RESEARCH ARTICLE

High-resolution genome assembly and linkage mapping in *Meloidogyne hapla* reveal non-canonical telomere repeats and recombination hotspots associated with effector proteins

Pallavi Shakya[1], Muhammad I. Maulana[2,3], Etienne G. J. Danchin[4], M. Laurens Voogt[2], Stefan J. S. van de Ruitenbeek[2], Jacinta Gimeno[1], Adam P. Taranto[1,5], Alison C. Blundell[1], Evelin Despot-Slade[6], Nevenka Meštrović[6], Ana Zotta Mota[4], Dadong Dai[7], Valerie M. Williamson[1], Mark G. Sterken[2]*, Shahid Siddique[7]*

1 Department of Plant Pathology, University of California, Davis, California, United States of America, **2** Laboratory of Nematology, Droevendaalsesteeg PB, Wageningen University and Research, Wageningen, The Netherlands, **3** Department of Plant Protection, Faculty of Agriculture, Universitas Gadjah Mada, Yogyakarta, Indonesia, **4** INRAE, Université Côte d'Azur, CNRS, Sophia-Antipolis, France, **5** School of BioSciences, The University of Melbourne, Melbourne, Victoria, Australia, **6** Ruđer Bošković Institute, Zagreb, Croatia, **7** Department of Entomology and Nematology, University of California, Davis, California, United States of America

* mark.sterken@wur.nl (MGS); ssiddique@ucdavis.edu (SS)

Abstract

Root-knot nematodes (*Meloidogyne spp.*) are among the most destructive agricultural pests that cause significant yield losses across a wide range of crops. *Meloidogyne hapla* is a valuable model for studying root-knot nematodes due to its parasitic diversity, small diploid genome, and a reproductive strategy that facilitates genetic analysis. Here, we report the most contiguous genome assembly to date for any plant-parasitic nematode built using PacBio HiFi, Oxford Nanopore, Illumina, and Hi-C sequencing. Genetic linkage analysis of F2 populations derived from crosses between *M. hapla* strains validated the assembly but also revealed anomalies indicating chromosome structure differences between parental isolates such as fissions, fusions, and rearrangements. Strikingly, we identified sharply delimited zones with extraordinarily high recombination on most chromosomes. Notably, several of these high recombination zones were significantly enriched for genes encoding secreted proteins, many of which contribute to parasitism. These findings suggest that meiotic recombination facilitates effector diversification and offer insight into how these parasites diversify their effector protein repertoire to change or expand their extraordinary host range. We further report the discovery of a novel 16-nucleotide tandem repeat and lack of canonical telomere repeats at chromosome ends. The localization of this 16-nt repeat at chromosome ends highlights a potentially divergent mechanism of chromosome-end maintenance in this nematode group. Overall, our study integrates high-resolution structural genomics, genetic mapping, and functional inference to

uncover links between genome architecture, recombination landscapes, and host–parasite interactions.

Author summary

Root-knot Nematodes (RKNs) are major agricultural pests, causing widespread crop losses globally. Among them, *Meloidogyne hapla* is a particularly valuable research model due to its compact diploid genome and reproductive flexibility that enables genetic studies. In this study, we generated a complete and ac-curate chromosome-scale genome assembly of *M. hapla* using state-of-the-art sequencing technologies. We used genetic maps to validate the accuracy of the genome and found that some *M. hapla* strains have structural differences in their genome wherein the chromosomes have fused or broken apart. We also discovered that *M. hapla* has zones with extraordinarily high recombination rates on most of its chromosomes and these zones are enriched in predicted secreted peptides that may contribute to parasitism. This suggests that recombination may help these nematodes to evolve new ways of overcoming plant defenses. Finally, we identified an unexpected 16-nucleotide long repeat at chromosomal ends instead of a typical telomere sequence hinting to an alternative strategy of chromosomal maintenance. Overall, our study provides the fundamental genom-ic resources on *M. hapla* and reveals how the genome structure and recombina-tion have shaped the parasitism in this organism.

Background

Nematodes are among the most abundant and diverse animal phyla, containing an estimated 1–10 million species [1]. This remarkable diversity is reflected in their eco-logical niches—ranging from free-living forms in soil and water to parasitic species infecting plants and animals. Among the approximately 4,100 species identified as plant parasites, species in the genus *Meloidogyne* (commonly known as root-knot nematodes, or RKNs) stand out as the most economically damaging [2,3]. The four most damaging species, *M. arenaria*, *M. incognita*, *M. javanica*, and *M. hapla*, pos-sess an extraordinary host range spanning a diverse array of crop species [4]. Three of these species (*M. arenaria*, *M. incognita*, *M. javanica*) are closely related (*Meloid-ogyne* Clade I) and are globally distributed in tropical and subtropical regions. In con-trast, *Meloidogyne hapla* (Clade II) is generally found in more temperate climates and parasitizes a broad range of host plants, although isolates differ in their host range, pathogenicity, and behavior [5–8].

The RKN lifecycle generally spans about a month and encompasses six distinct stages consisting of embryo, four juvenile stages (J1-J4) and an adult stage [9,10]. As obligate sedentary endoparasites, RKNs spend most of their life cycle feeding from a permanent site within the root vascular system. These feeding sites are

characterized by nematode-induced multinucleated giant cells that act as nutrient sinks for the nematodes, and by distinctive root galls or "knots", a hallmark of RKN infestation [11,12]. How RKNs establish these specialized feeding sites in such a broad range of plant species, spanning monocots, dicots, annuals, and perennials is a question of both scientific and practical interest. Nematode genes responsible for establishing feeding sites or contributing to differences in host range are largely unknown. However, several proteins secreted by plant-parasitic nematodes have been demonstrated to contribute to parasitism and are commonly referred to as effectors [13–16]. These effectors include genes likely acquired via horizontal transfers as well as many encoding pioneer proteins with no known protein motifs.

Meloidogyne spp. exhibit a wide range of karyotypes and reproductive mechanisms [17]. Whereas most Clade I species reproduce asexually without meiosis and carry genomes with various degrees of polyploidy, most isolates of *M. hapla* are diploid and reproduce by facultative meiotic parthenogenesis [5,18]. Sexual reproduction occurs when females are fertilized by migratory males, and sperm-oocyte fusion occurs to generate offspring [18,19]. However, cytological studies have shown that, in the absence of males, sister chromatids of meiosis II are rejoined to restore diploidy asexually [18,19]. This reproductive mechanism has facilitated controlled crosses between strains and the production of F2 lines. Molecular marker analysis of F2 lines from a cross of two different strains showed these lines are largely homozygous for sequence polymorphisms and thus resemble recombinant inbred lines (RILs) [19]. Additionally, these F2 lines have been used to produce a marker-based genetic map and to identify genetic loci affecting interactions with host and/or behavior [20–23].

Our study focuses on *M. hapla* for its genetic tractability and diploid genome, which make it a favorable model and therefore reference organism for RKN genetics and genomics. A previous draft genome sequence of the inbred *M. hapla* strain VW9 has been produced, and a DNA-based linkage map was generated based on segregation of polymorphisms in RIL-like F2 lines from a cross between strains VW9 and VW8 [24]. However, this genome assembly is highly fragmented, limiting both the localization of genes responsible for phenotypic traits and synteny analyses with other RKNs. Recent advances in sequencing technologies have led to highly contiguous assemblies for several diploid and polyploid *Meloidogyne* species [25–28]. In these studies, the canonical nematode telomeric repeats $(TTAGGC)_n$ were not detected at chromosome ends or anywhere else in the assemblies. Instead, in three Clade I species, *M. incognita*, *M. arenaria* and *M. javanica,* long arrays of species-specific complex tandem repeats were found, mostly enriched at one scaffold end [26,27]. Furthermore, telomere associated proteins such as telomerase and shelterin complexes that are evolutionarily conserved in other nematode clades were not identified suggesting an alternative mechanism for chromosome end maintenance may be at play in RKN [27].

Here, we generated a *de novo* chromosome-level assembly of diploid *M. hapla* using complementary long-read sequencing strategies. We characterized another non-canonical repeat sequences at the end of chromosome-length scaffolds. Additionally, we utilized data from previously generated F2 lines to compare scaffold structure with genetic linkage groups. This analysis revealed a recombination landscape characterized by extraordinary recombination rates mostly on chromosome arms and provided evidence for differences in chromosome structure/behavior between isolates. To the best of our knowledge, this is the first chromosome-scale RKN genome assembly validated by genetic linkage analysis. Furthermore, we examine the chromosomal distribution of genes encoding secreted proteins (effector candidates) and provide evidence for their enrichment in the high recombination zones.

Results

Chromosome-level genome assembly of *Meloidogyne hapla* strain VW9

We produced a high-quality, chromosome-level genome assembly for *Meloidogyne hapla* strain VW9, the same strain used in a previous assembly by Opperman et al. [24] (**PRJNA29083**). We used a combination of PacBio HiFi sequencing (56x coverage), Oxford Nanopore Technologies (ONT) long-read sequencing (143x coverage), Illumina short read sequencing (138x coverage), and Hi-C Chromatin conformation capture (263x coverage). Using Hifiasm with the HiFi,

ONT, and Hi-C datasets, we produced an initial assembly with 36 contigs. This assembly was assessed for contaminants with Blobtools, where all 36 contigs were assigned the taxonomy Nematoda (S1 Fig). This assembly was further polished with the Illumina reads and subsequently scaffolded using Juicer and 3D-DNA with the Hi-C data (Fig 1A). Additionally, the mitochondrial genome was assembled (S2 Fig). The resulting assembly consisted of 16 scaffolds for a total length of 59.2 Mb with an N50 of 3.8 Mb. The assembly size closely matched the 61.6 Mb genome size estimated by k-mer analysis (S3A Fig) and was consistent with flow cytometry data, which estimated the diploid nuclear DNA content of *M. hapla* to be 121 ± 3 Mb, corresponding to a haploid genome size of approximately 60 Mb [29]. Analysis of the Illumina data supported a diploid genome structure for *M. hapla* and almost entirely homozygous (S3B Fig). The high integrity of this assembly was further supported by the Merqury k-mer plot, which indicated that almost all the information present in Hi-Fi reads was captured in the haploid assembly (S3C Fig). Overall, this new *M. hapla* genome assembly represents a significant improvement over the previous version and is the most contiguous so far for a RKN.

In the Hi-C contact map, we observed patterns suggesting physical proximity between scaffold tips of different chromosomes (**Fig 1A**). To assess whether these patterns were an assembly artifact, we checked the quality of raw Hi-C reads and built Hi-C contact maps with different draft assemblies (S4 Fig). This analysis supports that the patterns are due to inter-chromosomal tip proximity.

To define the scaffold ends in the *M. hapla* assembly we analyzed the terminal 24 kb of each scaffold for tandem repeats using Tandem Repeat Finder (TRF) [30]. This analysis revealed arrays of a 16-mer with the consensus sequence (CCCAAGGTTTAAAAGG) at 17 scaffold ends and a variant 16-mer detected near the end of Scaffold 10. This 16-mer repeat was the only significant repeat detected in the terminal regions by TRF. Nine scaffolds (S3, S4, S5, S6, S7, S12, S13, S15, S16) have this tandem repeat at only one end, while four scaffolds (S2, S8, S9, and S14) show the same tandem repeat at both ends (**Fig 1B**). The length profile of these arrays ranges from approximately 1,289 bp, to 8,912 bp. For S10, a variant tandem repeat with consensus TTATAAAGGAAGTGGG is present starting 4 kb from one end. A search for the 16-mer tandem repeat array within scaffolds identified only three arrays at the center of the S1 scaffold (Fig 1B and S1 Table). The internal arrays in S1 form a large palindrome with over 300 copies of the repeat followed by over 300 copies of the complementary sequence. Among the scaffolds, only S11 entirely lacked either repeat array.

In addition, to validate whether these candidate telomeric repeats indeed localize to the chromosome ends, we performed fluorescence in situ hybridization (FISH) analysis on elongated chromosomes of *M. hapla* (**Fig 2**) using the 16-mer repeat found at 17 scaffold ends. Due to the tiny size of the chromosomes and the difficult biological material, we could not obtain complete chromosome complement spreads. Nevertheless, for the intact chromosomes that were discernible, the 16-mer repeats predominantly localized at or near the termini of multiple chromosomes (**Fig 2A and 2B**). Several chromosomes showed signal at only one end, while others displayed no signal (**Fig 2B**). The signal intensity varied between chromosomes, indicating differences in repeat array lengths. Furthermore, we detected a signal internally for at least one chromosome (**Fig 2A**), corroborating the presence of an internal repeat array as observed in S1. Overall, FISH results confirm the 16-mer repeat is enriched predominantly at chromosome ends and most likely constitute specific telomeric sequences in *M. hapla*.

After identifying and validating the primary 16-mer repeat, we searched for additional repeat motifs at scaffold ends using MEME suite [31]. The analysis revealed two distinct motifs: one at the ends of chromosomes 3, 10, 13 and 15, and another at ends of chromosomes 1 and 11 (S5 Fig and S1 Table). These motifs remain to be independently confirmed through FISH validation. Additionally, we searched for these repeats in the raw ONT reads using TelosearchLR [32]. We identified the primary 16-mer repeat as the top-ranked motif. Other motifs were either variants of the 16-mer or non-terminal repeats not enriched at chromosome ends. These results validate the 16-mer as the dominant telomere-associated repeat in *M. hapla* VW9.

Within the nematode order Rhabditida, the availability of improved chromosome-length assemblies led to the identification of seven ancient linkage blocks, known as Nigon elements [33]. These elements corresponded to six
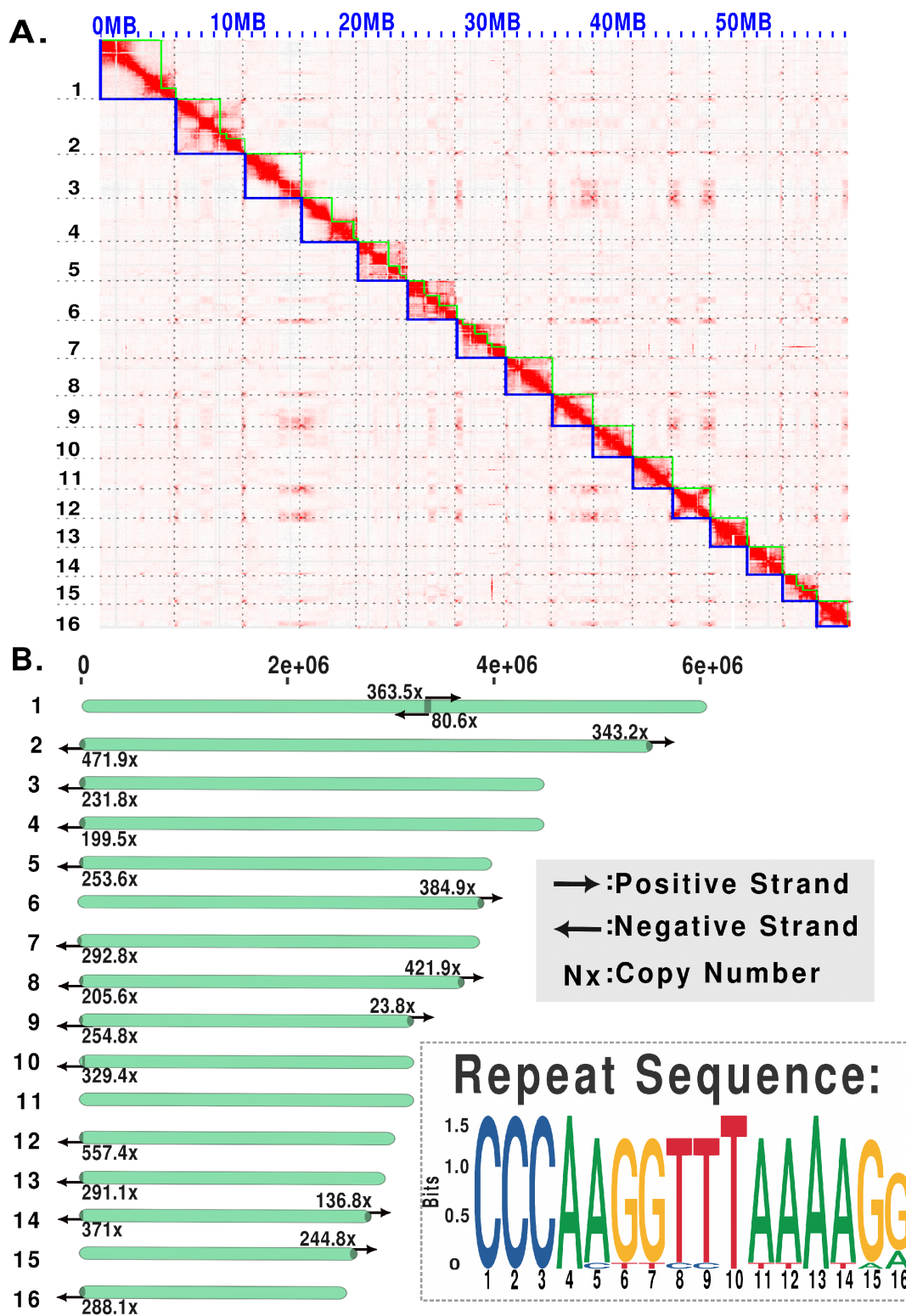
Fig 1. Chromosome-scale genome assembly and repeat structure of *Meloidogyne hapla* strain VW9. A. HiC contact map of *M. hapla* showing 16 chromosome-scale scaffolds. Green lines denote the edges of contigs, and blue lines denote the edges of scaffolds. B. Distribution of the 16-mer repeats across chromosome-scale scaffolds. Each horizontal bar represents a scaffold, with arrows indicating repeat orientation (rightward: positive strand; leftward: negative strand) and numbers showing repeat copy number per scaffold. The repeat at the end of Scaffold 10 is a variant of the others shown.
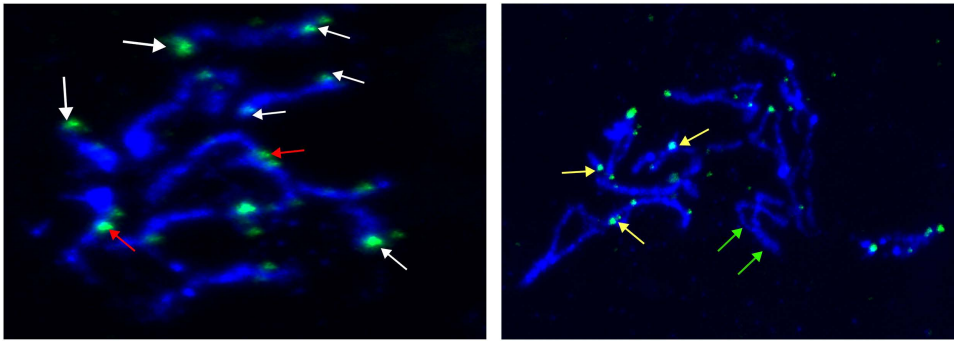
Fig 2. DNA FISH with the 16 bp tandem repeat probe on *M. hapla* chromosomes in different chromosome condensation stages (A and B). Probe is labeled with FITC (green) and chromosomes are counterstained with DAPI (blue). Arrows point to hybridization signals localized at one (white arrows) or both (yellow arrows) chromosome ends. Red arrow indicates a chromosome where tandem repeats appear to be located internally. Green arrows point to chromosomes without hybridization signals. Size bar = 5 μm.

https://doi.org/10.1371/journal.ppat.1013706.g002

autosomes (A, B, C, D, E, N) and a sex chromosome (Nigon X), based on the co-localization of orthologous gene groups. These Nigon elements have been identified in multiple nematode species, including *Caenorhabditis elegans*, *Pristionchus pacificus*, and *Auanema rhodensis* [34–36]. Using the previous *M. hapla* assembly, no clear Nigon assignments were identified, likely due to the highly fragmented nature of the assembly [33]. However, analysis of the current *M. hapla* assembly identified fragments of Nigon elements, with detectable, although highly fragmented, chromosomal segments. Nigon X, the ancestral sex chromosome, was not clearly identifiable (**Fig 3**). This pattern of Nigon elements suggests that extensive chromosomal rearrangements have occurred since *M. hapla* diverged from the ancestral 7-chromosome rhabditid species.

Genetic linkage map supports physical assembly and implies structural variation between nematode isolates

To assess whether our chromosome-length scaffolds correspond to true chromosomes, we produced a scaffold-based linkage map utilizing expressed sequence SNPs (eSNPs) from F2 lines derived from a cross between *M. hapla* strains VW9 and LM [20]. Using the new assembly, we identified 789 SNP alleles flanking crossover events in each of 84 F2 lines (**Fig 4**). The genome-wide recombination rate was high – 14.7 cM/Mb – compared to the 2.9 cM/Mb found in *C. elegans* [37]. The recombination frequency profiles (Marey maps) for twelve scaffolds (S2, S3, S5, S6, S8, S9, S10, S11, S13, S14, S15, S16) are characterized by a high recombination rate in the chromosome arms and low recombination in the chromosome centers or tips (**Fig 5**). This recombination pattern is similar to those described for *C. elegans*, *C. briggsae*, and *P. pacificus* [37,38]. However, the remaining chromosomes showed intriguing deviations from the typical recombination profile. Most notably, S1 showed high recombination through the central region in addition to chromosomal arms, while S4 showed no recombination and S7, S12 and S13 showed low recombination. Intriguingly, segregation correlation in F2 lines suggested that S1 consisted of two linkage groups and that S4 and S12 alleles were on the same linkage group (**S6 Fig**). In addition, S13 showed highly skewed distribution of alleles in F2 lines favoring LM (**Fig 4**).

We used change point analysis to further define the distribution and boundaries between regions with high and low recombination. This revealed sharp differentiation along the scaffolds between high recombination zones (HRZs), where rates ranged from 11 to 139 cM per Mb, and low recombination zones (LRZs) where rates were <5 cM/Mb (**Fig 5** and **S2 Table**).

To test whether scaffold mis assembly was contributing to the idiosyncrasies that we observed in the scaffold-based linkage map, we produced a *de novo* classical genetic linkage map using segregation data from 458 SNPs in 93 RIL-like F2 lines from the previously described cross between *M. hapla* strains VW9 and LM [20]. The resulting map contained 19
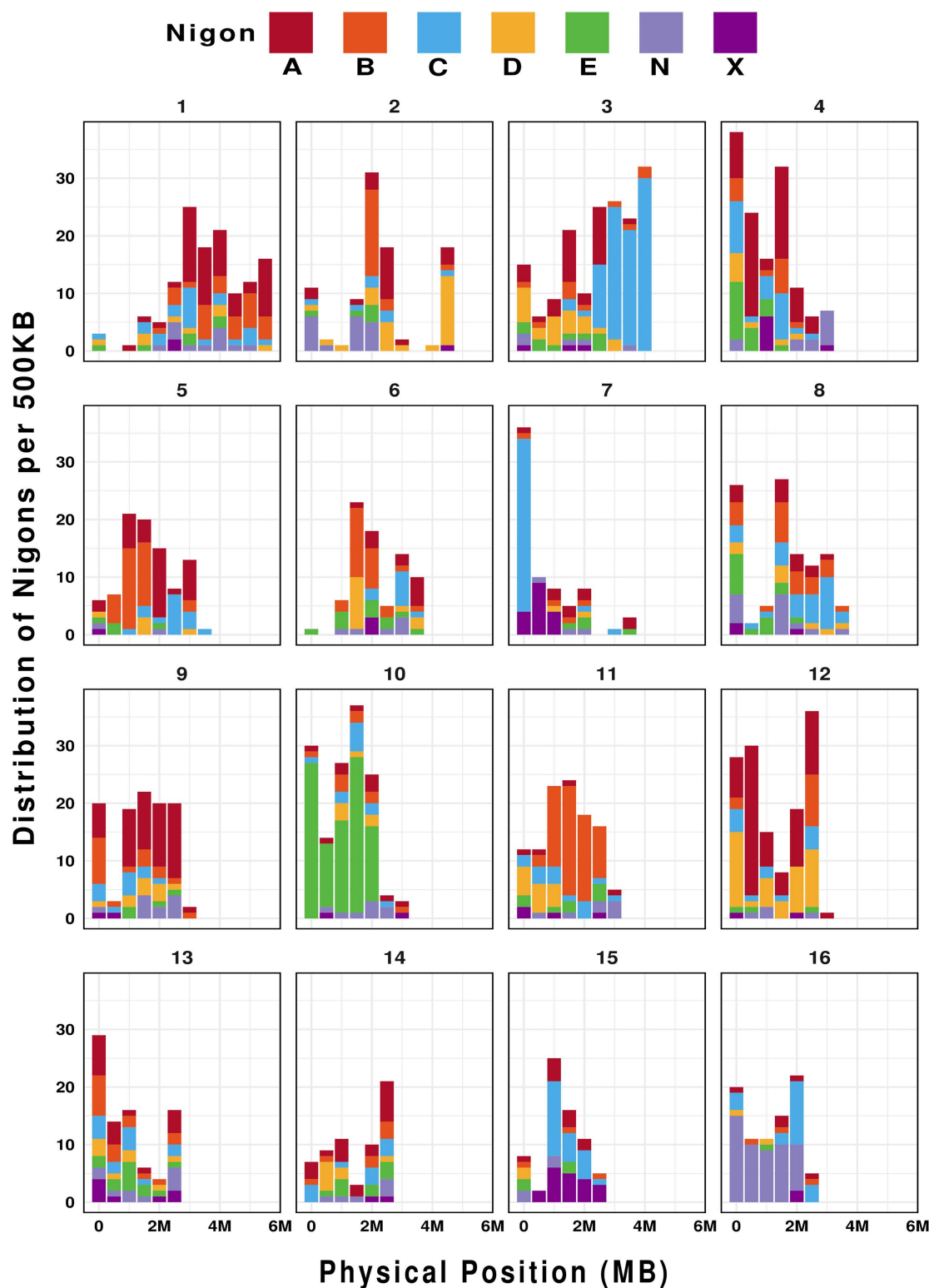
Fig 3. Distribution of Nigon elements along the length of *M. hapla* scaffolds. The X-axis represents the physical length of each scaffold, and the Y-axis represents the Nigon-defining loci per 500 Kb non-overlapping windows. The legend shows the color key for each Nigon Element from A through X.
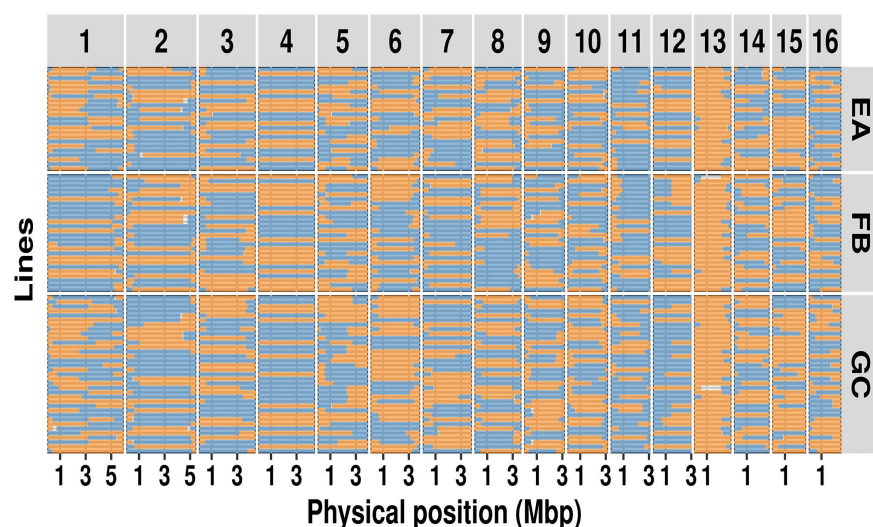
Fig 4. Recombination profile based on chromosome scaffolds. Allele present in each F_2 line on scaffolds 1-16. Colors represent regions with VW9 (blue) and LM (orange) alleles, whereas white areas indicate insufficient data. Panels EA, FB, and GC on the right indicate the female from which the F_2 lines were derived.

https://doi.org/10.1371/journal.ppat.1013706.g004
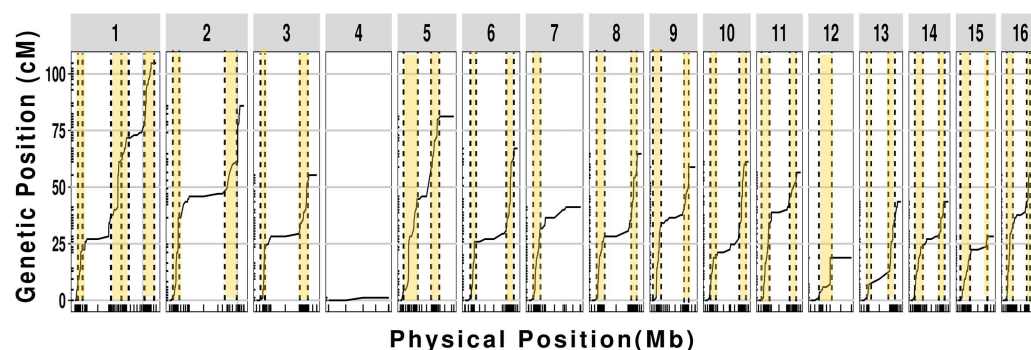
Fig 5. Recombination profile for each scaffold. Tick marks on the x axis indicate physical position on the scaffold and those on the y axis are the corresponding genetic positions based on SNP segregation in F_2 lines. High recombination zones (HRZs) are highlighted in yellow. The number of scaffolds is depicted on top of each box.

https://doi.org/10.1371/journal.ppat.1013706.g005

linkage groups (LGs). We then located the mapped SNPs on scaffolds in the new genome assembly to compare chromosome scaffolds to the genetic linkage groups (S3 Table). For thirteen linkage groups, all SNPs were assigned to a single scaffold, supporting that these scaffolds represented full length chromosomes (Table 1). However, scaffolds S1, S2 and S13 were each divided into two genetic linkage groups. The most striking discrepancy was S1, which formed two linkage groups in the classical map.

As noted above, S1 showed anomalously high recombination in the center as well as on arms. Interestingly, the break between the linkage groups in the VW9xLM cross and the high central recombination occurred in the region of the internal palindrome corresponding to the candidate telomeric repeat (**Fig 1B**). To determine whether this joining was a scaffolding error, we examined nanopore reads and identified 2902 raw ultralong nanopore reads greater than 60 kb spanning the palindromic region of S1. Hence, the existence of a single molecule spanning this region is strongly supported. Additionally,

Table 1. Alignment of *Meloidogyne hapla* VW9 scaffolds with genetic linkage groups.

Scaffold	Genetic Linkage Group	
	VW9 X LM[a]	VW8 X VW9[b]
S1	LG1a, LG1c	LG1
S2	LG2a, LG2c.15	LG2, LG15
S3	LG4	LG4
S4	LG14	LG14, LG17
S5	LG10.12	LG10, LG12
S6	LG2b	LG2
S7	LG6	LG6
S8	LG3	LG3
S9	LG1b	LG1
S10	LG16	LG16
S11	LG5	LG5
S12	LG13.17	LG13
S13	LG9a, LG9b	LG9
S14	LG7	LG7
S15	LG8	LG8
S16	LG11	LG11

[a]LG names are based on those assigned in previous work [20,22].

https://doi.org/10.1371/journal.ppat.1013706.t001

we noted a difference between the recombination patterns in progeny from female FB, one of the three F1 females from which the RIL-like F2 lines derive showed no recombinants in the central region of S1 chromosome whereas progeny from the other two females (EA and GC) showed high apparent segregation in this region.

To further investigate the anomalies between scaffolds and linkage groups, we utilized a published genetic linkage map generated with segregation data of DNA polymorphisms in 183 F2 lines from a cross between *M. hapla* strains VW8 and VW9 [21]. We then located the mapped SNPs onto scaffolds in our new genome assembly. Ten linkage groups corresponded to single scaffolds (**Table 1**). However, even though VW9 was a parent in both crosses (VW9xLM and VW8xVW9), there were differences in the alignment of chromosome scaffolds to linkage groups for the two crosses. Notably, LG1 spanned all SNPs of S1 and S9, predicting it to be a very long linkage group. Also, LG2 contained all markers for scaffolds S2 and S6. Interestingly, the highly skewed marker segregation in favor of LM alleles seen for S13 in the VW9xLM cross was not observed in the VW8xVW9 cross. In addition, in the VW8x9 linkage map, S4 and S12 did not show repressed recombination. Since both genetic crosses used VW9 as one parent but a different second parent, the disparity between the chromosome scaffolds and linkage maps is likely due to structural differences between genomes of the three strains. For example, inversions, translocations or other genomic rearrangements (chromosome fusions, breakage) could result in the distorted segregation patterns that we observed. Other intriguing discrepancies between scaffolds and LGs remain to be explained but would likely require producing chromosome-length sequences of multiple *M. hapla* strains.

Together these results suggest that the differences between linkage maps and the new chromosomal *M. hapla* assembly are due to differences between the genome structure of VW9 and other parental strains. Therefore, the current assembly likely corresponds to a largely accurate representation of the chromosomal DNA molecules in strain VW9.

Characterization of predicted secreted protein gene repertoire of *M. hapla*. To annotate the newly generated VW9 genome, we utilized Iso-Seq data from mixed developmental stages, including eggs, J2, and females. This yielded 4,117,943 reads, which were clustered into 240,273 high-quality isoforms with an N50 of 2,018 bp. Additionally, RNA-seq

data were obtained from nematode-infected roots of *Solanum lycopersicum cv* Moneymaker and *S. pimpinellifolium* cv. G1.1554 at five post-inoculation time points (5, 7, 10, 12, and 14 days after inoculation). The integration of these two datasets using BRAKER3 produced a high-quality structural annotation comprising 11,229 protein-coding genes. This gene count is lower than the 14,700 genes reported in the 2008 assembly [24], which was likely due to the reliance in the previous study on ab-initio predictions, which can overestimate gene numbers. The new annotation as assessed with BUSCO using Eukaryota_odb10 [39] showed 88.2% completeness for coding sequences with only 2.4% fragmentation (S4 Table**).** Similarly, the annotation assessed using BUSCO using Nematoda_odb10 database showed 63.8% completeness for coding sequences with only 2.7% fragmentation. This represents a notable improvement over the 2008 assembly, which showed 78.5% completeness and 10.2% fragmentation with eukaryota_odb10, and 50.5% completeness with 4.1% fragmentation with nematoda_odb10 databases.

To identify genes likely involved in parasitism (candidate effectors), we screened the 11,229 predicted proteins for the presence of secretion signal sequences (SignalP6.0; [40] and the absence of transmembrane domains (DeepTMHMM; [41]. This resulted in the identification of 1,258 genes encoding predicted secreted proteins (PSPs). We then used the full set of 1,258 genes to perform ortholog searches across 71 nematode species using OrthoFinder. Of the 71 nematode species, 33 were plant parasitic nematodes and the rest were either free living or animal parasitic nematodes. This analysis revealed that 1,172 of these PSPs had orthologs in at least one other species while 86 were unique to *M. hapla;* 675 *M. hapla* PSPs were conserved across all 71 nematode species; 401 were present only in RKNs; 3 conserved only with other PPNs; and 1 was shared with only one other nematode species (Fig 6A and 6B and S7 Table). Among the *M. hapla*-specific PSPs, 56 were single-copy genes (singletons) and 30 present in multiple copies.

Of the 1,258 genes encoding PSPs in *M. hapla*, 540 contained known functional domains in Interpro and EggNOG databases, while the remaining 718 did not and are referred to as "pioneer PSPs" (S5 and S6 Tables). For the 1,172 PSPs that had orthologs in other nematodes, 536 contained known domains and 636 were pioneers (Fig 6C and S7 Table). Most of the PSP genes present only in *M. hapla* were pioneers. Of the single-copy PSPs unique to *M. hapla*, 53 had no known functional domains. Three encoded proteins with identifiable features, including a glycine-rich domain, a collagen triple helix repeat, and an SH3 domain. Similarly, among the 30 PSP genes unique to *M. hapla*, but present as multiple copies, only one had a known domain (a protein kinase).

Within the 540 PSP families containing a known functional domain, we identified multiple gene families previously shown to play a role in parasitism in other nematodes (S8 Table). These include several families of CAZymes–31 Glycoside Hydrolases (GH), 17 Glycosyl Transferases (GT), 7 Carbohydrate Esterases (CE), and 12 Pectate Lyases (PL) [42] (S7 Fig and S9 Table). We further verified the annotation of these CAZymes through the dbCAN database [43] (S10 Table). Previous studies suggest that horizontal gene transfers contributed to the acquisition of CAZymes in plant-parasitic nematodes [42]. Using AvP (Alienness vs Predictor) pipeline analysis, we found that 20 CAZymes including 7 GHs, 1 CE and all PLs have evidence of horizontal gene transfer (HGT) (S10 Table). Similarly, other PSPs (Alginate Lyases, Lytic transglycosylases, Fungal chitosanase, Protein Kinase domain containing protein, and Trypsin) also showed evidence of horizontal transfer (S10 Table). Furthermore, consistent with previous phylogenetic studies that suggest pectate lyases expansion through gene duplication during *Meloidogyne* evolution [21,24], our analyses revealed patterns of clade separation and subsequent expan-sion across all four CAZymes families (S8–S10 Figs).

In our structural annotation pipeline, default filtering criteria were employed, which typically exclude very short open reading frames. In addition, the RNA-seq and Iso-seq datasets used in the annotation process were not optimal for identifying short transcripts, as the sequencing read lengths were often insufficient to detect and reconstruct these smaller genes with confidence. As a result, proteins shorter than 66 amino acids were likely not annotated, potentially omitting small, secreted proteins that may function as plant peptide mimics. To more comprehensively identify such candidates, we reanalyzed the 2008 genome annotation and identified 19 small PSPs. Among these, five had functional annotations, including a C2H2 zinc finger domain protein, a fructosyltransferase, a phosphotransferase system component, a
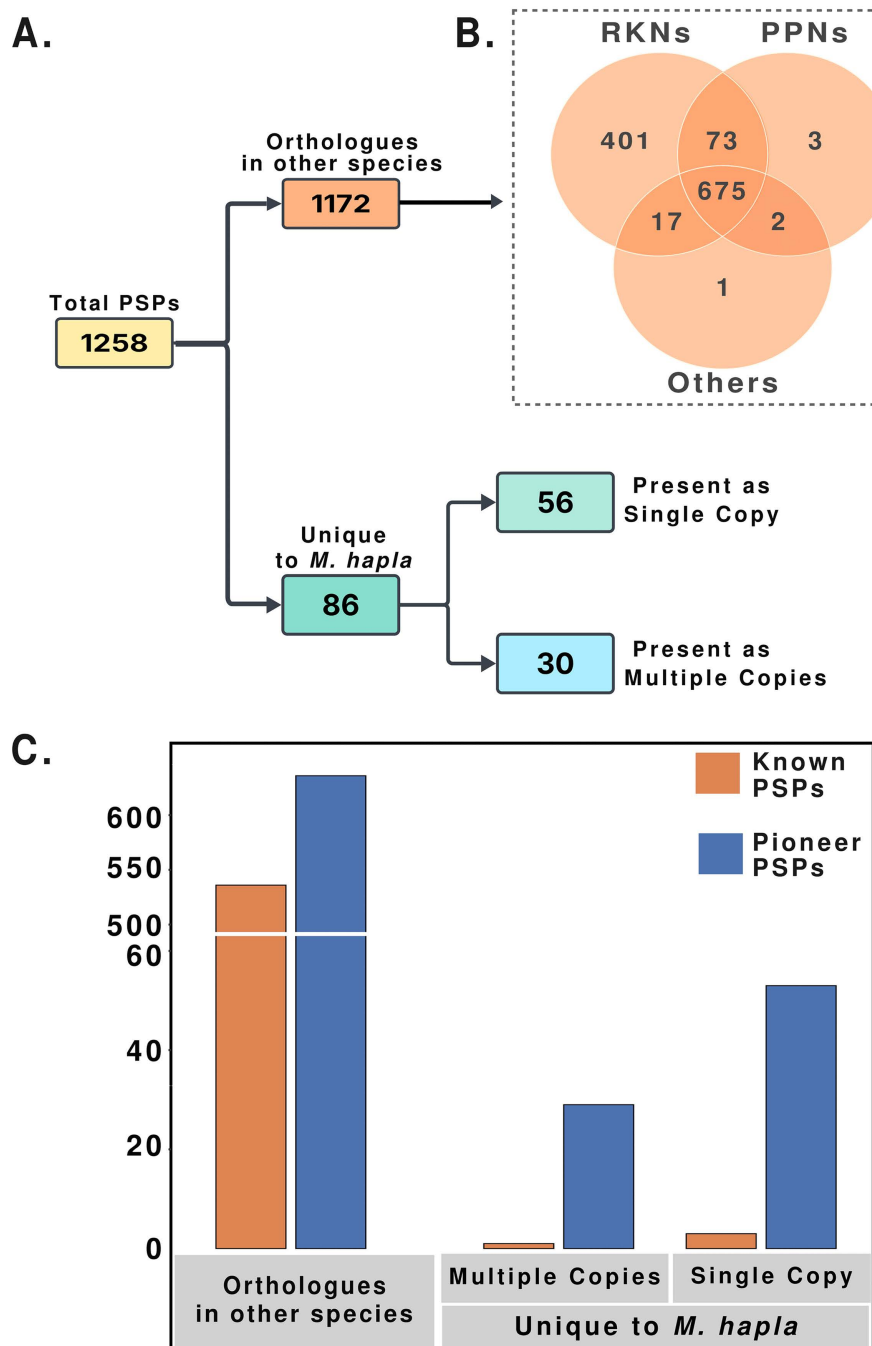
Fig 6. Classification of *M. hapla* PSPs. A. Total PSPs were characterized by the presence or absence of orthologs in a set of 71 nematode species. The identification of orthologs was done using Orthofinder. B. Venn diagram shows the PSPs that are shared across RKNs, PPNs and other nematode species. C. The distribution of PSPs into Known and Pioneer PSPs is shown for PSPs with orthologs in other nematode species and for those unique to M. hapla. (i.e., those unique to *M. hapla* having single or multiple copies.

https://doi.org/10.1371/journal.ppat.1013706.g006

bifunctional nuclease-like protein, and a tyrosinase copper-binding protein. The remaining 14 proteins lacked known functional annotations (S11 Table). When mapped onto the updated *M. hapla* genome assembly, these PSPs were found to be distributed across multiple genomic regions (S11 Table).

Next, we conducted a BLAST search for known plant peptide mimics, including C-terminally Encoded Peptides (CEPs), Inflorescence Deficient in Abscission (IDA) and Rapid Alkalinization Factors (RALF)—all previously described to modulate parasitism related responses in plants [44–46]. We found that all 12 CEPs were located on S13, while IDA1 and IDA2 were located on S7 and S9, respectively. The three RALFs were located on S1, S5 and S10 (S12 Table). No PSY peptide genes were detected, consistent with prior findings that these genes are restricted to Clade I *Meloidogyne* species [47].

Genes encoding predicted secreted proteins (PSPs) are enriched in high recombination zones

To examine the genomic distribution of protein-coding genes, we compared their empirical cumulative distribution to a theoretical uniform distribution. Overall, genes appear to be evenly distributed along the chromosomes (S11 Fig). However, certain scaffolds (SF1, SF4, SF6, SF7, and SF13) show noticeable,—though often modest—deviations from this uniform pattern. In these regions, uneven gene distribution is sometimes, but not always, associated with unusual recombination patterns. For instance, S1 separates into two genetic linkage groups in the VW9 × LM cross, while S4 and S12 appear to be physically linked. In contrast, S7 contains only a single region with a markedly higher recombination rate. This pattern of uniform gene distribution is consistent with observations in other holocentric species, such as spider mite (*Tetranychus urticae*) and *C. elegans*, where genes are generally evenly distributed along the chromosomes [48,49].

Given the rapid evolution of effectors under selection pressure in various plant pathogens [50,51], we assessed whether PSP genes cluster in HRZs or LRZs (S2 Table). A hypergeometric test for enrichment found that PSP genes were significantly enriched in HRZs ($p = 2.5*10^{-8}$; S13 Table). A scaffold-specific analysis found PSP enrichment on S2, S3, S14, and S16 while other scaffolds did not show significant differences in PSP distribution between HRZs and LRZs (Fig 7 and S14 Table). In addition, annotated PSP families–Cysteine-rich secretory protein, Papain cysteine proteases, Aspartyl proteases, C-type lectin, peptidases, CAZymes, Catalases, Peroxidases, Astacins, Lipases and SXP/RAL2 family–were predominantly localized within HRZs (Fig 8). One notable case is HRZ of S3, which exhibits the highest recombination rate in the genome at 139 cM/Mb (Fig 9A and S2 Table). This HRZ spans 340 KB, and contains 60 genes, 23 (~38%) of which encode PSPs (Figs 9B and 8C). Among these, 16 are pioneer PSPs, while the remaining include known PSPs such as lysozyme, carboxylesterases, glycoside hydrolases, Galectin and Calycin (Lipocalin) (Fig 9C and 9D).

In *C. elegans*, evolutionarily conserved genes are enriched in low recombination regions of chromosomes [52]. To assess whether *M. hapla* shows a similar pattern, we identified 5610 *C. elegans* orthologs from the Orthofinder analysis. Of these, 4195 map to the LRZ and 1415 to the HRZ. An enrichment analysis revealed no overrepresentation of *C. elegans* orthologs in HRZs; however, these orthologs were significantly enriched LRZs ($p = 1.9*10^{-5}$; S13 Table).

Discussion

Like other *Meloidogyne* species, *M. hapla* lacks telomerase and canonical telomere-associated proteins [27], suggesting that it relies on an alternative mechanism to maintain chromosome ends. In this study, we generated a chromosome-scale genome assembly for the *M. hapla* strain VW9. This assembly consists of 16 scaffolds, corresponding to the previously reported chromosome number for this strain. One notable finding of this study is the identification of tandem repeats consisting of a consensus 16-mer sequence (CCCAAGGTTTAAAAGG) at one or both ends of most scaffolds, and intriguingly, also in an interior region of Scaffold 1. These repeats show no recognizable sequence similarity to the more complex terminal repeats recently identified in Clade I *Meloidogyne* species [26,27] or to the terminal repeats in the diploid species *M. graminicola* [26]. This lack of homology suggests that chromosome end repeat motifs have significantly diversified among RKN species. Notably, most of the 16-mer repeats in our assembly occur at one end of the scaffold, a finding further supported by FISH analysis, which detects these repeats at a single chromosome end. At the ends of the scaffolds

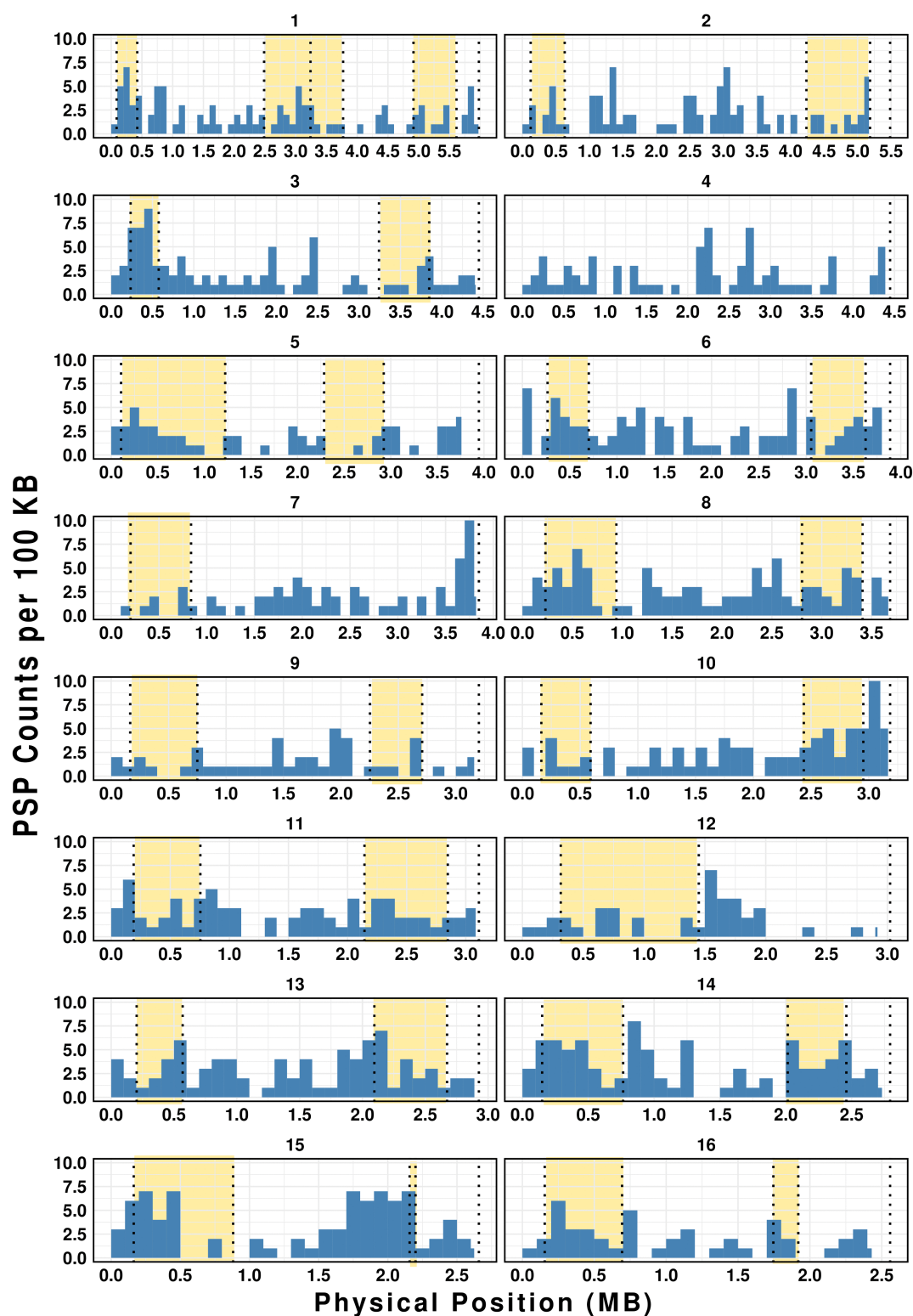**Fig 7. Distribution of PSP genes in the High Recombination Zones (HRZs).** Each of the 16 scaffolds is divided into 100 kb bins (X-axis). Y-axis represents the number of PSPs per bin. HRZ regions are highlighted in yellow. Genome-wide enrichment analysis shows significant enrichment of PSPs in the HRZs (Hypergeometric test, *p-value = 2.5*10^-8*).

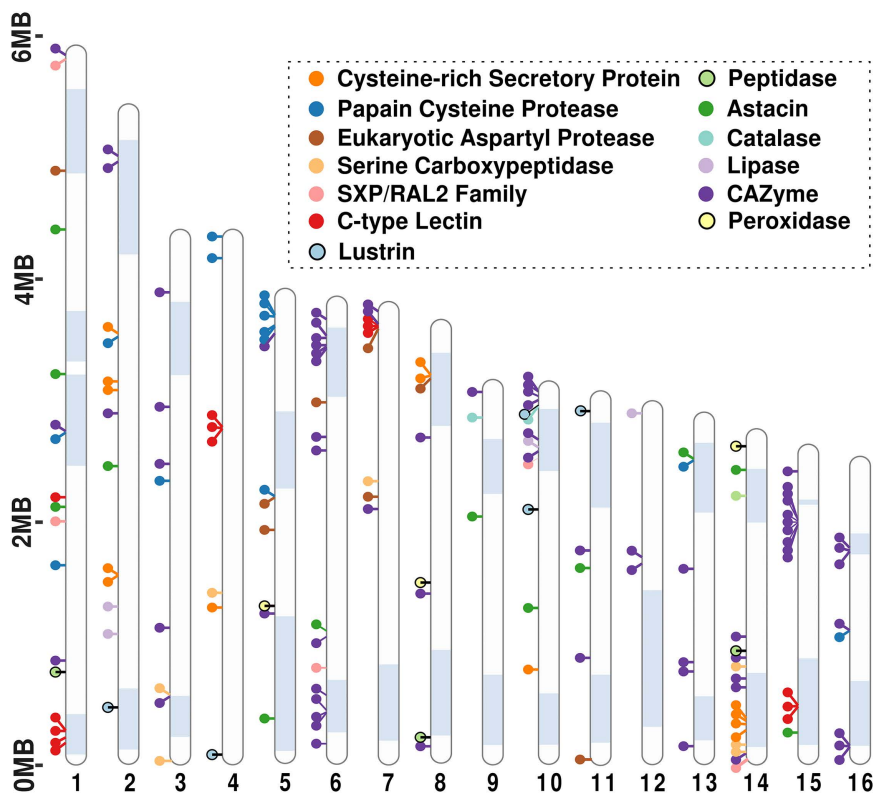https://doi.org/10.1371/journal.ppat.1013706.g007

Fig 8. Distribution of PSPs with conserved domains along the genome of *M. hapla*: Ideogram of PSPs with known functional domains along the scaffolds of *M. hapla* genome. Each circle represents the PSPs, and they are pinned with respect to their positions on the scaffolds. The highlighted blue regions denote the HRZs.

https://doi.org/10.1371/journal.ppat.1013706.g008

where the consensus 16-mer repeat was not detected, we identified two distinct repeat sequences. However, their chromosomal locations have not yet been verified by FISH analysis. Additionally, we cannot rule out the possibility that our assembly is incomplete and that some terminal sequences remain unresolved. The presence of repeats only at one end of the chromosome is reminiscent of *C. elegans* pairing centers, which facilitate homologous pairing and crossover during meiosis [53]. Whether the terminal regions of *M. hapla* chromosomes play a role in telomere maintenance, meiotic pairing, or both remains to be investigated.

M. hapla is capable of facultative sexual reproduction and therefore, we can perform controlled crosses and track allele segregation in RIL-like F2 lines. These data provide a valuable resource for studying chromosome structure and parasitism-related adaptations in plant-parasitic nematodes. The recombination patterns of individual F2 lines along the majority of scaffolds in *M. hapla* resemble those observed in *C. elegans*, with crossovers primarily localized to the arms. However, the overall recombination rate in *M. hapla* is roughly five times higher than that in *C. elegans*, whose meiotic crossovers are restricted to one per chromosome per meiosis [37,54,55]. The *M. hapla* genetic map suggests that the one crossover per-chromosome limit may also apply to this species. Therefore, its higher recombination rate could be primarily due to its higher chromosome number and smaller genome size. Recombination in *M. hapla* is localized to more narrowly defined regions with sharp boundaries, resulting in very high local recombination rates. Intriguingly, electron microscopy studies have described the presence of synaptonemal complexes and distinctive electron-dense "recombination nodules" in oocytes of a meiotic race of *M. hapla* [56].
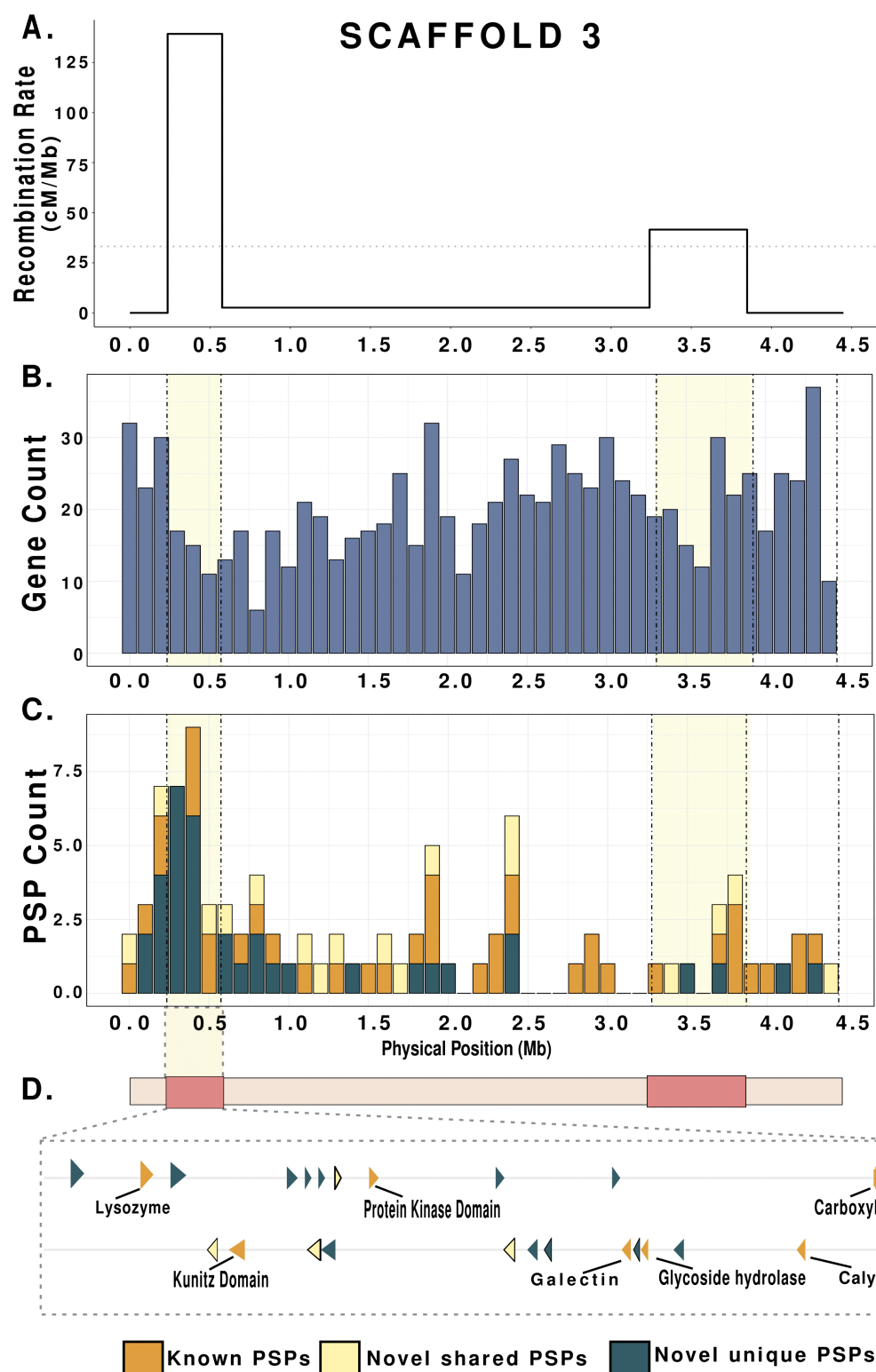
Fig 9. High recombination zones and gene distribution in Scaffold 3. A. Recombination rate (cM/Mb) of Scaffold 3 along its physical position (Mb) showing sharply defined HRZs. The dotted horizontal line indicates the average recombination rate for the whole chromosome. B. Histogram of Gene Count per 100 Kb along Scaffold 3. The highlighted regions denote HRZs. C. Stacked bar graph of PSP Count per 100 Kb along the physical positions of scaffold 3 where the highlighted regions denote HRZs. Each color represents functionally annotated PSPs, pioneer PSPs shared with other *Meloidogyne*

species and pioneer PSPs that are unique to *M. hapla*. D. Highlights of the PSP genes found in HRZ of Scaffold 3. They are colored according to the ones shown in C. The arrows represent positive and negative strands. The known PSPs are labelled with their respective functional annotations.

Comparing the genome assembly scaffolds for *M. hapla* strain VW9 and the genetic maps derived from F$_2$ crosses (VW9×LM and VW8×VW9) revealed differences in genome rearrangements among the three nematode strains (**Table 1**). The parental strains were derived from geographically distinct field isolates: VW9 from California (U.S.A), VW8 from the Netherlands, and LM from France [6,23]. The reduced recombination observed in S4, S7, and S12 in the VW9xLM cross (**Figs 4 and 5**) is typical of crosses between parents with genomic differences, such as translocations and inversions. S1 exhibited particularly intriguing segregation patterns. Both our assembly and ultra-long sequencing reads support the notion that S1 is a single chromosome. However, F2 progeny from the VW9×LM cross displayed high recombination rates in the central region of this chromosome for progeny from two of three F1 females (females EA and GC), whereas progeny from the third female (female FB) exhibited no recombination (**Fig 4**). This finding suggests that S1 may exist as a single chromosome in female FB but as two separate chromosome-like entities in females EA and GC. Notably, the classical linkage map from the VW8×VW9 cross indicates that scaffold S1 is linked to another scaffold (S9), forming a single chromosome. Such variability aligns with earlier global surveys that identified differences in chromosome number and modes of reproduction among *M. hapla* field isolates [17]. Indeed, a cross between females with 15 chromosomes and males with 17 chromosomes produced viable progeny possessing more than 15 chromosomes, suggesting that the chromosome number is fluid within this species [18]. Holocentric chromosomes, such as those found in Nematoda, might be particularly tolerant of fragmentation and fusion events compared to monocentric chromosomes. Additionally, a non-canonical mechanism for chromosome end maintenance might further facilitate this genomic plasticity.

Some, but not all, segregation anomalies that we observed in F2 lines are consistent with inversions, translocations, or other differences in chromosome structure between the parental lines. Intriguingly, for F2 lines of the VW9xLM cross, alleles on Scaffold 13 were strongly skewed toward strain LM (**Fig 4**), but no such bias was observed in VW8xVW9 F2 lines. This strongly biased segregation pattern matches the patterns for toxin-antidote elements observed in multiple species, including the nematode species *C. elegans* and *P. pacificus* [57–60]. The occurrence of such elements hampers overall reproductive success and strongly favors the presence of the toxin-antidote allele in the offspring. In this case, the data are consistent with the presence of a toxin-antidote element in strain LM. Further analysis and additional genetic crosses should provide insight into other observed segregation anomalies.

In our annotation analysis, we focused on PSP genes, as they likely include most of the genes involved in parasitism. Most PSP genes in *M. hapla* had orthologs in other *Meloidogyne* species (**Fig 6**). Additionally, most PSP genes that encode proteins with known domains including all genes with signatures of acquisition via HGT are shared across RKN species. This pattern suggests that the HGT events that introduced these genes occurred before the speciation of these nematodes. Furthermore, over half of the PSP genes that *M. hapla* shares with other RKN species have no identified functional motifs, and nearly all PSP genes that are unique to *M. hapla* are pioneers. Perhaps these unique pioneer PSP genes originated via *de novo* gene birth or represent orphan genes derived from unknown mechanisms [61].

The availability of genetic segregation data provides a unique opportunity to investigate the relationship between recombination rates and the genomic distribution of parasitism genes in meiotic RKN species. In some plant pathogenic fungi, effector genes are often enriched in chromosomal regions associated with high recombination rates. For example, recombination hotspots harbor effector genes in the fungal pathogen *Blumeria graminis* [62]. Similarly, in *Fusarium graminearum* and *Zymoseptoria tritici*, effector genes are enriched in chromosome regions near subtelomeres or on accessory chromosomes; these regions are characterized by higher recombination rates and are described as having a "two-speed" genome architecture [63,64]. Our genome-wide analysis of PSP gene distribution found that these genes are enriched in HRZs (**Fig 7**). By contrast, conserved orthologs shared with C. elegans were enriched in LRZs. Thus, enrichment in HRZs appears to be specific to PSP genes rather than a general feature of gene localization in *M. hapla*.

Double-strand DNA breaks produced during meiosis have been shown to promote ge-nomic variation in humans [65]. In line with this finding, we suggest that the en-richment of PSP genes in HRZs may facilitate diversification, enabling this pest to evade host recognition or expand host range. Interestingly, Clade I *Meloidogyne* species, which reproduce asexually and lack meiosis, and therefore meiotic recombination, also exhibit broad host ranges and can evolve to avoid host resistance. These ameiotic species are of hybrid origin and carry two divergent subgenomes [26,29,66], sug-gesting that their ability to acquire new specificity may be due to interactions between the subge-nomes [66]. Thus, the diploid, meiotic *M. hapla* and the asexual Clade I spe-cies likely achieve adaptability through fundamentally different genomic mecha-nisms. Under-standing how recombination and genome architecture drive PSP evolution could provide new insights into host–parasite co-evolution and support the development of control strategies for nematode infestations. Future studies will focus on comparative genomics and functional analysis to identify the evolutionary processes underlying PSP diversifica-tion in these important plant pathogens.

Materials and methods

High molecular weight DNA extraction

Meloidogyne hapla strain VW9 [6] was propagated on tomato (*Solanum lycopersicum*) cultivar VFNT cherry in a UC Davis Greenhouse facility. Nematode eggs (~1 million) were collected from roots then cleaned by sucrose floatation method and flash frozen in liquid nitrogen. High molecular weight (HMW) gDNA extraction was carried out at UC Davis Genome Center. Two ml of lysis buffer containing 100 mM NaCl, 10 mM Tris-HCl pH 8.0, 25 mM EDTA, 0.5% (w/v) SDS and 100 µg/ml Proteinase K was added to the tube containing frozen eggs. The samples were mixed with gentle pipetting and homog-enized at room temperature overnight. The lysate was then treated with 20 µg/ml RNAse at 37 °C for 30 minutes. The lysate was cleaned with equal volumes of phenol/chloroform using phase lock gels (Quantabio Cat # 2302830). The DNA was precipitated from cleaned lysate by adding 0.4X volume of 5M ammonium acetate and 3X volume of ice-cold ethanol. The DNA pellet was washed with 70% ethanol twice and resuspended in an elution buffer (10mM Tris, pH 8.0). The purity of gDNA was assessed using NanoDrop spectrophotometer and 260/280 ratio of 2.0 and 260/230 of 2.29 were observed. For PacBio HiFi, DNA yield was 8 µg as quantified by Qubit 2.0 Fluorometer (Thermo Fisher Scientific, MA). The integrity of the HMW gDNA was verified on a Femto pulse system (Agilent Technologies, CA) where 68% of the DNA was found in fragments above 100 Kb.

PacBio HiFi sequencing for genome assembly

The HiFi SMRTbell library was constructed using the SMRTbell prep kit 3.0 (Pacific Biosciences, Menlo Park, CA, Cat. #102-182-700) according to the manufacturer's instructions. HMW gDNA was sheared to a target DNA size distribution between 15–20 kb using Diagenode's Megaruptor 3 system (Diagenode, Belgium; cat. B06010003). The sheared gDNA was concentrated using 1X of SMRTbell cleanup beads provided in the SMRTbell prep kit 3.0 for the repair and a-tailing incubation at 37°C for 30 minutes and 65°C for 5 minutes, followed by ligation of overhang adapters at 20°C for 30 min-utes, cleanup using 1X SMRTbell cleanup beads, and nuclease treatment at 37°C for 15 minutes. The SMRTbell library was size selected using 3.1X of 35% v/v diluted AMPure PB beads (Pacific Biosciences, Menlo Park, CA; Cat. #100-265-900) to progressively remove SMRTbell templates <5kb. The 15 – 20 kb average HiFi SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA) using one 8M SMRT cell (Pacific Biosciences, Menlo Park, CA; Cat #101-389-001), Sequel II sequencing chemistry 2.0, and 30-hour movies on a PacBio Sequel II sequencer. This way, we generated 232,679 sequences with an average length of 14,391 base pairs.

Nanopore sequencing for genome assembly

For Nanopore sequencing, the sequencing libraries were prepared from 1.5µg of high molecular weight DNA using a ligation sequencing kit SQK-LSK114 (Oxford Nanopore Technologies, Oxford, UK). The manufacturer's library preparation

protocol was followed apart from extended incubation times for DNA damage repair, end repair, ligation and bead elution. For sequencing, the PromethION device P24 was used. Thirty fmol of the final library from the sample was first loaded on the PromethION flow cell R10.4.1 (Oxford Nanopore Technologies, Oxford, UK) and run was set up using PromethION MinKNOW 22.12.5 for 17 hours. Data was base called live during sequencing with super-high accuracy mode using ONT-guppy-for-promethion 6.4.6. This way, we generated 584,625 reads with an average length of 27,317 base pairs.

RNA extraction for Iso-Seq sequencing

Total RNA was extracted from three life stages of *M. hapla*: eggs, newly hatched juveniles, and female nematodes dissected from tomato roots 21 days post infection. RNAeasy mini kit was used as per the manufacturer's instruction for RNA extraction (Qiagen, catalog nr. 74106). To eliminate genomic DNA from RNA samples, TURBO DNAase (Life Technologies, AM1907) treatment was administered. The concentration and purity of the extracted RNA were determined using Nanodrop One Microvolume UV-Vis Spectrophotometer, and RNA integrity and quality were evaluated using bioanalyzer.

Iso-Seq sequencing for genome annotation

cDNA and SMRTbell library were constructed using SMRTbell prep kit 3.0 (PN 102-396-000 REV02 APR2022) according to manufacturer's instructions. The cDNA was synthesized using NEBNext Single Cell/Low Input cDNA Synthesis & Amplification Module (New England Biolabs, Ipswich, MA, Cat. #E6421L). After fifteen cycles of PCR for cDNA amplification, the cDNA was purified using 0.86X SMRTbell cleanup beads. Subsequently, the cDNA libraries derived from each sample were pooled and used for SMRTbell library construction using the SMRTbell prep kit 3.0 (Pacific Biosciences, Menlo Park, CA, Cat. #102-182-700) with the following enzymatic steps: repair and a-tailing, ligation of overhang adapters, and nuclease treatment. The Iso-Seq SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA) using one 8M SMRT cell (Pacific Biosciences, Menlo Park, CA; Cat #101-389-001), Sequel II sequencing chemistry 2.0, and 24-hour movies on a PacBio Sequel II sequencer.

RNA extraction for short-read time-series RNA-seq

Total RNA from the early stage of nematode infection was extracted from galls of tomato-infected roots at five different time points post inoculation. Fourteen days after sowing, tomato plants of *Solanum lycopersicum* cv. Money Maker and *Solanum pimpinellifolium* were inoculated in-vitro with either 0 or 200 surface-sterilized J2s of *M. hapla* strain VW9 (as described in [67]. At 5, 7-, 10-, 12-, and 14-days post inoculation (dpi), all galls present in a single square dish-plate were dissected and flash frozen in liquid nitrogen for RNA-seq. In non-infected plants, to reduce the variation in development, roots segments adjacent to the roots of infected plants were dissected as control. Per time points, two to three technical replicates were taken per plant per nematode treatments.

Liquid nitrogen flash-frozen root galls were smashed and homogenized using Tissuelyzer (Qiagen, Hilden). RNA extraction was performed using the Maxwell 16 LEV-plant RNA kit (Promega) following the manufacturer's instruction. After isolation, RNA concentration and purification was evaluated using NanoPhotometer and spectrophotometer (IMPLEN, CA, United States). RNA integrity checking, library preparations, RNA sequencing and quality filtering was done using BGISEQ-500 at BGI TECH SOLUTIONS (Hongkong) with at least 50 million clean paired-end reads of 150 bp per sample.

HiC sequencing for scaffolding

Chromatin conformation capture data was generated using a Phase Genomics (Seattle, WA) Proximo Hi-C 4.5 kit, which is a commercially available version of Hi-C Protocol [68]. For library preparation, approximately 500 mg tissue was finely chopped and then crosslinked for 15 minutes at room temperature with end-over-end mixing in 1 mL of Proximo

crosslinking solution. Crosslinking reaction was terminated with a quenching solution for 20 minutes at room temperature again with end-over-end mixing. Quenched tissue was rinsed once with 1X Chromatin Rinse Buffer (CRB). The tissue was transferred to a liquid nitrogen-cooled mortar and ground to a fine powder. Powder was resuspended in 700 μL Proximo Lysis Buffer 1 and incubated for 20 minutes with end-over-end mixing. A low-speed spin was used to clear the large debris and the chromatin containing supernatant transferred to a new tube. Following a second higher speed spin, the supernatant was removed and the pellet containing the nuclear fraction of the lysate was washed with 1X CRB. After removing 1X CRB wash, the pellet was resuspended in 100 μL Proximo Lysis Buffer 2 and incubated at 65°C for 15 minutes.

Chromatin was bound to Recovery Beads for 10 minutes at room temperature, placed on a magnetic stand, and washed with 200 μL of 1X CRB. Chromatin bound on beads was resuspended in 150 μL of Proximo fragmentation buffer and 2.5 μL of Proximo fragmentation enzyme added and incubated for 1 hour at 37°C. Reaction was cooled to 12°C and incubated with 2.5 μL of finishing enzyme for 30 minutes. Following the addition of 6 μL of Stop Solution, the beads were washed with 1X CRB and resuspended in 100 μL of Proximo Ligation Buffer supplemented with 5 μL of Proximity ligation enzyme. The reaction was incubated at room temperature for 4 hours with end-over-end mixing. To this volume, 5 μL of Reverse Crosslinks enzyme was added and the reaction incubated at 65°C for 1 hour. After reversing crosslinks, the free DNA was purified with Recovery Beads and Hi-C junctions were bound to streptavidin beads and washed to remove unbound DNA. Washed beads were used to prepare paired-end deep sequencing libraries using the Proximo Library preparation reagents. HiC sequencing was performed on Illumina (Novaseq) which generated over 52 million read pairs.

Illumina DNAseq for polishing

DNA extraction from eggs was performed using the CTAB method as described in [26]. KAPA HyperPrep Kit was used according to the manufacturer's protocol for library preparation. This involved processes such as enzymatic fragmentation, end repair, A-tailing, and Illumina-compatible adapter ligation followed by PCR amplification and cleanup using magnetic beads. The final libraries were quantified using Qubit and assessed for size distribution using an Agilent Bioanalyzer. Consequently, the libraries were sequenced on the Illumina NovaSeq X platform using paired-end 150 bp reads. The library preparation and sequencing were performed by Novogene, USA. This way we generated 27 million read pairs.

Genome profiling of PacBio HiFi reads

The genome profiling of the raw reads generated with PacBio HiFi was conducted using Jellyfish v.2.3.0, Genomescope2 and Smudgeplot0.2.5 [69,70]. K-mer frequencies for the raw reads were determined using Jellyfish with parameters set at "-m 21 -s 1000000000". To deduce the lower and upper coverage range for these reads, "smudgeplot.py cutoff" was utilized. Subsequently, Smudgeplot was generated by using these lower and upper coverage values. Genomescope analysis was performed using default settings with the histogram data generated by Jellyfish as an input.

Generation of draft assemblies

Multiple draft assemblies of the *M. hapla* genome were made using a combination of sequencing reads using various assembly software. For the final draft assembly that was used to make the final chromosome-scale assembly, we used Hifiasm v.2.0.0 [71]. First, we used Canu to correct the ONT reads to ensure the higher accuracy of ONT reads [72]. After this, we incorporated PacBio HiFi, Canu corrected ONT reads, and raw Hi-C reads into Hifiasm to make a haplotype resolved assembly. The primary assembly generated using this method was polished using Illumina sequencing data in five cycles with Pilon v.1.23 using a Snakemake pipeline (See GitHub) [73]. This corrected 222 bases, introduced 154 insertions, and removed 255 bases. Further, we used Meryl v.1.0 and Merqury v.1.3 to evaluate the quality of the genome assembly and Blobtools2 to check for any contamination in the genome [74–76]. For Blobtools, we followed the protocol outlined in this Github repository in which, we used reference proteomes (release:2024_04) from the UNIPROT database

for diamond BLASTx, and BLAST nt database (2.14), in addition to minimap2 v.2.26-r115 and Samtools v.1.17 [77–80]. For BUSCO analysis, we used eukaryota_odb10 and nematoda_odb10 to check the quality of genome, CDS and protein sequences.

Chromosome-scale assembly

We used HiC sequencing of VW9 strain to scaffold the draft genome into a chromosome-scale assembly of the *M. hapla* genome. HiC data were generated by Phase Genomics using their Proximo HiC Kit Protocol with four cutter restriction enzymes: DpnII, DDel, Hinfl and Msel. Accordingly, we obtained the restriction sites and ligation junctions of these enzymes and generated restriction enzyme cut sites using the *generate_site_positions.py* script within the Juicer v.1.6. The information of the restriction sites was input within the Juicer v.1.6, and it was run with the early exit parameter [81]. The HiC reads were aligned with the draft genome assembly with bwa alignment v.0.7.17-r1188 [79]. This was followed by quality control and filtering where low quality and chimeric reads were filtered out. The filtered aligned reads were merged into merged_sort.bam file. After the removal of duplicates from the merged file, merged_nodups.txt file was obtained which consisted of interaction matrices. The information obtained from Juicer was used to produce scaffolded assembly using 3D-DNA v.180114 [82]. 3D-DNA uses the interaction matrices to group the contigs into scaffolds. The initial scaffolded output was used to produce a draft contact map. This contact map was manually curated in Juicebox v.1.11.08 [81]. The manual curation involved minor correction of contig positions according to the contact map. The manually curated assembly was then reviewed and finalized using run-asm-pipeline-post-review.sh script from 3D-DNA pipeline and final HiC contact map was made with the sizes of scaffolds in descending order.

Nigon elements

The output of BUSCO run 'full_table.tsv' was used as an input to visualize the Nigon elements using vis_ALG (https://github.com/pgonzale60/vis_ALG) [33].

Structural and functional annotation

Before structurally annotating the genome, the draft chromosome-scale assembly was masked using RepeatModeler v.2.0.5 and RepeatMasker v4.1.5 [83,84]. For Iso-Seq pre-processing, we used isoseq3 v.4.0.0 which used ppbam v.2.4.99, pbcopper v.2.3.99, pbmm2 v.1.11.9, minimap2 v.2.15, parasail v.2.1.3, boost v.1.77, htslib v.1.17 and zlib v.1.2.13 [85]. First, we clustered the demultiplexed Full-Length-Non-Chimeric (FLNC) reads using Iso-Seq cluster tool. The high-quality clustered and polished reads were then mapped to the genome using pbmm2 align with preset Iso-Seq parameter. After this, the reads were collapsed using isoseq3 collapse with parameters *–do-not-collapse-extra-5exons, –max-5p-diff5' and –max-3p-diff5'*. These parameters prevent collapsing of isoforms that have extra 5' exons and allow up to 5 base pairs difference at both 5' and 3' ends of the sequences ensuring that transcript diversity was preserved.

For structural genome annotation, we used BRAKER3 installed from its long-read branch [86]. In the first run, BRAKER3 incorporated short-read RNA-seq reads to guide the annotation. In the second run, BRAKER3's long-read protocol was applied during which GeneMarkS-T predicted the protein-coding regions from the long-read transcripts. Finally, TSEBRA was used to combine the data from both short-read and long-read evidence to produce a finalized structural annotation. For functional annotation of the predicted protein-coding genes, we used BLAST2GO from OmicsBox v.3.3.2, InterproScan v.5.72-103.0 and EggNOG v.2.1.12 with eggnog database v.5.02 and novel family database v.1.0.1 [87,88].

Genetic map based on eSNPs of *M. hapla*

We produced a scaffold-based recombination profile using expressed sequence SNPs (eSNPs) previously identified between 98 RIL-like F lines of *M. hapla* produced by crossing strains VW9 and LM [20]. The transcriptome data in that

study was produced using RNA extracted from *Medicago truncatula* root galls infected with each RIL-like line and was made available from GEO under accession numbers PRJNA229407 and SRP078507. We used hisat2 (v. 2.2.1) to align the reads to the newly assembled chromosome-level genome. Thereafter reads were sorted, duplicate reads were removed, and the reads were indexed using Samtools (v. 1.14). Quality and alignment statistics were generated using fastqc (v. 0.12.1), Samtools, and mosdepth (v. 0.3.3). We used multiqc (v. 1.14) to inspect the qualities per sample. Next, variants were called using bcftools (v. 1.16) with standard settings except (ploidy = 2, keep_alts = true, min_mq = 20, min_bq = 13, min_idp = 5, max_idp = 500, min_qual = 20, min_ad = 2). This generated a file with 58,114 variants, which were further quality filtered using custom scripts in R (v. 4.2.2).

Next, we filtered the variants based on occurrence over the genotypes. We expressed variant occurrence as alternative allele frequencies (fraction of alt calls over total calls). When filtering for at least one clear alternative allele call we found 15,248 sites with at least one alt call. We then removed sites with more than 30% of samples uncalled, leaving 8,455 variants that were evenly distributed over the chromosomes. We called crossovers based on change-point analysis using the change-point package (v. 2.2.4) in R. Per RIL per chromosome, we first interpolated the uncalled sites based on the median genotype of the adjacent five variants before and after the uncalled variant. Subsequently, we used cpt.mean with the method BinSegto identify crossovers. The crossovers were visually inspected for accuracy.

Guo et al. [20] reported that some of the RIL-like lines appeared to be heterozygous. We confirmed this and therefore removed data from EA18, EA30, FB15, FB22, GC4, GC5, GC6, GC31, and GC45 before further analysis. Furthermore, for five lines the coverage by RNA-seq was very low, requiring too much imputation to be of use (EA16, FB19, 12_GC14, GC36, and GC46). This left us with genetic data for 84 RIL-like lines. The set of markers for these lines was pruned for informative markers (markers at either side of a recombination event). The genotype at the ends of the chromosomes were interpolated from the first and last called marker per RIL. In total, this yielded a set of 789 markers for the 84 RILs (S15 Table).

For recombination analysis we first calculated the genetic distances in centimorgans using a custom R script. We analysed physical versus the genetic distance using change-point package in R, we then identified transition points between the domains for each linkage group by performing a change-point analysis using the binary segmentation method [89].

Integration of genetic linkage maps with chromosome-scale assemblies

To produce a framework genetic linkage map for chromosome-scale assembly, we used segregation data from 458 SNPs in 93 RIL-like F2 lines produced from previously described cross between *M. hapla* strains VW9 and LM [20]. Previously, this segregation data was used to produce a *de novo* genetic map using the previous genome assembly of *M. hapla* obtaining 19 linkage groups [21,24]. To identify the SNP positions in our genome assembly, we extracted sequences spanning 250 bp both upstream and downstream of each of the 458 SNPs using Bedtools v.2.31 [90]. These sequences were subjected to a BLAST against chromosome-scale assembly. We then manually located the positions of the SNPs in the chromosome-scale assembly which was used to align chromosome-length scaffolds to the genetic linkage groups. The same strategy was used to anchor 182 genetic markers developed from an earlier genetic cross in which *M. hapla* strain VW8 as female parent was crossed with VW9 to the current genome assembly [21].

Terminal repeat regions identification

We ran Tandem Repeats Finder (TRF v4.09) on the terminal 24 kb of each chromosome and recovered a single 16-bp tandem repeat at the 3′ ends of 17 scaffolds [30]. Notably, this was the only significant repeat detected in terminal regions. We then inspected scaffold S1 and identified a similar, but more degenerate, copy of the same repeat near its centre. Thereafter, we scanned the entire set of scaffolds to search for the occurrence of this conserved repeat. However, only S1 contained the 16mer repeat at its center.

To identify additional end-associated repeats, we extracted the first and last 2kb of the 16 scaffolds and used these as input to MEME (classic mode, any-number-of-repetitions), employing a first-order background model to account for the low GC content [31]. We requested up to five frequently occurring motifs with lengths of 30–60 nt. After the MEME run, we analysed motif locations and frequencies with MAST (default parameters) and selected the subset of sequences that contained the most frequent motif. We then performed a second MEME run on this subset, allowing retrieval of a single motif and adjusting the motif length according to the first run to refine the model. Sequences corresponding to occurrences of this refined motif were aligned with MAFFT (–auto), an HMM profile was built with hmmbuild (HMMER), and the profile was searched against the whole genome with nhmmer to verify specific enrichment of repeat arrays at scaffold ends. We iterated this workflow on the remaining sequences (32 – n) until no further runs were necessary and all retained motifs were verified to form repeat arrays at scaffold termini.

Finally, we validated chromosome-end repeats directly in raw Oxford Nanopore reads using TeloSearchLR [32]. We analysed base-called FASTQ files with default parameters except for the k-mer limits (10–40 nt) and terminal window sizes, which we set to 1 kb and 5 kb in separate runs. This analysis confirmed the 16-mer as the top-ranked terminal motif; other motifs detected were either variants of the 16-mer or non-terminal repeats not enriched at chromosome ends. The TeloSearchLR results were used to verify the terminal motifs identified by TRF and MEME.

Predicted secreted proteins (PSPs) filtering

We used SignalP v.6.0h on slow sequential and DeepTMHMM v.1.0.42 to predict the signal peptides and transmembrane domains in the protein-coding genes respectively [40,41]. We used custom R scripts employed in R v.4.4.1 and RStudio build 421 and python scripts employed in python v.3.10.8 to filter out the genes with Signal Peptide and without any transmembrane domains. The custom R and python scripts can be obtained in this GitHub repository.

Ortholog analysis

Orthofinder v. 2.5.4 was used to find orthologs of all annotated proteins in *M. hapla* against 72 nematode species with Tardigrade as outgroup [91]. We used the option -msa to produce multiple sequence alignments and phylogenetic trees for all OrthoGroups. We started from a previous OrthoFinder study [27] and we replaced the former *M. hapla* set of predicted proteins by this new one.

Functional annotation of CAZymes

For functional annotation of CAZymes (GH, PL and GT), we used dbCAN3 using DIAMOND, HMMER via CAZY and dbCAN-sub [92] and we further verified the annotations with InterproScan and EggNOG mapper. Carbohydrate Esterases (CEs) were not annotated by dbCAN3, so we used the predictions from BLAST2GO, InterproScan and EggNOGmapper to annotate these proteins.

Horizontal gene transfer (HGT) analysis

For HGT analysis, we used the AvP (Alienness vs Predictor) pipeline with a database constructed from UNIREF90 to select, extract and detect the HGT candidates among the predicted secreted proteins [93].

Phylogenetic trees for CAZymes

For construction of phylogenetic trees, we first used Clustal Omega for the multiple sequence alignment of each CAZyme family with their respective outgroups. Then we used Iqtree to construct the Maximum Likelihood tree under the best-fit model with ultrafast bootstrap that generated 1000 pseudo-replicates to assess support for every internal node. The command used was: iqtree -s input_alignment.phylip -B 1000. Finally, the tree was uploaded to iTOL to make further customizations.

Hypergeometric test

This analysis was done using R (v4.2.2) using tidyverse (v1.3.2) and purr (0.3.4) packages (Github link). The chromosomal domain content ("HRZ" and "LRZ") was based on the genetic linkage map (S2 Table). Gene start and end positions, obtained from genome annotation, were matched to chromosomal domains by ensuring that a gene's entire length falls within a single domain interval. Genes overlapping multiple domains were assigned to each matching domain, and genes not directly overlapping any domain were assigned to the nearest domain based on the minimum distance between gene boundaries and domain boundaries. We then calculated the total number of genes in each domain (N_domain) and total number genes (N_total). After this, we extracted PSPs per domain (k_domain) and total number of PSPs genomewide (k_total). We followed this with an enrichment test, where we tested enrichment of PSPs in each domain using a one-tailed hypergeometric test (R's *phyper* function), with the null hypothesis that PSPs are randomly distributed across domains. For a give domain (LRZ/HRZ), the test conducted using significance threshold of $\alpha = 0.05$

$$P(X \geq k\_domain) = phyper(k\_domain - 1, N\_domain, N\_total - N\_domain, k\_total, lower.tail = FALSE)$$

where, N_domain = number of genes in domain
   k_domain = number of PSPs in domain
   N_total = total number of genes
   k_total = total number of PSPs. Same process was repeated for *C. elegans* orthologs.

Mitochondrial genome assembly and annotation

In the draft assembly, Contig 35l, with a coverage of 1750x was identified as mitochondrial DNA. After manually removing the tandem duplications at the contig ends and rotating it, we annotated this genome using MITOS2 from web based Galaxy v.24.14 [94]. However, we were unable to fully annotate the mitochondrial genome as the program could not detect multiple core and tRNA genes. Hence, to assemble and annotate a linear mitochondrial DNA, we used MitoHiFi [95] and MITOS2. Again, multiple tRNAs were not identified. Finally, we used MitoZ v.3.6 with Illumina DNA sequences [96]. We used Mitoz all commands with clade option Nematoda and genetic code 5 followed by annotation using the command Mitoz annotate. Although this process gave us the core genes in the mitochondrial genome, it also failed to annotate most of the tRNA genes. We then used MITOS2 to reannotate the linear mitochondrial genome [94]. MITOS2 was able to annotate 10 out of 22 tRNAs. tRNAscan-SE [97] also to detect the remaining tRNAs. We then downloaded the mitochondrial tRNA genes from *M. graminicola* from GenBank accession number KJ139963.1 and blasted these sequences against the linear mitochondrial genome of *M. hapla*. This approach identified 17 out of 22 tRNA sequences. Non-canonical tRNA structures (e.g., armless tRNAs) which are common in nematodes [98] and likely contributed to our difficulties in identifying all the expected tRNA sequences. We then employed Circos to produce a diagram of the mitochondrial genome.

DNA fluorescence in situ hybridization (FISH) of the candidate terminal repeat

M. hapla genomic DNA was isolated from 50 collected egg sacs using Qiagen Blood and Tissue kit (Qiagen) according to the manufacturer's procedure with slight modifications. Tissue was homogenized for 30 sec using electric homogenizer and pestle and incubated overnight at 56 °C. Sample was then treated with RNase A for 5 min at RT and final elution was in 100 µL of the elution buffer. For PCR procedures obtained DNA was diluted to concentration of 0,02 ng/µL. To produce probes corresponding to the 16 bp repeat (GTTTAAAAGGCCCAAG) identified at scaffold ends, we extended primers to encompass more than one monomer to avoid primer dimers as much as possible (Mhap_tel_4R AAGATTTAAAAGGC-CCAAGATTTAAAAGAC Mhap_tel_4L CTTTTAAACCTTGGGTCTTTTAAACCTTG). FISH probes were prepared according to the previously described procedure [99] with 61.7 °C as annealing temperature. Prepared probes were cleaned

using QIAquick PCR Purification Kit (Qiagen) and eluted in 50 μL of nuclease-free water. They were tested on 1% agarose gel to check their concentration and length (S12 Fig).

Whole *M. hapla* females were squashed onto slides and different fixatives were tested to optimize chromosome morphology. Ethanol and acetic acid fixative proved to be optimal for both preservation of chromosome morphology and FISH analyses and slide preparation was done as described [100]. The FISH procedure was done according to established protocols [27,99] except for skipping the pretreatment step of specimens in 45% acetic acid as here the slides were fixed in a mixture already containing acetic acid. Similarly, as *M. incognita*, it was not possible to count chromosome numbers exactly and prometaphases/metaphases are rare. Telomeric signals had different intensity from very large to very discrete and consequently it was hard to count the precise number of certain signals. Best evaluation was obtained on elongated chromosomes (prometaphase).

Supporting information

S1 Fig. Blobplot/Contamination plot of draft assembly.
(TIF)

S2 Fig. Mitochondrial genome of Meloidogyne hapla strain VW9.
(TIF)

S3 Fig. Genome Profile of Meloidogyne hapla strain VW9.
(TIF)

S4 Fig. HiC quality check and HiC assembly with ONT reads.
(TIF)

S5 Fig. Three groups of chromosome-end repeats found in M. hapla genome.
(TIF)

S6 Fig. Genetic correlation between and within scaffolds.
(TIF)

S7 Fig. Distribution of CAZymes in the genome of M. hapla.
(TIF)

S8 Fig. Phylogenetic tree and distribution of GH proteins.
(TIF)

S9 Fig. Phylogenetic trees and distribution of CE and PL proteins.
(TIF)

S10 Fig. Phylogenetic tree and distribution of GT proteins.
(TIF)

S11 Fig. Distribution of genes across the genome of M. hapla.
(TIF)

S12 Fig. Agarose gel image of prepared FISH probe with λ DNA for concentration assessment of markers.
(TIF)

S1 Table. Chromosome-end repeats, their positions and copy numbers in each scaffold.
(XLSX)

S2 Table. Recombination Data per Scaffold.
(XLSX)

S3 Table. Marker positions from two crosses: VW8 X VW9 and VW9 X LM in VW9 scaffolds.
(XLSX)

S4 Table. BUSCO scores for genome, coding sequence (CDS) and protein coding genes.
(XLSX)

S5 Table. Functional annotation of Predicted Secreted Peptides (PSPs) using EggNOG and InterProScan.
(XLSX)

S6 Table. Functionally un-annotated PSPs and their positions in M. hapla genome.
(XLSX)

S7 Table. List of orthologs and unique genes in M. hapla (Orthofinder Results).
(XLSX)

S8 Table. Classification of major known PSPs according to their functional annotation.
(XLSX)

S9 Table. Ideogram data used for CAZymes.
(XLSX)

S10 Table. Annotation of CAZymes from dbCAN3 server and their HGT evidence according to AvP tool.
(XLSX)

S11 Table. Small PSPs from [24] assembly and their positions in the new genome.
(XLSX)

S12 Table. Known plant peptide mimics and their distribution in M. hapla genome.
(XLSX)

S13 Table. Hypergeometric test for enrichment of PSPs and C. elegans orthologs in high and low recombination zones of M. hapla scaffolds.
(XLSX)

S14 Table. Hypergeometric test for enrichment of PSPs in each scaffold of M. hapla genome.
(XLSX)

S15 Table. Genotypes of RILs and corresponding markers.
(XLSX)

Acknowledgments

Author contributions

Conceptualization: Pallavi Shakya, Etienne G. J. Danchin, Evelin Despot-Slade, Nevenka Meštrović, Valerie M. Williamson, Mark G. Sterken, Shahid Siddique.

Data curation: Pallavi Shakya, Mark G. Sterken, Shahid Siddique.

Formal analysis: Pallavi Shakya, Etienne G. J. Danchin, M. Laurens Voogt, Stefan J. S. van de Ruitenbeek, Adam P. Taranto, Ana Zotta Mota, Mark G. Sterken.

Funding acquisition: Mark G. Sterken, Shahid Siddique.

Investigation: Pallavi Shakya, Muhammad I. Maulana, Jacinta Gimeno, Alison C. Blundell, Evelin Despot-Slade, Nevenka Meštrović.

Methodology: Pallavi Shakya, Muhammad I. Maulana, Etienne G. J. Danchin, Jacinta Gimeno, Adam P. Taranto, Alison C. Blundell, Evelin Despot-Slade, Ana Zotta Mota, Dadong Dai, Valerie M. Williamson, Mark G. Sterken, Shahid Siddique.

Project administration: Mark G. Sterken, Shahid Siddique.

Resources: Valerie M. Williamson, Shahid Siddique.

Supervision: Valerie M. Williamson, Shahid Siddique.

Validation: Pallavi Shakya, Stefan J. S. van de Ruitenbeek.

Visualization: Pallavi Shakya, Evelin Despot-Slade, Nevenka Meštrović.

Writing – original draft: Pallavi Shakya, Mark G. Sterken, Shahid Siddique.

Writing – review & editing: Pallavi Shakya, Muhammad I. Maulana, Etienne G. J. Danchin, M. Laurens Voogt, Stefan J. S. van de Ruitenbeek, Jacinta Gimeno, Adam P. Taranto, Alison C. Blundell, Evelin Despot-Slade, Nevenka Meštrović, Ana Zotta Mota, Dadong Dai, Valerie M. Williamson, Mark G. Sterken, Shahid Siddique.

References

1. Hugot J-P, Baujard P, Morand S. Biodiversity in helminths and nematodes as a field of study: an overview. Nematol. 2001;3(3):199–208. https://doi.org/10.1163/156854101750413270

2. Bird DM, Williamson VM, Abad P, McCarter J, Danchin EGJ, Castagnone-Sereno P, et al. The genomes of root-knot nematodes. Annu Rev Phytopathol. 2009;47:333–51. https://doi.org/10.1146/annurev-phyto-080508-081839 PMID: 19400640

3. Trudgill DL, Blok VC. Apomictic, polyphagous root-knot nematodes: exceptionally successful and damaging biotrophic root pathogens. Annu Rev Phytopathol. 2001;39:53–77. https://doi.org/10.1146/annurev.phyto.39.1.53 PMID: 11701859

4. 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(9):946–61. https://doi.org/10.1111/mpp.12057 PMID: 23809086

5. van der Beek JG, Los JA, Pijnacker LP. Cytology of parthenogenesis of five Meloidogyne species. Fundament Appl Nematol. 1998;21(4):393–9.

6. Liu QL, Williamson VM. Host-specific pathogenicity and genome differences between inbred strains of Meloidogyne hapla. J Nematol. 2006;38(1):158–64. PMID: 19259441

7. Melakeberhan H, Douches D, Wang W. Interactions of selected potato cultivars and populations of Meloidogyne hapla adapted to the Midwest U.S. Soils. Crop Sci. 2012;52(3):1132–7. https://doi.org/10.2135/cropsci2011.08.0409

8. Mitkowski NA, Abawi GS. Root-knot nematodes. Plant Health Instructor. 2003. https://doi.org/10.1094/phi-i-2003-0917-01

9. Escobar C, Barcala M, Cabrera J, Fenoll C. Chapter One—Overview of Root-Knot Nematodes and Giant Cells. In: Escobar C, Fenoll C, editors. Advances in Botanical Research. Academic Press; 2015. pp. 1–32.

10. Páez AR. Plant-parasitic nematodes and their management: A focus on new nematicides. Nematodes—ecology, adaptation and parasitism. IntechOpen; 2023.

11. Lin C-J, Siddique S. Parasitic nematodes: dietary habits and their implications. Trends Parasitol. 2024;40(3):230–40. https://doi.org/10.1016/j.pt.2023.12.013 PMID: 38262837

12. Rutter WB, Franco J, Gleason C. Rooting out the mechanisms of root-knot nematode-plant interactions. Annu Rev Phytopathol. 2022;60:43–76. https://doi.org/10.1146/annurev-phyto-021621-120943 PMID: 35316614

13. Bali S, Gleason C. Unveiling the diversity: plant parasitic nematode effectors and their plant interaction partners. Mol Plant Microbe Interact. 2024;37(3):179–89. https://doi.org/10.1094/MPMI-09-23-0124-FI PMID: 37870371

14. Molloy B, Shin DS, Long J, Pellegrin C, Senatori B, Vieira P, et al. The origin, deployment, and evolution of a plant-parasitic nematode effectorome. PLoS Pathog. 2024;20(7):e1012395. https://doi.org/10.1371/journal.ppat.1012395 PMID: 39074142

15. Rehman S, Gupta VK, Goyal AK. Identification and functional analysis of secreted effectors from phytoparasitic nematodes. BMC Microbiol. 2016;16:48. https://doi.org/10.1186/s12866-016-0632-8 PMID: 27001199

16. Siddique S, Radakovic ZS, Hiltl C, Pellegrin C, Baum TJ, Beasley H, et al. The genome and lifestage-specific transcriptomes of a plant-parasitic nematode and its host reveal susceptibility genes involved in trans-kingdom synthesis of vitamin B5. Nat Commun. 2022;13(1):6190. https://doi.org/10.1038/s41467-022-33769-w PMID: 36261416

17. Triantaphyllou AC, Hirschmann H. Cytogenetics and morphology in relation to evolution and speciation of plant-parasitic nematodes. Annu Rev Phytopathol. 1980;18(1):333–59. https://doi.org/10.1146/annurev.py.18.090180.002001

18. Triantaphyllou AC. Polyploidy and reproductive patterns in the root-knot nematode Meloidogyne hapla. J Morphol. 1966;118(3):403–13. https://doi.org/10.1002/jmor.1051180308 PMID: 5927453

19. Liu QL, Thomas VP, Williamson VM. Meiotic parthenogenesis in a root-knot nematode results in rapid genomic homozygosity. Genetics. 2007;176(3):1483–90. https://doi.org/10.1534/genetics.107.071134 PMID: 17483427

20. Guo Y, Fudali S, Gimeno J, DiGennaro P, Chang S, Williamson VM, et al. Networks underpinning symbiosis revealed through cross-species eQTL mapping. Genetics. 2017;206(4):2175–84. https://doi.org/10.1534/genetics.117.202531 PMID: 28642272

21. Thomas VP, Fudali SL, Schaff JE, Liu Q, Scholl EH, Opperman CH, et al. A sequence-anchored linkage map of the plant-parasitic nematode Meloidogyne hapla reveals exceptionally high genome-wide recombination. G3 (Bethesda). 2012;2(7):815–24. https://doi.org/10.1534/g3.112.002261 PMID: 22870404

22. Thomas VP, Williamson VM. Segregation and mapping in the root-knot nematode Meloidogyne hapla of quantitatively inherited traits affecting parasitism. Phytopathology. 2013;103(9):935–40. https://doi.org/10.1094/PHYTO-03-13-0074-R PMID: 23758293

23. Wang C, Lower S, Thomas VP, Williamson VM. Root-knot nematodes exhibit strain-specific clumping behavior that is inherited as a simple genetic trait. PLoS One. 2010;5(12):e15148. https://doi.org/10.1371/journal.pone.0015148 PMID: 21151553

24. 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(39):14802–7. https://doi.org/10.1073/pnas.0805946105 PMID: 18809916

25. 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. Mol Plant Microbe Interact. 2021;34(8):981–6. https://doi.org/10.1094/MPMI-12-20-0337-A PMID: 33779267

26. 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. 2023;14(1):7156. https://doi.org/10.1038/s41467-023-42700-w PMID: 37935661

27. Mota APZ, Koutsovoulos GD, Perfus-Barbeoch L, Despot-Slade E, Labadie K, Aury J-M, et al. Unzipped assemblies of polyploid root-knot nematodes genomes reveal new kinds of unilateral composite telomeric repeats. Nat Commun. 2023;15(1):773. https://doi.org/10.1038/s41467-024-44914-y

28. 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. PLoS One. 2024;19(6):e0302506. https://doi.org/10.1371/journal.pone.0302506 PMID: 38843263

29. Blanc-Mathieu R, Perfus-Barbeoch L, Aury J-M, 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(6):e1006777. https://doi.org/10.1371/journal.pgen.1006777 PMID: 28594822

30. Benson G. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 1999;27(2):573–80. https://doi.org/10.1093/nar/27.2.573 PMID: 9862982

31. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, et al. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res. 2009;37:W202–8. https://doi.org/10.1093/nar/gkp335 PMID: 19458158

32. Chung G, Piano F, Gunsalus KC. TeloSearchLR: an algorithm to detect novel telomere repeat motifs using long sequencing reads. G3 (Bethesda). 2025;15(6):jkaf062. https://doi.org/10.1093/g3journal/jkaf062 PMID: 40169380

33. Gonzalez de la Rosa PM, Thomson M, Trivedi U, Tracey A, Tandonnet S, Blaxter M. A telomere-to-telomere assembly of Oscheius tipulae and the evolution of rhabditid nematode chromosomes. G3 (Bethesda). 2021;11(1):jkaa020. https://doi.org/10.1093/g3journal/jkaa020 PMID: 33561231

34. Prabh N, Rödelsperger C. Multiple Pristionchus pacificus genomes reveal distinct evolutionary dynamics between de novo candidates and duplicated genes. Genome Res. 2022;32(7):1315–27. https://doi.org/10.1101/gr.276431.121 PMID: 35618417

35. Rödelsperger C. Comparative genomics of sex, chromosomes, and sex chromosomes in Caenorhabditis elegans and other nematodes. Methods Mol Biol. 2024;2802:455–72. https://doi.org/10.1007/978-1-0716-3838-5_15 PMID: 38819568

36. Tandonnet S, Koutsovoulos GD, Adams S, Cloarec D, Parihar M, Blaxter ML, et al. Chromosome-wide evolution and sex determination in the three-sexed nematode Auanema rhodensis. G3 (Bethesda). 2019;9(4):1211–30. https://doi.org/10.1534/g3.119.0011 PMID: 30770412

37. Rockman MV, Kruglyak L. Recombinational landscape and population genomics of Caenorhabditis elegans. PLoS Genet. 2009;5(3):e1000419. https://doi.org/10.1371/journal.pgen.1000419 PMID: 19283065

38. Ross JA, Koboldt DC, Staisch JE, Chamberlin HM, Gupta BP, Miller RD, et al. Caenorhabditis briggsae recombinant inbred line genotypes reveal inter-strain incompatibility and the evolution of recombination. PLoS Genet. 2011;7(7):e1002174. https://doi.org/10.1371/journal.pgen.1002174 PMID: 21779179

39. Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. Curr Protoc. 2021;1(12):e323. https://doi.org/10.1002/cpz1.323 PMID: 34936221

40. Teufel F, Almagro Armenteros JJ, Johansen AR, Gíslason MH, Pihl SI, Tsirigos KD, et al. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol. 2022;40(7):1023–5. https://doi.org/10.1038/s41587-021-01156-3 PMID: 34980915

41. Hallgren J, Tsirigos KD, Pedersen MD, Armenteros JJA, Marcatili P, Nielsen H, et al. DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural networks. bioRxiv. 2022. pp. 2022.04.08.487609.

42. Haegeman A, Jones JT, Danchin EGJ. Horizontal gene transfer in nematodes: a catalyst for plant parasitism? Mol Plant Microbe Interact. 2011;24(8):879–87. https://doi.org/10.1094/MPMI-03-11-0055 PMID: 21539433

43. Yin Y, Mao X, Yang J, Chen X, Mao F, Xu Y. dbCAN: a web resource for automated carbohydrate-active enzyme annotation. Nucleic Acids Res. 2012;40:W445–51. https://doi.org/10.1093/nar/gks479 PMID: 22645317

44. Kim J, Yang R, Chang C, Park Y, Tucker ML. The root-knot nematode Meloidogyne incognita produces a functional mimic of the Arabidopsis inflorescence deficient in abscission signaling peptide. J Exp Bot. 2018;69(12):3009–21. https://doi.org/10.1093/jxb/ery135 PMID: 29648636

45. Mishra S, Hu W, DiGennaro P. Root-knot-nematode-encoded CEPs increase nitrogen assimilation. Life (Basel). 2023;13(10):2020. https://doi.org/10.3390/life13102020 PMID: 37895402

46. Zhang X, Peng H, Zhu S, Xing J, Li X, Zhu Z, et al. Nematode-Encoded RALF Peptide Mimics Facilitate Parasitism of Plants through the FERONIA Receptor Kinase. Mol Plant. 2020;13(10):1434–54. https://doi.org/10.1016/j.molp.2020.08.014 PMID: 32896643

47. Yimer HZ, Luu DD, Coomer Blundell A, Ercoli MF, Vieira P, Williamson VM, et al. Root-knot nematodes produce functional mimics of tyrosine-sulfated plant peptides. Proc Natl Acad Sci U S A. 2023;120(29):e2304612120. https://doi.org/10.1073/pnas.2304612120 PMID: 37428936

48. Grbić M, Van Leeuwen T, Clark RM, Rombauts S, Rouzé P, Grbić V, et al. The genome of Tetranychus urticae reveals herbivorous pest adaptations. Nature. 2011;479(7374):487–92. https://doi.org/10.1038/nature10640 PMID: 22113690

49. Spieth J, Lawson D, Davis P, Williams G, Howe K. Overview of gene structure in C. elegans. WormBook: The Online Review of C. elegans Biology. WormBook; 2018.

50. Fouché S, Plissonneau C, Croll D. The birth and death of effectors in rapidly evolving filamentous pathogen genomes. Curr Opin Microbiol. 2018;46:34–42. https://doi.org/10.1016/j.mib.2018.01.020 PMID: 29455143

51. Jagdale S, Rao U, Giri AP. Effectors of root-knot nematodes: an arsenal for successful parasitism. Front Plant Sci. 2021;12:800030. https://doi.org/10.3389/fpls.2021.800030 PMID: 35003188

52. Hillier LW, Miller RD, Baird SE, Chinwalla A, Fulton LA, Koboldt DC, et al. Comparison of C. elegans and C. briggsae genome sequences reveals extensive conservation of chromosome organization and synteny. PLoS Biol. 2007;5(7):e167. https://doi.org/10.1371/journal.pbio.0050167 PMID: 17608563

53. Li M, Zhu C, Xu Z, Xu M, Kuang Y, Hou X, et al. Structural basis for C. elegans pairing center DNA binding specificity by the ZIM/HIM-8 family proteins. Nat Commun. 2024;15(1):10355. https://doi.org/10.1038/s41467-024-54548-9 PMID: 39609407

54. Hillers KJ, Villeneuve AM. Chromosome-wide control of meiotic crossing over in C. elegans. Curr Biol. 2003;13(18):1641–7. https://doi.org/10.1016/j.cub.2003.08.026 PMID: 13678597

55. Yu Z, Kim Y, Dernburg AF. Meiotic recombination and the crossover assurance checkpoint in Caenorhabditis elegans. Semin Cell Dev Biol. 2016;54:106–16. https://doi.org/10.1016/j.semcdb.2016.03.014 PMID: 27013114

56. Goldstein P, Triantaphyllou AC. Occurrence of synaptonemal complexes and recombination nodules in a meiotic race of Meloidogyne hapla and their absence in a mitotic race. Chromosoma. 1978;68(1):91–100. https://doi.org/10.1007/bf00330375

57. Burga A, Ben-David E, Lemus Vergara T, Boocock J, Kruglyak L. Fast genetic mapping of complex traits in C. elegans using millions of individuals in bulk. Nat Commun. 2019;10(1):2680. https://doi.org/10.1038/s41467-019-10636-9 PMID: 31213597

58. Rillo-Bohn R, Adilardi R, Mitros T, Avşaroğlu B, Stevens L, Köhler S, et al. Analysis of meiosis in Pristionchus pacificus reveals plasticity in homolog pairing and synapsis in the nematode lineage. Elife. 2021;10:e70990. https://doi.org/10.7554/eLife.70990 PMID: 34427184

59. Seidel HS, Rockman MV, Kruglyak L. Widespread genetic incompatibility in C. elegans maintained by balancing selection. Science. 2008;319(5863):589–94. https://doi.org/10.1126/science.1151107 PMID: 18187622

60. Zdraljevic S, Walter-McNeill L, Bruni G, Bloom JS, Leighton DHW, Marquez H, et al. A maternal-effect toxin-antidote element causes larval arrest in C. elegans. bioRxiv. 2024. pp. 2024.04.26.591160.

61. Fakhar AZ, Liu J, Pajerowska-Mukhtar KM, Mukhtar MS. The lost and found: unraveling the functions of orphan genes. J Dev Biol. 2023;11(2):27. https://doi.org/10.3390/jdb11020027 PMID: 37367481

62. Müller MC, Praz CR, Sotiropoulos AG, Menardo F, Kunz L, Schudel S, et al. A chromosome-scale genome assembly reveals a highly dynamic effector repertoire of wheat powdery mildew. New Phytol. 2019;221(4):2176–89. https://doi.org/10.1111/nph.15529 PMID: 30388298

63. Croll D, Lendenmann MH, Stewart E, McDonald BA. The impact of recombination hotspots on genome evolution of a fungal plant pathogen. Genetics. 2015;201(3):1213–28. https://doi.org/10.1534/genetics.115.180968 PMID: 26392286

64. Laurent B, Palaiokostas C, Spataro C, Moinard M, Zehraoui E, Houston RD, et al. High-resolution mapping of the recombination landscape of the phytopathogen Fusarium graminearum suggests two-speed genome evolution. Mol Plant Pathol. 2018;19(2):341–54. https://doi.org/10.1111/mpp.12524 PMID: 27998012

65. Hinch R, Donnelly P, Hinch AG. Meiotic DNA breaks drive multifaceted mutagenesis in the human germ line. Science. 2023;382(6674):eadh2531. https://doi.org/10.1126/science.adh2531 PMID: 38033082

66. 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(10):2844–61. https://doi.org/10.1093/gbe/evx201 PMID: 29036290

67. Verhoeven A, Finkers-Tomczak A, Prins P, Valkenburg-van Raaij DR, van Schaik CC, Overmars H, et al. The root-knot nematode effector MiMSP32 targets host 12-oxophytodienoate reductase 2 to regulate plant susceptibility. New Phytol. 2023;237(6):2360–74. https://doi.org/10.1111/nph.18653 PMID: 36457296

68. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science. 2009;326(5950):289–93. https://doi.org/10.1126/science.1181369 PMID: 19815776

69. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011;27(6):764–70. https://doi.org/10.1093/bioinformatics/btr011 PMID: 21217122

70. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. Nat Commun. 2020;11(1):1432. https://doi.org/10.1038/s41467-020-14998-3 PMID: 32188846

71. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021;18(2):170–5. https://doi.org/10.1038/s41592-020-01056-5 PMID: 33526886

72. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722–36. https://doi.org/10.1101/gr.215087.116 PMID: 28298431

73. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11):e112963. https://doi.org/10.1371/journal.pone.0112963 PMID: 25409509

74. Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit - interactive quality assessment of genome assemblies. G3 (Bethesda). 2020;10(4):1361–74. https://doi.org/10.1534/g3.119.400908 PMID: 32071071

75. Miller JR, Delcher AL, Koren S, Venter E, Walenz BP, Brownley A, et al. Aggressive assembly of pyrosequencing reads with mates. Bioinformatics. 2008;24(24):2818–24. https://doi.org/10.1093/bioinformatics/btn548 PMID: 18952627

76. Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020;21(1):245. https://doi.org/10.1186/s13059-020-02134-9 PMID: 32928274

77. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021;18(4):366–8. https://doi.org/10.1038/s41592-021-01101-x PMID: 33828273

78. Griffin S. 2023. Available from: https://github.com/sogriffin98/Contamination_of_Genomes_using_BlobToolKit

79. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997. arXiv. 2013.

80. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34(18):3094–100. https://doi.org/10.1093/bioinformatics/bty191 PMID: 29750242

81. Durand NC, Shamim MS, Machol I, Rao SSP, Huntley MH, Lander ES, et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016;3(1):95–8. https://doi.org/10.1016/j.cels.2016.07.002 PMID: 27467249

82. Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, et al. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. Science. 2017;356(6333):92–5. https://doi.org/10.1126/science.aal3327 PMID: 28336562

83. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117(17):9451–7. https://doi.org/10.1073/pnas.1921046117 PMID: 32300014

84. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2021.

85. Ren Y, Tseng E, Smith TPL, Hiendleder S, Williams JL, Low WY. Long read isoform sequencing reveals hidden transcriptional complexity between cattle subspecies. BMC Genomics. 2023;24(1):108. https://doi.org/10.1186/s12864-023-09212-9 PMID: 36915055

86. Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, et al. BRAKER3: Fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS and TSEBRA. bioRxiv. 2024. https://doi.org/10.1101/2023.06.10.544449

87. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021;38(12):5825–9. https://doi.org/10.1093/molbev/msab293 PMID: 34597405

88. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40. https://doi.org/10.1093/bioinformatics/btu031 PMID: 24451626

89. Killick R, Eckley IA. changepoint: AnRPackage for Changepoint Analysis. J Stat Soft. 2014;58(3). https://doi.org/10.18637/jss.v058.i03

90. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–2. https://doi.org/10.1093/bioinformatics/btq033 PMID: 20110278

91. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20(1):238. https://doi.org/10.1186/s13059-019-1832-y PMID: 31727128

92. Zheng J, Ge Q, Yan Y, Zhang X, Huang L, Yin Y. dbCAN3: automated carbohydrate-active enzyme and substrate annotation. Nucleic Acids Res. 2023;51(W1):W115–21. https://doi.org/10.1093/nar/gkad328 PMID: 37125649

93. Koutsovoulos GD, Granjeon Noriot S, Bailly-Bechet M, Danchin EGJ, Rancurel C. AvP: A software package for automatic phylogenetic detection of candidate horizontal gene transfers. PLoS Comput Biol. 2022;18(11):e1010686. https://doi.org/10.1371/journal.pcbi.1010686 PMID: 36350852

94. Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G, et al. MITOS: improved de novo metazoan mitochondrial genome annotation. Mol Phylogenet Evol. 2013;69(2):313–9. https://doi.org/10.1016/j.ympev.2012.08.023 PMID: 22982435

95. Uliano-Silva M, Ferreira JGRN, Krasheninnikova K, Darwin Tree of Life Consortium, Formenti G, Abueg L, et al. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. BMC Bioinformatics. 2023;24(1):288. https://doi.org/10.1186/s12859-023-05385-y PMID: 37464285

96. Meng G, Li Y, Yang C, Liu S. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. Nucleic Acids Res. 2019;47(11):e63. https://doi.org/10.1093/nar/gkz173 PMID: 30864657

97. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. Nucleic Acids Res. 2021;49(16):9077–96. https://doi.org/10.1093/nar/gkab688 PMID: 34417604

98. Jühling F, Pütz J, Florentz C, Stadler PF. Armless mitochondrial tRNAs in Enoplea (Nematoda). RNA Biol. 2012;9(9):1161–6. https://doi.org/10.4161/rna.21630 PMID: 23018779

99. 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(5):1943–65. https://doi.org/10.1093/molbev/msaa336 PMID: 33399875

100. Gržan T, Despot-Slade E, Meštrović N, Plohl M, Mravinac B. CenH3 distribution reveals extended centromeres in the model beetle Tribolium castaneum. PLoS Genet. 2020;16(10):e1009115. https://doi.org/10.1371/journal.pgen.1009115 PMID: 33125365