Genome Biology

https://doi.org/10.1186/s13059-026-04122-x

Article in Press

Reproductive mode and ploidy are associated with distinct organization of clustered centromeric regions in Meloidogyne nematodes

Received: 07 Apr 2025

Accepted: 20 May 2026

Published online: 27 May 2026

Cite this article as: Despot-Slade, E., Volarić, M., Veseljak, D. *et al.* Reproductive mode and ploidy are associated with distinct organization of clustered centromeric regions in Meloidogyne nematodes. *Genome Biol* (2026). https://doi.org/10.1186/s13059-026-04122-x

Evelin Despot-Slade, Marin Volarić, Damira Veseljak, Lucija Horvat, Lana Semenić, Ana Mota, Marine Poullet, Etienne Danchin, Brankica Mravinac & Nevenka Meštrović

Reproductive mode and ploidy are associated with distinct organization of clustered centromeric regions in Meloidogyne nematodes

Evelin Despot-Slade [#1], Marin Volarić [#1], Damira Veseljak[1], Lucija Horvat[1], Lana Semenić[1], Ana Paula Zotta Mota[2], Marine Poullet[2], Etienne G. J. Danchin [2], Brankica Mravinac [1], Nevenka Meštrović *[1]

[1]Division of Molecular Biology, Ruđer Bošković Institute, Bijenička cesta 54, 10000, Zagreb, Croatia
[2]Institut Sophia Agrobiotech, INRAE, Université Côte d'Azur, CNRS, 400 routes des Chappes, 06903, Sophia-Antipolis, France
#These authors contributed equally to this work.

*Corresponding author: nevenka@irb.hr

Abstract:

Background: In most eukaryotes, chromosomes are monocentric, each containing a single centromeric region. However, plant-parasitic nematodes of the genus Meloidogyne display a unique, repeat-based and clustered centromere organization. This unusual centromere structure, together with their asexual reproductive strategies, provides an opportunity to investigate the diversity and evolution of centromere organization.

Results: Using high-quality genome assemblies, we examine the genome organization and evolutionary patterns of centromeres in six Meloidogyne species differing in reproductive strategies (mitotic parthenogenesis vs. meiotic reproduction) and ploidy (polyploid vs. diploid). All species possess conserved CenH3 and 19-bp box-containing tandem repeats in their centromeres. Centromeric regions in three closely related and one distantly related mitotic/polyploid species exhibit a complex organization, predominantly consisting of higher-order arrangements of centromeric arrays associated with both autonomous and non-autonomous Helitron transposons, leading to longer and more dispersed centromeric regions. In contrast, meiotic/diploid species have simpler, less expanded centromeres that lack Helitron elements, highlighting the role of Helitron-mediated amplification and spread of centromeres in mitotic/polyploid species.

Conclusions: These findings suggest that reproductive strategy and/or genome ploidy may influence centromere organization, although it remains to be determined whether these factors are causative or simply correlated with differences in centromere organization. We suggest that the expansion-prone centromeric regions in mitotic/polyploid Meloidogyne may drive rapid chromosome evolution, potentially influencing their adaptability to environmental changes.

Keywords:

Clustered centromeres, nematodes, mitotic and meiotic species, transposons

Background

Centromeres are complex chromosomal loci playing key roles in the accurate process of chromosome segregation. Abnormalities in these regions cause many disorders such as aneuploidy, infertility, cancer or even cell death [1]. Regarding centromere architecture, the majority of animal and plant species have monocentric chromosomes characterized by primary constriction with a single regional centromere at metaphase. The centromere identity is defined epigenetically and relies on the presence of CenH3 proteins, centromere-specific histone H3 variants. This type of centromeres is often enriched in repetitive DNA, mainly megabase-sized satellite DNAs (satDNAs). Due to their abundance and repetitiveness, these regions are usually not included in even the most complete chromosome assemblies based on ultra-long reads [2]. The only exceptions are the recently published Telomere-to-Telomere assemblies of prominent model organisms such as human and *Arabidopsis* that contain the complete (peri)centromere [3–5].

In contrast to the most frequently occurring monocentric chromosomes, holocentric chromosomes have been found in some nematode, insect and plant species, in which the centromere function is distributed in several places along the length of the chromosome [6–8]. As in monocentric chromosomes, the identity of holocentromeres is generally defined by the epigenetic determinant centromeric histone H3 (CenH3). However, it was shown that the sequences and organization of holocentromeres vary greatly among species. In the nematode model *Caenorhabditis elegans*, centromere-specific sequences were not identified [9] but they coincide with binding sites for transcription factors [10]. In contrast, one satDNA has been confirmed as the underlying centromere sequence in the holocentric plant *Rhynchospora* [11]. Further diversity in holocentric architecture was revealed in four insects, which lack both the CenH3 and specific centromeric sequences [12]. A recent comparative study between holocentromeric plants and their closest monocentric relative, demonstrated that the transition to holocentricity affected 3D genome architecture [13]. Meta-polycentromeres are a newly discovered centromere transitional form that occupy 50% of the chromosome length, found in plants [14,15] and insects [16,17].

Root-knot nematodes of the genus *Meloidogyne* are major global pests, responsible for ~5% of agricultural damage worldwide [18]. They reproduce through obligate or facultative sexual and fully asexual mode. The most widespread and economically important species, *Meloidogyne incognita* (Minc), *M. javanica* (Mjav) and *M. arenaria* (Mare), are obligate mitotic parthenogenetic species that group within MIG species complex [19]. These species have the widest geographical distribution and range of hosts among *Meloidogyne* species [20]. Their polyploid genomes (3n to 4n), consist

subgenomes which resulting most likely from inter-species hybridization events [21,22]. They are very closely related at the mitochondrial level, suggesting recent hybridizations from closely related female lineages [22,23]. Recently, another obligate mitotic and triploid species, *M. enterolobii* (Ment), has gained attention due to its unique ability, to overcome plant resistance, leading to quarantine status [24]. However, in contrast to mitotic and polyploidy MIG and Ment species, *M. hapla* (Mhap) is diploid parthenogenetic species that undergoes meiosis and facultative sexual reproduction and also represent widespread species. Apart from Mhap, other meiotic parthenogenetic species tend to have a narrow host range. Obligatory mitotic parthenogenetic *Meloidogyne* species originate from facultatively parthenogenetic meiotic ancestors through the suppression of meiosis during oocyte maturation. Despite their asexual reproduction mode, which is typically associated with limited genome variation due to the absence of genetic exchange, *Meloidogyne* species, exhibit rapid adaptability to unfavorable conditions [24,25]. Although the mechanisms underlying this adaptability have yet to be elucidated, it has recently been shown that transposable elements contribute to genome plasticity and species diversification in these asexual species [26].

Recently, *Meloidogyne* species have been recognized as important model organisms for centromere study due to their unique centromere chromosome organization [27]. In contrast to previous studies conducted on the centromere of various species [28–30], demonstrated the exceptional divergence of CenH3 among species, our previous study showed that one specific CenH3 (αCenH3), whose association with tubulin has been experimentally confirmed, is highly preserved as nearly identical in closely related mitotic *Meloidogyne* species [27]. In addition, we found αCenH3-associated short arrays of tandem repeat sequences (TRs), containing a 19-bp conserved box despite highly diversified centromeric TR sequences. We also identified a unique pattern of uneven distribution along the chromosomes, appearing as clusters in *M. incognita.*

Given that genomic studies of non-monocentric centromeres, especially in animal species, are still very limited and prompted by the unique cluster organization of the centromere in *Meloidogyne*, this study aimed to perform a comparative analysis of the sequence organization and evolution of clustered centromere across different *Meloidogyne* species. To achieve this goal, six Meloidogyne species were selected. Four of these are mitotic and polyploid: *M. incognita* (Minc), M. *javanica* (Mjav), *M. arenaria* (Mare), and *M. enterolobii* (Ment). Among them, Minc, Mare, and Mjav are closely related forming the MIG species complex within Clade I, with Ment as an outgroup [19,31,32]. The selection of these species was intended to analyze organization of centromeric regions under conditions of exclusive mitosis and polyploidy, facilitating also investigation at the subgenome level. The other two species selected for

analyses of centromeric regions are meiotic and diploid: *M. hapla* (Mhap) and *M. chitwoodi* (Mchi). Despite both being meiotic and diploid, phylogenetic evidence indicates they are distantly related; Mhap belongs to Clade II, while Mchi is classified within Clade III. Moreover, phylogenetic markers reveal that Mhap is equidistant from Clade I (mitotic MIG and Ment) and Clade III (meiotic Mhi) [19,31,32]. This research was made possible by the recent availability of high-quality genome assemblies [23,24,33–36], including some at chromosome-scale resolution [23,36], which enable studies that were previously limited by highly fragmented assemblies [22,37,38]. Ultimately, we compare the organization of centromeric regions in four mitotic and polyploid species from Clade I with two meiotic species from Clades II and III.

Here, we present the first comprehensive analyzes of previously inaccessible clustered centromeric organization in asexual parthenogenetic *Meloidogyne* species exhibiting diverse reproductive modes and genome ploidy levels. We demonstrate that mitotic and polyploid *Meloidogyne* species possess a unique organization of centromeric regions, distinct from previously described, characterized by a higher-order organization of different 19-bp box-containing tandem repeats associated with non-autonomous and autonomous transposon elements. In contrast, the clustered centromere of meiotic and diploid species displays a simpler organization composed of monomeric arrays lacking transposable elements and showing a reduced potential for dispersion along the chromosomes.

Results

The relationships among six *Meloidogyne* species selected for centromere analysis are shown as a phylogenetic tree in Fig. 1A, based on various phylogenetic markers [19,31,32]. In our recent study, the Minc (3n), Mjav (4n), and Mare (4n) (MIG species) genomes have been assembled in 291, 364, and 377 contigs, respectively [33]. This represents a significant improvement compared to the previous reference assemblies, which consisted of around 12,000–31,000 scaffolds [22,37] (Additional file 1: Table S1). The new assembly sizes for tree MIG species are consistent with previous flow cytometry estimates [33]. Although chromosome-level assemblies of MIG species have previously been published [23], the known genomic variability observed among populations [39] prompted us to analyses organization of centromeric regions using high-quality genome assemblies [33]. These assemblies, in contrast to those provided by Dai et al. (2023) were generated from the same MIG populations used for the experimental centromere studies in Despot-Slade et al. (2021). To compare the organization trends in other populations, we also performed the same comparative analyses of centromeric regions on the

chromosome-level assemblies of MIG from Dai et al. (2023) and report these findings in a separate section.

The genome assembly generated for the mitotic Ment, is also consistent with flow cytometry and constitutes a useful platform for analyses of organization centromeric regions in mitotic/polyploid species [24]. In Clade II, a new, genome assembly comprising 16 chromosomes is available for the meiotic species Mhap [36], while for Clade III, the Mchi genome has 30 contigs almost reaching chromosome-level [34].

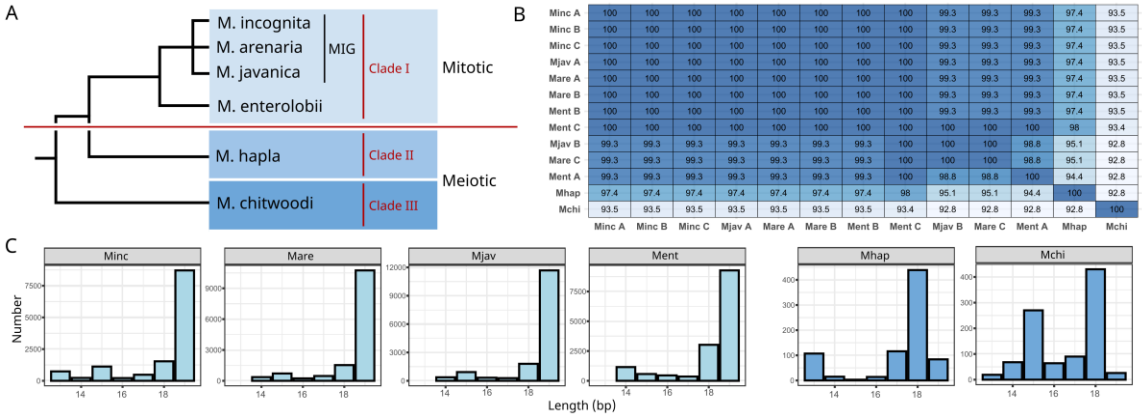

Fig. 1. CenH3 and DNA centromere determinants in six *Meloidogyne* species.
A) *Meloidogyne* phylogeny based on different phylogenetic markers [19,31,32] Clade I mitotic and polyploid species of *M. incognita* (Minc), *M. javanica* (Mjav), *M. arenaria* (Mare) (MIG species group) and the sister to this group, *M. enterolobi* (Ment); meiotic and diploid Clade II (*M. hapla;* Mhap) and Clade III (*M. chitwoodi;* Mchi) species. [24]. B) Identity matrix of all αCenH3 proteins (A, B and C represent an αCenH3 gene variant due to genome polyploidy) found in all six *Meloidogyne* species analysed based on their multiple sequence alignment (Additional file 2: Fig. S1). C) Conservation of the 19-bp box in the six *Meloidogyne* species studied: the number of 19-bp boxes in each species that have 0, 1, 2, 3, 4 or 5 mismatches compared to the canonical 19-bp box.

First, to confirm the presence of conserved centromeric H3 histone (CenH3) candidates in six selected *Meloidogyne* species, Minc αCenH3 sequence was used as a query for tBLASTn analysis against assembly for each species (Additional file 2: Fig. S1). As expected, the polyploid mitotic species show the presence of multiple proteins with an exceptionally high degree of intra- and interspecific identity, ranging from 98.8% to 100% while diploid meiotic Mhap and Mchi have one αCenH3 per genome (Fig. 1B). Besides

the expected very high similarity among αCenH3 sequences in mitotic species, Mhap also shows significant similarity (95–97%) compared to mitotic Clade I species. The most distant, but still highly similar, is Mchi αCenH3, with the similarity of 93% to both meiotic Mhap (Clade II) and mitotic species (Clade I). Our previous ChIP-Seq analysis revealed that the αCenH3-associated DNA in Minc consists predominantly of tandem repeats (TRs) arrays usually composed of divergent monomers. However, despite of the diversity of centromeric TRs, all these αCenH3-enriched repeats share a completely conserved 19-bp box which has been considered as a centromeric sequence determinant in Minc [27]. In the same work, the colocalization of 90–95% between αCenH3 regions and 19-bp box-containing TRs was also demonstrated in two other MIG species, Mare and Mjav.

The next step was the detection of putative centromere candidates in genome assemblies using the 19-bp box. We mapped the genome assemblies of all selected species using the 19-bp box as the query sequence, allowing up to six mismatches. The aim was to test the presence and level of conservation of the 19-bp box in these species. The mapping results showed significant conservation of the 19-bp box in all analyzed species (Fig. 1C). In Minc, Mare, and Mjav 86% of all 19-bp box hits were completely conserved and 13% had one mismatch. Similarly, in Ment, 54% of 19-bp box sequences were completely conserved while 24% has one mismatch. Although the 19-bp box in the meiotic species Mhap and Mchi is somewhat altered, a dominant fraction of about 40% is still either completely conserved or has only a single mutation.

To confirm the colocalization of CenH3 regions and the 19-bp box in meiotic Mhap, we employed the same approach previously used for Mare and Mjav in Despot-Slade et al. (2021). Specifically, we performed combined immunofluorescence (IF) using anti-αCenH3 antibodies alongside the primed in situ labelling (PRINS) assay targeting the 19-bp box on Mhap chromosome slides (Additional file 2: Fig. S2). The results demonstrated, consistent with findings in Minc, Mare, and Mjav, a high degree of overlap between αCenH3 and 19-bp box signals, indicating that αCenH3 is deposited on Mhap chromosomes and is associated with the 19-bp box (Additional file 2: Fig. S2). This supports the hypothesis that the 19-bp box is linked to the centromere in this distantly related meiotic/diploid Mhap species.

Unfortunately, experimental confirmation for mitotic/polyploid Ment and meiotic/diploid Mchi was not possible due to the unavailability of nematode material. However, the experimental confirmation of centromere association with 19-bp box in four species –three mitotic/polyploid (Minc, Mjav, and Mare) and one meiotic/diploid (Mhap)—together with the high conservation of both, αCenH3 histone and the 19-bp box, in Ment and Mchi, suggests that the 19-bp box could also be centromeric in these species.

This prompted us to proceed with characterizing αCenH3-associated centromeric sequences using 19-bp box-containing sequences in the genome assemblies of these six selected species.

Characterization of sequences in centromeric regions in mitotic MIG species group

To define the centromeric sequences in three closely related mitotic MIG species, Minc, Mare and Mjav, the 19-bp box was re-mapped allowing only one mismatch on the three new high-quality assemblies [33].

The results showed the presence of 10,631, 12,831 and 13,997 of these 19-bp box in Minc, Mare and Mjav, respectively (Fig. 2A). The further aim was to use the most effective method for detecting the vast majority of 19-bp box-containing tandem repeats (TRs) in these three genome assemblies. Therefore, we extracted all TRs that began with the 19-bp box and ended before the next 19-bp box. All extracted 19-bp box-containing TRs were then cross-referenced with the consensus sequences of the six previously described TRs (16-32, CL16g1, CL16g2, CL16g3, CL25, CL32) [27]. Of all extracted TR sequences, 78.8% corresponded to the six previously described TRs. The remaining 19-bp box-containing TRs, (accounting for 21.2%), were extracted and grouped according to sequence similarity into eight new TRs (G1-G8). To summarized, the results of these analyses revealed 14 different 19-bp-box-containing TRs in MIG genomes; six previously found with ChIP-seq and eight novel TRs (Fig. 2A and Additional file 1: Table S2). Validation of our approach to characterize the complete set of centromeric TRs in genome assemblies was performed by comparing the total number of detected 19-bp boxes with the total number of detected TRs. The results show approximately 95% congruence between detected 19-bp boxes and extracted centromeric TRs for all three species (Fig. 2A), demonstrating a high-fidelity association between the detected 19bp box and TRs. Consequently, there is only about 5% of 19-bp box-containing sequences that are not incorporated into TRs but are scattered throughout the genome. Given the variability of TR monomers in the genome, this method proved to be the most accurate for defining an almost complete set of 19-bp box-containing TR sequences in the genome.

Sequence analyses of the consensus sequences generated from all extracted TRs (using a majority rule with >80% nucleotide identity) revealed that all TR families are GC-rich (>50%) and monomeric units, with the exception of G2, are of about 40 bp to 90 bp long (Additional file 2: Fig. S3). Although the TRs share the conserved 19-bp box, their overall mutual sequence identity is relatively low (Fig. 2C). Only two pairs of the TRs exhibit >75% sequence identity (CL16g3 with 16_32 and CL25 with CL16g2). In contrary, the interspecific identity of monomeric units for particular TRs is high, ranging from 87.4% for G6 to 97.5% for G2 (Additional file 1: Table S2). The annotations of centromeric TRs in all three genomes

show that centromeric TRs account for 0.34% (0.67 Mb), 0.27% (0.8 Mb), and 0.24% (0.73 Mb) of the Minc, Mjav, and Mare assemblies, respectively. The numbers of annotated copies for each TR in the three genomes are shown in Fig. 2D. As expected, the six previously defined abundant centromeric TRs (CLx) obtained by ChIP-seq analysis in Minc are the most represented in all three genomes. However, eight new TRs containing 19-bp box with low to medium abundance were also detected (Gn) across all three genomes, representing approximately 20% of all centromeric TRs. The detection of these newly identified centromeric TRs aligns with our previous findings, which demonstrated through cytological analysis, combining IF (with anti CenH3) and FISH (with TRs identified by ChIP) that these TRs overlapped approximately 80% of CenH3 domains, leaving about 20% of CenH3 domains uncovered [27]. However, in the same study, immunofluorescence (with anti-CenH3) and PRINS (using the 19-bp box) demonstrated approximately 95% overlap between the CenH3 regions and the 19-bp box, suggesting the presence of additional centromeric TRs not captured by ChIP. The most likely candidates within these regions are the newly identified 19-bp box-containing TRs, G1–G8.

Regarding the amount of centromeric TRs in related genomes, analyzes show that the genomes share all detected centromeric TRs, but differ in genome abundance (Fig. 2D). The most abundant TR families such as CL16g2 and CL16g3 are highly abundant in Mare and Mjav, whereas the third most abundant TRs, CL25, exhibits higher abundance in Minc compared to Mare and Mjav. The low- and medium-abundance TRs are relatively consistent in all three species.

To gain better insight into the sequence relationships between the monomer units of each TRs, a PCA was performed. The PCA was performed using FactoMineR on the primary sequence feature vectors. The resulting PCA plots for the orthologous TRs showed extensive mixing of repeat units from all three species (Fig. 2E), which is consistent with interspecific sequence identity (Additional file 1: Table S2). Minor segregation of some species-specific variants was detected in G4 (Mare subcluster) and CL32 (Mjav subcluster) (Fig. 2E). The comprehensive identification of the centromeric TRs in the analyzed genomes provides an excellent starting point for subsequent analyses of the linear organization of centromeres in these closely related MIG genomes.

A

	19-bp box	14 TRs	% overlap
M. inc	10,631	10,101	**95.0**
M. are	12,831	12,201	**95.1**
M. jav	13,997	13,218	**94.4**

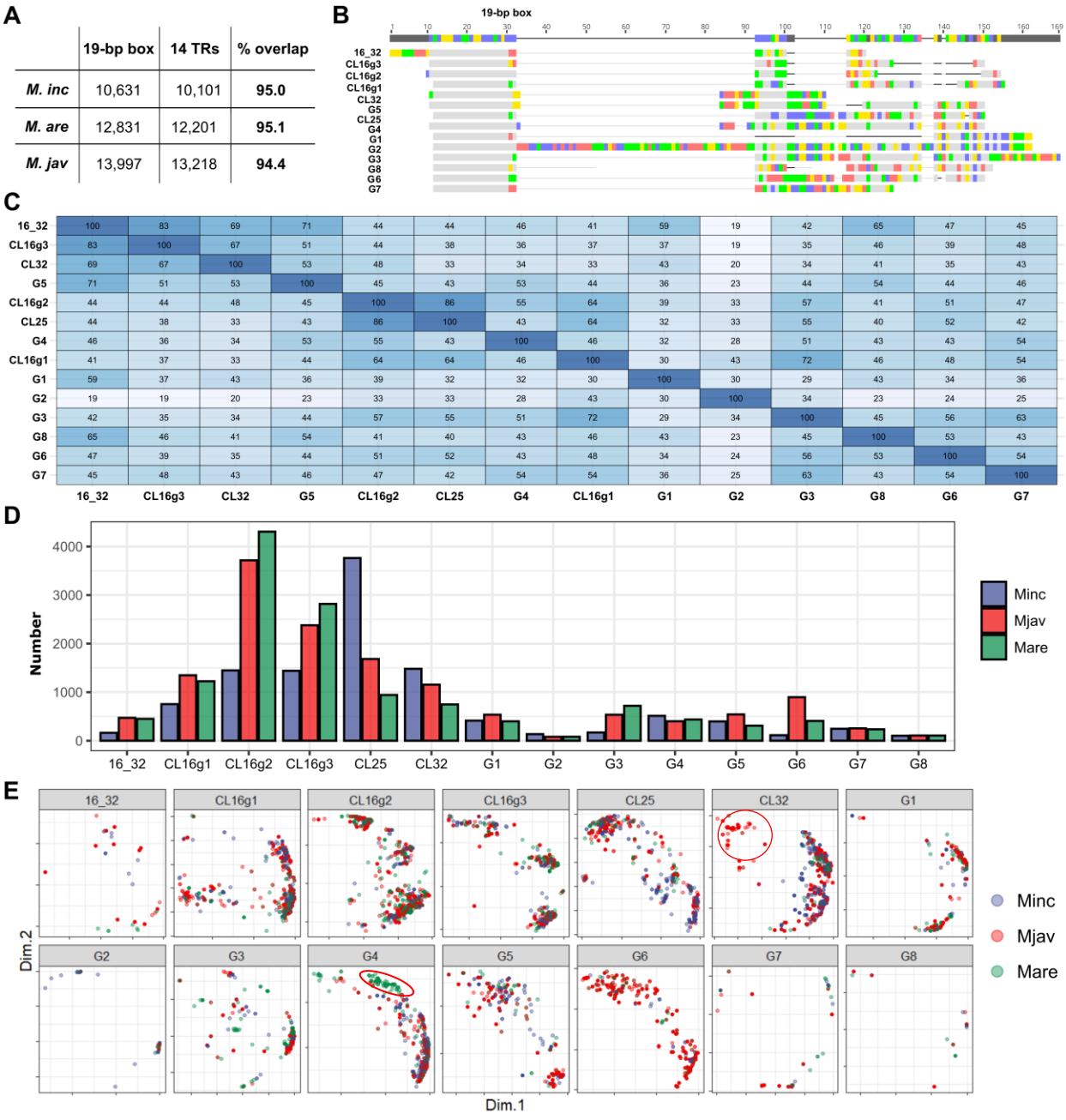

Fig. 2. Characterization of centromeric TRs in the three MIG species group (Minc , Mare and Mjav)

A) The number of 19-bp boxes and 14 TRs as well as percentage of their positional overlap in the three MIG assemblies. B) MAFFT alignment of 14 different TRs containing the 19-bp box. All sequences overlap on their conserved 19-bp box and the rest of sequence shows low sequence similarity. C) Pairwise sequence similarity matrices among consensus sequence of 14 centromeric TRs. Only 2 pairs of the TRs show a >75% sequence similarity represented with darker blue color. D) The number of 14 centromeric TRs in Minc, Mare and Mjav. The most abundant TR family is CL16g2 with 9471 total copies across the three genomes. E) PCA plots of the distance matrices generated from the alignments of TR monomers for each TR families from all three species. As proposed

by the alignments, a high degree of admixture between the three species is visible, presented as large clusters of points representing all species. The G4 and CL32 subclusters are circled.

Characterization of centromeric arrays in mitotic MIG species group

To understand the composition of centromeres at a higher resolution, we conducted a comprehensive search for centromeric arrays in all three MIG genomes. The first step was to map the centromeric arrays composed of 14 different centromeric TRs. Our previous ChIP data indicated that the organization of centromeric arrays is fundamentally complex [27]. Therefore, it was important to develop a strategy that ensures the annotation and extraction of the vast majority of centromeric arrays in the assemblies, considering the sequence variability of the TRs and the complexity of the array composition. For that reason, BLAST annotation was performed using consensus sequences of the 14 TRs monomer units (Fig. 2B) with parameters of sequence coverage >70% and a percentage identity >70% to ensure proper mapping centromeric arrays.

The annotation of 14 TRs in the three genomes revealed 2928 arrays in total, with 779, 985 and 1164 arrays found in the Minc, Mare and Mjav genomes, respectively. After a database of centromeric arrays was created, PCA analyzes of all arrays obtained from the three MIG species were performed (Fig. 3A). The distance and interconnectivity between dots correlate with the similarity between arrays, with closer dots indicating higher similarity between different arrays. Following the PCA analysis, the arrays were clustered into 16 classes (AR_N), contain >20 arrays using DBSCAN. The naming convention was based on DBSCAN clusters, with each cluster assigned a number for accurate tracking in the PCA plot. These numerical identifiers were used consistently in all subsequent analyses, allowing efficient monitoring of both intra- and inter-array relationships. These array clusters were arbitrarily divided into two main groups: Group1 and Group2. The aim of this analysis was to categorize arrays to facilitate the description of their structural features, and to identify their prominent common characteristics (Fig. 3B). Ungrouped clusters such as AR_67, AR_71 and AR_81 (Fig. 3A) was also considered and, according to their specific organization, included in the array class “Other” (Fig. 3B). The array clusters were further examined, and it was found that the monomeric CL16g2 arrays in Group 2 were scattered into several smaller clusters due to differences in array length. Therefore, these arrays were merged into a single cluster labelled AR_CL16g2. These 17 array clusters (AR_N plus AR_CL16g2) comprise many of the detected arrays (72.8%; 2132 out of 2928 arrays). The class defined as NA includes 764 arrays scattered in 60 clusters with a small number of arrays (<20), as well as arrays that were not clustered.

To validate the representation of individual TRs in the 17 most abundant array clusters, we annotated all 14 TR families to the 17 array clusters defined in all three MIG genomes and found significant deviation

in the representation of four TR families: CL32, G4, G5 and G6 (Additional file 2: Fig. S4). The representation of these TRs in the 17 array groups was less than 50%. Detailed analyses which combined manual inspection, genome annotation, and adjustment of parameters to ensure accurate detection and classification of TRs, revealed that underrepresented TRs, were scattered in multiple clusters of <20 arrays due to their organization. Specifically, TRs CL32, G4 and G6 were organized as homogeneous monomeric arrays with different lengths, whereas G5 was found at the ends of monomeric CL32 arrays. The difference in length was the main reason for splitting them into multiple low-abundance clusters. Therefore, we combined the related low-abundant clusters and included them in the analyses as AR_32, AR_G4 and AR_G6. Ultimately, the 20 array clusters identified through PCA analyses and manual curation exhibited substantial representation of individual TRs (Additional file 2: Fig. S5), confirming their credibility for further downstream analyses.

In addition, to assess whether array clusters contain homogenous group of arrays, we performed a genetic distance analysis for 16 non-monomeric array clusters (Additional file 2: Fig. S6). The heatmap visualization of the similarity among arrays within each cluster shows high sequence homogeneity in almost all array clusters, confirming the high accuracy of the mapping and clustering. Some clusters contain nearly identical sequences (e.g. AR_42, AR_47), while some array clusters such as AR_22 and AR_74 include array variants. However, their mutual sequence divergence is up to 0.3 distance score, corresponding to a maximum divergence of 30% between arrays.

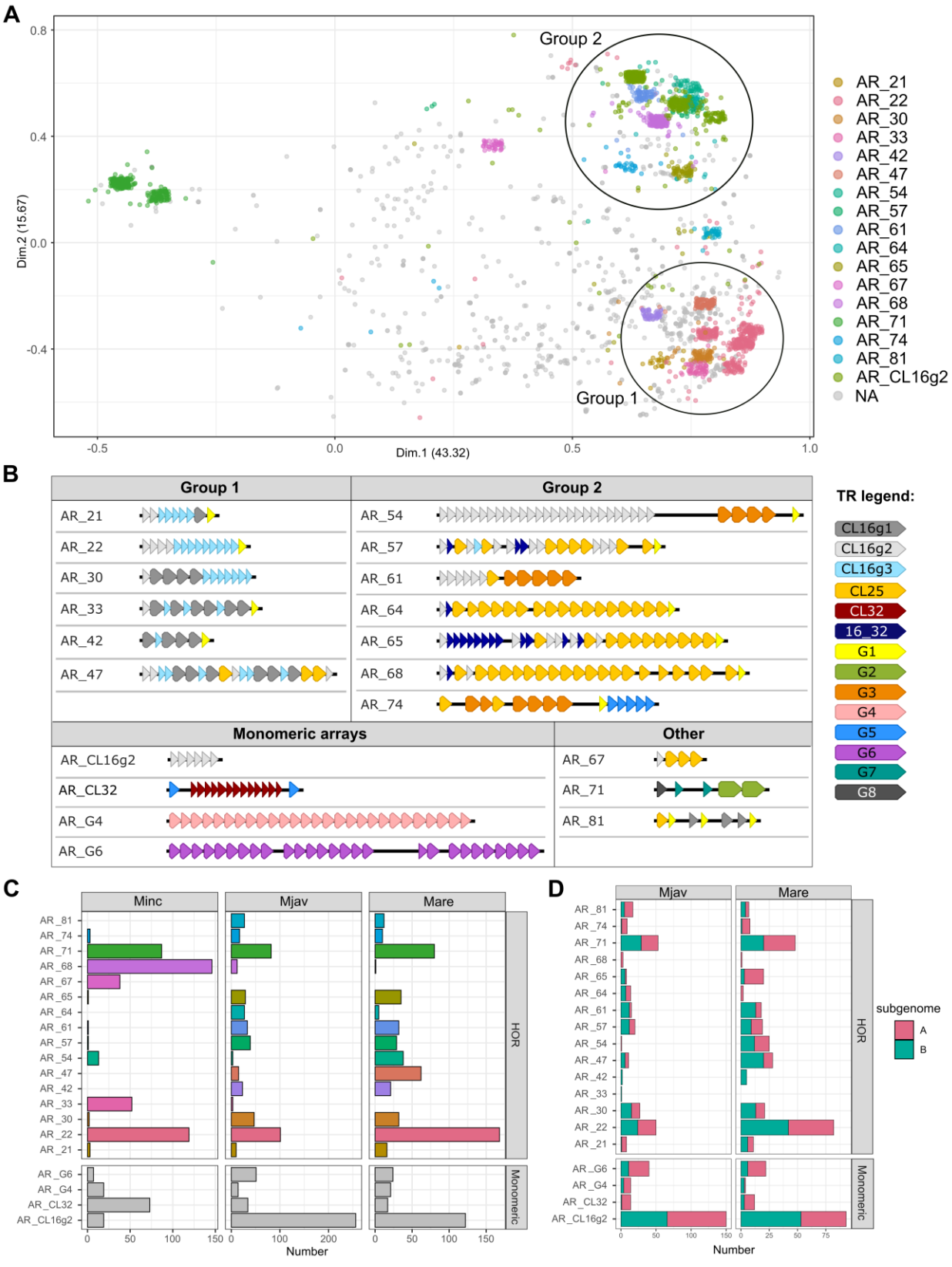
A
Group 2
Group 1
Dim.2 (15.67)
Dim.1 (43.32)
AR_21
AR_22
AR_30
AR_33
AR_42
AR_47
AR_54
AR_57
AR_61
AR_64
AR_65
AR_67
AR_68
AR_71
AR_74
AR_81
AR_CL16g2
NA
B
Group 1
Group 2
Monomeric arrays
Other
TR legend:
CL16g1
CL16g2
CL16g3
CL25
CL32
16_32
G1
G2
G3
G4
G5
G6
G7
G8
C
Minc
Mjav
Mare
HOR
Monomeric
Number
D
subgenome
A
B

Fig. 3. Characterization and definition of centromeric array types in three MIG species
A) PCA analysis of the distance matrix of all centromeric TR containing arrays across the three MIG species followed by DBSCAN clustering of the first two PCA dimensions. The largest 17 clusters are colored, while the low abundance clusters and unclassified arrays are presented as NA. From the graph it is possible to create two distinct group of array sequences clusters, that are circled and marked as Group 1 and Group 2. **B)** Schematic representations of 20 most abundant classified array types in the three MIG genomes. Arrays are sorted in two major groups (Group 1 and Group 2) based on PCA clustering results (Fig. 3A) representing higher-order-repeat (HOR) organization. In addition, homogenous arrays (Monomeric arrays) and array types that do not belong to any of the previous (Other) are presented. **C)** Number of the main array types in three MIG species: HOR arrays from Group 1, Group 2 including Other and Monomeric arrays. Number of array types in Minc genome when compared to the Mare and Mjav genomes show a significant difference (Chi-square = 1118.5, df = 38, p-value < 2.2e-16). **D)** Subgenome distribution of 20 array types between subgenomes AABB in Mjav and Mare species.

The represented array type was extracted, retaining the name of the corresponding cluster from PCA, along with monomeric arrays identified through additional analyses (Fig. 3B). The two dominant groups (Group 1 and Group 2) contain arrays with complex high-order-repeat (HOR) structures. Group 1 comprises six HOR array types, mainly consisting of three TRs (CL16g1, CL16g2, and CL16g3) in various combinations, with a relatively short array length (up to 1500 bp) (Additional file 2: Fig. S7A). Group 2 includes seven HOR array types, characterized by the presence of CL25 TR in almost all array types in combination with other TRs such as 16_32, CL16g2, and G5. Unlike Group 1, Group 2 exhibits much greater variation in the length of array types, particularly evident in AR_65 and AR_74 (Additional file 2: Fig. S7A). More detailed inspection of array variants in Group 2 frequently shows variation in the length of array segments composed of CL25. The third group of array types, referred to as monomeric arrays, consists mainly of one type of TR. These array types have been manually curated, and their names correspond to the TRs they comprise. Their length varies much more than in the HOR array types (Additional file 2: Fig. S7A). The fourth group, classified as 'Other' consists of unrelated and relatively short HOR array types that do not belong to any other group but are well represented in the genomes. Detailed information on array statistics in all three MIG is presented in Additional file 1: Table S3.

A comparative analysis of array lengths across all three genomes revealed that arrays longer than 2500 bp are rare (Additional file 2: Fig. S7B). An analysis of the total array lengths shows that the overall profiles are more similar in Mare and Mjav than in Minc. This is particularly evident at the 2000 bp length in the Minc genome, which has more frequent arrays of this length compared to Mjav and Mare. The number and abundance of array types in each genome (Fig. 3C and Additional file 2: Fig. S8) show a significant difference between the Minc and the Mare and Mjav genomes (Chi-square = 1118.5, df = 38, p-value < 2.2e-16). Specifically, the Minc genome has five HOR array types (AR_71, AR_68, AR_67, AR_33 and AR_22) and one monomeric array type (AR_Cl32), which are dominantly represented in both number and abundance. In addition, the HOR type AR_68 is dominant in Minc compared to Mare and

Mjav, while AR_67 and AR_33 are Minc-specific array types. Mare and Mjav contain a greater number of different HOR array types, most of which have a moderate genome abundance. Some of these are found exclusively in the Mare and Mjav genomes. In contrast, the abundance of monomeric arrays does not follow the distribution pattern of HOR array types in these two genomes. Monomeric arrays are generally more abundant in the Mjav genome compared to Mare (Additional file 2: Fig. S8).

Given that MIG species are polyploid and likely contain subgenomes resulting from inter-species hybridization events [40], we investigated whether subgenome-specific subsets of array types are retained in genomes. We selected Mare and Mjav for comparison, as both have tetraploid genomes with equal representation of A and B subgenomes (AABB). In our previous work [33], we unambiguously assigned 64 contigs to the A subgenome and 56 to the B subgenome in Mare, and 58 contigs to the A subgenome and 62 to the B subgenome in Mjav, accounting for almost half of the total contigs in the assemblies (Additional file 1: Table S4). The pattern of array type distribution across subgenomes is qualitatively similar in both species, with no apparent restriction of particular array types to specific subgenomes (Fig. 3D). Moreover, the most prominent array types are distributed in similar proportions across the subgenomes.

Linear organization of centromeric regions in mitotic MIG species

Previous experimental analyzes have shown that the *Meloidogyne* centromeres exhibit cluster organization, which implies the accumulation of centromeric sequences in the form of clusters of different sizes along the chromosomes [27]. In support, the mapping of centromeric arrays in the genome assembly in this work revealed their grouping and prompted the investigation of their internal structure. The stretch of multiple centromere arrays appearing in contiguous blocks which are separated by no more than 10 kb was defined as holo-unit (see Fig. 7). Therefore, the first step was to determine the sequences lying between the centromeric arrays in holo-units which we designated as intermediate sequences (ISs). We identified, and extracted 1877 ISs located between arrays within holo-units across all three genomes. These ISs were subjected to PCA analysis, and based on their relationship, they were divided into 11 groups containing more than 10 ISs each. These 11 groups (IS_N) were subsequently classified as intermediate sequence types (Fig. 4A). Analysis of the genetic distance between sequences in each IS cluster showed a relatively high degree of sequence identity (Additional file 2: Fig. S9) with the value scores lower than 0.3, corresponding to ~30% difference between the sequences of the groups. Only parts of three IS groups (IS_5, IS_7 and IS_10) showed greater divergence indicating the presence of distinct variants.

Similar to the centromeric array types, the IS abundance patterns showed the difference between MIG species (Additional file 2: Fig. S10). Minc has four represented IS classes, two species-specific (IS_1 and IS_3) and two shared with Mare and Mjav (IS_2 and IS_4). Mare and Mjav are also dominated by IS_2 and IS_4 and IS_8 are moderately abundant, while the other IS classes are only weakly represented in both species. To investigate whether IS types that occur in more than one species exhibit species-specific subgrouping, we performed additional PCA analyses of the three dominant IS types (IS_2, IS_4 and IS_8) (Fig. 5A). The results show that the subgrouping of these ISs is not species-specific, but that the genomes share different variants of IS types.

A comparative analysis of the length distribution of extracted ISs for all three genomes is presented in Fig. 4B and Additional file 1: Table S5. Minc displays a profile with prominent peaks in the length ranges of approximately 1500 bp and 3500 bp, correspond to the lengths of the most prominent ISs (IS_1, IS_2, IS_3 and IS_4) (Additional file 2: Fig. S11). The profiles of Mjav and Mare follow a pattern characterized by accumulation of ISs in several regions of different lengths, with prominent peaks around 1500 bp and 3000 bp. Detailed examination of the IS alignments showed that insertions larger than one kilobase, contribute to length variability and thus lead to IS sequence heterogeneity for some IS classes, as shown in Additional file 2: Fig. S9. However, variants with long insertions constitute the minor fraction of the total IS sequences. If the insertions are excluded and only the shared part of the sequence is compared, a high IS sequence similarity is found in each group, ranging from 79% to 99% sequence identity (Additional file 1: Table S6).

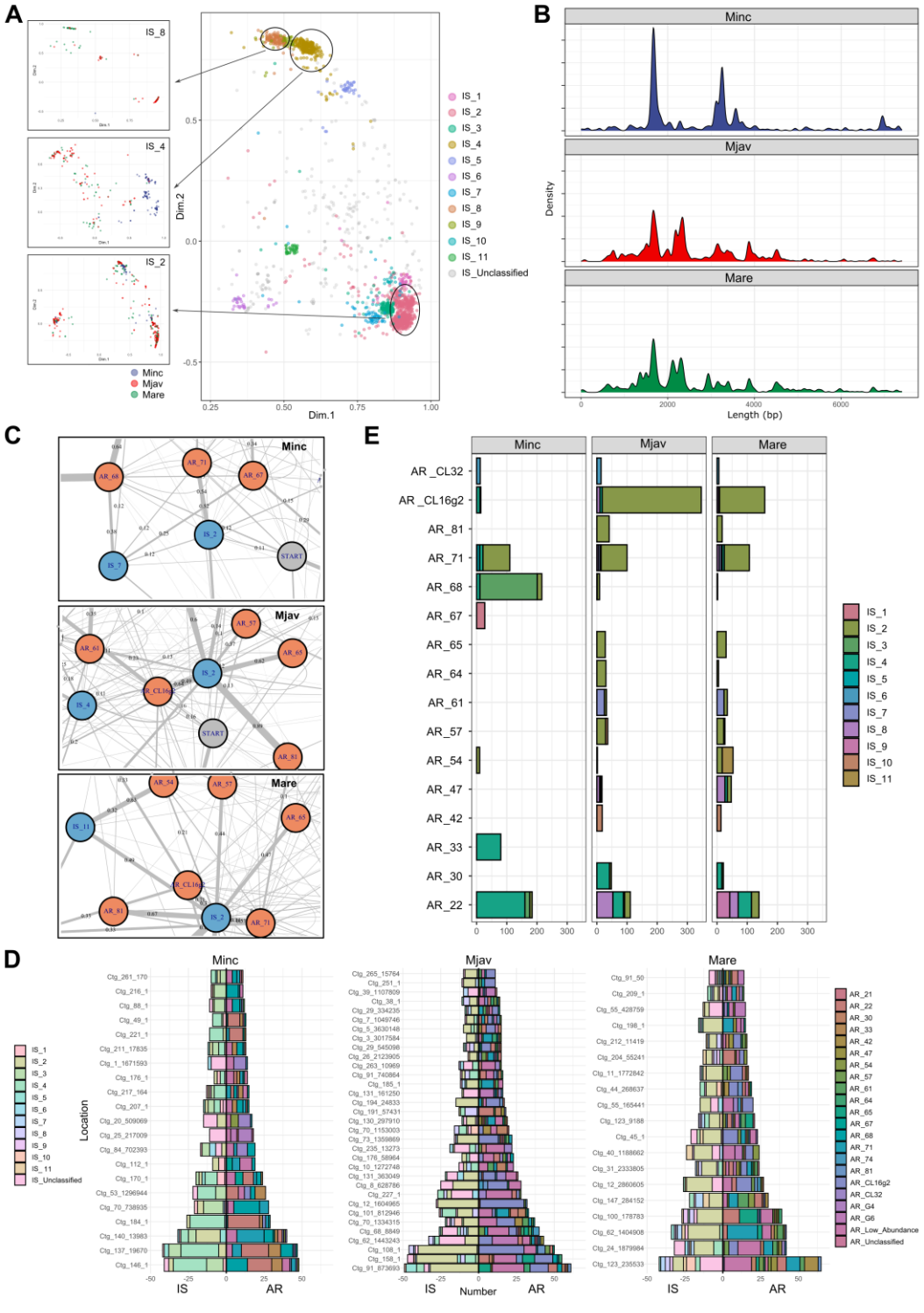
A
IS_8
IS_4
IS_2
Minc
Mjav
Mare
Dim.1
Dim.2
IS_1
IS_2
IS_3
IS_4
IS_5
IS_6
IS_7
IS_8
IS_9
IS_10
IS_11
IS_Unclassified
B
Minc
Mjav
Mare
Density
Length (bp)
C
Minc
Mjav
Mare
E
Minc
Mjav
Mare
AR_CL32
AR_CL16g2
AR_81
AR_71
AR_68
AR_67
AR_65
AR_64
AR_61
AR_57
AR_54
AR_47
AR_42
AR_33
AR_30
AR_22
D
Minc
Mjav
Mare
Location
IS
AR
Number
AR_21
AR_22
AR_30
AR_33
AR_42
AR_47
AR_54
AR_57
AR_61
AR_64
AR_65
AR_67
AR_68
AR_71
AR_74
AR_81
AR_CL16g2
AR_CL32
AR_G4
AR_G6
AR_Low_Abundance
AR_Unclassified

Fig. 4. Intermediate sequence (IS) definition and organization of arrays and IS regions
A) PCA analysis of all intermediate sequences (ISs). The most prominent IS_2 (pink), IS_4 (yellow) and IS_8 (orange) clusters were further analyzed by PCA and colored by species (corresponding PCA boxes are denoted by arrows). The ISs that were not classified or grouped in clusters of >10 sequences are represented as 'Unclassified' B) The length distribution of ISs in Minc, Mare and Mjav. C) Extractions of the most frequent Markov chain model estimates for transition probabilities considering the combination of different array types (orange) – IS classes (blue). The visualization only shows transition probabilities > 0.05. D) Number of array type and IS class in large holo-units (N>20) across the three MIG species. Colocalization of array types (AR) and intermediate sequences (IS) in Minc, Mjav and Mare presents as large blocks E) Summarized graph of most common AR type – IS class colocalizations across the three MIG species using Markov chain model.

The next step to gain insight into the linear organization was to disclosed the combinations of array types (ARs) with the intermediate sequence classes (ISs). For the first time in centromere research, we applied Markov chain analysis to systematically examine the most frequently co-occurring AR types and IS classes within the linear structure of centromeres. Specifically, we developed transition matrices to estimate the probability that a specific AR type is associated with a particular IS, thereby providing a probabilistic framework to identify the predominant arrangements among AR types and IS classes. This methodology enabled us to go beyond simple co-occurrence counts and uncover the dominant patterns of AR–IS associations that define the higher-order linear organization of centromeres (Fig. 4C). These complex data sets also provide an overview of the centromeric organization scattered across approximately 300 contigs of these still fragmented MIG assemblies. The summary analyzes of Markov chains for AR types and ISs organization within each species were visualized accordingly, with the dominant holo-units, those containing more than 20 ARs and IS in continuous arrangement, are shown at Fig. 4D. This genome-wide assessment showed that the Minc and Mare genomes each have approximately 20 contigs featuring long holo-units (>20 ARs and ISs in continuity), whereas the Mjav genome contains as many as 33 such contigs. These analyses also showed a clearly different IS profile between Minc and Mare/Mjav (Fig. 4D, left side). Specifically, IS_4 and IS_3 are predominant in Minc, while IS_2 is predominant in Mare/Mjav genomes. Examples of the most prominent long-range organization of holo-units obtained by these analyses are presented in Additional file 2: Fig. S12. The simple holo-units exhibit an organization in which one type ARs (HOR or monomeric) alternates with specific IS type. These holo-units usually consist of about 10 elements (ARs + ISs) and are about 30 kb in length. The other group of holo-units, called composite holo-unit, is more complex, consisting of different HOR array types in combination with various IS classes. In some cases, these composed holo-units consist of more than 30 elements (ARs + ISs) and can be up to 100 kb long. Detailed colocalization profiling of specific AR types and IS classes are presented in Fig. 4E and Additional file 2: Fig. S13. The

most prominent AR type and IS class combination in Minc are AR_68/IS_3 and AR_22/IS_4. Mare and Majv showed a higher correlation in the pattern of AR/IS combinations (Pearson coefficient 0.29) compared to Minc (Pearson coefficient -0.23 for Minc-Mare, -0.27 for Minc-Mjav), characterized by the frequent AR_CL16g2/IS_2 combination (Additional file 2: Fig. S13). In addition to the most common AR/IS combinations present in genome, an analysis also showed tendency that specific IS class combine with specific AR type; for example, IS_3 combined frequently with AR_22 and AR_33 in Minc, and IS_ 2 combined frequently with AR_CL16g2 and AR_71 in both, Mare and Mjav (Fig. 4E). Conversely, a certain AR type can be combined with several ISs, and this is most obvious for AR_22, which in Minc combines predominantly with IS_4, while in Mare/Mjav is associated with four different ISs.

Comparative analyses of linear organization of centromeric regions in mitotic species including distantly related *M. enterolobii*

The mitotic parthenogenetic *M. enterolobii* (Ment), although belonging to the same clade as MIG species (Clade I), is an outgroup to the MIG species according to phylogenetic analyses (Fig. 1A). The detection of conserved αCenH3 and 19-bp box sequences, along with the recently available high-quality genome assembly for Ment [24] prompted us to investigate the organization of the centromeric regions in this species. As preliminary analyses showed that the 19-bp-containing indicated TRs in Ment are composed of similar but more divergent monomers compared to MIG, we first conducted a detailed analysis of Ment and then compared it with analyses of the centromeric regions in MIG. First, the Ment genome was mapped with all centromeric TRs found in MIG species using relaxed parameters with sequence coverage > 70% and percent identity > 70%. The validation of our approach to characterize the 19-bp box containing TRs in Ment genome assembly was checked by comparing the total number of detected TRs (11974), using MIG centromeric TRs as query, with the total number of detected 19-bp boxes (14380) in Ment. The results show about 81% agreement between 19-bp boxes and detected centromeric TRs for Ment. This coverage was considered as a representative set of the 19-bp box containing TRs from Ment, serving as a basis for comparative genomic analyzes of arrays and ISs organization as well as holo-units in relation to MIG. Analysis of all 19-bp box containing TRs, showed that Ment has an average sequence identity of 86%, compared to TRs of MIG species (Fig. 5A), consistent with phylogenetic distance of Ment from MIG species. Regarding the presence of Ment-specific variants of TRs in comparison to MIG, the analyzes show that all TRs detected in Ment are also present in MIG species (Fig. 5B), albeit with lower sequence identities. However, the most prominent difference is observed in TRs abundance. For example, ECL32 (CL32-like) and EG5 (G5-like) TRs, which

are more abundant in Ment than in MIG, while ECL16g2 (CL16g2-like) and ECL25 (CL25-like) TRs show the opposite pattern with lower abundance in Ment compared to MIG (Fig. 5B and Additional file 1: Table S7).

In addition, using the same approach as in MIG species, we performed annotation of Ment specific TRs and revealed 1129 arrays in total. PCA analysis of all arrays showed that, in contrast to the MIG genomes, the array clusters in Ment are much more divergent, with a mean centroid divergence along PCA (accounting for 44% of the variance) of 0.11 for the MIG group and 0.17 for Ment (Fig. 5C and Additional file 2: Fig. S14). Since the same approach was used to generate array clusters in the MIG species, the lack of homogeneity in Ment is further supported by the absence of distinct clusters in the PCA plot. The observed divergence within the arrays' groups in Ment is supported by the analysis of arrays' lengths (Fig. 5D). While distributions of array lengths pattern are the most ordered in Minc, with few highly-frequent lengths, the arrays in Mare and Mjav and especially in Ment show a less organized array structure (Fig. 5D). Despite divergence among arrays within a group, mainly due to variation in array length, the 12 representative array types could be distinguished in Ment (Additional file 2: Fig. S15). Similar to the MIG species, the Ment arrays also show an HOR organization characterized by two or more TRs as building units, but monomeric arrays with only one type of TR (e.g. AR2) were also found as in MIG species (Additional file 2: Fig. S15). In addition, we defined ISs from Ment genome and compared with ISs of MIG species. In contrast to relatively divergent array clusters due to oscillations in the array lengths, ISs proved to be similar compared to MIG. Indeed, PCA analysis shows that ISs group in two main clusters similar to MIG (Fig. 5E), and that their length follows the pattern of Mare and Mjav species (Fig. 5F).

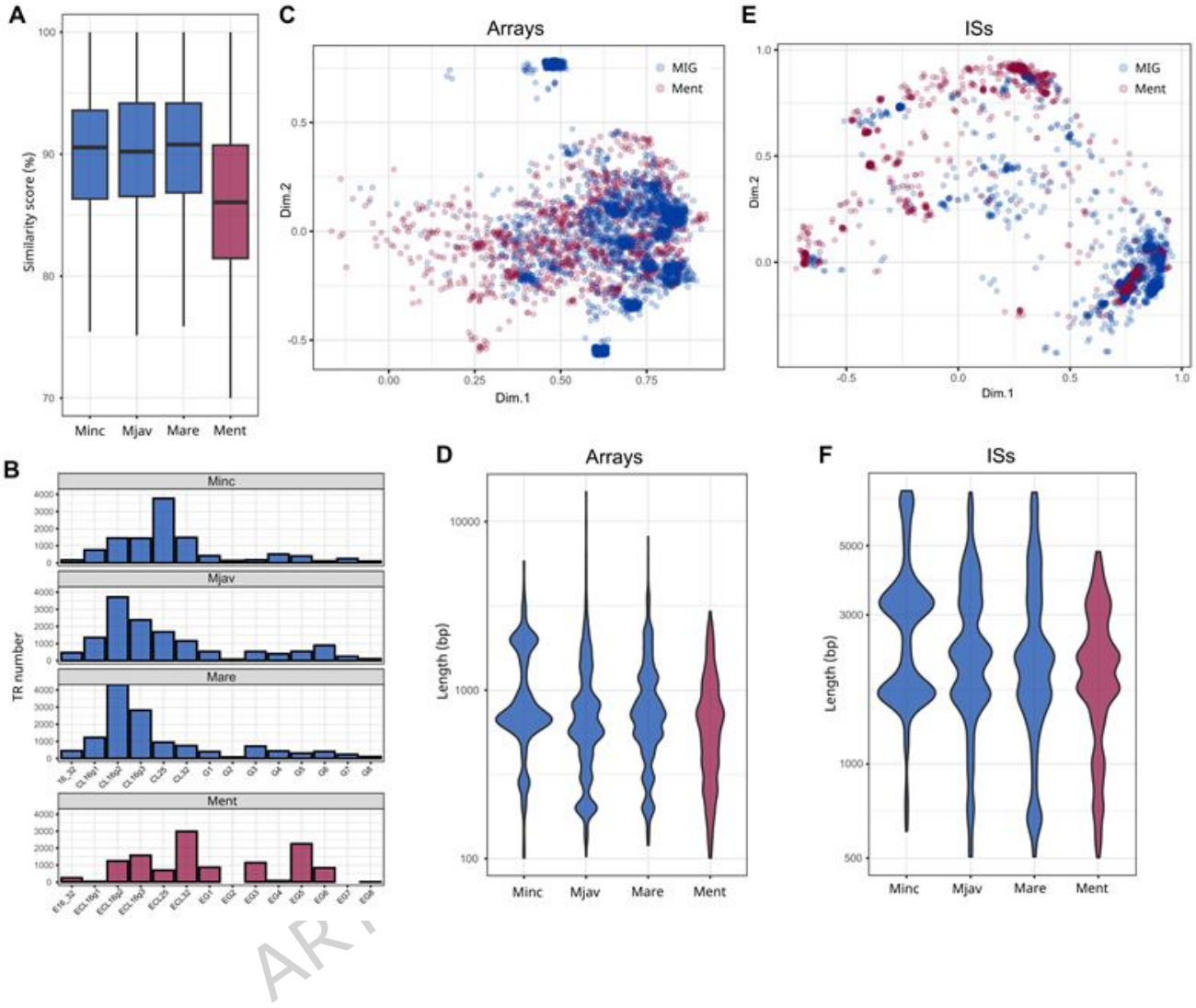

Fig. 5. Analysis of centromeric regions in mitotic Clade I species *M. enterolobii*

A) Similarity scores for all TR sequences found in Ment and the three MIG species (Minc, Mjav and Mare). MIG don't have statistically significant differences (Welch two sample t-test, Minc – Mare p.value = 0.2274, Minc – Mjav = 0.203, Mare - Mjav p.value = 0.9574), while Ment shows statistically significant difference (Welch two sample t-test *p* value 0.011, FDR adjusted *p* value 0.0454) B) Comparison of the number of different TRs in Ment and three MIG species. Ment has most of the TRs found in the MIG group species, albeit with altered sequence identities, as shown in panel A. The TRs variants of Ment are labelled with a prefix "E". C) PCA analysis of all arrays. The red dots represent the arrays of Ment and the blue dots represent the arrays of three MIG species. Highly conserved MIG array groups are visible, while the less length-restricted Ment arrays are scattered. D) Array length profiles of MIG and Ment. Ment arrays have less pronounced peaks. E) The PCA of all ISs. Although there are Ment-specific IS sequences, most IS groups overlap with MIG ISs. F) Length profiles of all ISs. While the arrays lengths are less conserved in Ment than in the MIG group species, the ISs show a more similarity in their length profile across all mitotic species.

Helitrons* are associated with centromeric regions in mitotic *Meloidogyne

The above analyses confirmed that the four investigated mitotic/polyploid species, including the distantly related Ment, showed that the sequences located between the centromeric arrays, the intermediate sequences (ISs), exhibit a high degree of length conservation and sequence similarity. The most probable candidates are dispersed repetitive sequence. Given that *Meloidogyne* genomes contain transposable elements (TEs) as the dominant fraction of dispersed repetitive sequences [26], the first step was to search the repeat annotation database generated with EarlGrey, a fully automated TE annotation pipeline [41], using all extracted ISs from the four mitotic/polyploid species. The results show that a significant proportion of IS annotations mainly overlap with the class of DNA transposons, RC/Helitron elements, in closely related MIGs and distantly related Ment (Fig. 6A). To quantify this, we performed a Chi-square Goodness-of-Fit test for each species, which confirmed that the distribution of TE families is significantly non-uniform ($p < 0.0001$ in all cases). RC/Helitron represents the primary driver of this variation, with the highest statistical enrichment observed in Mare (Chi-sq = 4498.9, p approx 0) and Mjav (Chi-sq = 2589.9, p approx 0) while being lower but still significant in Minc (Chi-sq = 631.7, $p = 3.33 \times 10^{-130}$) and Ment (Chi-sq = 467.6, $p = 7.36 \times 10^{-97}$). In addition to the prevalence of RC/Helitron elements in all species, only a small subset of ISs correlate with SINE/5S and DNA/Zisupton in Mare and Ment, respectively.

Helitrons are transposons that exist in two forms: autonomous and non-autonomous elements [42]. Autonomous Helitrons contain open reading frames (ORFs) that encode the Rep/Helicase (RepHel) transposase, whereas non-autonomous Helitrons lack this capability. Both types exhibit unique structural features, including 5′-TC and 3′-CTRR termini, with a tendency to favor 3′-CTAG sequences. Additionally, Helitrons often harbor a short palindromic sequence of 16 to 20 nucleotides near the 3′ end, capable of forming a hairpin structure. Heliano is a recently developed tool for detecting autonomous and non-autonomous Helitrons in complex genomes [43].

Since it has been shown that IS sequences do not correspond to RepHel, our objective has been to assess whether the structural features of RC/Helitron non-autonomous elements are retained within ISs. To investigate this, we first searched for palindromic sequences with a stem of at least 5 base pairs and a loop of 4 nucleotides, as these have previously been identified as critical features of highly divergent RC/Helitrons [43]. The analyses of total 2722 ISs from all four mitotic species show a presence of palindromic sequences in ISs, specifically 1486 with more than 100 hits across the ISs from all four studied species (Additional file 2: Fig. S16). To analyze their mutual similarities, we performed PCA

analyses and disclosed a wide range of sequence similarity among extracted palindromes (Fig. 6B). For further analyses we selected the 13 most prominent palindromes with at least 1500 hits; (Additional file 2: Fig. S16) which also represent mutual sequence identity (Fig. 6B left). These 13 of most abundant palindromic motifs can be divided into four subgroups, each subgroup consisting of palindromes of high sequence identity but with variation in length (Fig. 6B right). To confirm that these palindromes are indeed key determinants of RC/Helitron we analyzed the distribution of these palindromes along an IS sequence (Fig. 6C). The results of mapping 13 palindrome sequences on the entire set of ISs showed that six palindromes which belong to Groups 1 and 4 are almost exclusively found at the 3' ends of ISs (red in Fig. 6C). Considering the position of these six palindromes on ISs, their high degree of sequence conservation and the potential to form a hairpin structure, they are most likely candidates for the 3'-terminal hairpin structures of non-autonomous Helitron elements. The mapping of other seven palindromes showed their localization in the first part of ISs (blue in Fig. 6C). Detailed inspection showed that these palindromes frequently form tandemly repeated minisatellite-like loci (Additional file 2: Fig. S17). These palindromes/microsatellites contain the TC dinucleotide and serve as another determinant of Helitron, as they typically begin with a 5' T (C/T) replication start. Mapping of the replication-terminal Helitron-specific CTAG motif on the entire ISs also revealed the highest density in the 3' regions of the ISs (Additional file 2: Fig. S18). These three observed structural features (Fig. 6C) suggest that ISs in mitotic species correspond to non-autonomous Helitron element, specifically HLE1 elements [43] embedded in different centromeric arrays.

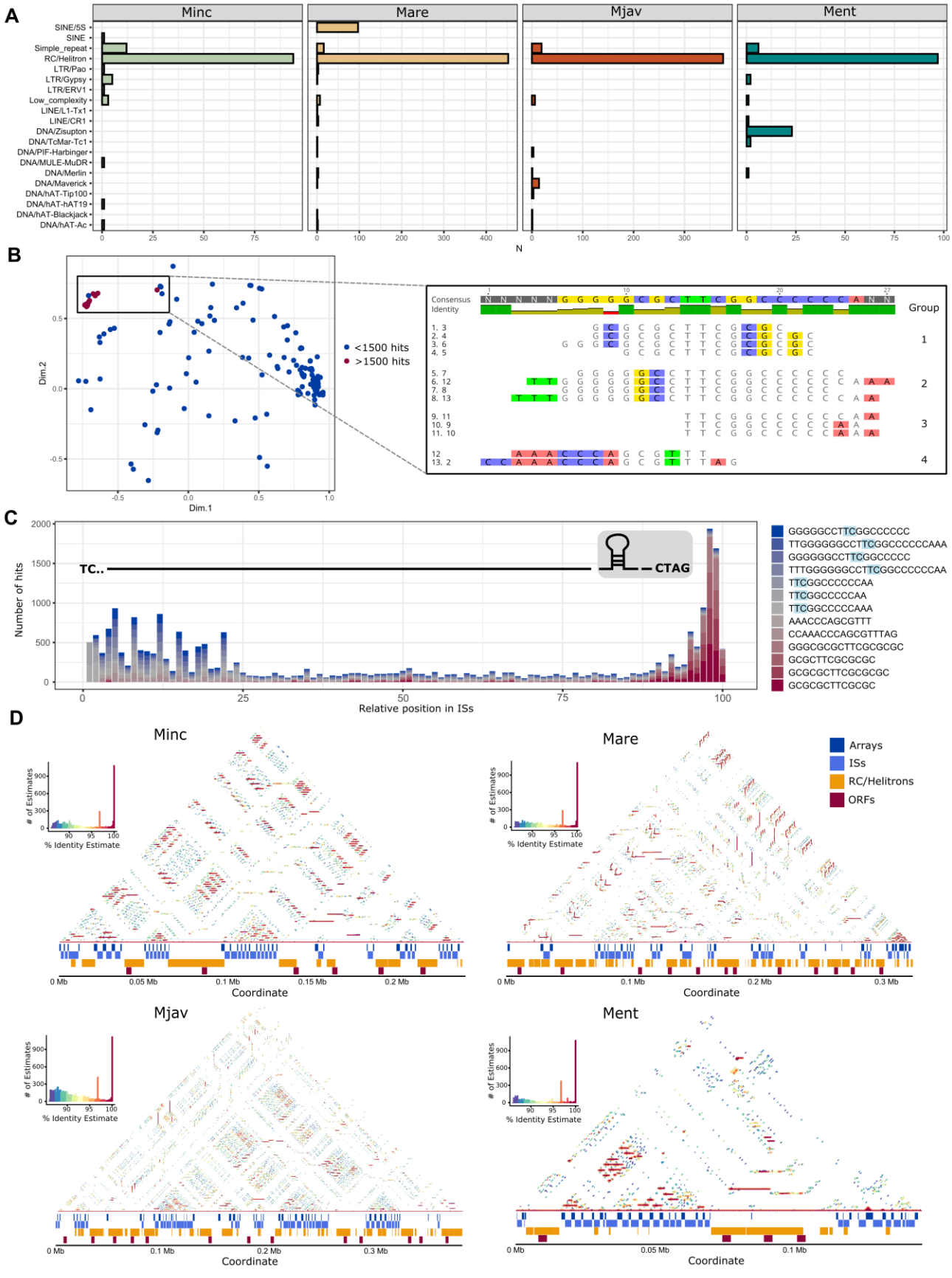
A
Minc
Mare
Mjav
Ment
SINE/5S
SINE
Simple_repeat
RC/Helitron
LTR/Pao
LTR/Gypsy
LTR/ERV1
Low_complexity
LINE/L1-Tx1
LINE/CR1
DNA/Zisupton
DNA/TcMar-Tc1
DNA/PIF-Harbinger
DNA/MULE-MuDR
DNA/Merlin
DNA/Maverick
DNA/hAT-Tip100
DNA/hAT-hAT19
DNA/hAT-Blackjack
DNA/hAT-Ac
N
B
Dim.1
Dim.2
<1500 hits
>1500 hits
Consensus
Identity
Group
C
Number of hits
Relative position in ISs
TC..
CTAG
GGGGGCCTTCGGCCCCCC
TTGGGGGGCCTTCGGCCCCCCAAA
GGGGGGCCTTCGGCCCCC
TTTGGGGGGCCTTCGGCCCCCCAA
TTCGGCCCCCCAA
TTCGGCCCCCAA
TTCGGCCCCCAAA
AAACCCAGCGTTT
CCAAACCCAGCGTTTAG
GGGCGCGCTTCGCGCGC
GCGCTTCGCGCGC
GCGCGCTTCGCGCGC
GCGCGCTTCGCGC
D
Minc
Mare
Mjav
Ment
# of Estimates
% Identity Estimate
Coordinate
Arrays
ISs
RC/Helitrons
ORFs

Fig. 6. Analyses of Helitron elements in centromeric regions of mitotic/polyploid species
A) Number of genome wide overlaps of EarlGrey TE annotations with arrays/ISs sequences in four mitotic species. The most abundant class is the RC/Helitron in all four species. B) PCA analysis of palindromic sequences mapped within ISs from all four mitotic species. The zoomed-in part represents the most abundant (N>1500) palindromic sequences of all species together with their alignment on the right. Alignment of the 13 most abundant palindromic sequences which are grouped based on sequence identity. C) Distribution of the 13 selected palindromes within ISs across the four species. The four palindromes in red are most often found on the 5' end of IS sequences. The TC containing palindromes are denoted in light blue on the legend, and they form most of the 3' end of the sequences. D) ModDotPlot of selected long holo-clusters of the three closely related mitotic MIG species and distant mitotic Ment. Non-autonomous Helitrons are present within ISs. Autonomous Helitron elements characterized by HLE1-ORFs coding region detected using Heliano are embedded in the regions between holo-units.

Further, we aimed to investigate whether centromeric regions possess an autonomous version alongside to the identified non-autonomous Helitron elements within ISs, and to understand their relationship to the centromere. To achieve this goal, given that autonomous Helitron can be over 10 kb in length, we first defined larger centromeric regions. Regarding the terminology of linear organization, ARs in contiguous blocks separated by ISs no more than 10 kb in length were defined previously as holo-unit while groups of holo-units separated by regions less than 50 kb in size represent the holo-cluster. The total number of holo-clusters determined according to these parameters ranged from 144 in Minc to 210 in Mare. As anticipated, the overall number of holo-clusters is higher in Mare and Mjav consistent with their larger genome sizes (Additional file 2: Fig. S19A). For further analysis of Helitron, we selected only long holo-clusters, defined as those with total length greater than 50 kb. Consistent with genome size, the highest number of long holo-clusters (about 40) were also observed in Mjav and Mare while Minc contain 34 long holo-clusters (Additional file 2: Fig. S19B). The lowest number of long holo-clusters observed in Ment (19) might be due to the genome's property of generating shorter holo-clusters. Further, we conducted a Helitron search on these long holo-cluster database obtained from these four species using Heliano [43]. The analysis revealed that the number of holo-clusters correlates with the number of HLE1-ORFs in all mitotic species, including the distant Ment indicating the occurrence of at least one HLE1-ORF within long holo-clusters (Additional file 2: Fig. S19B and Additional file 1: Table S8).
To clearly depict long range organization, we performed a ModDotPlot analysis and showed one of the largest centromeric regions (0.2 Mb-0.3 Mb long) for each mitotic species in Fig. 6D. Interestingly, we observed a conserved linear organization of the centromere in all mitotic/polyploid species. This complex linear organization is characterized by holo-units composed of different array types arranged with IS-associated non-autonomous HLE1 Helitrons, while autonomous Helitron elements, HLE ORFs, are present in regions (10-50kb long) between the holo-units (Fig. 6D).

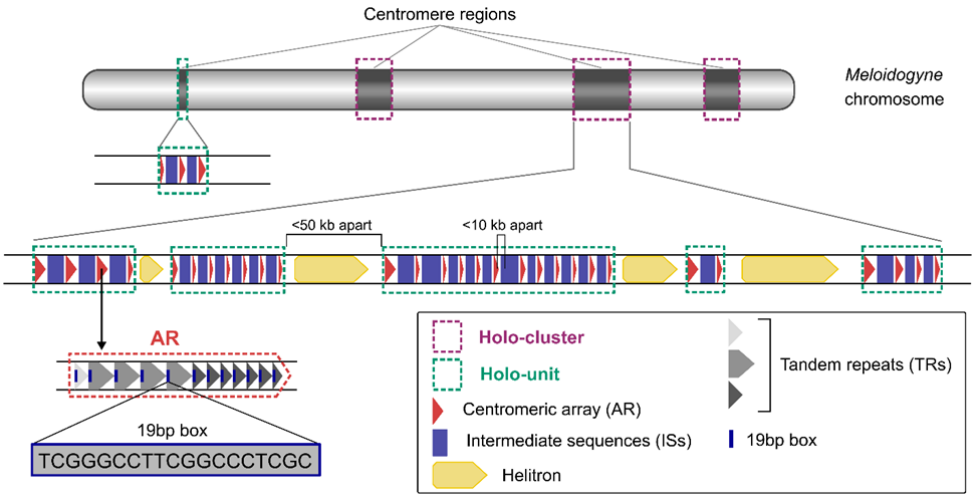

Fig. 7. Schematic presentation of centromeric regions found in mitotic and polyploid *Meloidogyne* species. Holo-units consist of centromeric arrays and ISs. ISs represent non-autonomous Helitron elements located between centromeric arrays. Holo-clusters consist of holo-units and autonomous Helitron elements between them.

The long-range organization of centromeric regions in mitotic/polyploid *Meloidogyne* species is presented as schematic representation in Fig. 7. The basic centromere unit in these species is a 19-bp box containing tandem repeats (TRs). These four species share a 14 different families of 19-bp box-containing TRs, with preferential monomer lengths from 40 bp to 90 bp (Fig. 2). Centromeric arrays are composed of different 19-bp box-containing TRs (Fig. 7). There are 20 different array types, with common lengths up to 2 kb, representing predominantly complex higher-order repeat (HOR) centromeric arrays organization (Fig. 3). Holo-units are composed of centromeric arrays (ARs) and intermediate sequences (ISs), where ARs are in contiguous blocks separated by ISs no more than 10 kb in length (Fig. 7). ISs represent non-autonomous Helitron elements between centromeric arrays, mainly about 1.5 kb and 3.5 kb in length, with sequence identity ranging from 79% to 99% (Fig. 4 and Additional file 1: Table S6) and (Fig. 6C). Holo-units, with rare exceptions when found scattered independently, are most often grouped into holo-clusters distributed along the chromosomes. A holo-cluster is defined as a group of holo-units separated by less than 50 kb with autonomous Helitron elements embedded in the regions between holo-units (Fig. 6D). For illustration, the number of holo-clusters ranges from 144 in Minc to 210 in Mare; among them, about 40/genome have a total length greater than 50 kb in MIG.

The comparison of long-range organization of centromeric regions in other populations of mitotic/polyploid MIG species

Multiple previous studies across MIG populations have demonstrated substantial karyotypic evolution, with chromosome numbers varying widely; for example, Minc populations ranges from 32 to 46 chromosomes [39]. As previously noted, the long-range organization of centromeric regions in MIG species was initially characterized using assemblies generated from the same populations involved in experimental centromere studies [27]. However, considering the availability of chromosome-level assemblies from additional MIG populations [23] and accounting for the genomic diversity observed between MIG populations, we also examined organization of centromeric regions across these assemblies and compared our findings with previously reported analyses using Zotta Mota et al. 2024 assemblies. MIG assemblies from Dai et al. (2023) include 39 chromosomes (Ch1–Ch39) plus 157 contigs for Minc, 47 chromosomes (Ch1–Ch47) plus 288 contigs for Mjav, and 52 chromosomes plus 429 contigs for Mare (Additional file 2: Fig. S20). These chromosome assemblies are comparable in size to those reported by Mota et al. (2024) and both are consistent with previous flow cytometry estimates.

Analysis of the distribution of 14 TRs containing the 19-bp box across both assembly groups revealed similar level of abundance for almost all TRs (Additional file 2: Fig. S21). Only Cl25 TR shows the discrepancy between Minc assemblies/populations, with more abundance in Dai et al. (2023) compared to Mota et al. (2024). Using the same approach, centromere arrays (ARs) and intermediate sequences (IS) has been extracted and compare between assemblies. The distribution of 20 array types (AR_CL_N) showed consistent patterns between assemblies, with exceptions in AR_CL_68 (primarily composed of Cl25 TRs) with more abundance in the Minc genome, while AR_CL_16g2 was less abundant in the Mjav genome from Dai et al. (2023) (Additional file 2: Fig. S22A). The distribution pattern of ISs among assemblies is found still more conserved (Additional file 2: Fig. S22B). In addition, the length profiles of TR arrays and IS sequences were largely similar between assembly groups (Additional file 2: Fig. S23). The only minor variations can be seen around 800 and 2000 bp in array lengths between Minc assemblies/populations (Additional file 2: Fig. S23A). Regarding centromeric organizations, the largest holo-clusters from Dai et al. (2023) MIG assemblies/populations are presented in Additional file 2: Fig. S24. Holo-units consist of various AR types alongside IS sequences (non-autonomous Helitron elements) while holo-clusters comprise different holo-units associated with autonomous Helitrons such as observed in Mota et al. (2024) assemblies/populations. In summary, comparative analyses of centromeres between two populations of MIG species reveal the same composition and general trends

in complex linear organization of centromere. The minor differences in TR and array proportions across assemblies are likely reflecting population-specific dynamics.

Thanks to the chromosome-level assemblies, we also examined linear organization of centromere at the chromosome scale for all three species, Minc, Mjav and Mare (Additional file 2: Figs. S25-27). The results of mapping confirmed a previously suggested chromosomal pattern of centromere investigated by experimental approach: unevenly distributed centromere regions along the chromosomes, supporting the concept of cluster-centromeres rather than true holocentromeres. Interestingly, despite the genomes being polyploid (3x or 4x), indicating the presence of chromosome triplets or quartets, no similar pattern of centromeric cluster distribution could be identified among any chromosomes through visual inspection.

Diverse linear organization of centromeric regions in meiotic and diploid *versus* mitotic and polyploid *Meloidogyne* species

To investigate organization centromeric organization in meiotic and diploid species, we focused on Mhap and Mchi, belong to the Clade II and Clade III respectively (Fig. 1A) As presented above, these two species exhibit well-conserved αCenH3 and 19-bp box (Fig. 1B and 1C) as well as highly-contiguous genome assemblies [34,36]. These features made them ideal candidates for comparative analyzes with mitotic species to uncover possible differences in linear organization of centromeric regions. Our first goal was to identify and characterize all TRs with a 19-bp box in Mhap and Mchi, allowing two mismatches according to the findings from Fig. 1C. The alignment of extracted TRs-containing 19-bp box from both meiotic species was subjected to PCA (Fig. 8A). The distinct separation of TRs containing the 19-bp box between the two meiotic species, each forming separate species-specific clusters, can be observed. Mchi exhibited only two dominant TR clusters (MG1 and MG7), whereas Mhap showed a more complex organization with five distinct TRs clusters (MG2-MG6) (Additional file 2: Fig. S28). The consensus sequences of these TRs found in both meiotic/diploid species, Mhap and Mchi, have monomer lengths similar to those in four mitotic/polyploid species (Minc, Mjav, Mare and Ment), ranging from 40-80 bp (Additional file 2: Fig. S29A). Moreover, TRs found in two meiotic/diploid species, like those identified in four mitotic/polyploid species, are short sequences that, apart from a 19-bp box, have relatively low mutual sequence identity (Additional file 2: Fig. S29B). Furthermore, using the same approach as for mitotic species, we performed annotation of specific TRs and uncovered arrays in Mhap and Mchi. The length distributions of arrays in Mhap and Mchi were similar to those observed in mitotic species, with 87.6% of arrays in mitotic and 93.94% of arrays in meiotic/diploid Mhap and Mchi being shorter than 1500 base pairs (Fig. 8B, left). However, the most striking difference in composition of

centromeric regions between these two groups of species was the total number of TRs and arrays per genome. Both the number of TRs and the number of arrays were more than 2.5 times lower in meiotic/diploid species compared to mitotic/polyploid species when adjusted to total genome size (Fig. 8B, right). To compare the intrinsic organization of arrays between these two groups, we mapped different TRs on arrays database and selected arrays according to the number of different TRs per array in all six species (Fig. 7C). Meiotic/diploid species have fewer different TRs per array compared to mitotic/polyploid species. Moreover, arrays are preferentially dominated by one or two TRs in meiotic/diploid, suggesting a monomeric array trend, whereas mitotic/polyploid species show that half of the analyzed arrays consist of three or four TRs, indicating a high proportion of HOR array organization, as shown previously. Moreover, Mchi does not have any arrays that contain more than two different TRs. Another notable difference in composition of centromeric regions involves the sequences between the centromeric arrays, known as intermediate sequences (ISs). The comparison of ISs showed that these elements are three times less abundant in meiotic genomes; 74 in mitotic genomes versus 24 in meiotic genomes, standardized by genome size (Fig. 7D). In addition, ISs are generally shorter in length in meiotic (less than 1500pb) compared to mitotic (up to 7500bp). To compare the complexity of organization between these two groups, holo-clusters (group of holo-units separated by less than 50 kb) were extracted from the genome (Fig. 7E). Although both group of species have many shorter holo-clusters (below 5000 bp), mitotic/polyploid species have significantly larger number of long holo-clusters (Wilcoxon rank sum test, $W = 78332$, $p = 1.98 \times 10^{-15}$). This was further corroborated by constructing a linear model of the number of holo-clusters versus genome size across the six species (Additional file 2: Fig. S30). Although the number of holo-clusters correlates almost perfectly with genome size, their mean length differs dramatically between the two groups. The average length of holo-clusters in mitotic/polyploid species is 29 kb, whereas in meiotic/diploid species it is only 6.1 kb, roughly holo-clusters are five times longer in mitotic/polyploid species. This difference in organization becomes even clearer when holo-clusters over 50 kb in length were analyzed. While there are between 30 and 40 of holo-clusters over 50 kb per genome in mitotic/polyploid species, there is only one in meiotic/diploid Mhap, and none in Mchi (Fig. 8E). Furthermore, the significance of the proportion of large clusters above 100 kb is also observed in mitotic species compared to meiotic, with the largest in Mjav, measuring 374 kb in length. Since the results on mitotic/polyploid species showed integrated autonomous and non-autonomous Helitrons in the centromeric regions, to determine whether Mhap and Mchi show similar trends, we annotated their centromeric regions and compared with Minc, Mare, Mjav and Ment annotations (Fig. 8F). The results show that meiotic/polyploid species (Mhap and Mchi)

possess more than five times fewer non-autonomous Helitron elements in their intermediate sequences (ISs) compared to mitotic species (Minc, Mare, Mjav and Ment); 20 versus 110, when standardized by genome size. Additionally, mapping analyses of Helitron coding HLE1-ORFs reveal an almost complete absence of autonomous Helitron elements within the centromeric regions of both Mhap and Mchi (Additional file 2: Fig. S31). Beyond their absence in the centromeres, genome-wide analyses of autonomous Helitrons in these meiotic/diploid genomes show that are nearly entirely missing.

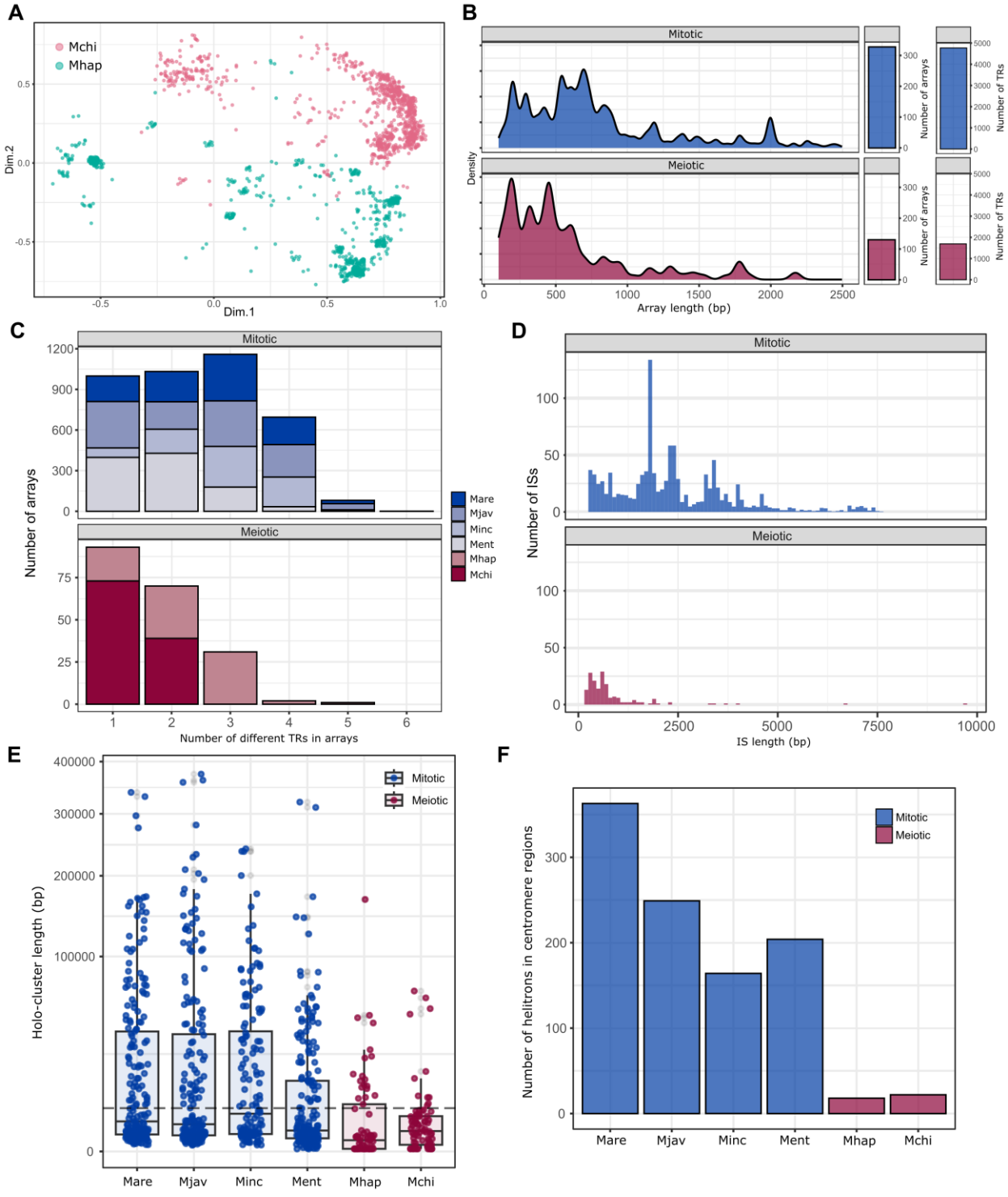

Fig. 8. Comparative analyses of centromeric regions between two meiotic/diploid species (Mhap and Mchi) and four mitotic/polyploid (Minc, Mare, Mjav and Ment)

A) PCA of TRs extracted between 19bp boxes in *M. hapla* (Mhap) and *M. chitwoodi* (Mchi). Clear separation between the two meiotic species that belong to two distant phylogenetic Clades (II and III) can be observed. B) Standardized length distribution of TR array length in mitotic and meiotic species along with their average number per genome and the total number of detected TRs. While the profile of array lengths remains similar, the overall number is 2.52x smaller (328 Mb *vs* 130 Mb genome size standardized) and the total number of TRs is 2.82x larger

when standardized according to genome sizes in mitotic species. C) Number of distinct TRs per array in all six studied species. Meiotic species have less distinct TRs per array (Wilcoxon rank sum test, W=19.5, p=0.00563) when compared to mitotic species, with retaining the similar albeit shorter length profile. D) Distribution of ISs length in mitotic and meiotic species, standardized to the array number ratio. The smaller number and the lack of specific IS sizes indicate different organization of the centromere in meiotic species. E) Centromere region length profile across meiotic and mitotic species. Although both classes have many shorter centromere regions, mitotic species have significantly more ($p = 1.98 \times 10^{-15}$, Wilcoxon rank sum test) larger centromere regions with a 5x higher mean length (29 kb for mitotic vs. 6.1 kb for meiotic). The underlying distribution (1Q, mean and 3Q) of centromere region length per species is provided by the boxplots. F) Per species overlap of Helitron and centromere region annotations. Helitron annotations were reduced to non-overlapping regions (since many Helitron annotations provided by Heliano overlap especially in the larger centromere regions of mitotic species). Meiotic species have significantly less Helitrons overlapping with centromere region annotations.

We also analyzed the linear organization of centromeric regions at the chromosome level in Mhap and Mchi (Additional file 2: Fig. S32). This mapping confirmed a previously proposed clustered centric pattern, characterized by unevenly distributed centromere regions along and among the chromosomes, similar to that observed in mitotic/polyploid species. However, Mhap and Mchi exhibit simple holo-clusters primarily consisting of monomeric array types and lacking, both autonomous and non-autonomous Helitron elements, unlike mitotic/polyploid species, which possess centromeres composed of complex holo-clusters. A single complex holo-cluster containing Helitron elements was identified in Mhap, whereas Mchi displays a purely simple organization of centromere without Helitron elements (Additional file 2: Fig. S30).

Discussion

Our previous research on the mitotic species Minc showed that the centromeric histone αCenH3 colocalize with α-tubulin, supporting presence of the CenH3 epigenetic marker within the Minc centromeric regions [27]. Additionally, ChIP data demonstrated that the αCenH3 regions are associated to different tandem repeats, positioning Minc among the species with recently repeat-based centromere [44,45]. To date, only a few cases of repeated-based centromeres have been characterized; notably, aside from Minc, all have been found in plants [46]. Regarding chromosome organization, the centromeric regions in Minc display an unusual pattern characterized by an uneven, clustered distribution of αCenH3 along the chromosomes. The discovery of this unique organization of centromeric regions in animal species, combined with the parthenogenetic reproduction of *Meloidogyne* species, prompts us to investigate its long-range and genome-wide organization. Consequently, centromeric regions of four obligate mitotic, polyploid species and two facultative meiotic, diploid *Meloidogyne* species have been analyzed using high-quality reference genome assemblies [24,33,34].

These recently assembled genomes provided a valuable resource for comparative analyzes of centromeric regions and studies of different centromere pattern.

First, we found that the six *Meloidogyne* species examined possess highly conserved centromeric H3 histone, CenH3s, regardless of species phylogenetic distance. Considering the rapid evolution of CenH3 observed in most animal and plant species studied to date [47,48], the *Meloidogyne* specific CenH3s that remain almost conserved in species, separated by over 50 million years represent an extreme case. The sequence conservation of *Meloidogyne* CenH3s across distant relatives, contrasted with the absence of CenH3s in the holocentromeres of certain insects together with some CenH3-free holocentromere regions in the plant *Cuscuta*, [12,49] highlights the overall high structural and epigenetic flexibility of holocentromeres [47,48].

The centromeric regions of repeat-based centromeres, whether mono- or holo-, are typically enriched with various repetitive DNA families, predominantly megabase-sized satellite DNAs, which tend to evolve rapidly and often display significant differences even between closely related species [47]. However, our previous experimental study involving three related mitotic/polyploid MIG species (Minc, Mare, and Mjav) revealed that the DNA associated with CenH3 consists of different tandem repeats (TRs), all sharing a completely conserved 19-bp box. Furthermore, the experimental results confirmed that these 19-bp box-containing TRs which represent the centromeric region in all three closely related MIG species: Minc (confirmed by ChIP data), Mare and Mjav (confirmed by IF/PRINS) [27]. In this study, we also demonstrate that, beyond MIG species, mitotic/polyploid Ment species and the more distantly related meiotic/diploid species, Mhap and Mchi, exhibit also highly similar 19-bp box sequences associated with divergent TRs. In this study, we also confirmed the association of 19-bp box with centromeric regions in meiotic/diploid Mhap, using combined immunofluorescence using anti-αCenH3 antibodies and PRINS assays targeting the 19-bp box on Mhap chromosomes. The results confirmed the association of 19-bp box with centromeric regions also in this distantly related meiotic species. Furthermore, the presence of the highly conserved αCenH3 centromeric protein alongside the conserved 19-bp box within TRs in two other species, mitotic/polyploid Ment and meiotic/diploid Mchi species strongly suggests role of these sequences in centromere as well. The presence of conserved 19-bp box in highly divergent centromeric TRs in all these species suggests that centromeres in *Meloidogyne* are at least partially genetically determined. The observed long-term conservation of both genetic (19-bp box) and epigenetic determinants (CenH3) across *Meloidogyne* species further support our previous hypothesis that the conservation of centromere components in these species could be result of sub-specialization of CenH3 variant exclusively for mitosis [27].

The presence of shared centromeric sequence features among these species prompted us to investigate the long-range organization of centromeric regions, with the aim of comparing it across these six *Meloidogyne* species and identifying its evolutionary patterns. We revealed, that the TR arrays, organized as complex higher order organization (HOR) composed of different TRs families or monomeric units build the basic units of centromeric regions. However, the intrinsic organization of the arrays shows a different pattern in mitotic/polyploid compared to meiotic/diploid species. Mitotic/poliploid species tend to exhibit a HOR of arrays consisting of two or more highly divergent TRs in different and complex combinations. In contrast, meiotic/diploid species, especially Mchi, have a simpler organization with arrays consisting of one or at most two consecutive TRs. However, despite difference in composition, TRs arrays maintain a preferred length profile, characterized by short arrays ranging from 300bp-1kb in all analyzed species. The difference was also observed in centromere abundance; the number of arrays, standardized by genome size, was found to be three times higher compared to meiotic/diploid *Meloidogyne*.

In previous studies, the HOR organization within centromeric arrays has been confirmed in monocentromeres of many organisms, including those in humans and *Arabidopsis*. However, in these species, HORs are based on variants of a single satellite DNA, alpha in humans and CEN180 in *Arabidopsis* [5,50]. In contrast, HOR organization, observed in mitotic/polyploid *Meloidogyne* has not been detected in either holocentromeres or monocentromeres. Regarding, the abundance of centromere regions, a previous study on repeat-based holocentromeres in plant *Rhynchospora* revealed that each chromosome contains between 448 and 727 centromeric regions, composed of Tyba repeats spanning 15–25 kb [11]. Conversely, each chromosome of plant *Chionographis* features only 7–11 extremely long centromeric units, ranging from 0.24 to 4.46 Mb in length, thus approximating the length of a monocentromere [45]. Similarily, recent research on humans and *Arabidopsis* suggests that monocentromeres, similar to repeat-based holocentromeres, are not continuous sequences, but are composed of multiple neighboring centromere units [5,50]. A recently discovered specific type of repeat-based centromeres are meta-polycentromeres, characterized by multiple domains of centromere satellite DNAs within extended chromosome constriction [14–17]. Another important question that arises with repeat-based, non-monomeric centromeres, including holocentric and clustered centromeres, is the extent to which CenH3 regions actually correlate with centromere function. The extreme case of the plant Cuscuta, where a CenH3-free mechanism exists, suggests a dynamic and flexible organisation of these non-monomeric centromeres. In *Meloidogyne*, although the link between CenH3 regions and tubulin has been experimentally confirmed [27], it is obvious that CenH3 regions do

not fully overlap with kinetochore components. This leads to the hypothesis that CenH3 regions may be redundant and that only a fraction of CenH3 represents the functional centromere core, a phenomenon also observed in monocentric humans and yeast [51,52].

Considering current knowledge of repeat-based centromeres, including observations from *Meloidogyne*, it can be concluded that, while DNA component of these centromeres are built upon the fundamental unit of TRs, they exhibit significant diversity in their organization, abundance, and distribution.

In addition to examining general trends in the composition of centromeric units in *Meloidogyne*, our analysis also focused on the overall organization of these regions and compared mitotic/polyploid with meiotic/diploid *Meloidogyne* species to identify putative differences in centromere organization [5,53]. We first discovered the association of centromeric arrays with several different intermediate sequences (IS) in mitotic/polyploid species. Furthermore, various combinations of array types and different IS classes form holo-units, revealing an additional layer of complexity in the organization of higher-order repeats in four mitotic/polyploid species. This higher-order organization of holo-units exhibits a very similar pattern in these species, particularly among Mare and Mjav probably as a consequence of their genome tetraploidy (AABB subgenomes), in contrast to AAB subgenomes of the triploid Minc [22,24][22,24]. Additionally, the more distantly related Ment exhibits a slightly less organized array structure, primarily due to length variations, however, the arrangement of ISs and holo-units follows similar patterns to those observed in MIG species. Notably, even in distantly related Ment, whose genome originated through hybridization from parental species other than MIG, the organization of the centromeric regions remains fundamentally highly similar, regardless of their phylogenetic distance and differing parental origins.

Another significant finding is, that ISs embedded between the arrays in all mitotic/polyploid species represent non-autonomous transposon elements belonging to Helitron group. In addition, we identified a third layer of centromere organization in these species, manifested as holo-clusters composed of groups of holo-units with autonomous Helitron elements in the regions between them. We found that holo-clusters (Fig. 7) in four analyzed mitotic/polyplod species could encompassing even about 350kb long stretch of centromere which belong to extremely long centromeric regions in holocentric species.

In contrast to the complex organization of centromeric regions observed in mitotic/polyploid species, meiotic/diploid Mhap and Mchi exhibit a much simpler arrangement. Specifically, beyond the simple organization of the arrays themselves, these two meiotic/diploid species lack both non-autonomous and autonomous Helitrons within their centromeres. Notably, our study demonstrates that the absence of both classes of Helitron elements in holocentromeres of meiotic/diploid species correlates with a

significantly lower number of holo-clusters per genome compared to mitotic/polyploid species. This suggests a reduced level of amplification and spread of both, holo-units and holo-clusters. This reduction in abundance of centromeric regions and complexity in meiotic/diploid genomes highlights a fundamental difference in the organization of centromere between species these two groups of *Meloidogyne*, mitotic/polyplod versus meiotic/diploid.

From our study, it appears that non-autonomous elements may be involved in the amplification of arrays in holo-units, while autonomous Helitrons have an impact on the dramatic amplification and distribution of holo-clusters across chromosomes of mitotic/poliploid species. In contrast, meiotic species show lower distribution and less complex organization likely due to the absence of centromere-associated Helitron elements. Helitrons belongs to a superfamily of DNA transposons that are widely distributed in most eukaryotic genomes. It is assumed that they are mobilized by a “rolling cycle” [53], which includes extrachromosomal DNA (eccDNA). Autonomous Helitron elements encode all proteins required for their transposition, and non-autonomous Helitrons require enzymatic activity encoded by an autonomous element for transposition to occur [54]. On this basis, the burst and spreading of holo-clusters through eccDNA and subsequent reintegration into a new site along the chromosome is the most likely mechanism for the propagation of holocentromere mediated by Helitrons in mitotic *Meloidogyne* species. Previous studies have shown that Helitrons are a universal feature of plant centromeres. In wheat and rice, a large number of tandem-repetitive Helitrons are found in the centromeres of all chromosomes, and the authors emphasize that Helitrons likely contribute to centromere formation and centromere plasticity [55,56]. The most prominent example of putative Helitron involvement in centromere organization is the presence of a non-autonomous TCR1 Helitron associated with a centromeric Tyba tandem repeat in the holocentromeric plant *Rhynchospora [13]*.

Our comparative studies of centromeric regions in mitotic/polyploid and meiotic/diploid *Meloidogyne* species reveal correlations between the associations of non-autonomous and autonomous Helitron elements and the extent of their amplification and distribution, thus demonstrating a direct influence of these elements on centromere genomic dynamics. Although, similarity in organizational patterns among three closely related MIG species may be influenced by species relatedness; this correlation is not consistent across all species. For example, mitotic/polyploid Ment, although belonging to the same Clade but being distant from other MIG species, displays highly similar organization of centromeric regions. Additionally, meiotic/diploid Mhap and Mchi, which are distantly related and belong to Clades II and III respectively, both exhibit limited dispersion of centromeric regions probably due to the absence of Helitron. Moreover, Mhap (Clade II) is equally distant from species from Clade I (both MIG and Ment)

and Mchi (Clade III), and Mchi is considered closer to the ancestral species of the genus [39]. This relationship is further supported by the sequence similarity of CenH3 among these species, analyzed in this study, where CenH3 of meiotic/diploid Mhap showed greater similarity to mitotic/polyploid species from Clade I (MIG and Ment) than with meiotic/diploid Mchi from Clade III. These data suggest that phylogenetic relationships alone cannot explain the observed differences in organization of centromeric regions between these group of species.

Although all analyzed species are parthenogenetic, the species Mhap and Mchi undergo meiosis. In the germ cells the chromosome number is reduced by meiosis, and the somatic chromosome number is restored after the fusion of the second polar nucleus with the egg pronucleus [57]. The main difference between mitotic and meiotic *Meloidogyne* parthenogens is that mitotic do not pass any checkpoint involving bivalent homologous chromosomes. It has already been shown that meiotic/diploid species, in contrast to mitotic/polyploid, have far fewer repetitive elements, including transposons [37,58]. It can be assumed that meiosis in these species is less tolerant of the accumulation of repetitive elements such as centromeric TRs probably because of the risk for of dangerous rearrangements and genome reshuffling while mitotic/polyploid species probably have greater flexibility in genomic rearrangements [37,58]. In that line, the extensive tolerance to the amplification of holocentric TRs by Helitron elements in species that are exclusively mitotic, could be also due to their polyploidy which can enable higher genome flexibly to fragmentation, chromosomal rearrangements or even a loss of chromosome parts. The absence of a subgenome-specific subset of holo-clusters, along with the lack of similarity in holo-cluster distribution pattern among homologous chromosomes in mitotic/polyploid species, suggests that extensive amplification and distribution of centromeric regions occurred following species hybridization events. Almost all *Meloidogyne* species are parthenogenetic, and those with particularly high parasitic potential are mitotic/polyploid species [57]. The ancestors of *Meloidogyne* were sexual, parthenogenetic meiotic species evolved from sexual species, and obligate mitotic parthenogenetic species evolved from facultative meiotic parthenogenetic species after meiosis was suppressed during egg maturation [25]. In general, sexual reproduction, which allows the combination of genetic material, plays a key role in the evolutionary and adaptive potential of parasite species, while in contrast, parthenogenetic species show less adaptive capacity. However, this does not apply to mitotic/polyploid parthenogenetic *Meloidogyne* species, which appear to show great adaptability despite their asexuality [57]. It can be hypothesized that the abundant centromeric regions with a high potential for expansion, especially in the context of polyploidy, may drive rapid chromosome evolution, contributing to the adaptability and plasticity of

these mitotic/polyploid genomes, and thereby enhancing their capacity to respond to environmental changes.

It has been previously observed that holocentromeres can differ in their organization between mitosis and meiosis within the same species, as demonstrated in plants of the genus *Rhynchospora*. During mitosis, holocentromeres are typically "line-like," while in meiosis holocentromeres form "cluster-like" structures [59]. Therefore, a comprehensive analysis of mitotic and meiotic centromeres within the same *Meloidogyne* species would be the most direct approach to confirm our hypothesis. However, although the existence of Mhap mitotic populations has been well documented in earlier studies [60], these populations are unfortunately not available today. A single Helitron associated holo-cluster found in meiotic Mhap may reflect an ability of Mhap to switch between forms. In contrast, no holo-clusters with Helitron are found in Mchi, consistent with its status as an exclusively meiotic species [61].

Therefore, in the absence of phylogenetic replication, we cannot conclusively determine that different reproductive modes of hybridization events are the primary factors driving centromere diversity in *Meloidogyne* species. Nonetheless, it is plausible that factors such as mode of reproduction and species hybridization, which can lead to suppression of recombination in mitotic mode and polyploidy as a result of hybridization, may influence centromere organization. However, it remains uncertain whether these factors are directly causal or whether differences in centromere organization are merely passive outcomes of these processes. Finally, considering that new long-read chromatin profiling methods enable the study of chromatin features across centromeric regions, our further investigation will focus on analyzing the methylation patterns within these regions to identify the components that contribute to the functional centromere core.

Conclusions:

Through high-resolution genome assembly comparisons, we demonstrate that evolution of centromeric regions in mitotic and polyploid *Meloidogyne* species is driven by Helitron-mediated amplification and expansion. Because the emergence of mitotic and polyploid parthenogenesis in the *Meloidogyne* lineage examined here appears to have occurred only once, we cannot attribute this difference in centromeric regions, compared to meiotic lineages solely to differences in reproductive biology or to the polyploidization event. Even so, both reproductive mode and genome polyploidy remain plausible contributors to the observed variation in content and organization of centromeric regions in *Meloidogyne* species. Whether these factors actively shape these features of centromeric regions or the patterns we observe reflect secondary, non-causal outcomes, however, remains to be investigated. Our

findings have important implications for understanding evolution of DNA components of centromere and offer insights into genome evolution.

Methods

Genome assemblies

Long read-based genome assemblies of three *Meloidogyne* species, *M. incognita, M. javanica* and *M. arenaria* (populations from INRA collection) have been downloaded from the repository (https://doi.org/10.57745/RAZ8JS [62], https://doi.org/10.57745/50YQVI [63] and https://doi.org/10.57745/VBCSTD [64] respectively) [33]. The chromosome-level genome assemblies of these three species (populations were collected from farmlands in China) have been downloaded from the website (http://bmb.hzau.edu.cn/sjxz.htm) [23]. For *M. enterolobii* we used recently published genome assembly [24] with GenBank reference number GCA_963681835.1 downloaded from NCBI [65]. *M. chitwoodi* genome assembly [34] has been downloaded from NCBI under reference GCA_015183035.1 [66] and *M. hapla* genome assembly was made available through personal correspondence with dr. Shahid Siddique of UC Davis, published in Shakya et al. (2025).

αCENH3 Analysis

To identify *αCENH3* sequences in all 6 studied species, NCBI t*BLASTn* [67] was used to search for all *αCENH3* protein sequences previously described in Minc, Mare, Mjav and Mhap [27] within the genome assemblies. The retrieved sequences were manually extracted and aligned using MAFFT [68]. The resulting alignment distance matrix was visualized using the *pheatmap* package [69] in R.

Slides preparation

Reproductive tissues from *M. hapla* females were isolated and processed using the cytospin technique. Samples were collected in 10 µg/ml colcemid (Roche), pierced with a needle, and incubated at 4 °C for 1 hour to overnight. Tissues were then washed with PBS, homogenised with a micro-tube homogeniser for 30 seconds, and transferred to a Dounce homogeniser. The tissue was dissociated with 30 strokes of Pestle A, then strained through 100 µm and 40 µm cell strainers. The resulting suspension was adjusted with PBS to a final volume of 400 µL, suitable for loading into a Cytospin funnel, corresponding to 5–10 females per coated Cytoslide (Shandon, ThermoFisher Scientific). Slides were spun for 10 minutes at

1200 rpm using a Cytospin 4 centrifuge (Shandon, ThermoFisher Scientific). After drying, slides were fixed by immersion in ice-cold fixative for 20 minutes, then completely dried and stored. For fixation, samples were incubated in a methanol:acetone (1:1) solution at -20 °C for 20 minutes.

Immunodetection (IF) and Primed in situ labeling (PRINS)

Slides with cytosmears were blocked with 2.5% BSA in PBST and incubated with anti-CenH3 (1:400 dilution) overnight at 37 °C. After washing, slides were incubated with Alexa594 anti-rabbit secondary antibody (Abcam, ab150080, 1:1000) for 1 hour at 37 °C, then washed and processed further for PRINS. The pretreatment in 45% acetic acid for 10 minutes and in 2xSSC for 5 minutes were performed and slides were then treated with RNase A for 30 minutes at 37 °C, washed three times with PBS, and fixed with 1% formaldehyde in PBS containing 50 mM $MgCl_2$. Following dehydration in a cold ethanol series, denaturation was performed in 70% formamide in 2xSSC at 70 °C for 2 minutes. Slides were dehydrated again and air-dried. The reaction mixture was prepared in 50 µL containing 2 µM primer (19-bp box; TCGGGCCTTCGGCCCTCGC), 2,5 µM MgCl2, 150 µM each of dATP, dCTP, dGTP, 96 µM dTTP and 54 µM biotin-16-dUTP, 1 U of GoTaq® G2 DNA Polymerase (Promega) and 1X Colorless GoTaq® Reaction Buffer (Promega). On each prewarmed slide, 25 µL of prepared mixture was applied, covered with coverslip, sealed and continued to heat at appropriate annealing and elongation temperature of 65 °C for 30 min. Reaction was stopped by washing in 50 mM NaCl, 50 mM EDTA, pH 8 buffer for 5 min at 65 °C followed by 3x5 min washes in 4x SSC with 0,05% Tween 20. Immunodetection involved fluorescein avidin D and biotinylated anti-avidin D (Vector Laboratories). Slides were counterstained with DAPI, dried, and mounted in Mowiol 4-88 (Sigma-Aldrich). To reduce nonspecific staining, slides were incubated with Image-iT FX Signal Enhancer (Invitrogen) for 30 minutes. Microscopic images were recorded using confocal STED confocal microscope system (Abberior Instruments) in confocal mode equipped with a 100x NA 1.4 (oil) [UPLSAPO 100XO]. Each fluorochrome was capture separately and images were merged and analyzed using Image J and Adobe Photoshop software.

Detection of centromeric tandem repeats (TRs)

First, we determined the number of 19-bp boxes in the genome assemblies. The initial validation of 19bp box conservation across all six studied species was done using BLASTshort algorithm allowing from 0 to 5 mismatches (Fig. 1C). Next, the centromere-specific 19-bp boxes were searched on both the forward and reverse strands and detected using Geneious Prime 2023.2.1 (https://www.geneious.com)

and the Find Motif tool. To detect and extract all tandem repeats (TRs) that began with a 19-bp box and ended before the next 19-bp box for MIG and *M. enterolobii*, one mismatch was allowed, while two mismatches were permitted for both meiotic species, *M. hapla* and M. *chitwoodi*.

All extracted 19-bp box-containing TRs from MIG species were then cross-referenced with the consensus sequences of the six previously described 19-bp box-containing centromeric TRs (CL16g1, CL16g2, CL16g3, CL25, CL32 and 16_32) in MIG species [27]. Consensus sequences of these TRs were annotated on the genome assemblies using the BLAST algorithm and filtered using 70% sequence identity and 70% sequence coverage parameters. Next, all TRs detected using the six TR consensus sequences were overlapped with all detected 19-bp box containing TRs to identify potential new centromeric TRs that were not detected in previous ChIP data [27] due to their low genome abundance. These new TRs were extracted and grouped into 8 new TRs, named using the convention Gn, where G represents the specific group and ”n” the number (Fig. 1). Detection of TRs in Ment were done with consensus of TRs obtained from MIG species (>70% query coverage and >70% percentage identity) using SatAnnot tool. Subsequently, both the 19-bp box annotations and the detected Ment species-specific TRs were used to define arrays.

The TR families were annotated on the genomes with *SatAnnot (https://github.com/mvolar/SatAnnot)* package developed within *SatXplor* pipeline [70]. From the resulting BLAST table, the high-similarity hits (>70% query coverage and >70% percentage identity) were filtered and GFF3 annotation files were used for subsequent inspection. The annotations were further filtered on a “best-hit per location” basis to avoid overlapping annotations due to the high inter-similarity between some regions of TR families.

Identification and extraction of centromeric arrays

Arrays of TRs containing 19-bp box were identified using a custom R-based approach. To account for sequence variability and minor misannotations that could cause unnecessary array fragmentation, all known TR annotations were first extended by 20 base pairs upstream and downstream. This extension ensured the merging of closely positioned or partially overlapping annotations into continuous array regions, allowing for more accurate array definitions. Using this approach, arrays were delineated as the reduced set of merged TR blocks. TRs within a single array were consistently oriented in the same direction, allowing confident assign of strand orientation to each array. Extracted arrays were obtained directly from the reference genomes using the getSeq function from the BSgenome R package [71], enabling precise sequence retrieval for downstream analyses. Extracted were obtained directly from the

reference genomes using the getSeq function from the BSgenome R package [71], enabling precise sequence retrieval for downstream analyses.

Classification of arrays sequences using automatic clustering

The downstream analysis followed a dedicated pipeline [70,72]. Extracted array sequences were aligned using MAFFT [68] with --adjustdirection and --reorder parameters, and multiple sequence alignment MSA distance matrices were generated using the *dist.dna* function from the *ape* package [73] with F81 genetic distance model. PCA was then performed on the distance matrices using the *FactoMineR* package [74]. Afterwards, DBSCAN clustering algorithm [75] was applied with an epsilon value of 0.01 for arrays in R. Because arrays of different length but high sequence similarity were sometimes separated into multiple small clusters, the resulting clusters were further manually curated to include variable length monomeric arrays of G4 and G6, as well as CL16g2 and CL32g in their respective array types and singular groups because it was identified that their length was the reason for their separation in the DBSCAN clusters. Finally, the clustered array types were linked back to their genomic locations and their annotation tables were created and a thorough visual inspection of annotations was performed. The decided naming convention for automatically assigned array types was AR_N, where AR stands for array, and _N is the randomly assigned cluster number. Additionally, for each algorithmically created array cluster we generated genetic distance heatmap using the pheatmap package. Centroids of each array type were calculated to assess dispersion levels in the PCA visualizations.

Identification and extraction of IS sequences

Intermediate sequences (ISs) were defined as the genomic regions flanked by arrays that were larger than 200 bp but shorter than 10 kb, thus excluding short gaps. Unlike arrays, ISs lacked characteristic features or orientation signals, making it impossible to determine with certainty which array they were functionally or directionally associated with, and no reliable strand orientation could be assigned. ISs were also extracted directly from the reference genomes using the getSeq function. Downstream analyses of both arrays and ISs followed a similar pipeline: sequences were aligned using MAFFT [68] with --adjustdirection and --reorder, distance matrices were generated using dist.dna with the F81 model [73] , and PCA was performed using the FactoMineR package [74]. Distinct clusterings were detected using the DBSCAN algorithm [75] (epsilon = 0.011). The decided naming convention of automatically assigned IS types was IS_N, where IS stands for intermediate sequence, and _N is the

randomly assigned cluster number. Additionally, IS_2, IS_4 and IS_8 additional PCA analysis following the same procedure described above were performed due to their high presence in all three MIG species.

Creation of clusters and Markov chain analysis

Following the definition of array types and ISs, certain periodic groupings of these sequences were discovered scattered throughout the genomes, following the cluster-like holocentromere organization in *Meloidogyne*. The first level of this organisational pattern was the holo-unit, i.e. the basic organisational blocks of holocentromeres, consisting of groups of arrays and IS sequences less than 10 kb apart. The second level was the holo-cluster, representing groups of holo-units separated by larger interspersed regions up to 50 kb in size. To further investigate this complex centromere organisation along contigs, we applied Markov chain analysis for the first time in the study of long-range holocentromere organisation. This approach was chosen due to the complex organisation of the Meloidogyne centromere, characterised by multiple array types and IS classes represented within both levels of organisation (holo-units and holo-clusters). A Markov chain transition matrix was constructed, with each entry representing the probability of transition between a specific AR type and an adjacent IS class, including their positional relationships along contigs within larger holo-cluster regions. This provides a probabilistic framework for understanding how these elements are arranged and interact across the genome. For analysis of complex organization and Markov chain model fitting, we filtered only holo-cluster containing >10 array types/ISs regions. The arrays and ISs contained in these clusters were then discretized and enumerated which allowed for creation of transition state vectors for each holo-clusters. Next Markov chain model fitting was performed on species-specific transition state vectors using markovchain R package [76] for all three species separately, and the transition matrices were analyzed and plotted using *igraph* R package [77].

Detection of Helitron Elements and Potential Palindromes

We used EarlGrey [41] with default parameters to annotate transposable elements and other repeats in the genomes of *M. enterolobii*, *M. incognita*, *M. javanica*, and *M. arenaria*. The annotations of all detected elements were overlapped with the holo-cluster annotations using foverlaps function from R data.table package, with an additional filtering step of satellite annotations provided by EarlGrey.

Since the genome assemblies of meiotic species *M. chitwoodi* and *M. hapla* lacked accurate repeat annotations and we required more precise Helitron annotations, a state-of-the-art *de novo* detection approach was applied using *Heliano*, using 15000 bp window size, 2500 bp max distance between HUH and Helicase domains and allowing multiple terminal sequences due to the nature of ISs (multiple palindromic sequences within a single IS) whilst lowering the false-positive cutoff of LTS and RTS sequences to 0.7 from default 0.5, which enables comprehensive Helitron identification along with the detection of HLE1 open reading frames (ORFs) [43].

Palindromes within ISs were detected using the *findPalindromes* function from the *Biostrings* package, with the following parameters: *min.armlength* = 4, *max.looplength* = 4, *mismatch* = 1. Post-detection, the palindromes were filtered to retain only those present in more than 1,500 copies across all four genomes. The CTAG motif was identified using the *vmatchPattern* function from *Biostrings*, allowing zero mismatches. To further characterize the distribution of palindromes and CTAG motifs, each IS sequence was divided into 100 bins based on length, and the relative density of detected palindromes and motifs in each of 100 bins was visualized.

Holo-cluster visualization

The main array types were visualized using Geneious Prime software and subsequently exported for further processing and classification. Holocentromere organization was examined using *ModDotPlot*, and track annotation was performed with the *ggbio* package to facilitate comparative genomic analyses of holo-cluster organization in the four mitotic species. For visualization purposes (Fig. 6D) we used EarlGrey Helitron annotations for non-autonomous elements and ORF annotations as discovered by Heliano indicating autonomous Helitron elements.

Data processing and statistical analyses

All genomic data processing and analysis were conducted in R, utilizing the *data.table* package for efficient data handling, filtering, and transformation. Graphical representations were generated primarily using the *ggplot2* package [78] for scatter plots and bar charts, and the *pheatmap* package [69] for heatmap visualizations, providing detailed representations of clustered data.

Genomic motif visualizations, including 19-bp box sequences and TR alignments, were performed in Geneious Prime, which facilitated extraction and graphical output preparation. Statistical analysis was performed in R using the base R package and the built in tests where applicable, t.test() for parametric

data, wilcox.test() for non-parametric data, lm.fit for linear regression, cov() for Pearson correlation and finally chisq.test() for Chi-square testing where applicable.

Peer review information

Andrew Cosgrove and Alison Monroe were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team. The peer-review history is available in the online version of this article.

Author contributions

NM designed the project and supervised the study. NM Funding acquisition. EDS and MV carried out most bioinformatic analyses. MV and EDS prepared figures, tables, and supporting information. MV, EDS and NM interpreted the findings. LS contributed to basic analysis. DV conducted the combined IF/PRINS analysis. LH performed the confocal microscopy. APZM and MP performed EarlGrey annotations. BM and DEGJ provided resources. NM wrote the manuscript with the input of EDS and MV. MV, EDS, BM, DEGJ and NM improved and revised the manuscript. All authors read and approved the final manuscript.

Data availability

Code and Supplementary data necessary for running all analysis is available as Supplementary Material at FigShare (https://doi.org/10.6084/m9.figshare.28715339) [79]. Genome assemblies used in this study include published data [33] for *M. incognita* [62], *M. javanica* [63], and *M. arenaria* [64], downloaded from the INRAE repository (https://entrepot.recherche.data.gouv.fr/dataverse/assemblies). Published genome assemblies for these three species, but from populations in China [23], were downloaded from the website (http://bmb.hzau.edu.cn/sjxz.htm). Published genome assemblies of *M. enterolobii* [24] (assembly accession GCA_963681835.1 [65]) and *M. chitwoodi* [34] (assembly accession GCA_015183035.1 [66]) were downloaded from NCBI. The *M. hapla* genome assembly was obtained through personal correspondence with dr. Shahid Siddique of UC Davis and is now published [36] and available with assembly accession GCA_051171035.1 [80].

Funding

This work has been fully supported by Croatian Science Foundation under the project IP-2024-05-2675.

Acknowledgments

We thank Shahid Siddique and Pallavi Shakya for early access to the *M. hapla* data prior to publication

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Additional files

Additional file 1: Tables S1-S8. Table S1. Used genome assemblies and their general statistics. Table S2. Pairwise similarities across the three genomes for all extracted monomers sequences for each centromeric family. Table S3. General statistics of detected array types in 3 MIG species. Table S4. Contigs with assigned subgenomes in M. arenaria and M. javanica. Table S5. General statistics of detected IS types in 3 MIG species. Table S6. IS consensus lengths and their relative fragment conservation scores determined using BLAST. Table S7. General statistics of detected 19bp box containing TRs in 3 MIG species and M. enterolobii. Table S8. Overview of Helitron and ORF statistics in analyzed genomes.

Additional file 2: Figures S1-S32. Fig. S1. Alignment of centromeric H3 histone, αCenH3 found in all six *Meloidogyne* species; Minc, Mare, Mjav, Ment, Mhap and Mchi. Fig. S2. The four cytosmears of *M. hapla* chromosomes with double immunofluorescence (IF). Fig. S3. Consensus sequences of the 14 TR families. Fig. S4. Representation of individual TRs in arrays derived from PCA and subsequent DBSCAN clustering. Fig. S5. Representation of individual TRs within defined arrays following manual curation. Fig. S6. Pairwise genetic distances within each array cluster. Fig. S7. A) Individual array lengths for each class of MIG (Minc, Mjav and Mare) arrays (AR_N). B) Overall length profile for all arrays across the three genomes (Minc, Mjav and Mare). Fig. S8. Genomic abundances (in kb) of the array types (AR_N) in the three MIG species. Fig. S9. Pairwise genetic distances between ISs in clusters for each IS type. Fig. S10. The number of intermediate sequences (ISs) for each IS class (IS_N) in MIG (Minc, Mjav and Mare) genomes. Fig. S11. Individual lengths for each IS class in MIG (Minc, Mjav and Mare). Fig. S12. Examples of organization of holo-units in the three MIG species (Minc, Mare and Mjav). Fig. S13. Estimated number of array type – IS transitions in all three MIG species. Fig. S14. PCA analysis of the distance matrix of arrays (Ment_AR_N) composed of 19 bp box TRs in *M. enterolobii* (Ment). Fig. S15. Depiction of arrays in the Ment genome with more than 20 occurrences. Fig. S16. Number of hits vs. Palindrome

index discovered in the IS sequences of all 4 mitotic species (Minc, Mare, Mjav, Ment). Fig. S17. Minisatellite forming palindromic motifs. Fig. S18. Quantity of Helitron-characteristic CTAG motif in complete set of IS sequences from MIG and Ment species. Fig. S19. A) Linear correlation between genome size and number of holo-clusters in Minc, Mare, Mjav and Ment. B) Number of holo-clusters >50kb in length per species (blue) and number of holo-clusters with at least one HLE1 ORF within (many have multiple). Fig. S20. Comparison of genome assembly metrics for *M. incognita* (Minc), *M. javanica* (Mjav), and *M. arenaria* (Mjav). Fig. S21. Number (N) of 14 centromeric TRs in Minc, Mjav and Mare assemblies/populations. Fig. S22. A) The comparison of the number of A) the array types and B) the number of IS classes of Minc, Mjav and Mare between assemblies from Dai et al. 2023 and Zotta Mota et al. 2024. Fig. S23. A) The comparison of the array lengths in Minc, Mjav and Mare assemblies/populations from Dai et al. 2023 and Zotta Mota et al. 2024. B) The comparison of ISs lengths in Minc, Mjav and Mare assemblies/populations from Dai et al. 2023 and Zotta Mota et al. 2024. Fig. S24. Examples of organization of the largest holo-clusters from different chromosomes (Chr_n) in the three MIG species (Minc, Mjav and Mare) from Dai et al. 2023 assemblies. Fig. S25. Localization of the centromeric arrays and ISs on 39 chromosomes of Minc from Dai et al. 2023 assembly. Fig. S26. Localization of the centromeric arrays and ISs on 47 chromosomes of Mjav from Dai et al. 2023 assembly. Fig. S27. Localization of the centromeric arrays and ISs on 52 chromosomes of Mare from Dai et al. 2023 assembly. Fig. S28. PCA of centromeric TRs from meiotic species forming 7 main TR clusters. Fig. S29. A) Alignment and identity matrix of centromeric TRs found in two meiotic species Mchi (*M. chitwoodi*) and Mhap (*M. hapla*). Fig. S30. Linear correlation between genome size and number of centromeric holo-clusters in Minc, Mare, Mjav, Ment, Mhap and Mchi. Fig. S31. Number of HLE1 ORFs per genome in meiotic (Mhap and Mchi) and mitotic species (MIG and Ment) obtained with the Heliano tool. Fig. S32. Localization of the centromeric arrays on 16 chromosomes of Mhap (A) and 30 contigs of Mchi (B).

References

1. Thompson SL, Bakhoum SF, Compton DA. Mechanisms of Chromosomal Instability. Current Biology. Elsevier Ltd; 2010;20:R285–95. https://doi.org/10.1016/j.cub.2010.01.034

2. Talbert PB, Henikoff S. The genetics and epigenetics of satellite centromeres. Genome Res. 2022;32:608–15. https://doi.org/10.1101/gr.275351.121

3. Altemose N, Logsdon GA, Bzikadze A V., Sidhwani P, Langley SA, Caldas G V., et al. Complete genomic and epigenetic maps of human centromeres. Science (1979). 2022;376. https://doi.org/10.1126/science.abl4178

4. Miga KH, Koren S, Rhie A, Vollger MR, Gershman A, Bzikadze A, et al. Telomere-to-telomere assembly of a complete human X chromosome. Nature. Springer US; 2020;585:79–84. https://doi.org/10.1038/s41586-020-2547-7

5. Naish M, Alonge M, Wlodzimierz P, Tock AJ, Abramson BW, Schmücker A, et al. The genetic and epigenetic landscape of the *Arabidopsis* centromeres. Science (1979). 2021;374. https://doi.org/10.1126/science.abi7489

6. Dernburg AF. Here, There, and Everywhere: Kinetochore Function on Holocentric Chromosomes. J Cell Biol. 2001;153:33–8.

7. Guerra M, Cabral G, Cuacos M, González-García M, González-Sánchez M, Vega J, et al. Neocentrics and Holokinetics (Holocentrics): Chromosomes out of the Centromeric Rules. Cytogenet Genome Res. 2010;129:82–96. https://doi.org/10.1159/000314289

8. Melters DP, Paliulis L V., Korf IF, Chan SWL. Holocentric chromosomes: Convergent evolution, meiotic adaptations, and genomic analysis. Chromosome Research. 2012;20:579–93. https://doi.org/10.1007/s10577-012-9292-1

9. Gassmann R, Rechtsteiner A, Yuen KW, Muroyama A, Egelhofer T, Gaydos L, et al. An inverse relationship to germline transcription defines centromeric chromatin in *C. elegans*. Nature. 2012;484:534–7. https://doi.org/10.1038/nature10973

10. Steiner FA, Henikoff S. Holocentromeres are dispersed point centromeres localized at transcription factor hotspots. Elife. 2014;3:1–22. https://doi.org/10.7554/eLife.02025

11. Marques A, Ribeiro T, Neumann P, Macas J, Novák P, Schubert V, et al. Holocentromeres in *Rhynchospora* are associated with genome-wide centromere-specific repeat arrays interspersed among euchromatin. Proc Natl Acad Sci U S A. 2015;112:13633–8. https://doi.org/10.1073/pnas.1512255112

12. Drinnenberg IA, deYoung D, Henikoff S, Malik HS ingh. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. Elife. 2014;3:1–19. https://doi.org/10.7554/eLife.03676

13. Hofstatter PG, Thangavel G, Lux T, Neumann P, Vondrak T, Novak P, et al. Repeat-based holocentromeres influence genome architecture and karyotype evolution. Cell. 2022;185:3153-3168.e18. https://doi.org/10.1016/j.cell.2022.06.045

14. Neumann P, Navrátilová A, Schroeder-Reiter E, Koblížková A, Steinbauerová V, Chocholová E, et al. Stretching the Rules: Monocentric Chromosomes with Multiple Centromere Domains. PLoS Genet. 2012;8:e1002777. https://doi.org/10.1371/journal.pgen.1002777

15. Neumann P, Pavlíková Z, Koblížková A, Fuková I, Jedličková V, Novák P, et al. Centromeres Off the Hook: Massive Changes in Centromere Size and Structure Following Duplication of CenH3 Gene in *Fabeae* Species. Mol Biol Evol. 2015;32:1862–79. https://doi.org/10.1093/molbev/msv070

16. Gržan T, Despot-Slade E, Meštrović N, Plohl M, Mravinac B. CenH3 distribution reveals extended centromeres in the model beetle *Tribolium castaneum*. Malik HS, editor. PLoS Genet. 2020;16:e1009115. https://doi.org/10.1371/journal.pgen.1009115

17. Huang YC, Lee CC, Kao CY, Chang NC, Lin CC, Shoemaker D, et al. Evolution of long centromeres in fire ants. BMC Evol Biol. BMC Evolutionary Biology; 2016;16:1–14. https://doi.org/10.1186/s12862-016-0760-7

18. Jones JT, Haegeman A, Danchin EGJ, Gaur HS, Helder J, Jones MGK, et al. Top 10 plant-parasitic nematodes in molecular plant pathology. Mol Plant Pathol. 2013;14:946–61. https://doi.org/10.1111/mpp.12057

19. Álvarez-Ortega S, Brito JA, Subbotin SA. Multigene phylogeny of root-knot nematodes and molecular characterization of *Meloidogyne nataliei* Golden, Rose & Bird, 1981 (Nematoda: Tylenchida). Sci Rep. Nature Publishing Group; 2019;9. https://doi.org/10.1038/s41598-019-48195-0

20. Triantaphyllou AC. Environmental Sex Differentiation of Nematodes in Relation to Pest Management. Annu Rev Phytopathol. 1973;11:441–62. https://doi.org/10.1146/annurev.py.11.090173.002301

21. Lunt DH. Genetic tests of ancient asexuality in Root Knot Nematodes reveal recent hybrid origins. BMC Evol Biol. 2008;16:1–16. https://doi.org/10.1186/1471-2148-8-194

22. Blanc-Mathieu R, Perfus-Barbeoch L, Aury JM, Da Rocha M, Gouzy J, Sallet E, et al. Hybridization and polyploidy enable genomic plasticity without sex in the most devastating plant-parasitic nematodes. PLoS Genet. 2017;13:e1006777. https://doi.org/10.1371/journal.pgen.1006777

23. Dai D, Xie C, Zhou Y, Bo D, Zhang S, Mao S, et al. Unzipped chromosome-level genomes reveal allopolyploid nematode origin pattern as unreduced gamete hybridization. Nat Commun. Nature Research; 2023;14:7156. https://doi.org/10.1038/s41467-023-42700-w

24. Poullet M, Konigopal H, Rancurel C, Sallaberry M, Lopez-Roques C, Mota APZ, et al. High-fidelity annotated triploid genome of the quarantine root-knot nematode, *Meloidogyne enterolobii*. Sci Data. 2025;12:184. https://doi.org/10.1038/s41597-025-04434-w

25. Castagnone-Sereno P, Danchin EGJ. Parasitic success without sex – the nematode experience. J Evol Biol. 2014;27:1323–33. https://doi.org/10.1111/jeb.12337

26. Kozlowski DKL, Hassanaly-Goulamhoussen R, Da Rocha M, Koutsovoulos GD, Bailly-Bechet M, Danchin EGJ. Movements of transposable elements contribute to the genomic plasticity and species diversification in an asexually reproducing nematode pest. Evol Appl. John Wiley and Sons Inc; 2021;14:1844–66. https://doi.org/10.1111/eva.13246

27. Despot-Slade E, Mravinac B, Širca S, Castagnone-Sereno P, Plohl M, Meštrović N. The Centromere Histone Is Conserved and Associated with Tandem Repeats Sharing a Conserved 19-bp Box in the Holocentromere of *Meloidogyne* Nematodes. Mol Biol Evol. 2021;38:1943–65. https://doi.org/10.1093/molbev/msaa336

28. Ravi M, Kwong PN, Menorca RMG, Valencia JT, Ramahi JS, Stewart JL, et al. The Rapidly Evolving Centromere-Specific Histone Has Stringent Functional Requirements in *Arabidopsis thaliana*. Genetics. 2010;186:461–71. https://doi.org/10.1534/genetics.110.120337

29. Masonbrink RE, Gallagher JP, Jareczek JJ, Renny-Byfield S, Grover CE, Gong L, et al. CenH3 evolution in diploids and polyploids of three angiosperm genera. BMC Plant Biol. BioMed Central Ltd.; 2014;14:383. https://doi.org/10.1186/s12870-014-0383-3

30. Maheshwari S, Ishii T, Brown CT, Houben A, Comai L. Centromere location in *Arabidopsis* is unaltered by extreme divergence in CENH3 protein sequence. Genome Res. Cold Spring Harbor Laboratory Press; 2017;27:471–8. https://doi.org/10.1101/gr.214619.116

31. Bert W, Karssen G, Helder J. Phylogeny and Evolution of Nematodes. In: Jones J, Gheysen G, Fenoll C, editors. Genomics and Molecular Genetics of Plant-Nematode Interactions [Internet]. Dordrecht: Springer Netherlands; 2011. p. 45–59. https://doi.org/10.1007/978-94-007-0434-3_3

32. García LE, Sánchez-Puerta MV. Comparative and Evolutionary Analyses of *Meloidogyne* spp. Based on Mitochondrial Genome Sequences. Jones J, editor. PLoS One. 2015;10:e0121142. https://doi.org/10.1371/journal.pone.0121142

33. Zotta Mota AP, Koutsovoulos GD, Perfus-Barbeoch L, Despot-Slade E, Labadie K, Aury J-M, et al. Unzipped genome assemblies of polyploid root-knot nematodes reveal unusual and clade-specific telomeric repeats. Nat Commun. 2024;15:773. https://doi.org/10.1038/s41467-024-44914-y

34. Bali S, Hu S, Vining K, Brown C, Mojtahedi H, Zhang L, et al. Nematode genome announcement: Draft genome of *Meloidogyne chitwoodi*, an economically important pest of

potato in the pacific northwest. Molecular Plant-Microbe Interactions. American Phytopathological Society; 2021;34. https://doi.org/10.1094/MPMI-12-20-0337-A

35. Winter MR, Taranto AP, Yimer HZ, Coomer Blundell A, Siddique S, Williamson VM, et al. Phased chromosome-scale genome assembly of an asexual, allopolyploid root-knot nematode reveals complex subgenomic structure. James K, editor. PLoS One. Public Library of Science; 2024;19:e0302506. https://doi.org/10.1371/journal.pone.0302506

36. Shakya P, Maulana MI, Danchin EGJ, Voogt ML, van de Ruitenbeek SJS, Gimeno J, et al. High-resolution genome assembly and linkage mapping in *Meloidogyne hapla* reveal non-canonical telomere repeats and recombination hotspots associated with effector proteins. PLoS Pathog. Public Library of Science; 2025;21. https://doi.org/10.1371/journal.ppat.1013706

37. Abad P, Gouzy J, Aury J-M, Castagnone-Sereno P, Danchin EGJ, Deleury E, et al. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. Nat Biotechnol. 2008;26:909–15. https://doi.org/10.1038/nbt.1482

38. Szitenberg A, Salazar-Jaramillo L, Blok VC, Laetsch DR, Joseph S, Williamson VM, et al. Comparative Genomics of Apomictic Root-Knot Nematodes: Hybridization, Ploidy, and Dynamic Genome Change. Genome Biol Evol. 2017;9:2844–61. https://doi.org/10.1093/gbe/evx201

39. Castagnone-Sereno P, Danchin EGJ, Perfus-Barbeoch L, Abad P. Diversity and Evolution of Root-Knot Nematodes, Genus *Meloidogyne*: New Insights from the Genomic Era. Annu Rev Phytopathol. 2013;51:203–20. https://doi.org/10.1146/annurev-phyto-082712-102300

40. Lunt DH, Kumar S, Koutsovoulos G, Blaxter ML. The complex hybrid origins of the root knot nematodes revealed through comparative genomics. PeerJ. PeerJ Inc.; 2014;2:e356. https://doi.org/10.7717/peerj.356

41. Baril T, Galbraith J, Hayward A. Earl Grey: A Fully Automated User-Friendly Transposable Element Annotation and Analysis Pipeline. Arkhipova I, editor. Mol Biol Evol. Oxford University Press; 2024;41. https://doi.org/10.1093/molbev/msae068

42. Barro-Trastoy D, Köhler C. Helitrons: genomic parasites that generate developmental novelties. Trends in Genetics. Elsevier Ltd; 2024;40:437–48. https://doi.org/10.1016/j.tig.2024.02.002

43. Li Z, Gilbert C, Peng H, Pollet N. Discovery of numerous novel Helitron -like elements in eukaryote genomes using HELIANO. Nucleic Acids Res. 2024;52:e79–e79. https://doi.org/10.1093/nar/gkae679

44. Mata-Sucre Y, Krátká M, Oliveira L, Neumann P, Macas J, Schubert V, et al. Repeat-based holocentromeres of the woodrush *Luzula sylvatica* reveal insights into the evolutionary transition to holocentricity. Nat Commun. 2024;15:9565. https://doi.org/10.1038/s41467-024-53944-5

45. Kuo Y-T, Câmara AS, Schubert V, Neumann P, Macas J, Melzer M, et al. Holocentromeres can consist of merely a few megabase-sized satellite arrays. Nat Commun. 2023;14:3502. https://doi.org/10.1038/s41467-023-38922-7

46. Kuo Y, Schubert V, Marques A, Schubert I, Houben A. Centromere diversity: How different repeat-based holocentromeres may have evolved. BioEssays. 2024;46:1–8. https://doi.org/10.1002/bies.202400013

47. Henikoff S, Ahmad K, Malik HS. The centromere paradox: Stable inheritance with rapidly evolving DNA. Science (1979). 2001;293:1098–102. https://doi.org/10.1126/science.1062939

48. Malik HS. The Centromere-Drive Hypothesis : A Simple Basis for Centromere Complexity. Prog Mol Subcell Biol. 2009;48:33–52. https://doi.org/10.1007/978-3-642-00182-6

49. Oliveira L, Neumann P, Jang TS, Klemme S, Schubert V, Koblížková A, et al. Mitotic Spindle Attachment to the Holocentric Chromosomes of *Cuscuta europaea* Does Not Correlate With the Distribution of CENH3 Chromatin. Front Plant Sci. 2020;10:1–11. https://doi.org/10.3389/fpls.2019.01799

50. Logsdon GA, Rozanski AN, Ryabov F, Potapova T, Shepelev VA, Catacchio CR, et al. The variation and evolution of complete human centromeres. Nature. Springer US; 2024;629:136–45. https://doi.org/10.1038/s41586-024-07278-3

51. Bodor DL, Mata JF, Sergeev M, David AF, Salimian KJ, Panchenko T, et al. The quantitative architecture of centromeric chromatin. Elife. eLife Sciences Publications Ltd; 2014;2014. https://doi.org/10.7554/eLife.02137

52. Coffman VC, Wu P, Parthun MR, Wu JQ. CENP-A exceeds microtubule attachment sites in centromere clusters of both budding and fission yeast. Journal of Cell Biology. 2011;195:563–72. https://doi.org/10.1083/jcb.201106078

53. Kapitonov V V., Jurka J. Rolling-circle transposons in eukaryotes. Proceedings of the National Academy of Sciences. 2001;98:8714–9. https://doi.org/10.1073/pnas.151269298

54. Omole AD, Czuppon P. Maintenance of long-term transposable element activity through regulation by nonautonomous elements. Genetics. Oxford University Press; 2025;229:1–13. https://doi.org/10.1093/genetics/iyae209

55. Wang Z, Zhao G, Yang Q, Gao L, Liu C, Ru Z, et al. *Helitron* and CACTA DNA transposons actively reshape the common wheat - AK58 genome. Genomics. Elsevier Inc.; 2022;114:110288. https://doi.org/10.1016/j.ygeno.2022.110288

56. Xiong W, Dooner HK, Du C. Rolling-circle amplification of centromeric *Helitrons* in plant genomes. The Plant Journal. 2016;88:1038–45. https://doi.org/10.1111/tpj.13314

57. Castagnone-Sereno P. Genetic variability and adaptive evolution in parthenogenetic root-knot nematodes. Heredity (Edinb). 2006;96:282–9. https://doi.org/10.1038/sj.hdy.6800794

58. Opperman CH, Bird DM, Williamson VM, Rokhsar DS, Burke M, Cohn J, et al. Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. Proc Natl Acad Sci U S A. 2008;105:14802–7. https://doi.org/10.1073/pnas.0805946105

59. Marques A, Schubert V, Houben A, Pedrosa-Harand A. Restructuring of holocentric centromeres during meiosis in the plant *Rhynchospora pubera*. Genetics. Genetics; 2016;204:555–68. https://doi.org/10.1534/genetics.116.191213

60. Triantaphyllou AC. Polyploidy and reproductive patterns in the root-knot nematode *Meloidogyne hapla*. J Morphol. 1966;118:403–13. https://doi.org/10.1002/jmor.1051180308

61. Van Der Beek JG, Folkertsma R, Poleij LM, Van Koert PHG, Bakker J. Molecular evidence that *Meloidogyne hapla*, *M. chitwoodi*, and *M. fallax* are distinct biological entities. Fundam Appl Nematol. 1997;20:513–20.

62. Zotta Mota AP, Danchin EGJ. *Meloidogyne incognita* genome assembly. Recherche Data Gouv. 2023. https://doi.org/10.57745/RAZ8JS

63. Zotta Mota AP, Danchin E. *Meloidogyne javanica* genome assembly. Recherche Data Gouv. 2023. https://doi.org/10.57745/50YQVI

64. Zotta Mota AP, Danchin E. *Meloidogyne arenaria* genome assembly. Recherche Data Gouv. 2023; https://doi.org/10.57745/VBCSTD

65. Poullet M, Konigopal H, Rancurel C, Sallaberry M, Lopez-Roques C, Zotta Mota AP, et al. *Meloidogyne enterolobii* genomic data [Internet]. Datasets. NCBI Genbank. 2024. https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_963681835.1/

66. Bali S, Hu S, Vining K, Brown C, Mojtahedi H, Zhang L, et al. *Meloidogyne chitwoodi* genomic data [Internet]. Datasets, NCBI GenBank. 2020. https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_015183035.1/

67. Gertz EM, Yu Y-K, Agarwala R, Schäffer AA, Altschul SF. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. BMC Biol. 2006;4:41. https://doi.org/10.1186/1741-7007-4-41

68. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. https://doi.org/10.1093/molbev/mst010

69. Kolde R. pheatmap: Pretty Heatmaps [Internet]. 2019. https://cran.r-project.org/package=pheatmap

70. Volarić M, Meštrović N, Despot-Slade E. SatXplor—a comprehensive pipeline for satellite DNA analyses in complex genome assemblies. Brief Bioinform. 2025;26:bbae660. https://doi.org/10.1093/bib/bbae660

71. Pagès H. BSgenome: Software infrastructure for efficient representation of full genomes and their SNPs [Internet]. 2024. https://bioconductor.org/packages/BSgenome

72. Volarić M, Despot-Slade E, Veseljak D, Mravinac B, Meštrović N. Long-read genome assembly of the insect model organism *Tribolium castaneum* reveals spread of satellite DNA in gene-rich regions by recurrent burst events. Genome Res. 2024;1–17. https://doi.org/10.1101/gr.279225.124

73. Paradis E, Schliep K. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics. Oxford University Press; 2019;35:526–8. https://doi.org/10.1093/bioinformatics/bty633

74. Lê S, Josse J, Husson F. FactoMineR : An R Package for Multivariate Analysis. J Stat Softw. 2008;25:1–18. https://doi.org/10.18637/jss.v025.i01

75. Ester M, Kriegel H-P, Sander J, Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining. AAAI Press; 1996. p. 226–231.

76. Spedicato GA. Discrete Time Markov Chains with R. R J [Internet]. 2017;9:84–104. https://doi.org/10.32614/RJ-2017-036

77. Csardi G, Nepusz T. The igraph software package for complex network research [Internet]. InterJournal. 2006. p. 1695. https://igraph.org

78. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016. https://ggplot2.tidyverse.org

79. Despot-Slade E, Volarić M, Veseljak D, Horvat L, Mota APZ, Poullet M, et al. Supplementary Material for centromere analysis in *Meloidogyne* species. Figshare. 2025. https://doi.org/10.6084/m9.figshare.28715339

80. Shakya P, Maulana MI, Danchin EGJ, Voogt ML, van de Ruitenbeek SJS, Gimeno J, et al. *Meloidogyne hapla* genomic data [Internet]. Datasets, NCBI GenBank. 2025. https://www.ncbi.nlm.nih.gov/datasets/genome/GCA_051171035.1/