014-1199-x

Ferree, P. M., Barbash, D. A. 2009. Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. PloS Biol. 7, e1000234. doi: 10.1371/journal.pbio.1000234

Ferreira, M. E., & Grattapaglia, D. 1995. Introdução ao uso de marcadores RAPD e RFLP em análise genética. Brasília: Embrapa-Cenargen.

Fu, J., Zhang, H., Guo, F., Ma, L., Wu, J., Yue, M., ... & Li, L. 2019. Identification and characterization of abundant repetitive sequences in *Allium cepa*. Sci. Rep., 9(1), 1-7. https://doi.org/10.1038/s41598-019-43868-3

Galindo-González, L. & Deyholos, M. K. 2012. Identification, characterization and distribution of transposable elements in the flax (*Linum usitatissimum* L.) genome. BMC Genomics, 13(1), 1-17. https://doi.org/10.1186/1471-2164-13-644

Galindo-González, L., Mhiri, C., Deyholos, M. K., & Grandbastien, M. A. 2017. LTR-retrotransposons in plants: Engines of evolution. Gene, 626, 14-25. https://doi.org/10.1016/j.gene.2017.05.061

Garrido-Ramos, M. A. 2017. Satellite DNA: an evolving topic. Genes, 8(9), 230. https://doi.org/10.3390/genes8090230

Gernhard, T. 2008. The Conditioned Reconstructed Process. J. Theor. Biol. 253, 769–778. doi:10.1016/j.jtbi.2008.04.005

Guerra, M., Ribeiro, T., & Felix, L. P. 2019. Monocentric chromosomes in *Juncus* (Juncaceae) and implications for the chromosome evolution of the family. Bot J Linn Soc, 191(4), 475-483. https://doi.org/10.1093/botlinnean/boz021

Hartley, G., & O'Neill, R. J. 2019. Centromere repeats: hidden gems of the genome. Genes, 10(3), 223. DOI: 10.3390/genes10030223

Heckmann, S., Macas, J., Kumke, K., Fuchs, J., Schubert, V., Ma, L., ... & Houben, A. 2013. The holocentric species *Luzula* elegans shows interplay between centromere and large-scale genome organization. The Plant J., 73(4), 555-565. DOI: 10.1111/tpj.12057

Herklotz, V., Kovařík, A., Wissemann, V., Lunerová, J., Vozárová, R., Buschmann, S., ... & Ritz, C. M. 2021. Power and Weakness of Repetition-Evaluating the Phylogenetic Signal from Repeatomes in the Family Rosaceae With Two Case Studies From Genera Prone to Polyploidy and Hybridization (*Rosa* and *Fragaria*). Front. Plant Sci., 12, 738119. DOI: 10.3389/fpls.2021.738119

Hloušková, P., Mandáková, T., Pouch, M., Trávníček, P., & Lysak, M. A. 2019. The large genome size variation in the *Hesperis* clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. Ann. Bot., 124(1), 103-120. DOI: 10.1093/aob/mcz046

Hofstatter, P. G., Thangavel, G., Lux, T., Neumann, P., Vondrak, T., Novak, P., ... & Marques, A. 2022. Repeat-based holocentromeres influence genome architecture and karyotype evolution. Cell, 185(17), 3153-3168. DOI: 10.1016/j.cell.2021.09.039

Ibiapino, A., Báez, M., García, M. A., Costea, M., Stefanović, S., & Pedrosa-Harand, A. 2022. Karyotype asymmetry in *Cuscuta* L. subgenus Pachystigma reflects its repeat DNA composition. Chromosome Res., 30(1), 91-107.

Jones, E., Simpson, D. A., Hodkinson, T. R., Chase, M. W., & Parnell, J. A. 2007. The Juncaceae-Cyperaceae interface: a combined plastid sequence analysis. Aliso: J. Syst. Flor. Bot., 23(1), 55-61. doi: 10.5642/aliso.20072301.06

Jorgensen, C. A. 1958. The flowering plants of Greenland. A taxonomical and cytological survey. Kongel. Danske Vidensk. Selsk. Biol. Skr., 9, 1-172. Doi: 10.5281/zenodo.1449077

Katoh, K., & Standley, D. M. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol., 30(4), 772-780. Doi: 10.1093/molbev/mst010

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28(12), 1647-1649. Doi: 10.1093/bioinformatics/bts199

Kejnovsky, E., & Jedlicka, P. 2022. Nucleic acids movement and its relation to genome dynamics of repetitive DNA: Is cellular and intercellular movement of DNA and RNA

molecules related to the evolutionary dynamic genome components?. BioEssays, 44: 2100242. doi.org/10.1002/bies.202100242

Kirschner, J., Balslev, H., Češka, A., Swab, J. C., Edgar, E., Garcia-Herran, K., ... & Wilton, A. 2002. Juncaceae 1: *Rostkovia* to *Luzula*, Species Plantarum: Flora of the World Part 6: 1-237. ABRS, Canberra, Australia. Doi: 10.5962/bhl.title.17474

Kirschner J. (ed). 2002a. Juncaceae 2: *Juncus* subgenus *Juncus*. – In: Orchard A. E. (ed.), Species plantarum: Flora of the World. Part 7: 1–336. – ABRS, Canberra, Australia. Doi: 10.5962/bhl.title.14705

Kirschner J. (ed). 2002b. Juncaceae 3: *Juncus* subgenus *Agathryon*. – In: Orchard A. E. (ed.), Species plantarum: Flora of the World. Part 8: 1–192. – ABRS, Canberra, Australia. doi: 10.5962/bhl.title.14706

Kirschner, J., Novara, L. J., Novikov, V. S., Snogerup, S., & Kaplan, Z. 1999. Supraspecific division of the genus *Juncus* (Juncaceae). Folia Geobot., 34(3), 377-390. doi: 10.1007/BF02803151

Le Rouzic, A., Deceliere, G. 2005. Models of the population genetics of transposable elements. Genet. Res. 85, 171–181. doi: 10.1017/S0016672305007585

Leebens-Mack, J. H., Barker, M. S., Carpenter, E. J., Deyholos, M. K., Gitzendanner, M. A., Graham, S. W., ... & Li, X. 2019. One thousand plant transcriptomes and the phylogenomics of green plants. Nature, 574(7780), 679-685. doi: 10.1038/s41586-019-1693-2

Li, S. F., Su, T., Cheng, G. Q., Wang, B. X., Li, X., Deng, C. L., & Gao, W. J. 2017. Chromosome evolution in connection with repetitive sequences and epigenetics in plants. Genes, 8(10), 290. doi: 10.3390/genes8100290

Lisch, D. 2013. How important are transposons for plant evolution?. Nat. Rev. Genet., 14(1), 49-61. doi: 10.1038/nrg3374

Lockton, S., & Gaut, B. S. 2009. The contribution of transposable elements to expressed coding sequence in *Arabidopsis thaliana*. J. Mol. Evol., 68(1), 80-89. doi: 10.1007/s00239-008-9193-6

Loureiro, J., Rodriguez, E., Gomes, A., & Santos, C. 2007. Genome size estimations on *Ulmus minor* Mill., *Ulmus glabra* Huds., and *Celtis australis* L. using flow cytometry. Plant Biol., 9(4), 541-544. doi: 10.1055/s-2007-965433

Louzada, S., Lopes, M., Ferreira, D., Adega, F., Escudeiro, A., Gama-Carvalho, M., & Chaves, R. 2020. Decoding the role of satellite DNA in genome architecture and plasticity—An evolutionary and clinical affair. Genes, 11(1), 72. doi: 10.3390/genes11010072

Lu, M., Fang, Z., Sheng, F., Tong, X., & Han, R. 2021. Characterization and phylogenetic analysis of the complete chloroplast genome of *Juncus effusus* L. Mitochondrial DNA Part B, 6(5), 1612-1613. doi: 10.1080/23802359.2021.1921203

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.doi.org/10.1016/j.gene.2009.06.014

Márquez-Corro, J. I., Escudero, M., & Luceño, M. 2018. Do holocentric chromosomes represent an evolutionary advantage? A study of paired analyses of diversification rates of lineages with holocentric chromosomes and their monocentric closest relatives. Chromosome Res., 26(3), 139-152. doi: 10.1007/s10577-018-9572-z

Mata-Sucre, Y., Costa, L., Gagnon, E., Lewis, G. P., Leitch, I. J., & Souza, G. 2020a. Revisiting the cytomolecular evolution of the Caesalpinia group (Leguminosae): a broad sampling reveals new correlations between cytogenetic and environmental variables. Plant Syst. Evol., 306(2), 1-13. doi: 10.1007/s00606-020-01739-2

Mata-Sucre, Y., Sader, M., Van-Lume, B., Gagnon, E., Pedrosa-Harand, A., Leitch, I. J., ... & Souza, G. 2020b. How diverse is heterochromatin in the Caesalpinia group? Cytogenomic characterization of *Erythrostemon hughesii* Gagnon & GP Lewis (Leguminosae: Caesalpinioideae). Planta, 252(4), 1-14. doi: 10.1007/s00425-020-03478-3

Mehra, P. N., & Sachdeva, S. K. 1976. Cytological observations on some W. Himalayan monocots. II. Smilacaceae, Liliaceae and Trilliaceae. Cytologia, 41, 5-22.

Melters, D. P., Bradnam, K. R., Young, H. A., Telis, N., May, M. R., Ruby, J. G., ... & Chan, S. W. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol., 14(1), R10. doi: 10.1186/gb-2013-14-1-r10

Meštrović, N., Mravinac, B., Pavlek, M., Vojvoda-Zeljko, T., Šatović, E., & Plohl, M. 2015. Structural and functional liaisons between transposable elements and satellite DNAs. Chromosome Res., 23(3), 583-596. doi: 10.1007/s10577-015-9485-5

Mičieta, K. 1983. Contribution to the chromosome numbers of some species of the genus *Juncus* L. in Slovakia. Folia Geobot. Phytotax., 18, 195-198

Novikov, V. S. 1990. Konspekt sistemy roda *Luzula* DC (Juncaceae) (Survey of the system of the genus *Luzula*). Bjull. Mosk. Obšč. Ispyt. Prir. (5), 111-125.

Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P., & Macas, J. 2017. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Res., 45(12), e111. doi: 10.1093/nar/gkx257

Novák, P., Guignard, M. S., Neumann, P., Kelly, L. J., Mlinarec, J., Koblížková, A., ... & Leitch, A. R. 2020. Repeat-sequence turnover shifts fundamentally in species with large genomes. Nat. Plants, 6(11), 1325-1329. doi: 10.1038/s41477-020-00818-z

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., & Macas, J. 2013. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements

from next-generation sequence reads. Bioinformatics, 29(6), 792-793. doi: 10.1093/bioinformatics/btt054

Oksanen, J., Blanchet, F. G., Kindt, R., Legendre, P., Minchin, P. R., O'hara, R. B., ... & Wagner, H. 2013. Community ecology package. R package version, 2(0), 321-326.

Oliveira, M. A. S., Nunes, T., Dos Santos, M. A., Ferreira Gomes, D., Costa, I., Van-Lume, B., ... & Marques, A. 2021. High-Throughput Genomic Data Reveal Complex Phylogenetic Relationships in *Stylosanthes* Sw (Leguminosae). Front. Genet., 12, 722949. doi: 10.3389/fgene.2021.722949

O'Mahony, T. (2002). The comparative morphology of *Juncus conglomeratus* L. Compact Rush), *J. effusus* L. (Soft-rush) and their interspecific hybrid, IRISH BOTANICAL J, 9, 5-14.

Pagel, M. (1999). Inferring the historical patterns of biological evolution. Nature, 401, 877–884. doi: 10.1038/44766

Pedrosa, A., Sandal, N., Stougaard, J., Schweizer, D., & Bachmair, A. 2002. Chromosomal map of the model legume *Lotus japonicus*. Genetics, 161(4), 1661-1672. doi: 10.1093/genetics/161.4.1661

Pellicer, J., Hidalgo, O., Dodsworth, S., & Leitch, I. J. 2018. Genome size diversity and its impact on the evolution of land plants. Genes, 9(2), 88. doi: 10.3390/genes9020088

Pietzenuk B., Markus C., Gaubert H., Bagwan N., Merotto A., Bucher E., Pecinka A. 2016. Recurrent evolution of heat-responsiveness in Brassicaceae COPIA elements. Genome Biol. 17 (1), 209. DOI: 10.1186/s13059-016-1072-3

POWO. 2023. "Plants of the World Online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; http://www.plantsoftheworldonline.org/Retrieved 07 August 2023"

Price, M. N., Dehal, P. S., & Arkin, A. P. 2010. FastTree 2–approximately maximum-likelihood trees for large alignments. PloS one, 5(3), e9490. doi.org/10.1371/journal.pone.0009490

Raskina, O., Brodsky, L., Belyayev, A. 2011. Tandem repeats on an eco-geographical scale: outcomes from the genome of *Aegilops speltoides*. Chromosome Res. 19, 607–623. doi: 10.1007/s10577-011-9220-9

Revell, L. J. 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol., 3(2), 217-223. doi: 10.1111/j.2041-210X.2011.00169.x

Ribeiro, T., Marques, A., Novák, P., Schubert, V., Vanzela, A. L., Macas, J., ... & Pedrosa-Harand, A. 2017. Centromeric and non-centromeric satellite DNA organization differs in holocentric *Rhynchospora* species. Chromosoma, 126, 325-335. doi: 10.1007/s00412-016-0605-y

Ribeiro, T., Vaio, M., Félix, L. P., & Guerra, M. 2022. Satellite DNA probes of *Alstroemeria longistaminea* (Alstroemeriaceae) paint the heterochromatin and the B chromosome,

reveal a G-like banding pattern, and point to a strong structural karyotype conservation. Protoplasma, 1-14. doi: 10.1007/s00709-022-01795-7

Roalson, E. H. 2005. Phylogenetic relationships in the Juncaceae inferred from nuclear ribosomal DNA internal transcribed spacer sequence data. Int. J. Plant Sci., 166(3), 397-413. doi: 10.1086/427219

Rostoks, N., Park, Y. J., Ramakrishna, W., Ma, J., Druka, A., Shiloff, B. A., ... & Kleinhofs, A. 2002. Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Funct. Integr. Genomics, 2(1), 51-59. doi: 10.1007/s10142-001-0040-8

Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., & Camacho, J. P. M. 2016. High-throughput analysis of the satellitome illuminates satellite DNA evolution. Sci. Rep., 6(1), 28333. doi: 10.1038/srep28333

Sader, M., Vaio, M., Cauz-Santos, L. A., Dornelas, M. C., Vieira, M. L. C., Melo, N., & Pedrosa-Harand, A. 2021. Large vs small genomes in Passiflora: The influence of the mobilome and the satellitome. Planta, 253(4), 1-18. doi: 10.1007/s00425-021-03694-y

Satović, E., Vojvoda Zeljko, T., Luchetti, A., Mantovani, B., & Plohl, M. 2016. Adjacent sequences disclose potential for intra-genomic dispersal of satellite DNA repeats and suggest a complex network with transposable elements. BMC Genomics, 17(1), 1-12. doi: 10.1186/s12864-015-2350-y

Schley, R. J., Pellicer, J., Ge, X. J., Barrett, C., Bellot, S., Guignard, M. S., ... & Leitch, I. J. 2022. The ecology of palm genomes: repeat-associated genome size expansion is constrained by aridity. New Phytol., 236(2), 433-446. doi: 10.1111/nph.17680

Schönswetter, P., Suda, J., Popp, M., Weiss-Schneeweiss, H., & Brochmann, C. 2007. Circumpolar phylogeography of *Juncus biglumis* (Juncaceae) inferred from AFLP fingerprints, cpDNA sequences, nuclear DNA content and chromosome numbers. Mol. Phylogenet. Evol., 42(1), 92-103. doi: 10.1016/j.ympev.2006.06.010

Shannon, C.E. 1948. A mathematical theory of communication. Bell Syst. Tech. J., 27, 379–423. doi: 10.1002/j.1538-7305.1948.tb01338.x

Šimoníková, D., Čížková, J., Zoulová, V., Christelová, P., & Hřibová, E. 2022. Advances in the Molecular Cytogenetics of Bananas, Family Musaceae. Plants, 11(4), 482. doi: 10.3390/plants11040482

Šmarda, P., Knápek, O., Březinová, A., Horová, L., Grulich, V., Danihelka, J., ... & Bureš, P. 2019. Genome sizes and genomic guanine+cytosine (GC) contents of the Czech vascular flora with new estimates for 1700 species. Preslia, 91(2), 117-142. doi: 10.23855/preslia.2019.117

Smith, P. H. 2006. Revisiting *Juncus balticus* Willd. in England. Watsonia, 26 (1), 57-66.

Souza, G., Costa, L., Guignard, M. S., Van-Lume, B., Pellicer, J., Gagnon, E., ... & Wendt, T. 2022. Do tropical plants have smaller genomes? Correlation between genome size and

climatic variables in the Caesalpinia Group (Caesalpinioideae, Leguminosae). Perspect. Plant Ecol. Evol. Syst., 38, 13-23. doi: 10.1016/j.ppees.2022.01.003

Stitzer, M. C., Anderson, S. N., Springer, N. M., & Ross-Ibarra, J. 2021. The genomic ecosystem of transposable elements in maize. PLoS Genet., 17(10), e1009768. doi: 10.1371/journal.pgen.1009768

Talbert, P. B., & Henikoff, S. 2020. What makes a centromere? Exp. Cell Res., 389(2), 111895. doi: 10.1016/j.yexcr.2019.111895

Thakur, J., Packiaraj, J., & Henikoff, S. 2021. Sequence, chromatin and evolution of satellite DNA. Int. J. Mol. Sci., 22(9), 4309. doi: 10.3390/ijms22094309

Trávníček, P., Čertner, M., Ponert, J., Chumová, Z., Jersáková, J., & Suda, J. 2019. Diversity in genome size and GC content shows adaptive potential in orchids and is closely linked to partial endoreplication, plant life-history traits and climatic conditions. New Phytol., 224(4), 1642-1656. doi: 10.1111/nph.16166

Venner, S., Feschotte, C., & Biémont, C. 2009. Dynamics of transposable elements: towards a community ecology of the genome. Trends Genet., 25(7), 317-323. doi: 10.1016/j.tig.2009.05.003

Villanueva, R. A. M., & Chen, Z. J. 2019. ggplot2: elegant graphics for data analysis. Mol. Phylogenet. Evol., 147, 106766. doi: 10.1016/j.ympev.2020.106766

Vitales, D., Álvarez, I., Garcia, S., Hidalgo, O., Nieto Feliner, G., Pellicer, J., ... & Garnatje, T. 2020a. Genome size variation at constant chromosome number is not correlated with repetitive DNA dynamism in *Anacyclus* (Asteraceae). Ann. Bot., 125(4), 611-623. doi: 10.1093/aob/mcaa005

Vitales, D., Garcia, S., & Dodsworth, S. 2020b. Reconstructing phylogenetic relationships based on repeat sequence similarities. Mol. Phylogenet. Evol., 147, 106766. doi: 10.1016/j.ympev.2020.106766

Wang, D., Zheng, Z., Li, Y., Hu, H., Wang, Z., Du, X., ... & Yang, Y. 2021. Which factors contribute most to genome size variation within angiosperms? Ecol. Evol., 11(6), 2660-2668. doi: 10.1002/ece3.7163

Wendel, J. F., Lisch, D., Hu, G., & Mason, A. S. 2018. The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation. Curr. Opin. Genet. Dev., 49, 1-7. doi: 10.1016/j.gde.2018.01.001

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., ... & Schulman, A. H. 2007. A unified classification system for eukaryotic transposable elements. Nat. Rev. Genet., 8(12), 973-982. doi: 10.1038/nrg2165

Wicker, T., Stein, N., Albar, L., Feuillet, C., Schlagenhauf, E., & Keller, B. 2001. Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J., 26(3), 307-316. doi: 10.1046/j.1365-313x.2001.2641046.x

Wickham, H. 2016. Getting Started with ggplot2. In: ggplot2: Elegant graphics for data analysis, 11-31. doi: 10.1007/978-3-319-24277-4_2

Wilcox, M. 2011. Hybrid rushes in the UK–sterility and fertility. Deadline for contribution to News 117, 91- 21.

Zedek, F., & Bureš, P. 2018. Holocentric chromosomes: from tolerance to fragmentation to colonization of the land. Ann. Bot., 121(1), 9-16. doi: 10.1093/aob/mcx130

Zedek, F., Šmerda, J., Veselý, P., Horová, L., Kocmanová, J., & Bureš, P. 2021. Elevation-dependent endopolyploid response suggests that plants with holocentric chromosomes are less stressed by UV-B. Bot. J. Linn. Soc., 195(1), 106-113. doi: 10.1093/botlinnean/boaa114

Zonneveld, B. J. 2019. The DNA weights per nucleus (genome size) of more than 2350 species of the Flora of The Netherlands, of which 1370 are new to science, including the pattern of their DNA peaks. Forum Geobot., 8. doi: 10.3372/fg.2019.8

Zuo, S., Mandáková, T., Kubová, M., & Lysak, M. A. 2022. Genomes, repeatomes and interphase chromosome organization in the meadowfoam family (Limnanthaceae, Brassicales). Plant J. doi: 10.1111/tpj.15976

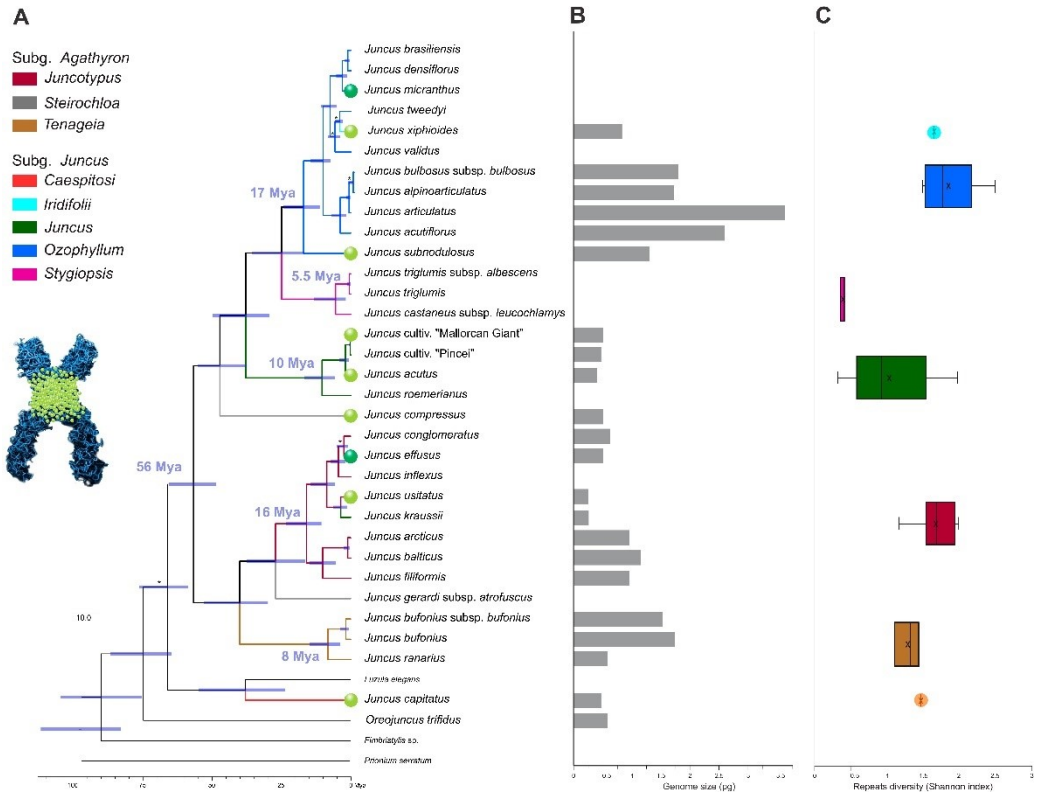


Fig. 1 Time-calibrated tree topology in the genus *Juncus* based on 7,031 bp rDNA alignment (A). (B) Genome size variation among the *Juncus* species is shown next to each species with nuclear genome size available in picograms (pg). (C) Shannon Index showing repeat diversity

for each clade are represented as boxplot. Green balls at the end of the phylogeny in A represent species analyzed cytologically here (light green) and in Guerra et al. (2019, dark green). Bootstrap support values (lower than 98) and the age of each clade (10,000,000 bootstrap replicates) have been provided in the tree nodes. Different sections within the genus are shown with colored bars. *Fimbristilis* sp. and *Prionium serratum* plastome were used as outgroup

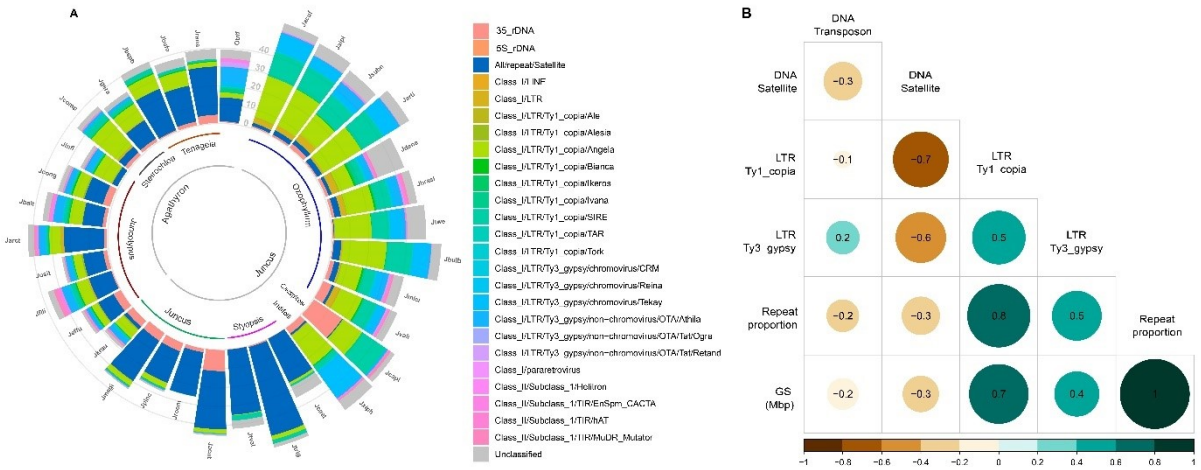


Fig. 2 Frequency of major classes of repeat DNA in the *Juncus* genome. All-to-all similarity comparison of individual relative abundance of total (A) and tandem (B) repeat DNA sequences in 33 *Juncus* taxa. Y-axis in A and B represent the proportion of repeats expressed as a percentage for each genome. (C) Heat map of the correlation matrix between nuclear genome size and genome traits: proportion of repeats, LTR-Ty1-copia element, LTR-Ty3-gypsy element, DNA satellite sequences and DNA transposons. Shades of green indicate increasing positive correlation coefficient; shades of brown indicate increasing negative correlation coefficient. Correlation values are within each ball. (D) Linear regression between the proportion of total repetitive DNA and nuclear genome size in *Juncus* species. The significance of the correlation coefficient (r) is shown in the graph ($p < 0.001$). Repeat coefficients were calculated by annotating clusters representing at least 0.01% of the genome in RepeatExplorer. Different colors in A and B represent different families of repetitive DNA sequences. See species name codes in Table 1

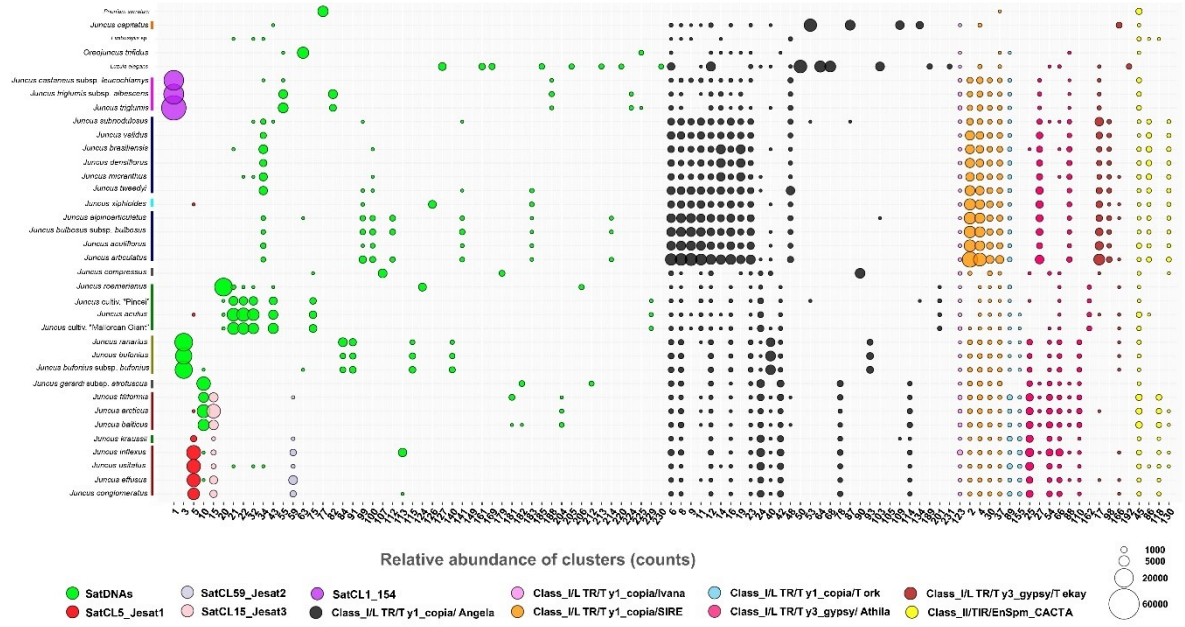


Fig. 3 Comparative analysis of the abundances of the main types of repetitive sequences shared among *Juncus* species (data in the figure is in Table S3). The size of the ball is proportional to the number of counts in that cluster for each species. The colors of the ball correspond to the different types of repetitive sequences. Colors next to the species correspond to the different sections

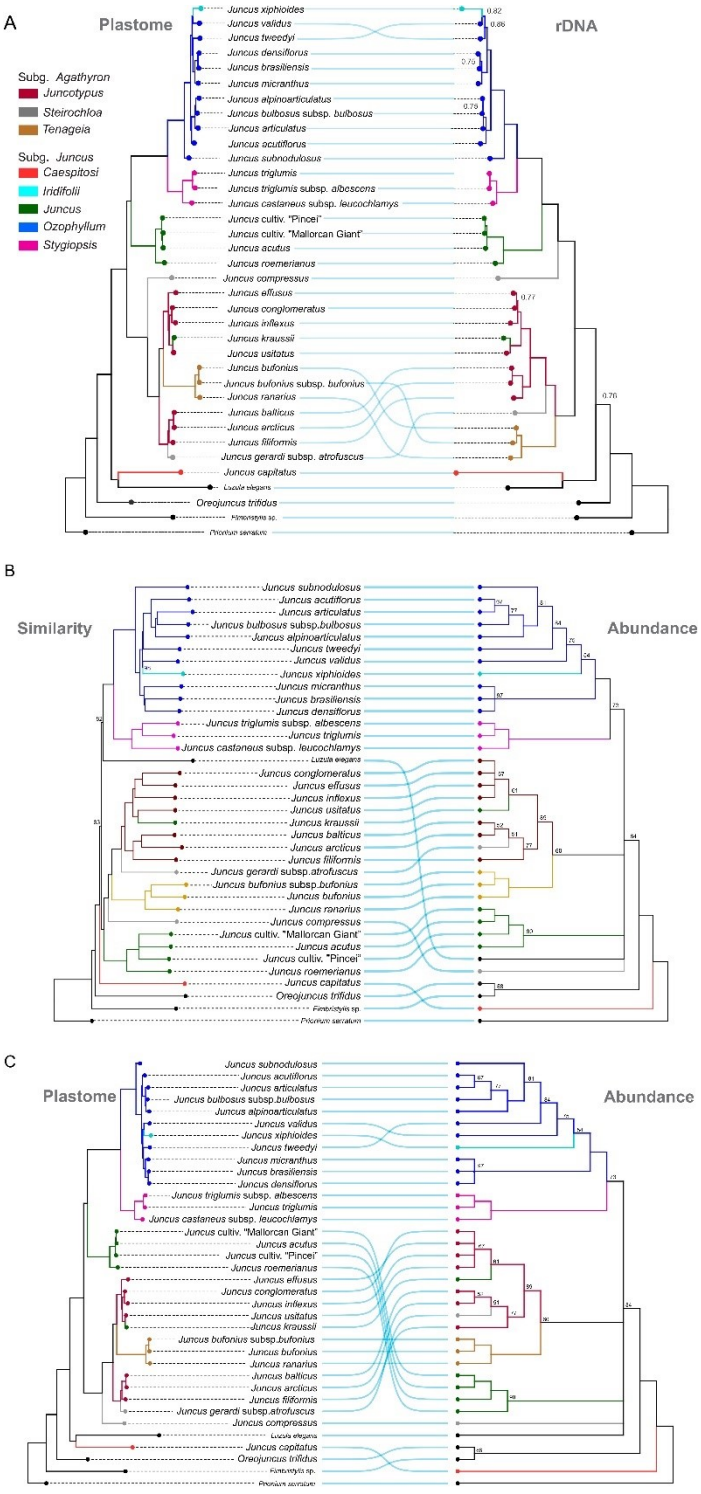


Fig. 4 (A) Comparative phylogenetic topologies of the similarity repeat-based tree (left) and the abundance repeat-based tree (right) among *Juncus* species. (B) Comparative tree topologies obtained from the plastome assembly tree (left) and the abundance repeat-based tree (right). The number above the branches represents the support of nodes less than 98%. Branch groupings by color represent sections within the genus. The blue lines in the center show the correspondence between species

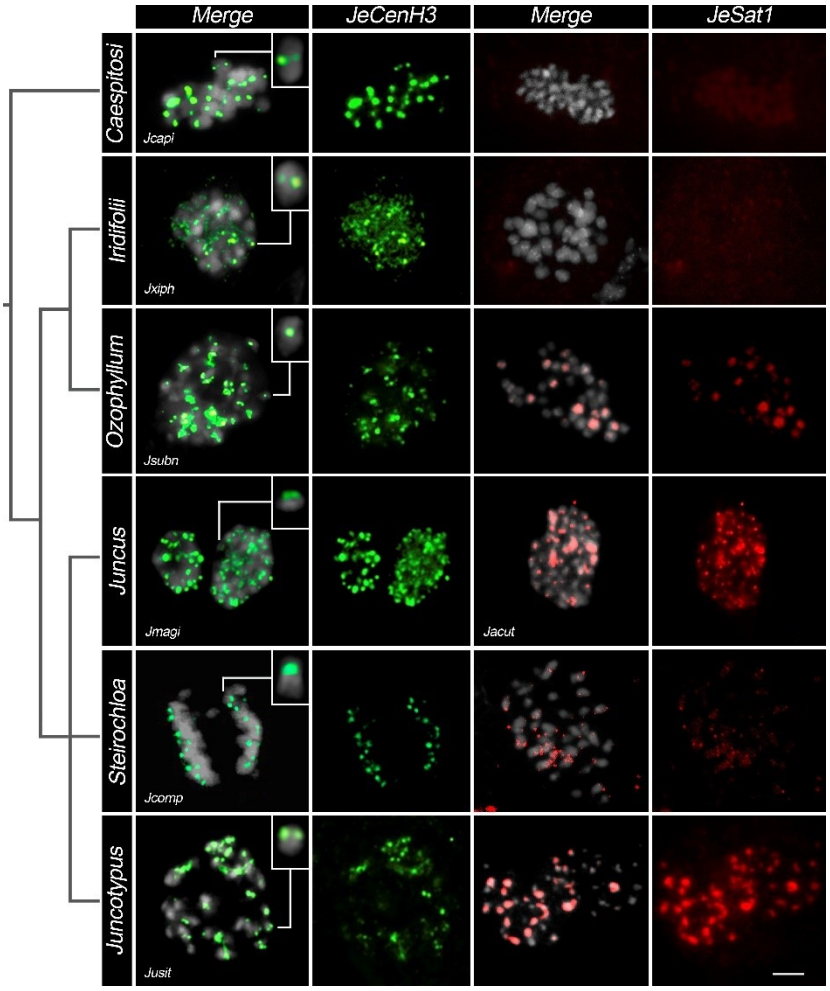


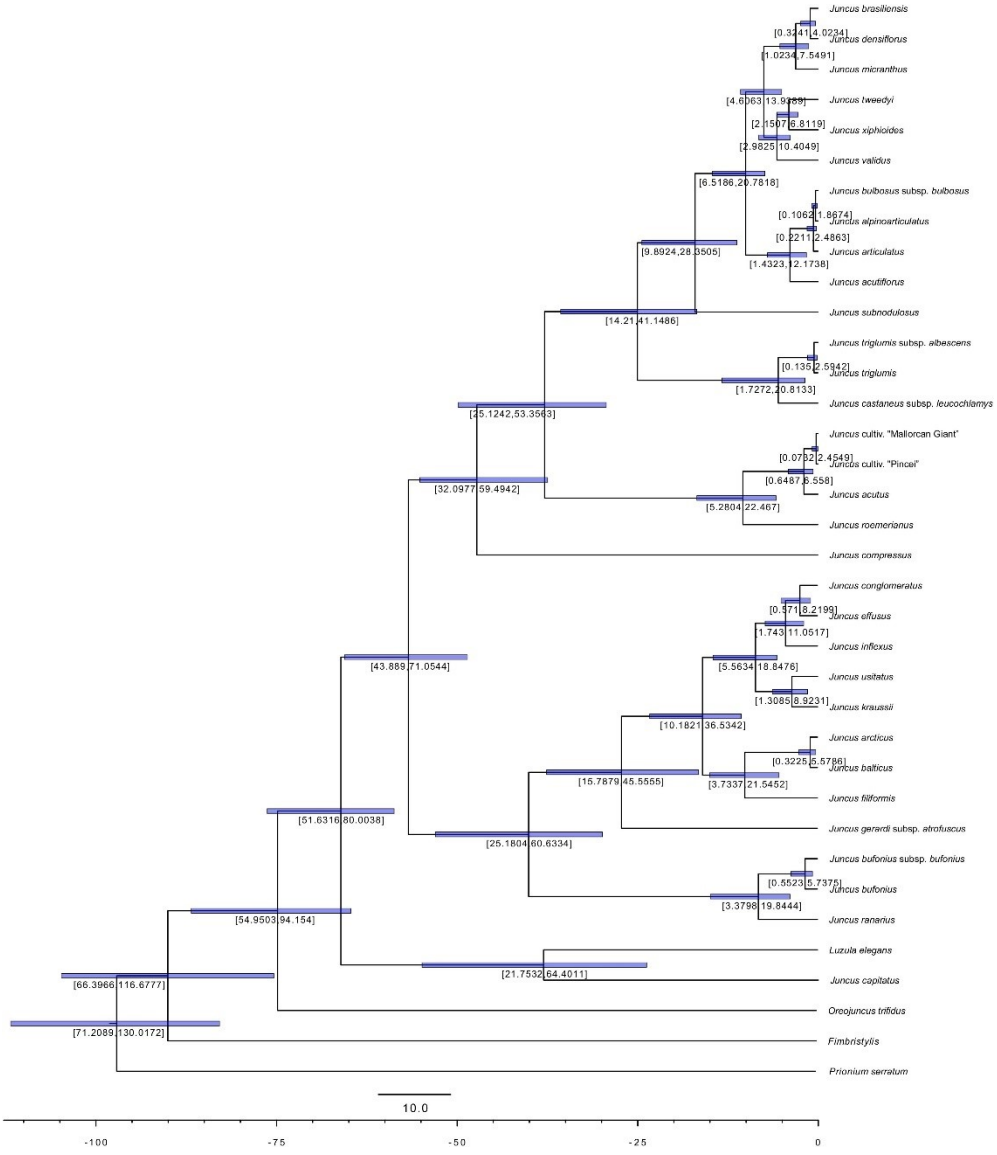
Fig. 5 Immuno-localization of the histone H3 variant CENH3 (JeCENH3) and centromeric satellite DNA JeSat1 on mitotic (pre) metaphase or anaphase chromosomes of *Juncus*. Column 1 and 3 shows merged images, column 2 shows CENH3 signals (green) and column 4 shows JeSat1 signals (red). One representative from sections *Caespitosi*, *Iridifoli*, *Juncus*, *Ozophyllum*, *Juncotypus* and *Steirochloa* were analyzed. Squares in the images of column 1 show chromosomes zoomed with CENH3 dot signals. Nuclei and chromosomes were counterstained with DAPI (blue). The bar corresponds to 2 µm

Supplementary Materials

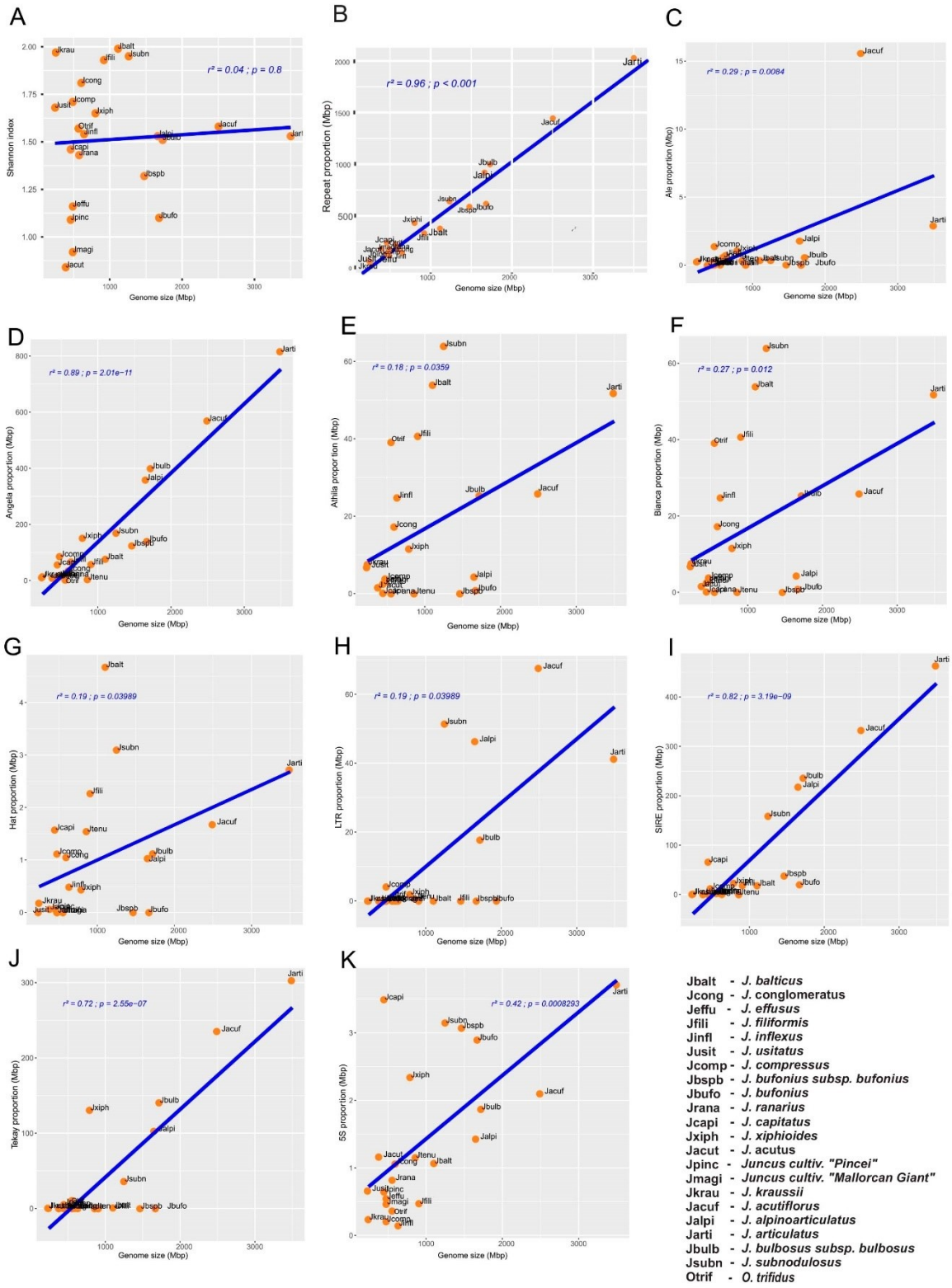
Supplementary Table S1 Genome proportions of different repeat classes and superfamilies in *Juncus* genome after individual RepeatExplorer analysis

Supplementary Table S2 Comparative abundance (read counts) of different repeat classes and superfamilies in *Juncus* species genome

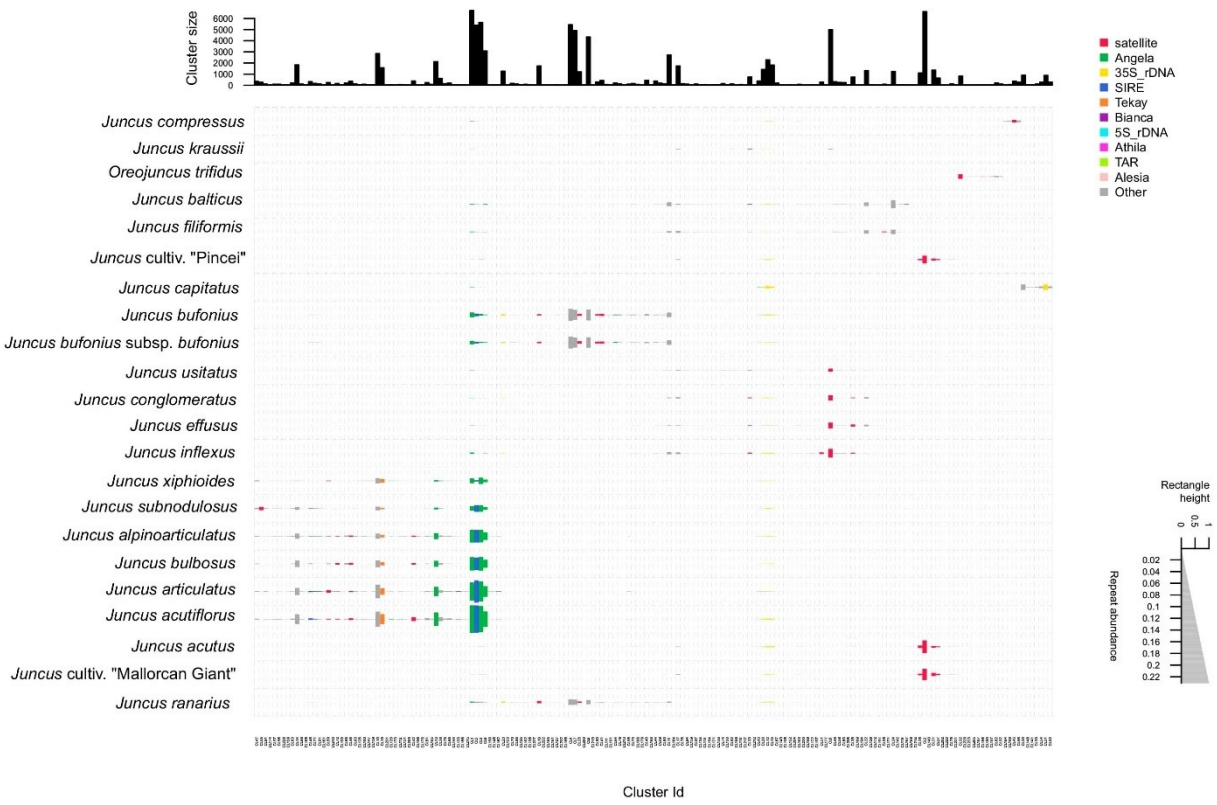
Supplementary Table S3 Cluster, monomer size and consensus sequence of the different DNA satellites found in *Juncus* genomes



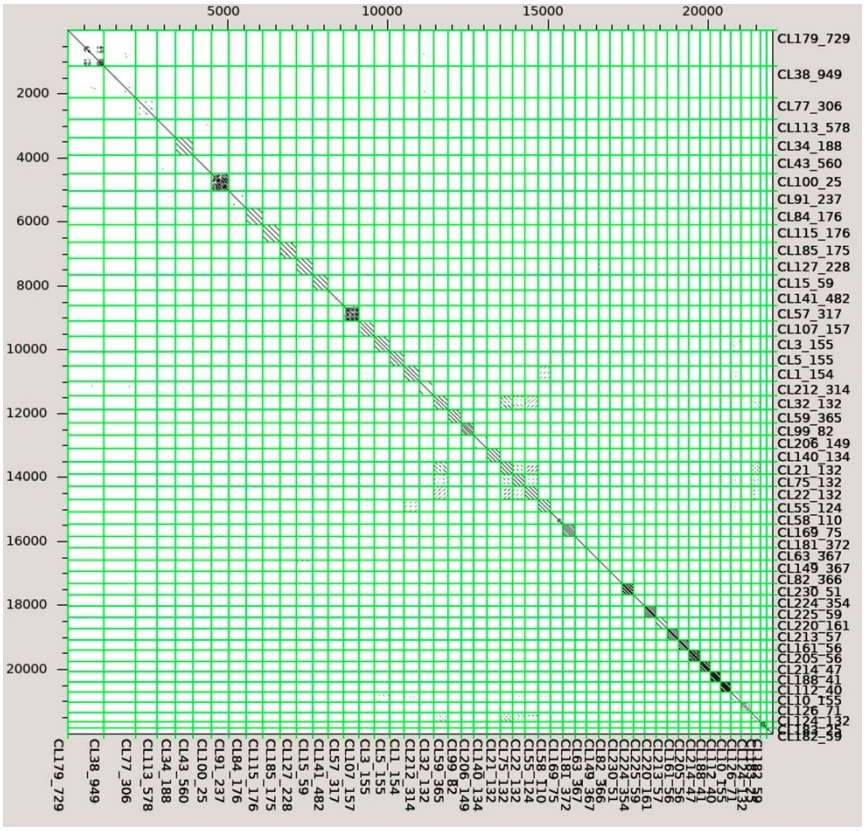
Supplementary Figure S1 Time-calibrated tree topology in the genus *Juncus* based on a 7,031 bp rDNA alignment, showing age intervals for each clade



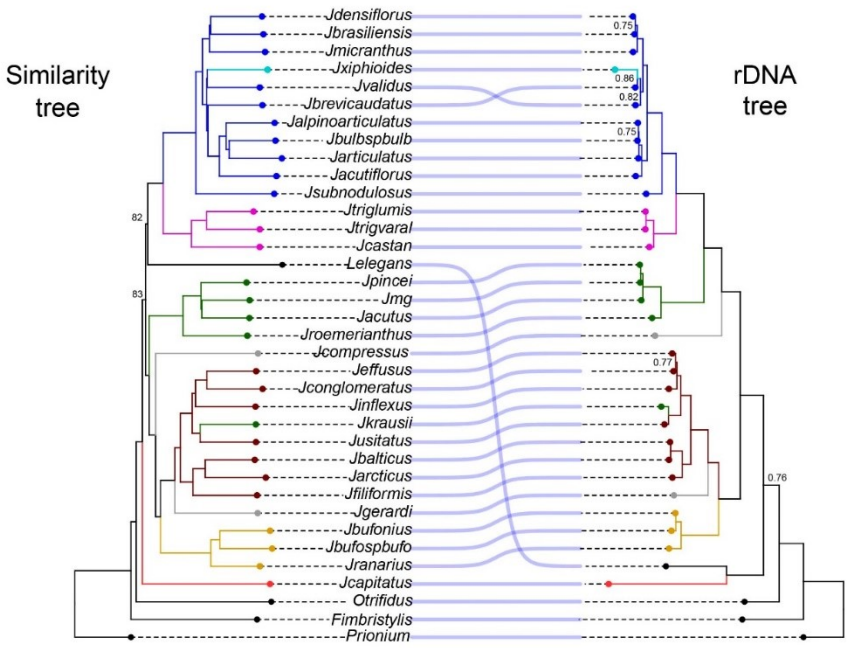
Supplementary Figure S2 Scatter plots showing the correlation between genome size and repeat diversity (A, Shannon index) and abundances of repeat lineages (B-J) used to assess repeat profiles and their contribution to genome size among *Juncus* species whose repeat compositions were analyzed with RepeatExplorer2. Data are expressed in Mbp



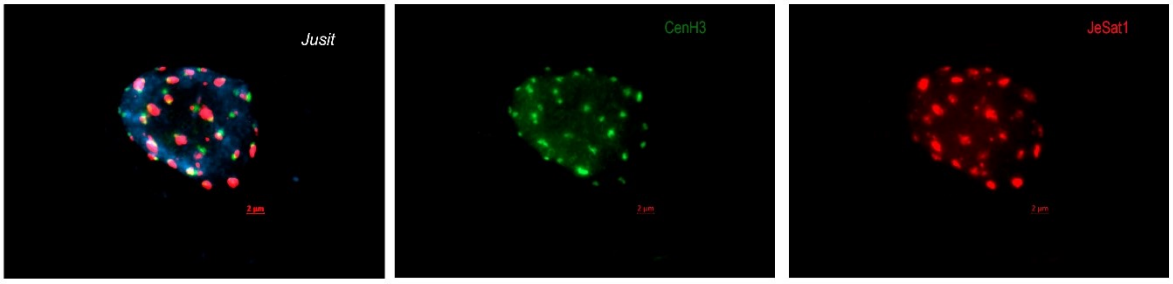
Supplementary Figure S3 All-to-all comparative analyses of the abundances of the main types of shared repetitive sequences in 0.1x genome covered for 21 *Juncus* species. The size of the rectangle is proportional to the number of reads in that cluster for each species The colors of the rectangle correspond to the different repetitive sequence types.



Supplementary Figure S4 Dot plot of all DNA satellites found within *Juncus* genomes showing the monomer size of each sequence, its name, tandem organization and sequence similarity



Supplementary Figure S5 Comparative tree topologies obtained from the repeat similarity and rDNA alignment. Branch groupings by color represent sections within the genus. The number on the branches represents the support of nodes below 90%. The blue lines in the center show the correspondence between species



Supplementary Figure S5 Co-localization of JeCENH3 (green) with the JeSat1 (red) *Juncus effusus* DNA satellite on interphase nuclei of *Juncus usitatus* species

ANEXO A – Normas de submissão da revista Botanical Journal of the Linnean Society - https://academic.oup.com/botlinnean/pages/General_Instructions

Preparation of Manuscript

Manuscript format and structure/style

Basic Formatting Guide

Authors should aim to communicate ideas and information clearly and concisely, in language suitable for the moderate specialist. Papers in languages other than English are not accepted unless invited. When a paper has joint authorship, one author must accept responsibility for all correspondence; the full postal address, telephone and fax numbers, and e-mail address of the author who is to check proofs (the corresponding author) should be provided. Although the Society does not specify the length of manuscripts, it is suggested that authors preparing long texts (20 000 words or more, including references, etc.) should consult the Editor before considering submission.

Please submit your manuscript in an editable format such as .doc, .docx or .rtf, prepared on A4, paginated, double spaced throughout (i.e. including references and quotations), with ample margins. If you submit your manuscript in a non-editable format such as PDF, this will slow the progress of your paper as we will have to contact you to request an editable copy.

Papers should conform to the following general layout:

A general style guide can be found here and a sample paper can be found here.

Article types

- Editorial
- Original Article
- Short Original Article
- Review
- Invited Review
- Comment

ANEXO B – Normas de submissão da revista Nature communications

https://www.nature.com/ncomms/submit/guide-to-authors

Guide to authors

Manuscript submissions

Nature Communications is an open access, multidisciplinary journal dedicated to publishing high-quality research in all areas of the biological, physical, chemical and Earth sciences. Papers published by the journal represent important advances of significance to specialists within each field.

Information about the common editorial policies of Nature journals, including *Nature Communications*, is available here.

Please see our brief guide to manuscript submission in PDF format for an overview of key information on submitting primary research for publication in *Nature Communications*.

As we are committed to making prompt and informed decisions about publication, we do not consider pre-submission enquiries. Instead, we prefer to make decisions with all relevant material at hand, and therefore encourage you to submit your full manuscript to us.