

## RESEARCH ARTICLE

# Phased chromosome-scale genome assembly of an asexual, allopolyploid root-knot nematode reveals complex subgenomic structure

Michael R. Winter<sup>1\*</sup>, Adam P. Taranto<sup>2</sup>, Henok Zemene Yimer<sup>3</sup>, Alison Coomer Blundell<sup>2</sup>, Shahid Siddique<sup>3</sup>, Valerie M. Williamson<sup>2</sup>, David H. Lunt<sup>1</sup>

**1** School of Natural Sciences, University of Hull, Hull, United Kingdom, **2** Department of Plant Pathology, University of California Davis, Davis, CA, United States of America, **3** Department of Entomology and Nematology, University of California Davis, Davis, CA, United States of America

\* [mrmrwinter@gmail.com](mailto:mrmrwinter@gmail.com)



## OPEN ACCESS

**Citation:** Winter MR, Taranto AP, Yimer HZ, Coomer Blundell A, Siddique S, Williamson VM, et al. (2024) Phased chromosome-scale genome assembly of an asexual, allopolyploid root-knot nematode reveals complex subgenomic structure. *PLoS ONE* 19(6): e0302506. <https://doi.org/10.1371/journal.pone.0302506>

**Editor:** Katherine James, Newcastle University, UNITED KINGDOM

**Received:** May 8, 2023

**Accepted:** April 5, 2024

**Published:** June 6, 2024

**Copyright:** © 2024 Winter et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** The data that support the findings of this study are publicly available from NCBI with the identifier(s) PRJNA939015 (Bioproject) and JARDRP000000000; GCA\_034785575.1 (Assembly). Scripts will be made available in a publicly accessible Zenodo repository.

**Funding:** This work was supported by USDA Multistate Research Project (SS; CA-D-ENM-2562-RR; United States Department of Agriculture

## Abstract

We present the chromosome-scale genome assembly of the allopolyploid root-knot nematode *Meloidogyne javanica*. We show that the *M. javanica* genome is predominantly allotetraploid, comprising two subgenomes, A and B, that most likely originated from hybridisation of two ancestral parental species. The assembly was annotated using full-length non-chimeric transcripts, comparison to reference databases, and *ab initio* prediction techniques, and the subgenomes were phased using ancestral k-mer spectral analysis. Subgenome B appears to show fission of chromosomal contigs, and while there is substantial synteny between subgenomes, we also identified regions lacking synteny that may have diverged in the ancestral genomes prior to or following hybridisation. This annotated and phased genome assembly forms a significant resource for understanding the origins and genetics of these globally important plant pathogens.

## Author summary

Root-knot nematodes represent one of the most significant crop parasites globally. Despite their agricultural importance, only limited genomic resources have been published to date, leaving a gap in the understanding of genetic mechanisms driving genome evolution and crop virulence. Here, we have used modern genomic and bioinformatic approaches to create a chromosome-scale reference assembly to investigate the origins and genomic composition of the root-knot nematode species *Meloidogyne javanica*. This species has an allopolyploid genome, reproduces by ameiotic parthenogenesis and is among the most damaging plant parasitic nematodes with a large and expanding plant host range.

Utilising modern long-read DNA sequencing and bioinformatics approaches, we successfully phased the assembly into its constituent subgenomes, a first for this agriculturally important clade. While we find the genomic landscape is mostly syntenic between

Multistate Research Project; <https://www.usda.gov/>), USDA NIFA Award (SS; 2021-6701-34736; United States Department of Agriculture National Institute of Food and Agriculture; <https://www.nifa.usda.gov/>) and a University of Hull THYME scholarship (MRW; University of Hull; <https://www.hull.ac.uk/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript."

**Competing interests:** The authors have declared that no competing interests exist.

subgenomes, we identified regions of minimal similarity, and highlight structural divergence between subgenomes.

## Introduction

### The assembly of allopolyploid genomes

Allopolyploidy is a genomic state characterised by more than two chromosomal complements, with one or more of these complements resulting from a hybridisation event leading to the presence of distinct (homoeologous) subgenomes within a single cell [1]. Allopolyploids may account for 11% of plant species including many model species and important crops [2]. Although not as frequent as in plants, genomic investigations are indicating that ancestral genome duplication, hybridisation, and complex genome arrangements are more widespread than previously recognized in animals [3, 4]. Assembly and analysis of allopolyploid genomes, however, is challenging for a number of reasons [5]. The increased number of alleles within an allopolyploid genome can interfere with algorithms used by many assemblers leading to the accumulation of switch errors; regions of the assembly where the sequence switches between haplotypes or homoeologs. In addition, the high amount of repeat content often found in allopolyploid genomes can result in fragmentation of the final assembly if sequenced reads fail to span the repeat [6, 7]. Another difficulty in accurate assembly of allopolyploid genomes has been ‘phasing’ i.e., the assignment of assembly contigs to the correct subgenome. Switch errors and misassemblies introduced during the assembly process can impair the signals required to successfully phase a scaffold, and potential crossover interactions between homoeologs can further complicate this signal [8, 9].

Assembly of allopolyploid genomes has become more feasible due to the advent of long-read sequencing technologies and better assembly algorithms. Most chromosome-scale allopolyploid assemblies in the literature are of agricultural plants [10–13], although a few chromosome-scale allopolyploid assemblies of animal genomes are now also available [14, 15].

### Root-knot nematodes

Root-knot nematodes (RKN)—genus *Meloidogyne*—are a group of obligate plant parasites that include species which severely reduce crop yield [16]. Second-stage juveniles (J2s) of RKNs hatch in the soil and are non-feeding, needing to invade a host plant root to complete their life cycle. Upon reaching the vascular cylinder, J2s induce the formation of a feeding site inside the root, characterised by formation of a gall (“root-knot”) and highly modified “giant cells” on which the nematode feeds [17]. Three closely related species within the *Meloidogyne* genus, *M. arenaria*, *M. incognita*, and *M. javanica*, which we refer to here as the *Meloidogyne incognita* group (MIG), have extremely broad host ranges spanning the majority of flowering plants [18, 19] and together are estimated to cost the agricultural industry tens of billions of US dollars a year [20–22]. Almost 100 RKN species have been described [23, 24], which differ in host range, pathogenicity, geographic range, morphology and reproductive mode.

Although RKN nematode species have diverse modes of reproduction including amphimixis, automixis, and obligate apomixis, cytological examination indicates that *M. javanica* and most other MIG species reproduce by mitotic parthenogenesis [25, 26]; that is, maturation of oocytes consists of a single mitotic division in which chromosomes remain univalent at metaphase. Phylogenomic analysis has revealed that each species possesses two divergent copies of many genes, that the three species likely originated from interspecific hybridisation, and that they share the same ancestors who have provided the A and B subgenomes [19, 27, 28].

Despite asexual reproduction, field isolates and greenhouse selections of MIG species are diverse and successful, differing in their ability to reproduce on specific crop species and varieties [20, 29–31]. A widely investigated example is acquisition of ability to reproduce on tomato with the resistance gene Mi-1, which confers effective resistance to MIG species and is widely deployed for nematode management in tomato [32]. Many independent studies have identified MIG populations that are able to break Mi-mediated resistance; these include both field isolates and greenhouse selections of isofemale lines [33, 34]. However, efforts to decipher the genetic mechanisms for these phenotypic variants have not so far been successful due in part to the lack of tractable genetics and limitations in genome assemblies. Current MIG genome assemblies are fragmented and the homoeologous subgenomes are mostly unphased [19, 28, 35, 36] making it difficult to compare homoeologous sequences, gain a true picture of diversity, or to understand the nature of functional variation.

Here we apply a combination of modern genomic and bioinformatic approaches to generate a highly contiguous, chromosome-level assembly of *M. javanica*, phased into two subgenomes, creating the first chromosome-scale genome assembly of an apomictic allopolyploid animal that we are aware of. This assembly should provide a very valuable framework for research into the diversity and functional divergence of plant pathogenic nematode species. In the wider research landscape, genomes such as those from within the MIG can aid in our understanding of adaptation, ploidy, and evolution of genomes following hybridisation events and loss of meiosis [37].

## Results

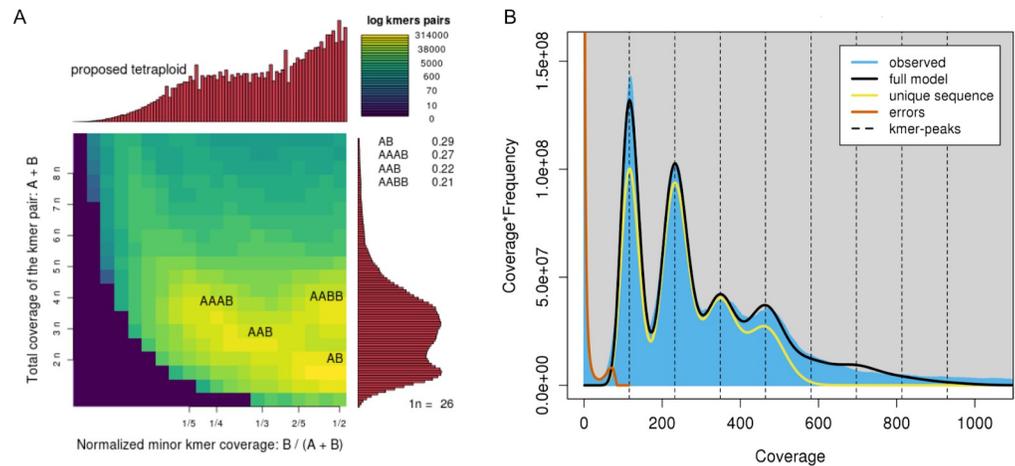
### Sequencing and profiling of read libraries

We used PacBio single-molecule real-time (SMRT) sequencing technology, Hi-C chromatin conformation capture, Nanopore long-read sequencing, and Iso-Seq RNA sequencing to generate a genome assembly of *Meloidogyne javanica* strain VW4. Following quality control and concatenation of two libraries, we obtained 2,255,922 PacBio HiFi reads totalling 35.39 gbp (S1 Fig in S2 File). After quality control and concatenation of Oxford Nanopore (ONT) PromethION data, we obtained 340,373 reads totalling 17.52 gbp (S2 Fig in S2 File). Our Hi-C library contained 375,330,537 read pairs, of which 26.20% were sufficiently unique and high quality for scaffolding with *Proximo* (Phase Genomics, WA). After demultiplexing of our Iso-Seq library, we obtained 2,506,897 full-length non-chimeric sequences which collapsed into 59,637 high quality isoforms.

Genome-wide k-mer profiling of concatenated PacBio HiFi read libraries with *smudgeplot* [38] indicated that 48% of the genome was tetraploid, 22% was triploid, and 29% was diploid (Fig 1A). However, this analysis likely underestimates the proportion of tetraploid regions, as crossover events or conversion between subgenomic copies can cause homogenisation. An incomplete or hypo-tetraploid state was also indicated using a k-mer spectra approach by *GenomeScope2*, predicting a haploid genome length of 68 mbp, and a duplication rate of 3.4 [38] (Fig 1B). This duplication rate is similar to the value seen for CEGMA genes in other assemblies of *M. javanica* (3.68) [28].

### Assembly and annotation

**Scaffolding and final assembly.** From our draft assemblies of the PacBio data, we carried forward an iteration assembled using *HiFiasm* [39] based on overall contiguity and comparison to expected diploid genome length. Following purging of duplicates from the PacBio assembly and scaffolding with Oxford Nanopore reads (S1 File) we obtained 37 contigs. Hi-C scaffolding with the *Proximo* pipeline identified 16 chromosome level clusters (Phase



**Fig 1. Genome profiling plots.** (A) Smudgeplot (left) proposing *M. javanica* as tetraploid, reporting the predicted percentages of ploidy levels in the genome as follows: tetraploid (48%), triploid (22%), or diploid (29%). (B) GenomeScope2 plot (right) showing four distinct peaks in both the predicted model of tetraploidy (black line) and in the observed k-mer spectra (blue fill). The amount of unique sequence falls to zero shortly after the fourth peak, indicating that k-mers at higher ploidy levels than four were mostly repetitive elements.

<https://doi.org/10.1371/journal.pone.0302506.g001>

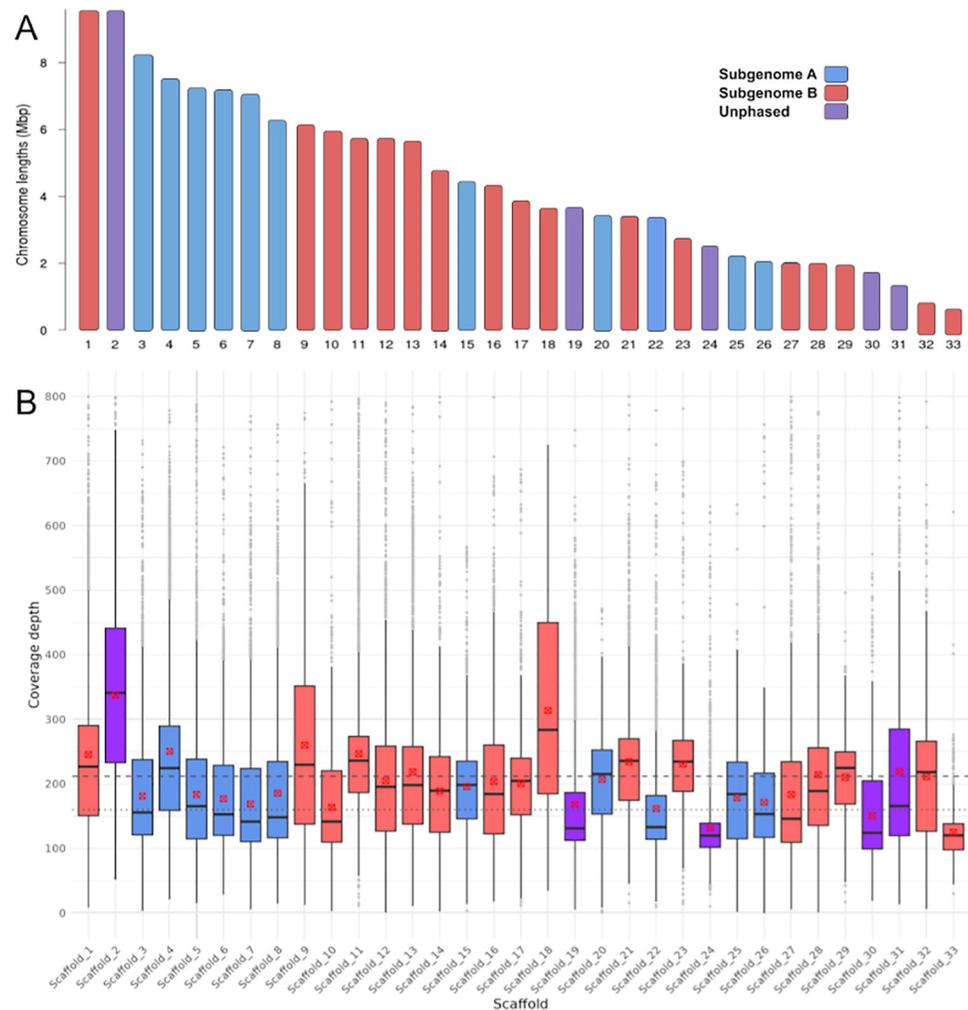
Genomics Ltd) but increased the total number of scaffolds from 37 to 66 (S3 Fig in S2 File). Following scaffolding with Hi-C, *samba* [40] joined some small scaffolds and fragmented the largest scaffold in the assembly, increasing the number of scaffolds from 66 to 69.

The final assembly scaffolds contain 150,545,692 bp with an N50 of 5,793,182 bp, at 30.11% GC content, and overall, 99.96% of reads in our combined PacBio HiFi library map successfully back to the assembly, indicating a high level of completeness (S1 Table in S3 File). Of the total assembly, 97.87% was contained in the longest 33 scaffolds (S4 Fig in S2 File), which ranged from 898 kbp to 9,595 kbp in length (Fig 2A; S2 Table in S3 File). These 33 scaffolds contained 99.85% of all transcribed gene models detected from our Iso-Seq data. Of the remaining 36 scaffolds, 4 were identified by *blobtools* [41] as likely contaminants (Arthropoda, Chordata, and Streptophyta; S5 Fig in S2 File). Nematode scaffolds are often misclassified as Arthropoda by *blobtools*, and so have been retained in the assembly. One further contig was the *M. javanica* mitochondrial genome (22,238 bp). The remaining 31 small contigs were all less than 215 kbp, with a mean length of 88.9 kbp, and since they contained few identifiably functional elements (0.19% of gene models), we excluded them from the final coverage and synteny analysis.

**Annotation.** In total 30.46% of our assembly was identified as repetitive elements, with 4.94% identified as retroelements, 3.77% as DNA transposons, while 16.88% remain unclassified repeats (S3 Table in S3 File). A total of 164,394 transcripts, representing 59,632 isoforms, were detected through mapping of our Iso-Seq library, of which 97% of reads mapped to the assembly. *MAKER3* detected a total of 22,433 genes, containing 227,617 exons, 10,044 5' and 5,253 3' UTRs (S4 Table in S3 File). When running BUSCO on transcriptome settings against the eukaryote\_odb10 database we find that 81.9% of genes are present (C:81.9% [S:43.5%, D:38.4%], F:8.6%, M:9.5%, n:255). When running against the nematoda\_odb10 we find 59.2% of BUSCOs (C:59.2% [S:27.5%, D:31.7%], F:2.7%, M:38.1%, n:3131).

**Core eukaryotic gene and single universal copy ortholog analysis.** Analysis of the final assembly with *CEGMA* [42] detected 233 of 248 CEGMA genes (93.95%). This is a higher level of completeness than the most recent *M. javanica* assembly, and comparable with contemporary *Meloidogyne* assemblies [19, 36, 43, 44] (S1 Table in S3 File).

The average number of orthologs for each complete CEGMA gene (a proxy for ploidy) is 1.88 indicating that only ~6% of the diploid genome is unassembled or not present in the



**Fig 2. Ideogram and coverage depth of longest 33 scaffolds.** Scaffolds are coloured according to phasing status; blue—subgenome A, red—subgenome B, purple—unphased. A, Ideogram of 33 largest scaffolds. These 33 scaffolds contain 98% the total length of the assembly, with the remaining 36 contigs being shorter than 250 kbp, containing few gene models (0.19%), and consisting of mostly repetitive elements. B, Boxplot displaying the distribution of coverage depth for each scaffold. Red points denote the mean of data in each box. Coverage has been limited to a maximum of 800x to exclude probable repetitive sites with anomalous coverage depth. Dashed line shows the overall mean for all coverage levels across 33 scaffolds. Dotted line shows the mode of all coverage across 33 scaffolds.

<https://doi.org/10.1371/journal.pone.0302506.g002>

biological chromosomes. 177 complete BUSCO genes were detected when using BUSCO's eukaryote database, representing 69.5% of genes in the database (C:69.5% [S:37.3%, D:32.2%], F:13.7%, M:16.8%,  $n = 255$ ). Of the BUSCOs identified, 32.2% were duplicated. When using BUSCOs nematode database, we find 1556 complete BUSCO genes, representing 49.7% of genes in the database (C:49.7% [S:22.5%, D:27.2%], F:3.9%, M:46.4%,  $n = 3131$ ).

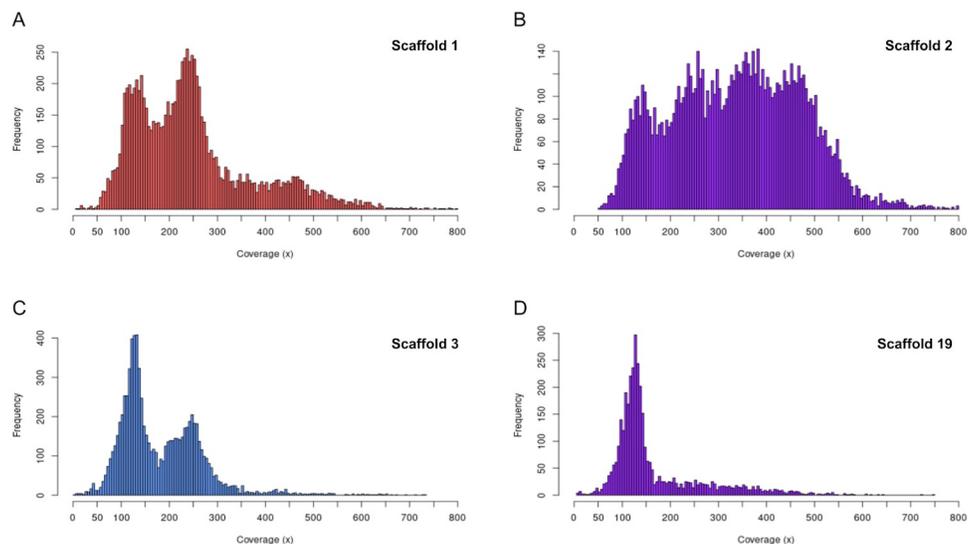
### Coverage and ploidy analysis

Mean coverage depth for all scaffolds—excluding mitochondrial—was 206.6x, falling to 206.1x for the longest 33 scaffolds (Fig 2B; S6 Fig in S2 File). For some scaffolds (2, 9 and 18) coverage depth was significantly in excess of this average, indicating collapsed regions; that is, these scaffolds represent three or four genomic copies, rather than two. Collapse is expected to arise in a

polyploid assembly when homoeologous sequences are similar enough at a nucleotide level that they are inferred to be multiple homologous alleles of the same region, which are then collapsed to form the reference assembly copy [8]. Other scaffolds (19, 22, 24, 33) exhibited lower mean coverage than the assembly-wide average suggesting that they may be present as a single copy.

We then examined the coverage depth frequency distribution for each scaffold (S7 Fig in S2 File). For a phased scaffold representing two identical homologs, the coverage depth distribution for that scaffold would be expected to have a single peak (240x). However, while predominantly single peaks were seen for some scaffolds with lower coverage—thought to be present as a single copy (19, 22, 24, 33)—we observed two peaks for the majority of scaffolds (Fig 3A & 3C; S7 Fig in S2 File). Two peaks would be expected if homologs are not identical and similarity is disrupted by indels, with the second peak representing the reduced coverage of the hemizygous region (~120x). Following this logic, number of peaks in the plot would indicate a respective number of biological alleles mapping to one copy in the genome assembly. For scaffold 2, which is over-represented in mean coverage depth, four clear peaks are present (Fig 3) indicating that more than two alleles map to the scaffold. This likely resulted from exclusion of the scaffold's homoeolog from the assembly, leading to an assembly collapse as noted above. Nevertheless, the presence of 4 peaks is consistent with the presence of polymorphisms between homologs as well as between homoeologous pairs.

To examine the ploidy distribution in another way, we plotted a sliding window of coverage across each scaffold (S8 Fig in S2 File). In support of the coverage depth distributions, the most frequent outcome was that scaffolds in the assembly show two layers of stratification in coverage at a constant proportional depth, indicating two copies distinguished by indels between them. This pattern was strongest for scaffolds with two coverage peaks, particularly phased homoeologs (discussed below, Fig 3A and 3B; S7 Fig in S2 File). Scaffold 2 and scaffold



**Fig 3. Coverage depth frequency distributions of scaffolds 1, 2, 3, and 19.** X-axis represents coverage depth and y-axis represents frequencies of coverage. Colour indicates phase status: red for subgenome B, blue for subgenome A, and purple for unphased scaffolds. Bin size = 5. (A) Scaffold 1, top left, displays two main peaks of coverage with a small tail suggesting short collapsed regions. (B) Scaffold 2, top right, displays four peaks of coverage, indicating that much of this scaffold is collapsed and four copies are mapping to it. (C) Scaffold 3, bottom left, shows two peaks and very little tail, indicating two copies mapping and little to no assembly collapse. (D) Scaffold 19, bottom right, shows only one peak at ~120x coverage, suggesting that only one copy maps to this scaffold.

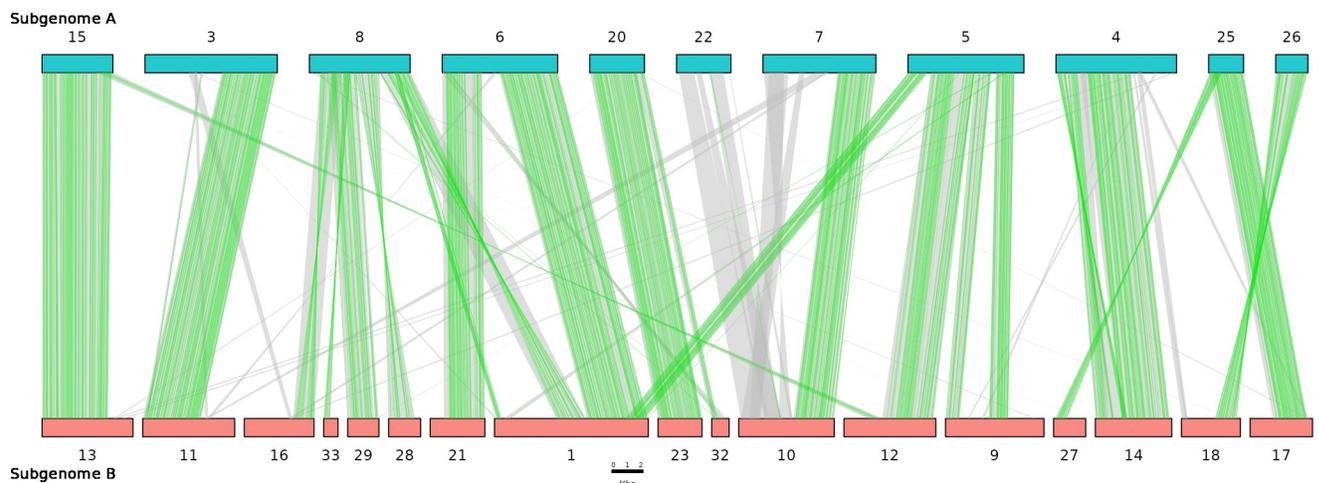
<https://doi.org/10.1371/journal.pone.0302506.g003>

18, which are over-represented in sequence depth, contain regions of four levels of coverage depth covering much of the length of both (Fig 3B; S8A and S8D Fig in S2 File). This increased amount of stratification at proportionally higher coverage depths (360x and 480x) reveals assembly collapse, where three or four copies, respectively, map to a single site. Together these coverage depth results suggest that the *M. javanica* assembly represents two homoeologous subgenomes (~85% of total length), with only 13% of the assembly unphased or collapsed.

### Identification of homoeologous pairs and phasing of subgenomes

Given the diploid nature of the assembly, which represents each subgenome as a single copy, we expected to find scaffolds from these subgenomes present in homoeologous pairs. Through detection of shared orthologs (S1 File) twenty pairings were identified between the longest 33 scaffolds of the assembly, each sharing between 20 and 351 CDS orthologs (Fig 4, S5 Table in S3 File). Alternative methods of identifying homoeologous pairs were corroborative (S6 Table in S3 File). Some pairs were not mutually exclusive, and exhibited CDS links to scaffolds outside of the primary pair, suggesting translocation and syntenic changes between them. All scaffolds that showed two depths of coverage were assigned as homoeologous pairs as expected if each subgenome was heterozygous. Four scaffolds were excluded from pairings including scaffold 2 which showed high amounts of collapse and scaffolds 19, 24, 30, and 31 which are relatively small and/or present as single copies (S7 and S9 Figs in S2 File; S2 Table in S3 File).

In order to assign contigs to a subgenome, we used a modified version of an ancestral k-mer spectra analysis [45] (S1 File). This approach is based on the premise that the allopolyploid's subgenomes possess repeats that diverged in the two parental genomes before hybridisation. This results in a distinguishable signature in each subgenome's k-mer spectrum and allows us to determine the parental species from which a given sequence descends. We successfully phased 85.39% of the assembly into A or B subgenomes (Fig 2A; S9 Fig in S2 File; S2 Table in S3 File). Scaffold 2, which exhibits extensive assembly collapse, did not phase using these methods. Some smaller scaffolds that did not phase by k-mer based methods were later assigned to a subgenome based on the phase status of their opposing homoeologous scaffold (S2 Table in S3 File). Of the total length of the final assembly, 39.15% was assigned to



**Fig 4. Macrosyntenic analysis between subgenomes.** Glyphs along top and bottom represent scaffolds assigned to either subgenome A (blue) or subgenome B (red). Green lines mark locations of syntenic regions identified from mapping of Iso-Seq sequences. Grey lines mark locations of syntenic regions identified through *MAKER3* gene prediction. Syntenic and collinearity were identified using the *MCSan* module of *JCVI* using Iso-Seq informed transcriptional annotation. Scaffolds that could not be assigned to a subgenome are not shown.

<https://doi.org/10.1371/journal.pone.0302506.g004>

subgenome A and 46.24% was assigned to subgenome B, leaving only 14.61% unassigned. Nucleotide similarity between the subgenomes was estimated at 86% for whole subgenomes and 91% between only CDS regions. Allelic divergence between alleles mapping to either subgenome was estimated, with subgenome A estimated at 97.1% and subgenome B at 97.7% (S1 File).

### Subgenomic synteny analysis

Comparison of annotation of scaffold pairs assigned to subgenomes A and B revealed both long regions of synteny and large structural differences between subgenomes (Fig 4). Scaffold 13 and scaffold 15 are syntenic along almost the entire length of the shorter homoeolog and share high nucleotide similarity throughout. Similarly, scaffold pairs 20 and 23, as well as 17 and 25, share long syntenic blocks of shared CDS with high nucleotide similarity.

Scaffolds 7 and 10 are syntenic for almost half of their length whilst the remaining sequence lengths share no synteny and have very low nucleotide similarity. Synteny analysis and scaffold comparison suggest that chromosomal fragmentation has occurred. For example, scaffold 8 of subgenome A shows long collinear blocks with scaffolds 16, 28, 29, and 33 of subgenome B. Re-examination of Hi-C and ONT long-read scaffolding as well as manual inspection of read mapping support this fragmentation. Scaffold 1 of subgenome B shares syntenic blocks with scaffolds 5, 6, and 8 of subgenome A, in an order that suggests chromosomal structural differences between the subgenomes. Many phased scaffolds also exhibit small amounts of extra-pair synteny indicating numerous small translocations throughout the genome.

## Discussion

### Assembly of the allopolyploid genome of *M. javanica*

We have used long-read sequencing and modern bioinformatic approaches to assemble and phase the allopolyploid genome of the plant pathogenic nematode *Meloidogyne javanica*. Our goal was to assemble a contiguous diploid assembly for *M. javanica*, representing the A and B subgenomes separately. Our current diploid assembly (150,545,692 bp) is, as expected for a tetraploid, approximately half the  $297 \pm 27$  Mb measured by flow cytometry [19, 28]. This total length is representative of both A and B subgenomes except for collapsed regions where sequence for both subgenomes A and B is considered together. Because some regions of the assembly have been shown by k-mer profiling and coverage analysis to be present in less than four copies (Fig 1; S8 Fig in S2 File), splitting the subgenomes into their component copies (A1, A2, B1, B2) to create a tetraploid assembly would not be expected to completely double the length.

**Annotation.** We annotated our assembly using both *ab initio* feature prediction algorithms and mapping of the full-length transcript sequences (S4 Table in S3 File). The total number of genes predicted—22,433—is within the range expected for a MIG species and is comparable to previous *M. javanica* assemblies [19, 28].

BUSCO scores for *Meloidogyne* are consistently lower than those of more widely studied organisms, and the number of genes we have detected (S4 Table in S3 File) in the assembly is consistent with what has been found for other *Meloidogyne* species [19, 28, 43, 46, 47] (S1 Table in S3 File). The paucity of established protein databases for less frequently investigated genera limits the accuracy of prediction-based annotation methods. With increased availability of sequence resources for plant parasitic nematodes, it should soon be possible to develop a more appropriate set of core genes.

**Evidence for a chromosome-scale assembly.** Previous genomic assemblies of *M. javanica* have contig counts numbering in the thousands [19, 28]. In our current assembly, more than 99% of reads in our PacBio HiFi libraries map to 33 large scaffolds (Fig 2B). We propose that

most of the 33 scaffolds represent full-length or nearly full-length chromosomes. Cytological examination in *M. javanica* indicates that the chromosome number ranges from 42–48 due to variation between isolates [23, 25]. Thus, we would expect 21–24 scaffolds in our assembly. The discrepancy between scaffold number and cytological observations could be due to an imperfect assembly or failure to identify very small chromosomes in the cytological studies. We have employed several independent scaffolding softwares, Hi-C chromatin contact mapping, and the manual examination of long-read mapping to contig termini, and see no evidence to support fusing additional contigs. Additional molecular and cytological studies may be required to resolve these differences.

Many chromosome-scale assemblies identify telomeres as defining the range of their scaffolds, yet we did not identify canonical telomeric repeats at scaffold termini. Additionally, we were also not able to identify a homolog of *C. elegans* telomerase (*trt-1*; ACC: NM\_001373211.4) in our assembly. This may suggest that non-standard telomere processes might be operating in root-knot nematodes as has been found for some other animals [48, 49].

### Genetic variation in *M. javanica* is dominated by indels

We present the assembly as a diploid representation. Read depth analysis of individual scaffolds indicates that homologs are not identical and indels are frequent. Additionally, some scaffolds appear to be present as a single copy suggesting that one of the homologs may have been lost. We identified many indels that delete or disrupt one or more copies of coding sequences, indicating that *M. javanica* is no longer genome-wide tetraploid either in copy number or functionality. A higher propensity for indel accumulation has been frequently seen in hybrid parthenogenetic species [50] and partial return to lower ploidy is a characteristic of many polyploids.

We find that 85.39% of our diploid assembly consists of one copy of each subgenome in homoeologous pairs, with two copies mapping to phased scaffolds and four copies mapping to collapsed regions. We successfully phased much of our assembly into subgenomes using k-mer signatures, enabling for the first time initial genome-wide comparison of MIG A and B subgenomes. Some scaffolds, notably scaffold 2 and some of the short scaffolds, could not be assigned to subgenomes. Scaffold 2 displays four levels of stratification in its coverage (S8 Fig in S2 File) and four peaks in its depth distribution (Fig 3B) indicating that the four copies (A1 + A2 and B1 + B2) are almost entirely collapsed into a single scaffold. We suggest that the homoeologous chromosomes represented by scaffold 2 may be too similar in sequence to assign to a subgenome by k-mer analysis due to a possible homogenization event. For the smaller scaffolds, the failure may have been due to their small size because they did not contain enough relevant k-mers. Some of the small scaffolds were later assigned to a subgenome using transcript alignments. No misassembly or erroneous scaffolding of the unphased sequences was detected through either programmatic or manual methods.

The allotetraploid genome of *M. Javanica* demonstrates extensive synteny between the A and B subgenomes with most genic regions represented in four copies, as detected from coverage stratification and ploidy profiling. We would expect however that there would be differences between the subgenomes, including indels and other structural variation, as this is typically observed between different species and the MIG have a hybrid origin. In accordance with this, we observed substantial structural variation between subgenomes A and B, including insertions, deletions, and translocations (Fig 4).

### Loss of synteny and fragmentation

We observe regions of the *M. javanica* genome where synteny between paired chromosomes is disrupted. One reason for this could be that the parental species (A and B) had genomes in

which the non-syntenic regions had diverged by translocation, insertion, or deletion. Upon hybridisation to create the allopolyploid MIG these diverged regions form the end of synteny blocks. An alternative explanation is that these changes happened after the hybridisation event in the tumultuous process of genome stabilisation immediately following it [51, 52]. Hybridisation, polyploidization, and the loss of meiosis are processes often associated with rapid genomic change [50, 53, 54] and the unique MIG species allopolyploid genomes we are currently studying may represent different balances between these forces. We observe 11 chromosome-scale scaffolds in subgenome A and 17 in subgenome B despite the clear synteny throughout these two subgenomes. Several scaffolds in subgenome A contain blocks of genes with regions that are syntenic to different subgenome B scaffolds. Similarly, there are cases where syntenic blocks in subgenome B are present on different scaffolds in subgenome A. Together these differences suggest that ancestral chromosomal fission, or fusion events or other types of exchange have occurred. *Meloidogyne* species, like other nematodes, have holocentric chromosomes [25]. Genomes with dispersed centromere structure are predicted to better tolerate chromosome fragmentation and fusion [55], as are species with ameiotic mechanisms of reproduction. The observed differences in copy number of chromosomes between isolates of *M. javanica* may be additional evidence for tolerance of chromosome fragmentation/fusion.

The majority of published assemblies of allopolyploids come from plants, where polyploidy might have shaped the genomes of around 70% of species [56]. Many allopolyploid plant genomes, however, show a higher level of synteny and structural conservation than we observe for *M. javanica* [13, 45, 57]. Similarly the few available chromosomal allopolyploid animal genome sequences available [14, 15] do not show extensive deletions and chromosomal fissions as does our genome assembly. Unlike other allopolyploid species with chromosomal genome sequences, *Meloidogyne javanica* reproduces by obligatory mitotic parthenogenesis (apomixis) and this lack of meiotic chromosome pairing may allow greater structural divergence and tolerance for the decay of synteny. We note that genomes from species not able to reproduce by meiosis are currently rare [50] and suggest that much more substantial genomic work on a range of species with different reproductive modes and ploidy levels will be required to reveal the diverse mechanisms shaping these genomes. It is apparent however that the changes surrounding allopolyploidy have contributed to the gene content, heterozygosity, and copy number throughout the *M. javanica* genome and these processes of locally fixed heterozygosity or rediploidization may contribute extensively to adaptive functional variation [58].

### Regional homogenization of subgenomes

Some regions of the genome assembly do not phase into A and B homoeologs due to very low divergence between the four gene copies. This could be explained either by the loss of whole chromosomes, compensated by the duplication of the remaining chromosome, or mitotic recombination (gene conversion) between homoeologs [59, 60]. It is unclear from this single genome how asexual recombination contributes to shaping the diversity of *M. javanica*, however this has been suggested in previous MIG genomic studies [19, 28] and may be further elucidated by our ongoing molecular evolution and population genomic studies.

It is possible that the initial tetraploidisation of *M. javanica* will buffer against deleterious phenotypic consequences of indels. It has been argued for angiosperms that this rediploidization process can contribute to adaptive species divergence by providing genomic and transcriptomic diversity [58, 61]. Other mechanisms of adaptive divergence may also operate at the same time. The increase in gene copy number created by polyploidization gives the potential for functional gene divergence by neo- or sub-functionalization, as well as adaptive phenotypes driven by copy number loss [62–64]. Adaptation by gene copy number variation has

already been reported in *M. incognita* [65] and it may be that genomic copy number variation more broadly is a major source of functional genetic variation in the MIG.

## A genomic framework for RKN functional and diversity studies

In this paper we present a highly contiguous, annotated, and phased genome assembly of the allotetraploid plant pathogenic nematode *M. javanica*. This genome assembly will provide many tools for diverse investigations by plant pathologists and nematologists in addition to aiding our understanding of the origins and diversity of *M. javanica*. It will also serve as a reference for investigating genome structure and pathogenicity in other *Meloidogyne* species. The contiguous nature of this genome and the high-quality annotation will facilitate RKN functional studies since transcripts can be mapped accurately to the annotated subgenome. This allows consideration of copy number variation, which may be an important component of functional variation in these species. Progress is being made by many groups in understanding the basis of nematode virulence and the key loci involved [66–68]. In such cases even light coverage sequencing of field isolates mapped to a high-quality genome assembly could give valuable information to commercial growers about the likely pathogenicity of those strains [69].

## Methods

### Reproducibility

Wherever possible, this study attempted to contain all bioinformatic processes in reproducible workflows or scripts, for the purpose of openness and enabling replication. Workflows and code are archived in a Zenodo repository along with final outputs (doi: [10.5281/zenodo.10784780](https://doi.org/10.5281/zenodo.10784780)). The raw reads plus final nuclear and mitochondrial assemblies are available on the International Nucleotide Sequence Databases (INSDC) (ACC: PRJNA939015) (ACC: GCA\_034785575.1).

### Biological material

*Meloidogyne javanica* strain VW4 was used for this work (Gleason et al, 2008; Szitenberg et al., 2017). Cultures of this strain have been maintained on tomato plants under greenhouse conditions for over 30 years [70]. Periodic transfers of single egg masses have been carried out to maintain uniformity. For DNA preparation, eggs were harvested from roots and cleaned by sucrose flotation as previously described [71] then flash-frozen in liquid N<sub>2</sub>. High molecular weight DNA (HMW DNA) isolation was carried out at UC Davis Genome Center (S1 File). Integrity of the HMW gDNA was verified on a Femto Pulse system (Agilent Technologies, CA) where majority of the DNA was found to be in fragments above 100 Kb.

Total RNA was isolated from three *M. javanica* life stages: eggs, freshly hatched juveniles, and females dissected from tomato roots 21 days after infection. This material was flash-frozen and RNA was extracted using an Rneasy Kit (Qiagen, USA) following the manufacturer's instructions. TURBO Dnase treatment was carried out to remove genomic DNA from total RNA samples (TURBO DNA-free Kit™, Ambion, USA). RNA concentration and purity were measured using a NanoDrop OneC Microvolume UV-Vis Spectrophotometer (Thermo Scientific, USA). RNA integrity and quality was assessed using a 2100 Bioanalyzer Agilent Technologies G2939BA.

### Sequencing and QC

**High fidelity (HiFi) long-read sequencing.** PacBio HiFi library preparation and sequencing of HMW DNA was performed by UC Davis DNA Technologies Core on a PacBio Sequel

II (S1 File). Data from the two generated libraries were pooled and only reads longer than 5000 bp and with a quality score over 15 were retained.

**Nanopore sequencing.** Nanopore sequencing of HMW DNA using Oxford Nanopore Technologies (ONT) systems was carried out by UC Davis DNA Technologies Core. The super-long-read DNA sequencing protocol (S1 File) yielded 23 gbp of data, which was then filtered to contain only reads longer than 25 kbp.

**Hi-C chromatin conformation capture.** DNA was prepared using Proximo Hi-C kit (Animal) as recommended by the manufacturer (Phase Genomics, Seattle, WA, USA). Library preparation and sequencing were carried out at the UC Davis DNA Technologies Core and scaffolding with Proximo was performed by Phase Genomics (S1 File).

**PacBio Iso-Seq.** Reads from one Sequel II SMRT cell were quality controlled and converted into clustered reads using the *IsoSeq3* pipeline with default parameters [72].

## Genome profiling

Profiling was performed by *Genomescope2* and *smudgeplot* [38]. All quality controlled PacBio HiFi libraries were combined and used as input for these programs. Both were run with default parameters, aside from *-ploidy* in *Genomescope2* set to 4, as this was the primary ploidy indicated by *smudgeplot* and suggested by visual analysis of mapped reads.

## Assembly

**Draft assembly.** Initial draft assemblies were generated using several different assemblers and appraised with *asmapp* [39, 73–77]. The most appropriate assembly based on length and contiguity was carried forward, after which haplotigs were identified and removed using *purge\_dups* [78].

**Scaffolding with Oxford Nanopore and Hi-C.** We used *SLR* [79] to scaffold the assembly with our trimmed and concatenated ONT reads. The assembly was then scaffolded with Hi-C reads using the *Proximo* pipeline by Phase Genomics Ltd. The resulting contact map was manually curated using *Juicebox* assembly tools [80] to produce the most likely assembly based on contact linkage information and TAD presence or absence. A third phase of scaffolding was performed by *samba*, to break potential misassemblies introduced during manual *Juicebox* scaffolding and scaffold sequences that were broken as debris [40].

## Annotation

**Repeat annotation.** *RepeatModeler* was applied to the assembly [81], with the resulting library of repeat models used as input for *RepeatMasker* to annotate repeat regions and generate a soft-masked version of the genome assembly [82].

**Gene annotation.** Iso-Seq reads were mapped to the assembly and collapsed using a *snakemake* workflow automating the *IsoSeq3* pipeline [39]. The annotation pipeline *MAKER3* was then employed to perform *ab initio* and predictive annotation of our assembly [83, 84]. Full methods of annotation iterations and parameters can be found in S1 File.

## Subgenome phasing

**Identification of homoeologous scaffold pairs.** Homoeologous pairs were detected through identification of orthologs shared between scaffolds (S1 File). Pairs sharing a large number of orthologs, and nucleotide similarity were considered homoeologous. This was validated with other methods, including MASH distance [85] and shared possession of duplicated BUSCO genes [86].

**Phasing of subgenomes.** Scaffolds were phased into A and B subgenomes using a k-mer based approach built on the approach taken by [45]. K-mers in the assembly were detected and counted using *jellyfish* [87], with only k-mers present more than 75 times in the assembly and represented at least twice as often in one subgenome than the other carried forward. These counts were then transformed into binomial distributions. Hierarchical clustering was performed on these sets, creating a dendrogram placing scaffolds into opposing clusters, each cluster representing a subgenome (S1 File).

### Synteny analysis

The *MCSan* (Python) [88] module of *JCVI* [89] was used to perform a synteny analysis between the two subgenomes. Custom scripts then extracted collinearity information and generated synteny plots (S1 File).

### Supporting information

**S1 File.**  
(PDF)

**S2 File.**  
(PDF)

**S3 File.**  
(PDF)

### Acknowledgments

We acknowledge the Viper High Performance Computing facility of the University of Hull and its support team, specifically Chris Collins, for their invaluable help in managing and deploying software, as well as the High Performance Computing team at UC Davis for making available their resources for part of our analysis.

### Author Contributions

**Conceptualization:** Michael R. Winter, Adam P. Taranto, Shahid Siddique, Valerie M. Williamson, David H. Lunt.

**Data curation:** Michael R. Winter, Adam P. Taranto.

**Formal analysis:** Michael R. Winter, Adam P. Taranto.

**Funding acquisition:** Shahid Siddique, David H. Lunt.

**Investigation:** Michael R. Winter, Adam P. Taranto, Henok Zemene Yimer, Alison Coomer Blundell, Valerie M. Williamson.

**Methodology:** Michael R. Winter, Adam P. Taranto, Henok Zemene Yimer, Shahid Siddique, Valerie M. Williamson, David H. Lunt.

**Project administration:** Michael R. Winter, Adam P. Taranto, David H. Lunt.

**Resources:** Shahid Siddique, Valerie M. Williamson, David H. Lunt.

**Software:** Michael R. Winter, Adam P. Taranto.

**Supervision:** Adam P. Taranto, Shahid Siddique.

**Validation:** Michael R. Winter, Adam P. Taranto, David H. Lunt.

**Visualization:** Michael R. Winter, Adam P. Taranto, Valerie M. Williamson.

**Writing – original draft:** Michael R. Winter, Valerie M. Williamson, David H. Lunt.

**Writing – review & editing:** Michael R. Winter, Shahid Siddique, Valerie M. Williamson, David H. Lunt.

## References

- Glover NM, Redestig H, Dessimoz C. Homoeologs: What Are They and How Do We Infer Them? *Trends Plant Sci.* 2016; 21: 609–621. <https://doi.org/10.1016/j.tplants.2016.02.005> PMID: 27021699
- Barker MS, Arrigo N, Baniaga AE, Li Z, Levin DA. On the relative abundance of autopolyploids and allopolyploids. *New Phytol.* 2015. <https://doi.org/10.1111/nph.13698> PMID: 26439879
- Schoenfelder KP, Fox DT. The expanding implications of polyploidy. *J Cell Biol.* 2015; 209: 485–491. <https://doi.org/10.1083/jcb.201502016> PMID: 26008741
- Session AM, Uno Y, Kwon T, Chapman JA, Toyoda A, Takahashi S, et al. Genome evolution in the allotetraploid frog *Xenopus laevis*. *Nature.* 2016; 538: 336–343. <https://doi.org/10.1038/nature19840> PMID: 27762356
- Ming R, Man Wai C. Assembling allopolyploid genomes: no longer formidable. *Genome Biol.* 2015; 16: 27. <https://doi.org/10.1186/s13059-015-0585-5> PMID: 25723730
- Kyriakidou M, Tai HH, Anglin NL, Ellis D, Strömviik MV. Current Strategies of Polyploid Plant Genome Sequence Assembly. *Front Plant Sci.* 2018; 9: 1660. <https://doi.org/10.3389/fpls.2018.01660> PMID: 30519250
- Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, et al. Towards complete and error-free genome assemblies of all vertebrate species. *Nature.* 2021; 592: 737–746. <https://doi.org/10.1038/s41586-021-03451-0> PMID: 33911273
- Zhang X, Wu R, Wang Y, Yu J, Tang H. Unzipping haplotypes in diploid and polyploid genomes. *Comput Struct Biotechnol J.* 2020; 18: 66–72. <https://doi.org/10.1016/j.csbj.2019.11.011> PMID: 31908732
- Saada OA, Friedrich A, Schacherer J. Towards accurate, contiguous and complete alignment-based polyploid phasing algorithms. *Genomics.* 2022; 114: 110369. <https://doi.org/10.1016/j.ygeno.2022.110369> PMID: 35483655
- Edger PP, Poorten TJ, VanBuren R, Hardigan MA, Colle M, McKain MR, et al. Origin and evolution of the octoploid strawberry genome. *Nat Genet.* 2019; 51: 541–547. <https://doi.org/10.1038/s41588-019-0356-4> PMID: 30804557
- Gan X, Li S, Zong Y, Cao D, Li Y, Liu R, et al. Chromosome-Level Genome Assembly Provides New Insights into Genome Evolution and Tuberos Root Formation of *Potentilla anserina*. *Genes.* 2021; 12. <https://doi.org/10.3390/genes12121993> PMID: 34946942
- Kolesnikova UK, Scott AD, Van de Velde JD, Burns R, Tikhomirov NP, Pfordt U, et al. Genome of selfing Siberian *Arabidopsis lyrata* explains establishment of allopolyploid *Arabidopsis kamchatica*. *bioRxiv.* 2022. p. 2022.06.24.497443. <https://doi.org/10.1101/2022.06.24.497443>
- Zheng Y, Yang D, Rong J, Chen L, Zhu Q, He T, et al. Allele-aware chromosome-scale assembly of the allopolyploid genome of hexaploid Ma bamboo (*Dendrocalamus latiflorus* Munro). *J Integr Plant Biol.* 2022; 64: 649–670. <https://doi.org/10.1111/jipb.13217> PMID: 34990066
- Du K, Stöck M, Kneitz S, Klopp C, Woltering JM, Adolphi MC, et al. The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. *Nature Ecology & Evolution.* 2020; 4: 841–852. <https://doi.org/10.1038/s41559-020-1166-x> PMID: 32231327
- Kuhl H, Du K, Scharf M, Kalous L, Stöck M, Lamatsch DK. Equilibrated evolution of the mixed auto-/allopolyploid haplotype-resolved genome of the invasive hexaploid Prussian carp. *Nat Commun.* 2022; 13: 4092. <https://doi.org/10.1038/s41467-022-31515-w> PMID: 35835759
- Perry RN, Moens M, Starr JL. *Root-knot Nematodes.* CABI; 2009.
- Williamson VM, Gleason CA. Plant–nematode interactions. *Curr Opin Plant Biol.* 2003; 6: 327–333. [https://doi.org/10.1016/s1369-5266\(03\)00059-1](https://doi.org/10.1016/s1369-5266(03)00059-1) PMID: 12873526
- 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
- Szitenberg A, Salazar-Jaramillo L, Blok VC, Laetsch DR, Joseph S, Williamson VM, et al. Comparative Genomics of Apomictic Root-Knot Nematodes: Hybridization, Ploidy, and Dynamic Genome Change. *Genome Biol Evol.* 2017; 9: 2844–2861. <https://doi.org/10.1093/gbe/evx201> PMID: 29036290

20. Wesemael W, Viaene N, Moens M. Root-knot nematodes (*Meloidogyne* spp.) in Europe. *Nematology*. 2011; 13: 3–16. <https://doi.org/10.1163/138855410X526831>
21. Jones JT, Haegeman A, Danchin EGJ, Gaur HS, Helder J, Jones MGK, et al. Top 10 plant-parasitic nematodes in molecular plant pathology. *Mol Plant Pathol*. 2013; 14: 946–961. <https://doi.org/10.1111/mpp.12057> PMID: 23809086
22. Bernard GC, Egnin M, Bonsi C. The impact of plant-parasitic nematodes on agriculture and methods of control. *Nematology-Concepts, Diagnosis and Control*. 2017; 10.
23. Eisenback JD, Triantaphyllou HH. Root-knot nematodes: *Meloidogyne* species and races. *Manual of agricultural nematology*. 1991; 1: 191–274.
24. Subbotin SA, Rius JEP, Castillo P. *Systematics of Root-knot Nematodes (Nematoda: Meloidogynidae)*. BRILL; 2021.
25. Triantaphyllou AC. Gametogenesis and the Chromosomes of *Meloidogyne nataliei*. Not Typical of Other Root-knot Nematodes. *J Nematol*. 1985; 17: 1–5.
26. 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: 333–351. <https://doi.org/10.1146/annurev-phyto-080508-081839> PMID: 19400640
27. Lunt DH. Genetic tests of ancient asexuality in root knot nematodes reveal recent hybrid origins. *BMC Evol Biol*. 2008; 8: 194. <https://doi.org/10.1186/1471-2148-8-194> PMID: 18606000
28. 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: e1006777. <https://doi.org/10.1371/journal.pgen.1006777> PMID: 28594822
29. Barker KR, Carter CC, Sasser JN. Identification of *Meloidogyne* species on the basis of differential host test and perineal pattern morphology. In: Hartman KM, Sasser JN. *An advanced treatise on Meloidogyne*. Raleigh: North Carolina State University Graphics; 1985; pp. 69–77.
30. Roberts PA, Thomason J. Variability in reproduction of isolates of *Meloidogyne incognita* and *M. javanica* on resistant tomato genotypes. *Plant Disease*. 1986; 70: 547–551.
31. Rammah A, Hirschmann H. Morphological Comparison of Three Host Races of *Meloidogyne javanica*. *J Nematol*. 1990; 22: 56–68.
32. Williamson VM, Kumar A. Nematode resistance in plants: the battle underground. *Trends Genet*. 2006; 22: 396–403. <https://doi.org/10.1016/j.tig.2006.05.003> PMID: 16723170
33. Gleason CA, Liu QL, Williamson VM. Silencing a candidate nematode effector gene corresponding to the tomato resistance gene Mi-1 leads to acquisition of virulence. *Mol Plant Microbe Interact*. 2008; 21: 576–585. <https://doi.org/10.1094/MPMI-21-5-0576> PMID: 18393617
34. Hajihassani A, Marquez J, Woldemeskel M, Hamidi N. Identification of Four Populations of *Meloidogyne incognita* in Georgia, United States, Capable of Parasitizing Tomato-Bearing Mi-1.2 Gene. *Plant Dis*. 2022; 106: 137–143. <https://doi.org/10.1094/PDIS-05-21-0902-RE> PMID: 34410860
35. Sato K, Kadota Y, Gan P, Bino T, Uehara T, Yamaguchi K, et al. High-Quality Genome Sequence of the Root-Knot Nematode *Meloidogyne arenaria* Genotype A2-O. *Genome Announc*. 2018; 6. <https://doi.org/10.1128/genomeA.00519-18> PMID: 29954888
36. Susič N, Koutsovoulos GD, Riccio C, Danchin EGJ, Blaxter ML, Lunt DH, et al. Genome sequence of the root-knot nematode *Meloidogyne luci*. *J Nematol*. 2020; 52: 1–5. <https://doi.org/10.21307/jofnem-2020-025> PMID: 32180388
37. Fox DT, Soltis DE, Soltis PS, Ashman T-L, Van de Peer Y. Polyploidy: A biological force from cells to ecosystems. *Trends Cell Biol*. 2020; 30: 688–694. <https://doi.org/10.1016/j.tcb.2020.06.006> PMID: 32646579
38. Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020; 11: 1432. <https://doi.org/10.1038/s41467-020-14998-3> PMID: 32188846
39. 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: 170–175. <https://doi.org/10.1038/s41592-020-01056-5> PMID: 33526886
40. Zimin AV, Salzberg SL. The SAMBA tool uses long reads to improve the contiguity of genome assemblies. *PLoS Comput Biol*. 2022; 18: e1009860. <https://doi.org/10.1371/journal.pcbi.1009860> PMID: 35120119
41. Laetsch DR, Blaxter ML. BlobTools: Interrogation of genome assemblies. *F1000Res*. 2017; 6: 1287. <https://doi.org/10.12688/f1000research.12232.1>

42. Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics*. 2007; 23: 1061–1067. <https://doi.org/10.1093/bioinformatics/btm071> PMID: 17332020
43. Koutsovoulos GD, Poulet M, Elashry A, Kozłowski DKL, Sallet E, Da Rocha M, et al. Genome assembly and annotation of *Meloidogyne enterolobii*, an emerging parthenogenetic root-knot nematode. *Scientific Data*. 2020; 7: 324. <https://doi.org/10.1038/s41597-020-00666-0> PMID: 33020495
44. Kozłowski DKL, Hassanaly-Goulamhousen R, Da Rocha M, Koutsovoulos GD, Bailly-Bechet M, Danchin EGJ. Movements of transposable elements contribute to the genomic plasticity and species diversification in an asexually reproducing nematode pest. *Evol Appl*. 2021; 14: 1844–1866. <https://doi.org/10.1111/eva.13246> PMID: 34295368
45. Cerca J, Petersen B, Lazaro-Guevara JM, Rivera-Colón A, Birkeland S, Vizueta J, et al. The genomic basis of the plant island syndrome in Darwin's giant daisies. *Nat Commun*. 2022; 13: 3729. <https://doi.org/10.1038/s41467-022-31280-w> PMID: 35764640
46. Koutsovoulos GD, Marques E, Arguel M, Duret L, Machado ACZ, Carneiro RMDG, et al. Population genomics supports clonal reproduction and multiple independent gains and losses of parasitic abilities in the most devastating nematode pest. *Evol Appl*. 2019; 26: 909. <https://doi.org/10.1111/eva.12881> PMID: 31993088
47. 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; MPMI12200337A. <https://doi.org/10.1094/MPMI-12-20-0337-A> PMID: 33779267
48. Mason JM, Reddy HM, Frydrychova RC. Telomere Maintenance in Organisms without Telomerase. In: Seligmann H, editor. *DNA Replication*. Rijeka: IntechOpen; 2011. <https://doi.org/10.5772/19348>
49. Pardue M-L, DeBaryshe PG. Retrotransposons that maintain chromosome ends. *Proc Natl Acad Sci U S A*. 2011; 108: 20317–20324. <https://doi.org/10.1073/pnas.1100278108> PMID: 21821789
50. Jaron KS, Bast J, Nowell RW, Ranallo-Benavidez TR, Robinson-Rechavi M, Schwander T. Genomic Features of Parthenogenetic Animals. *J Hered*. 2021; 112: 19–33. <https://doi.org/10.1093/jhered/esaa031> PMID: 32985658
51. Edger PP, McKain MR, Bird KA, VanBuren R. Subgenome assignment in allopolyploids: challenges and future directions. *Curr Opin Plant Biol*. 2018; 42: 76–80. <https://doi.org/10.1016/j.pbi.2018.03.006> PMID: 29649616
52. Emery M, Willis MMS, Hao Y, Barry K, Oakgrove K, Peng Y, et al. Preferential retention of genes from one parental genome after polyploidy illustrates the nature and scope of the genomic conflicts induced by hybridization. *PLoS Genet*. 2018; 14: e1007267. <https://doi.org/10.1371/journal.pgen.1007267> PMID: 29590103
53. Ma XF, Gustafson JP. Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenet Genome Res*. 2005; 109: 236–249. <https://doi.org/10.1159/000082406> PMID: 15753583
54. Balloux F, Lehmann L, de Meeùs T. The Population Genetics of Clonal and Partially Clonal Diploids. *Genetics*. 2003; 164: 1635–1644. <https://doi.org/10.1093/genetics/164.4.1635> PMID: 12930767
55. Carlton PM, Davis RE, Ahmed S. Nematode chromosomes. *Genetics*. 2022. <https://doi.org/10.1093/genetics/iyac014> PMID: 35323874
56. Masterson J. Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science*. 1994; 264: 421–424. <https://doi.org/10.1126/science.264.5157.421> PMID: 17836906
57. Shen Y, Li W, Zeng Y, Li Z, Chen Y, Zhang J, et al. Chromosome-level and haplotype-resolved genome provides insight into the tetraploid hybrid origin of patchouli. *Nat Commun*. 2022; 13: 1–15. <https://doi.org/10.1038/s41467-022-31121-w> PMID: 35717499
58. Dodsworth S, Chase MW, Leitch AR. Is post-polyploidization diploidization the key to the evolutionary success of angiosperms? *Bot J Linn Soc*. 2016; 180: 1–5. <https://doi.org/10.1111/boj.12357>
59. Harris S, Rudnicki KS, Haber JE. Gene conversions and crossing over during homologous and homeologous ectopic recombination in *Saccharomyces cerevisiae*. *Genetics*. 1993; 135: 5–16. <https://doi.org/10.1093/genetics/135.1.5> PMID: 8224827
60. Mansai SP, Innan H. The power of the methods for detecting interlocus gene conversion. *Genetics*. 2010; 184: 517–527. <https://doi.org/10.1534/genetics.109.111161> PMID: 19948889
61. Hollister JD. Polyploidy: adaptation to the genomic environment. *New Phytol*. 2015; 205: 1034–1039. <https://doi.org/10.1111/nph.12939> PMID: 25729801
62. Paquin C, Adams J. Frequency of fixation of adaptive mutations is higher in evolving diploid than haploid yeast populations. *Nature*. 1983; 302: 495–500. <https://doi.org/10.1038/302495a0> PMID: 6339947

63. Otto SP, Whitton J. Polyploid incidence and evolution. *Annu Rev Genet.* 2000; 34: 401–437. <https://doi.org/10.1146/annurev.genet.34.1.401> PMID: 11092833
64. Zörgö E, Chwialkowska K, Gjuvsland AB, Garré E, Sunnerhagen P, Liti G, et al. Ancient evolutionary trade-offs between yeast ploidy states. *PLoS Genet.* 2013; 9: e1003388. <https://doi.org/10.1371/journal.pgen.1003388> PMID: 23555297
65. Castagnone-Sereno P, Mulet K, Danchin EGJ, Koutsovoulos GD, Karaulic M, Da Rocha M, et al. Gene copy number variations as signatures of adaptive evolution in the parthenogenetic, plant-parasitic nematode *Meloidogyne incognita*. *Mol Ecol.* 2019; 26: 906. <https://doi.org/10.1111/mec.15095> PMID: 30964953
66. Pogorelko GV, Juvale PS, Rutter WB, Hütten M, Maier TR, Hewezi T, et al. Re-targeting of a plant defense protease by a cyst nematode effector. *Plant J.* 2019; 98: 1000–1014. <https://doi.org/10.1111/tpj.14295> PMID: 30801789
67. Kihika R, Tchouassi DP, Ng'ang'a MM, Hall DR, Beck JJ, Torto B. Compounds Associated with Infection by the Root-Knot Nematode, *Meloidogyne javanica*, Influence the Ability of Infective Juveniles to Recognize Host Plants. *J Agric Food Chem.* 2020; 68: 9100–9109. <https://doi.org/10.1021/acs.jafc.0c03386> PMID: 32786872
68. Song H, Lin B, Huang Q, Sun T, Wang W, Liao J, et al. The *Meloidogyne javanica* effector Mj2G02 interferes with jasmonic acid signalling to suppress cell death and promote parasitism in Arabidopsis. *Mol Plant Pathol.* 2021; 22: 1288–1301. <https://doi.org/10.1111/mpp.13111> PMID: 34339585
69. Sellers GS, Jeffares DC, Lawson B, Prior T, Lunt DH. Identification of individual root-knot nematodes using low coverage long-read sequencing. *PLoS One.* 2021; 16: e0253248. <https://doi.org/10.1371/journal.pone.0253248> PMID: 34851967
70. Yaghoobi J, Kaloshian I, Wen Y, Williamson VM. Mapping a new nematode resistance locus in *Lycopersicon peruvianum*. *Theor Appl Genet.* 1995; 91: 457–464. <https://doi.org/10.1007/BF00222973> PMID: 24169835
71. Branch C, Hwang C-F, Navarre DA, Williamson VM. Salicylic acid is part of the Mi-1-mediated defense response to root-knot nematode in tomato. *Mol Plant Microbe Interact.* 2004; 17: 351–356. <https://doi.org/10.1094/MPMI.2004.17.4.351> PMID: 15077667
72. PacificBiosciences. IsoSeq3. Github; 2022. Available: <https://github.com/PacificBiosciences/IsoSeq>
73. 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: 722–736. <https://doi.org/10.1101/gr.215087.116> PMID: 28298431
74. Shafin K, Pesout T, Lorig-Roach R, Haukness M, Olsen HE, Bosworth C, et al. Nanopore sequencing and the Shasta toolkit enable efficient de novo assembly of eleven human genomes. *Nat Biotechnol.* 2020; 38: 1044–1053. <https://doi.org/10.1038/s41587-020-0503-6> PMID: 32686750
75. Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods.* 2016; 13: 1050–1054. <https://doi.org/10.1038/nmeth.4035> PMID: 27749838
76. Biosciences P. pbipa: Improved Phased Assembler. Github; Available: <https://github.com/PacificBiosciences/pbipa>
77. Winter M. asmapp: ASMAPP assembly appraisal workflow. Built in snakemake, ASMAPP performs many basic and intermediate assembly appraisal tasks. Github; 2022. Available: <https://github.com/mrmrwinter/asmapp>
78. Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics.* 2020; 36: 2896–2898. <https://doi.org/10.1093/bioinformatics/btaa025> PMID: 31971576
79. Luo J, Lyu M, Chen R, Zhang X, Luo H, Yan C. SLR: a scaffolding algorithm based on long reads and contig classification. *BMC Bioinformatics.* 2019; 20: 1–11. <https://doi.org/10.1186/s12859-019-3114-9> PMID: 31666010
80. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* 2016; 3: 99–101. <https://doi.org/10.1016/j.cels.2015.07.012> PMID: 27467250
81. Smit AFA, Hubley R, Green P. RepeatModeler Open-1.0. 2008–2015. Seattle, USA: Institute for Systems Biology. Available from: <https://github.com/rmhubble/RepeatMasker>
82. Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013–2015. 2015.
83. Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics.* 2011; 12: 491. <https://doi.org/10.1186/1471-2105-12-491> PMID: 22192575

84. Campbell MS, Holt C, Moore B, Yandell M. Genome Annotation and Curation Using MAKER and MAKER-P. *Curr Protoc Bioinformatics*. 2014; 48: 4.11.1–39. <https://doi.org/10.1002/0471250953.bi0411s48> PMID: 25501943
85. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol*. 2016; 17: 132. <https://doi.org/10.1186/s13059-016-0997-x> PMID: 27323842
86. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015; 31: 3210–3212. <https://doi.org/10.1093/bioinformatics/btv351> PMID: 26059717
87. Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*. 2011; 27: 764–770. <https://doi.org/10.1093/bioinformatics/btr011> PMID: 21217122
88. Tang H, Bowers JE, Wang X, Ming R, Alam M, Paterson AH. Synteny and collinearity in plant genomes. *Science*. 2008; 320: 486–488. <https://doi.org/10.1126/science.1153917> PMID: 18436778
89. Tang H, Krishnakumar V, Li J. jvci: JCVI utility libraries. 2015. <https://doi.org/10.5281/zenodo.31631>