bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A telomere to telomere assembly of *Oscheius tipulae* and the evolution of rhabditid nematode chromosomes

Pablo Manuel Gonzalez de la Rosa*, Marian Thomson†, Urmi Trivedi†, Alan Tracey*, Sophie Tandonnet‡ and Mark Blaxter*

* Tree of Life, Wellcome Sanger Institute, Cambridge CB10 1SA, UK

† Edinburgh Genomics, School of Biology, University of Edinburgh, Edinburgh EH9 3JT UK

‡ Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo (USP), São Paulo, SP, Brazil

ORCID IDs

Pablo Manuel Gonzalez de la Rosa https://orcid.org/0000-0002-2009-1531

Marian Thomson

Urmi Trivedi

Alan Tracey https://orcid.org/0000-0002-4805-9058

Sophie Tandonnet https://orcid.org/0000-0001-7559-0154

Mark Blaxter https://orcid.org/0000-0003-2861-949X

Data deposited at INSDC under SRR12179520 and GCA_013425905.1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Rhabditid nematode chromosome evolution

Keywords: Nigon elements, telomere-to-telomere, nematode, evolution, Oscheius

Corresponding Author: Pablo Gonzalez pg17@sanger.ac.uk

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

A telomere to telomere assembly of Oscheius tipulae and the evolution of rhabditid nematode chromosomes **1**

ABSTRACT **7**

INTRODUCTION **8**

METHODS **12**

Nematode culture, DNA extraction and QC 12

Genomic sequencing on Oxford Nanopore PromethION 12

Genome assembly and polishing 13

Genome annotation 15

Comparison to genomes of other Rhabditida and identification of loci supporting Nigon elements 16

KEGG pathway enrichment analysis 17

Data availability 17

RESULTS **18**

A chromosomal assembly of Oscheius tipulae CEW1 from Oxford Nanopore PromethION data 18

Figure 1: Sub-telomeric extensions on the chromosomes of Oscheius tipulae CEW1 21

Figure 1: Sub-telomeric extensions on the chromosomes of Oscheius tipulae CEW1 23

Table 1: Telomeric extension sequences in Oscheius tipulae 24

Comparing chromosome structure in O. tipulae and other nematodes 25

Table 2: Metrics of the Oscheius tipulae CEW1 genome assembly 26

Figure 2: The Oscheius tipulae CEW1 genome 28

Figure 2: The Oscheius tipulae CEW1 genome 30

Figure 3: Local gene order comparison between Oscheius tipulae and Caenorhabditis elegans 31

Figure 3: Local gene order comparison between Oscheius tipulae and Caenorhabditis elegans 31

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosomal elements are conserved across the Rhabditida 33

Figure 4: Loci that define Nigon elements in rhabditid nematodes 34

Figure 4: Loci that define Nigon elements in rhabditid nematodes 35

Independent chromosomes in iB. Malayi and O. volvulus, and a set of Chromosome evolution and homology in the Rhabditida 36

Figure 5: Nigon painting of Rhabditid nematode chromosomes 40

Figure 5: Nigon painting of Rhabditid nematode chromosomes 44

Figure 6: A model of chromosome evolution in Rhabditida 45

Figure 6: A model of chromosome evolution in Rhabditida. 46

Dynamic evolution of rhabditid sex chromosomes 47

Table 3: Nigon elements and the X chromosomes of rhabditid nematodes 49

Functional coherence of Nigon element loci 50

DISCUSSION **51**

New technologies and complete genome sequencing of O. tipulae 51

Evolution of rhabditid nematode karyotypes 52

Figure 7: Sex chromosome evolution in the filarial nematodes (Spiruromorpha) 56

Figure 7: Sex chromosome evolution in the filarial nematodes (Spiruromorpha) 56

Functional coherence of loci that define Nigon elements 57

Outlook 57

ACKNOWLEDGEMENTS **59**

REFERENCES 60

Figure Legends **66**

Figure 1: Dual telomeres on the chromosomes of Oscheius tipulae CEW1 66

Figure 2: The Oscheius tipulae CEW1 genome 66

Figure 3: Local gene order comparison between Oscheius tipulae and Caenorhabditis elegans 67

Figure 4: Loci that define Nigon elements in rhabditid nematodes 67

Figure 5: Nigon painting of Rhabditid nematode chromosomes 68

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 6: A model of chromosome evolution in Rhabditida. 68

Figure 7: Sex chromosome evolution in the filarial nematodes (Spiruromorpha) 69

Supplementary Information **70**

Supplementary Text S1: Defining sets of orthologous loci in linkage through Rhabditid evolution 71

Supplementary Text S2: Nigon element analysis of Meloidogyne hapla 73

Figure S1: Raw data QC and read filtering 76

Figure S1. Raw data QC and read filtering 76

Figure S2: Distribution of SL and 5S RNA loci in the Oscheius tipulae genome 77

Figure S2: Distribution of SL and 5S RNA loci in the Oscheius tipulae genome 77

Figure S3: Clustering with different k values 78

Figure S3. Clustering with different k values 78

Figure S4: t-SNE of orthologous loci co-locations in 9 nematode species 79

Figure S4. t-SNE of orthologous loci co-locations in 9 nematode species 79

Figure S5: nOti 2.0 assembly aligned to nrOscTipu1.3 80

Figure S5: nOti 2.0 assembly aligned to nrOscTipu1.3 80

Figure S6: GC proportion plots of nematode chromosomes 81

Figure S6: GC proportion plots of nematode chromosomes 83

A B. malayi GC proportion in 10kb windows across the chromosomes. 83

Figure S7: Nigon painting of Caenorhabditis species 85

Figure S7: Nigon painting of Caenorhabditis species 87

Figure S8: Exploring the robustness of definition of Nigon element-defining sets of orthologues. 88

Figure S8: Exploring the robustness of definition of Nigon element-defining sets of orthologues. 88

Table S1: Chromosome numbers in Nematoda* 89

Table S2: Software used 90

Table S3: Raw data for genomic analysis of Oscheius tipulae 95

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S4. Sources of nematode assemblies included in the Kraken2 database for contamination assessment. 96

Table S5. Raw sequence allocation by Kraken grouped by Phylum 98

Table S6. Contiguity and BUSCO metrics of Oscheius tipulae assemblies 100

Table S7. Base fold coverage of the Oscheius tipulae assembly 102

Table S8: Sources and metrics of rhabditid nematode genome data 103

Table S9. BUSCO scores of assemblies and proteomes of rhabditid nematodes 105

Table S10: Nigon element-defining loci in fifteen nematode genomes 106

Table S11. KEGG categories enriched among Nigon defining loci 107

Table S12: Nigon-defining loci in the Meloidogyne hapla genome assembly 108

Supplementary references 111

Online data and code **116**

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Eukaryotic chromosomes have phylogenetic persistence. In many taxa, the number of chromosomes is related to the number of centromeres. However, in some groups, such as rhabditid nematodes, centromeric function is distributed across multiple sites on each chromosome. These holocentric chromosomes might, *a priori*, be expected to be permissive of large-scale chromosomal rearrangement, as chromosomal fragments could still partition correctly and fusions would not generate lethal conflict between multiple centromeres. Here, we explore the phylogenetic stability of nematode chromosomes using a new telomere-to-telomere assembly of the rhabditine nematode *Oscheius tipulae* generated from nanopore long reads. The 60 Mb *O. tipulae* genome is resolved into six chromosomal molecules. We find evidence of specific chromatin diminution at all telomeres. Comparing this chromosomal *O. tipulae* assembly with chromosomal assemblies of diverse rhabditid nematodes we identify seven ancestral chromosomal elements (Nigon elements), and present a model for the evolution of nematode chromosomes through rearrangement and fusion of these elements. We identify frequent fusion events involving NigonX, the element associated with the rhabditid X chromosome, and thus sex-chromosome associated gene sets differ markedly between species. Despite the karyotypic stability, gene order within chromosomes defined by Nigon elements is not conserved. Our model for nematode chromosome evolution provides a platform for investigation of the tensions between local genome rearrangement and karyotypic evolution in generating extant genome architectures.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

INTRODUCTION

Linear chromosomes are basic elements of the organisation of eukaryotic nuclear genomes. The number of chromosomes and the position of orthologous loci on them is generally conserved between closely-related species, and conserved karyotypic elements have been identified even between distantly related taxa (Nakatani et al., 2007; Putnam et al., 2008). The evolutionary trajectories of genes, in terms of rate of drift and efficiency of selection, are influenced by their chromosomal location. For example, genes on sex chromosomes will be exposed as haploid in the heterogametic sex (whether X0, XY or WZ), and their effective population size will be only 0.75 that of autosomal loci. More subtly, genes resident on longer chromosomes may be more affected by linked selection, as the number of recombination events is frequently limited to one per chromosome (Hammarlund et al., 2005) or chromosome arm, and the number of bases per centiMorgan will be larger in longer chromosomes. Gene evolution is also shaped by placement within chromosomes, with some regions, such as centromeres and subtelomeric regions experiencing higher rates of per-base and structural change (Rockman and Kruglyak, 2009). On longer timescales, genes that have travelled together on single chromosomes might evolve to share dependence on long-range regulatory landscapes, such as the three-dimensional topologically associated domains that characterise chromosomal organisation within interphase nuclei. For some sets of loci, such as HOX and paraHOX loci in most Metazoa, this constraint is evident between organisms that last shared common ancestors hundreds of millions of years ago (Krumlauf, 2018). Holocentric chromosomes offer a contrasting pattern of organisation to centromeric karyotypes, and this must affect the mode and tempo of genome and gene evolution.

Chromosome structural change is an important component of genome and species evolution (Sturtevant and Dobzhansky, 1936). Chromosomal elements, sets of loci that have been colocated on the same linkage group for long periods of evolutionary time, have been identified in many taxa, including mammals, Diptera (Bhutkar et al., 2008), Lepidoptera (d'Alençon et al., 2010) and Nematoda (Tandonnet et al., 2019). While many groups have deeply conserved karyotypes, species that have very different

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

numbers of chromosomes or synteny relationships to closely related taxa will allow exploration of the constraints that act to retain chromosome number and gene content, and also of the mechanisms that are involved in karyotypic evolution.

Most animals (Metazoa) have chromosomes with a defined centromere. However several groups have holocentric chromosomes, where centromeric function is distributed across each chromosome and there is no single pairing centre during meiosis (Albertson and Thomson, 1982; Melters et al., 2012). *A priori*, holocentric organisation might be thought to predispose a genome to increased rates of rearrangement both within and between chromosomes, as any chromosome fragment remaining after fission could still carry centromeric function, and fusions of chromosomes would not result in competing centromeres on the same molecule. Lepidoptera have holocentric chromosomes and generally conserved karyotypes (d'Alençon et al., 2010). The ancestral lepidopteran chromosome number is estimated to be 31, and while the genomes of some species that have fewer chromosomes, such as the genus *Heliconius* (where n=21) can be modelled through a series of simple fusions (d'Alençon et al., 2010), others, such as *Pieris napi* (the green-veined white butterfly; n=25) exhibit extensive rearrangement, including presumed ancestral linkage group fragmentation and fusion (Hill et al., 2019). Thus, to distinguish conservation of chromosome number *per se* from conservation of linkage groups, and to define the patterns and processes involved in changes in karyotype, complete, telomere-to-telomere chromosomal assemblies are needed (Hill et al., 2019).

Nematode chromosomes are also holocentric. The model nematode *Caenorhabditis elegans* (Rhabditomorpha, Rhabditina, Rhabditida; see (De Ley and Blaxter, 2002)) has n=6 and an X0 sex determination mechanism. In the order Rhabditida, n=6 is the commonest karyotype (Table S1) (Walton, 1959),, but n varies between 1 (e.g. *Diploscapter coronatus*, closely related to *Caenorhabditis*) (Fradin et al., 2017) and >50 (e.g. *Meloidogyne* polyploids) (Triantaphyllou, 1963). Conservation of karyotype implies that sets of genes will be colocated on the same chromosome for long evolutionary periods. The retention of syntenic groups of loci can be uncoupled from retention of large-scale gene order. In *Caenorhabditis* species (all with n=6) orthologous genes are overwhelmingly located on orthologous chromosomes but the order of genes is very

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

different between species due to rampant within-chromosome rearrangement (Stein et al., 2003; Stevens et al., 2020; Teterina et al., 2020). This pattern is also seen when comparing *Caenorhabditis* to other genera: linkage groups tend to be conserved but gene order is not (Doyle et al., 2019). In contrast, both linkage groups and gene order within each linkage group are conserved in the similarly holocentric Lepidoptera (Mongue et al., 2017).

Some nematodes have different karyotypes in their somatic cells compared to their germline. This process involves scission of germline chromosomes and loss of germline material, and is called chromatin diminution (Wang and Davis, 2014). Chromatin diminution has been observed in several metazoan taxa, including chordates (Kinsella et al., 2019) and arthropods, and is involved in the generation of the ciliate macronucleus (Rzeszutek et al., 2020). The process of diminution is best understood in *Ascaris suum* (Ascarididomorpha, Spirurina, Rhabditida), where the germline has n=24 chromosomes but somatic cells have n=36 (Wang et al., 2020). In *A. suum* the breakage events affect some but not all of the X chromosomes (*A. suum* has five X chromosomes) and autosomes, and breakage and neo-telomere addition happens in a defined area of the chromosome, but not at a precise base position. Related ascarididomorph nematodes also display diminution. Chromatin diminution has also been described in the tylenchomorph nematode *Strongyloides papillosus,* where loss of a specific internal fragment of one copy of the X chromosome generates a haploid region that is associated with males (i.e. sex determination in *S. papillosus* is effectively XX:X0, but the nullo-X is determined through specific deletion).

Previously we proposed the existence of seven ancestral chromosome elements in rhabditine nematodes, named Nigon elements, and used this model to understand chromosome evolution in a few genome-sequenced Rhabditina (Tandonnet et al., 2019). The lack of chromosomally-complete genomes limited the power of the model. Here we present an improved, chromosomal genome assembly of *Oscheius tipulae* (Rhabditomorpha, Rhabditina, Rhabditida). *O. tipulae* is a satellite genetic model organism that is used to understand the evolution of developmental systems such as the specification of the nematode vulva, and the genome sequence is required to underpin detailed genetic mapping (Besnard et al., 2017). The new

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

telomere-to-telomere assembly allowed us to identify unexpected features of chromatin diminution at the telomeres of each chromosome. We used the *O. tipulae* genome and other chromosomally-complete nematode genomes to fully define sets of orthologous genes associated with ancestral Nigon elements. This analysis allowed us to map ancient chromosomal fusions and scissions and identify a set of genes that is always associated with the X chromosome in both X0 and XY taxa in the order Rhabditida.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

METHODS

Nematode culture, DNA extraction and QC

Oscheius tipulae strain CEW1 (Evans et al., 1997) was obtained from Marie-Anne Félix (Institute of Biology of the Ecole Normale Supérieure, Paris), and cultivated at 20°C in 5 cm nematode growth medium lite plates seeded with *E. coli* HB101 (Stiernagle, 2006). Nematodes were washed from culture plates using an M9 buffer supplemented with 0.01% Tween 20. Nematodes were pelleted by low-speed centrifugation, and 100 µL samples transferred with minimal supernatant to 1.5 mL LoBind Eppendorf tubes. Nematodes were lysed by addition of 600 µL of Cell Lysis Solution (Qiagen) and 20 µL of proteinase K (20 µg/µL) and incubated at 56 °C with mixing at 300 rpm for 4 h. RNA was digested by adding 5 µL of RNAse Cocktail Enzyme Mix (Invitrogen) and incubating at 37°C for 1 h. Protein was precipitated by adding 200 µL of ice-cold Protein Precipitation Solution (Qiagen), gentle mixing and incubation on ice for 10 minutes. The precipitate was pelleted by centrifugation for 30 min at 4°C at 15,000 RPM. The supernatant was transferred to a LoBind tube and nucleic acids precipitated by addition of 600 µL of ice cold isopropanol, mixing by inversion, and incubation on ice for 10 min. Nucleic acids were pelleted by centrifugation at 4°C at 15,000 RPM. The supernatant was discarded and the pellet washed twice using 600 µL of 70% ethanol. The pellet was air dried for 5 min and resuspended in 20 µL of Elution Buffer. DNA recovery and quality was assessed by Qubit fluorimetry, Tapestation genomic Screentape (Agilent) and pulsed field gel electrophoresis using a Pippin Pulse instrument. The sample used for sequencing had a DNA concentration of 120 ng/µL, a DNA integrity number of 9 and an RNA concentration of 9 ng/µL.

Genomic sequencing on Oxford Nanopore PromethION

To generate fragments of a suitable size range for Oxford Nanopore PromethION sequencing, high molecular weight DNA was diluted to a concentration of 25 ng/uL and fragmented to an average peak size of 25 kb using a Megaruptor-2 instrument (Diagenode). Small fragments <1 kb were removed and the DNA concentrated using bead purification (0.4 x volumes of Ampure-XP beads. Two aliquots of 1 µg of sheared

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

O. tipulae DNA and control DNA (lambda 3,2 kb fragment) were subjected to DNA damage repair (NEBNext FFPE DNA Repair Mix; New England Biolabs) followed by DNA End Repair (NEBNext Ultra II End Repair/ dA-tailing Module; New England BioLabs). A second 0.4 x volume Ampure-XP bead clean up was carried out and DNA eluted in sterile distilled H_2O. Oxford Nanopore sequencing adapters were ligated to 750 ng of the recovered, end repaired DNA using the Ligation Sequencing kit (SQK-LSK-109; Oxford Nanopore) and NEBNext Ligation Module (New England BioLabs). Following a further 0.4 x volume Ampure-XP purification, the recovered DNA (16.25 fmol) was loaded onto a R9.4.1 PromethION flow cell following the manufacturer's instructions and a 60 hr sequencing run was initiated.

Raw reads were basecalled using Guppy (see Table S2 for software tools and settings used). The resulting dataset of 8.8 M reads spanned 108.4 Gb and had a read N50 of 19.1 kb (Figure S1 A; Table S3). To identify sequence contamination, we assembled a custom kraken2 database composed of bacteria, fungi, human, UniVec core and a selection of nematode genomes including the previous *O. tipulae* assembly (Table S4). The vast majority of reads (99.5%) were classified. One fifth (19.1% of the total bases) were classified as Nematoda, and the remainder as Proteobacteria (97.5% of these belonging to *Escherichia*; these likely derive from bacterial food). Minor human, fungal and other bacterial contamination was also present (Table S5). We removed reads classified as Bacteria, Chordata, Ascomycota, Basidiomycota or Microsporidia. The remaining data spanned 20.7 Gb in 2.8 million reads with a read N50 of 14.4 kb (an estimated ~340 fold coverage).

Genome assembly and polishing

Several different assembly strategies were explored (Table S6). Flye (Kolmogorov et al., 2019) in metagenome mode with the whole long read set yielded a chromosome level assembly of *O. tipulae* together with contaminant species (Figure S1 B). This assembly was polished using Racon (Vaser et al., 2017) and medaka using the decontaminated read set. For Pilon polishing (Walker et al., 2014), Illumina reads were trimmed with BBDuk (Bushnell, 2017) and aligned with BWA-MEM (Li and Durbin, 2009). We derived the chromosome assembly nOti 3.1 by stitching back the two

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sequences of chromosome I with RaGOO (Alonge et al., 2019) using the unpolished assembly as a reference. An alternate assembly was obtained using Flye in metagenome mode with only 40x Canu-corrected, decontaminated reads, followed by Racon and Pilon polishing. This assembly had higher BUSCO completeness than nOti 3.1, but was more fragmented. We derived a new consensus, nOti 3.2, from nOti 3.1 and the decontaminated-read Flye assembly using gap5 (Bonfield and Whitwham, 2010) giving the decontaminated read assembly a 100x relative weight. The resulting contigs were assigned chromosome names by the longest match to contigs in nOti 2.0, which were previously assigned to chromosomes (Besnard et al., 2017). Alignment of the raw reads against nOti 3.2 showed that the nuclear genome had an average per-base coverage of 334 fold (standard deviation of 198). The initial Oti_chrV sequence had the highest coverage and coverage heterogeneity (354 fold, SD 478) due to collapse of the ribosomal RNA cistron between positions 7413040 and 7440271 (see below; Table S7).

We curated the nOti 3.2 assembly by examining read coverage across the genome. A gap5 (Bonfield and Whitwham, 2010) database was built from a 200x sub-sample of the longest PromethION canu corrected reads. We noticed that all the chromosomes were characterised by a shorter majority sequence (80% of the average coverage depth) and a longer minority sequence (20% coverage depth). Both of these sequences terminated in telomeric repeat (long tandem repeats of TTAGGC). These alternate telomeric repeat addition sites appeared not to be artefacts because long reads supporting both versions were anchored in unique sequence. Previous Illumina short read data also identified similar major and minor components of the chromosome ends, and supported the same telomeric repeat addition sites. We manually extended all reads containing soft-clipped telomeric repeat sequence and then used the gap5 realign function to produce a new consensus from them. The left hand end of the Oti_chrIV sequence produced directly by the assembler was characterised by an artificial sequence as evidenced by a lack of reads that mapped to it and all reads being soft-clipped either side of it. This sequence was replaced with realigned, soft-clipped sequence from the adjacent mapped reads. Restoration of these soft-clipped data identified telomeric repeat at both ends of each chromosome. We

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

estimated the size of the highly collapsed ribosomal RNA cistron repeat on Oti_chrV. The rRNA cistron repeat was estimated to be 6.8 kb long and to be present in 117 copies, based on its coverage by Illumina short reads. The left hand side of the rRNA cistron repeat terminated in an obviously mispredicted sequence which was dealt with as for the Oti_chrIV telomere sequence. We extended the reads at the junctions into and out of the rRNA cistron repeat sequence to a minimum depth of 2 reads. We joined these flanks together using 760128 "N" characters to match 111 additional copies of the 6.8 kb rRNA cistron repeat. We polished this assembly via three rounds using freebayes through snippy (Seemann, 2014) with Illumina reads. We identified spliced leader RNA (SL) and 5S rRNA loci using Rfam models (Nawrocki et al., 2015) (Figure S2).

Genome annotation

We created a repeat library for our assembly following published protocols (Coghlan et al., 2018). Briefly, we identified repetitive sequences with RepeatModeler2 (Flynn et al., 2020), transposons with TransposonPSI (Haas, 2007) and Long Terminal Repeats (LTR) with LTRharvest (Ellinghaus et al., 2008). We discarded TransposonPSI predictions shorter than 50 bp. We filtered out LTRharvest predictions that lacked PFAM (Finn et al., 2016) and GyDB (Llorens et al., 2011) hidden Markov model domain hits using LTRdigest (Steinbiss et al., 2009). The three prediction sets were classified with RepeatClassifier and merged into a single library. Sequences were clustered if they had more than 80% identity using USEARCH (Edgar, 2010). This repeat library, together with the CONS-Dfam_3.1-rb20181026 database (Hubley et al., 2016), was used by RepeatMasker to annotate the repetitive regions in the assembly. We predicted protein coding genes using GeneMark-ES (Lomsadze et al., 2005). We explored helitron predictions using HelitronScanner (Xiong et al., 2014). For nucleotide and protein-coding gene comparisons in the telomere extensions we used BLAST+ (Camacho et al., 2009), ClustalW (Thompson et al., 2002) and JalView (Waterhouse et al., 2009). We identified genes and other sequence features of telomeric extensions using BLAST similarity searches and hidden Markov model searches using Dfam helitron models (Wheeler and Eddy, 2013). Images were generated using the circos toolkit (Krzywinski et al., 2009).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Comparison to genomes of other Rhabditida and identification of loci supporting Nigon elements

We assessed chromosomal assemblies and annotations of fifteen rhabditid nematode species (Table S8, Table S9). We extracted the longest protein per gene from genome annotation GFF3 files using AGAT (Dainat, 2020). Orthogroups were identified by Orthofinder (Emms and Kelly, 2019, 2015) using these proteomes. Orthogroups were filtered using Kinfin (Laetsch and Blaxter, 2017) to identify fuzzy single copy orthologues in at least 12 of the 15 species compared (see Supplementary Text S1).

We identified loci that are found on the same chromosomal unit in multiple species using a subset of nine genome assemblies that are resolved into chromosomes: *Auanema rhodensis, Brugia malayi, Haemonchus contortus, Meloidogyne hapla, Onchocerca volvulus, Oscheius tipulae, Pristionchus pacificus, Steinernema carpocapsae* and *Strongyloides ratti*. For *Meloidogyne hapla*, contigs were grouped into chromosomes according to the genetic linkage map (Opperman et al., 2008). For *Steinernema carpocapsae*, only the X chromosome was assembled to completeness, while the four autosomes were present as twelve unassigned scaffolds. To reduce phylogenetic bias we used a single *Caenorhabditis* species (*C. elegans*).

The chromosomal location of each single copy orthologue in each species was extracted from genome GFF3 files and collated in an orthologue-chromosomal allocation matrix by species. Scaffolds "a" and "b" of *O. volvulus* chromosome 1 were treated as a single chromosome, as were the *S. ratti* X chromosome scaffolds. We calculated the Dice distance (represented as 1-Dice distance) between all orthologue pairs based on the pattern of their chromosomal allocation between species. A pair of genes found on the same chromosome in all the species would have a similarity of 1, while a pair found on different chromosomes in all the species would have a similarity of 0. This matrix was clustered with CLARA (Clustering Large Applications) using an expected number of clusters, k, from 1 to 10 (Figure S3). CLARA identified the medoids (center of the clusters) by applying PAM (partitioning around medoids) to 5 independent subsamples of 10% of the orthologous loci. To identify the best number of clusters we assessed average silhouette values which will be higher when the average

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

loci is closer to members of the same clusters than to loci of other clusters. The highest silhouette value was found at k=7 when using CLARA or PAM clustering (Figure S3 A and D). CLARA was used because it reduced computation time. Similarly, 7 clusters were identified by the t-Distributed Stochastic Embedding (t-SNE) plot. These clusters were also found when using different perplexity values, which can be interpreted as the number of effective nearest neighbors, from 30 to 1000 (Figure S4). Orthologues were assigned to clusters as long as at least 7 taxa agreed on their colocation with other orthologues in the cluster. Cluster numbers were assigned to putative element groups when they contained more than 20 single copy orthologues, and allocated to Nigon element labels according to our previous classification (Tandonnet et al., 2019).

KEGG pathway enrichment analysis

We assessed functional enrichment of KEGG pathways among each Nigon defining loci set using the *C. elegans* representatives through gProfiler (Raudvere et al., 2019). We downloaded the *C. elegans* genes KEGG annotations using the R package KEGGREST (Tenenbaum, 2019). We used a Fisher exact test controlling the false discovery rate by Benjamini-Hochberg p-value correction to assess pathway enrichment using the 2175 *C. elegans* Nigon defining loci as comparator.

Data availability

The raw PromethION data are available in INSDC under accession SRR12179520 associated with the BioSample SAMN15480678. The genome assembly has been deposited in INSDC with accession number GCA_013425905.1. The data used to generate the Tables and Figures are available at https://docs.google.com/spreadsheets/d/1gO4j4jSgSYQ_Aofl59RdGbwHrz2L2loWfrDIyyRaohw/edit?usp=sharing. Scripts and intermediate files associated with this study are available in https://github.com/tolkit/otipu_chrom_assem (commit dc73c57) under a GPL-3.0 License.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

RESULTS

A chromosomal assembly of *Oscheius tipulae* CEW1 from Oxford Nanopore PromethION data

Our previous assembly of *O. tipulae* CEW1 had a contig N50 less than 1 Mb, and was resolved to chromosomes using genetic map data, which necessarily excluded contigs with no mapped loci and could not orientate contigs mapped through a single genetic locus (Besnard et al., 2017). To generate a telomere-to-telomere, chromosomally complete *O. tipulae* genome we resequenced CEW1 using the Oxford Nanopore Technologies PromethION platform, which generates large numbers of long reads from single molecules. After removing reads from bacterial contamination, we retained approximately 340 fold coverage of the expected 60 Mb genome (Table S5). We explored assembly options using a range of tools, and polished the assemblies to correct remaining errors with previously obtained Illumina short read data. Twelve assemblies were generated, with nematode sequence contiguities ranging from 7 to over 900 contigs (Table S6). The most contiguous assemblies had seven contigs, six with multi-megabase lengths and one corresponding to the mitochondrial genome. Each of the assemblies contained about 90% of the BUSCO set of conserved nematode orthologues. An assembly generated using Flye (Kolmogorov et al., 2019) in metagenome mode, combining data from decontaminated and full read sets, scored best using BUSCO and contiguity measures. We curated the genome by circularising the mitochondrion, and identifying and correcting two remaining issues in the nuclear sequence. The ribosomal RNA repeat cistron was present as a collapsed and jumbled sequence, as was the case until recently with the *C. elegans* complete genome sequence. A gap representing the estimated span of the repeat was inserted. While the 5S rRNA and spliced leader 1 RNA loci are present as a tandem repeat in *C. elegans* and related nematodes, the 5S and SL1 RNA genes were not clustered in *O. tipulae* (Figure S2).

The second issue concerned the chromosomal termini. Only four of the twelve ends of contigs in the initial assembly ended in telomeric repeat sequence ($[TTAGGC]_n$, the same repeat as found in *C. elegans*). We were able to extend each contig end into

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

telomeric repeat using long reads that overlapped the unique sequence at each end of each contig. The assembly was thus judged complete, telomere-to-telomere. The six longest contigs corresponded to the six genetically-defined linkage groups of *O. tipulae*, and we named these Otip_I, Otip_II, Otip_III, Otip_IV, Otip_V and Otip_X, following the previously defined chromosome nomenclature (Besnard et al., 2017). However, at each telomere we identified two independent and specific sites where there was transition from unique sequence to tandem hexamer telomere repeat: an internal site supported by 80% of the PromethION reads, and an external site, supported by a minority (20% of reads). The extension sequences were confirmed by relatively even, 60-fold coverage of PromethION reads and by matching reads in Illumina short-read data, which also showed a majority-short and minority-long distribution (Figure 1). These telomere repeat addition sites define three components on each chromosome: a central portion, present in all copies, and two sub-telomeric extension sequences, from the regions at the left and right ends, present in a minority of copies. In turn this implies that *O. tipulae* chromosomes are each found as long-form (carrying the extensions) and short form (lacking the extensions) versions. We cannot exclude the possibility that absence of the extensions on the left and right ends of each chromosome is determined independently, but the very similar proportional coverage of all extensions strongly suggests that chromosomes are either all short form or all long form in each cell.

The sub-telomeric extensions ranged from 4 kb to 133 kb (excluding the telomeric hexamer repeats), and totaled 349 kb (Table 1; Figure 1 B). The assembly included an additional 150 kb of telomeric repeat. The sub-telomeric extension sequences had an unremarkable GC proportion (45% - 49%) and contained unique sequence, including predicted protein coding loci that had support from uniquely-mapping transcript evidence (Besnard et al., 2017). Notably, the telomere extensions contained repeat families that were largely limited to the extensions (Table 1; Figure 1 B). One predicted protein coding gene overlapped the internal telomere addition site (gene 632t, encoding a neprilysin M13 metallopeptidase homologue, on Otip_V right end (Otip_V R)). This gene prediction appears to represent a full-length locus, as it aligns well with *C. elegans* orthologues, and would therefore be predicted to be non-functional in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

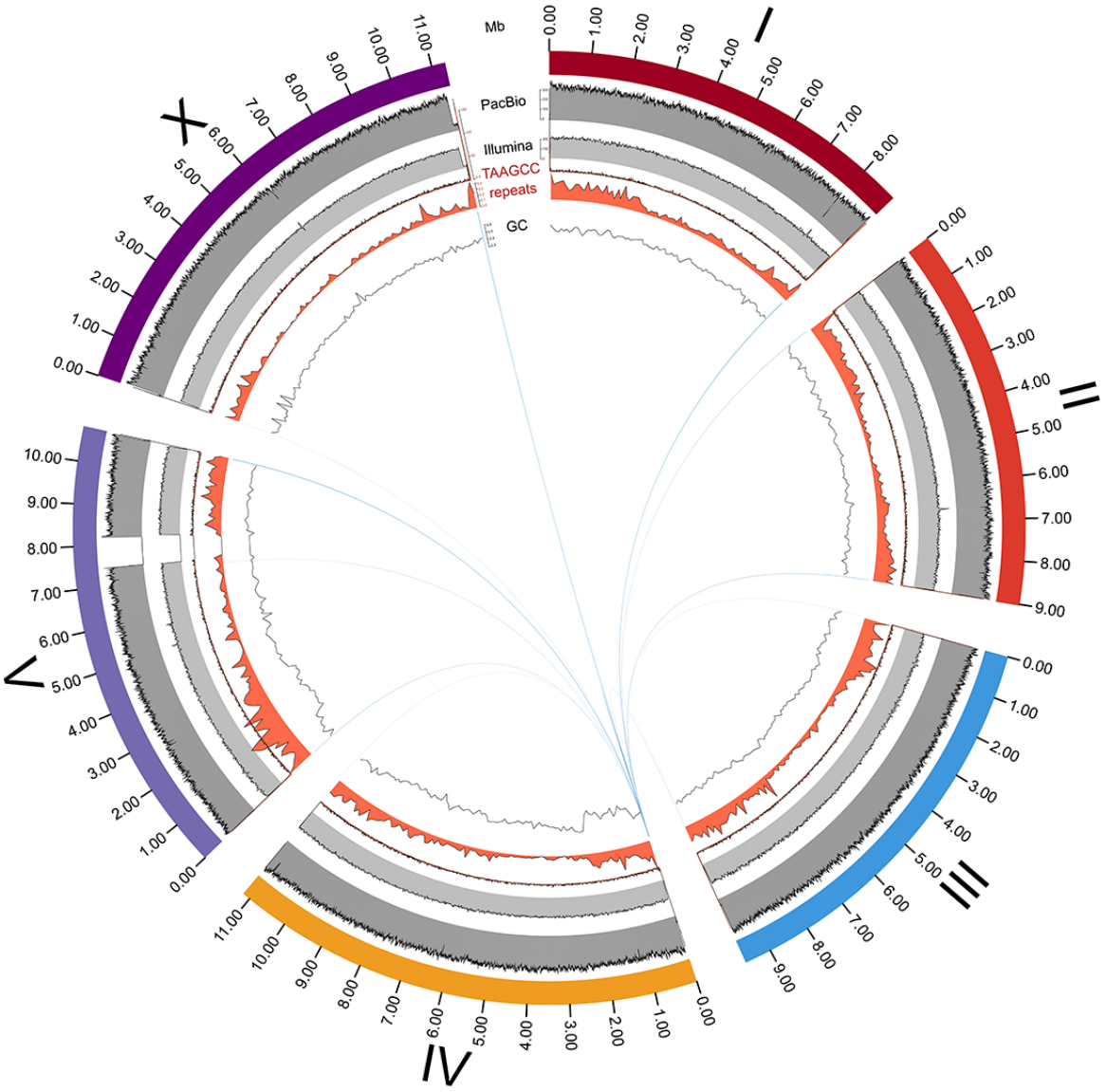
short-form chromosome. The repeat sequences unique to the sub-telomeric extensions were predicted to encode protein coding genes that had similarities to helitron transposon genes (Figure 1). The longest helitron-like predicted gene, 9690_t on Otip_IV left end (Otip_V L) (3985 amino acids) contains domain matches to Pif1-like, ATP-dependent DEAH box DNA helicases. Sequences similar to this putative helitron-like gene are present on eleven of the twelve telomere extensions (it was absent from the Otip_I L extension; Figure 1 A, Table 1). A single additional match was found on Otip_V, near the rRNA cistron repeat. Scanning of the genome for helitron-like sequences using RepeatFinder (Tarailo-Graovac and Chen, 2009) and HelitronScanner (Xiong et al., 2014) identified many additional, distinct helitron-like sequences, scattered across all the chromosomes. However, the telomere-associated helitron-like elements scored poorly in these searches, suggesting they are a new, perhaps distinct family.

The final nuclear genome assembly was a significant improvement over the previous reference in contiguity and completeness (Besnard et al., 2017) (Table 2). The 253 contigs of the previous assembly were super-scaffolded using genetic map information. All but eight of these previous assignments were replicated in the new assembly, and the data underlying the new assembly affirmed the eight new assignments. The longest contigs from the previous assembly tended to be found in the centres of the chromosomal contigs, and the ends of the new chromosomal contigs were represented by multiple shorter contigs in the previous assembly (Figure S5). The proportion of complete nematode BUSCO loci (nematoda_odb10) identified in both assemblies was 90%. Close analysis of the assemblies identified candidates for some of the 254 apparently missing orthologues. We note that similar low BUSCO completeness scores were recorded for the closely-related nematode *Auanema rhodensis* (Table S9).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

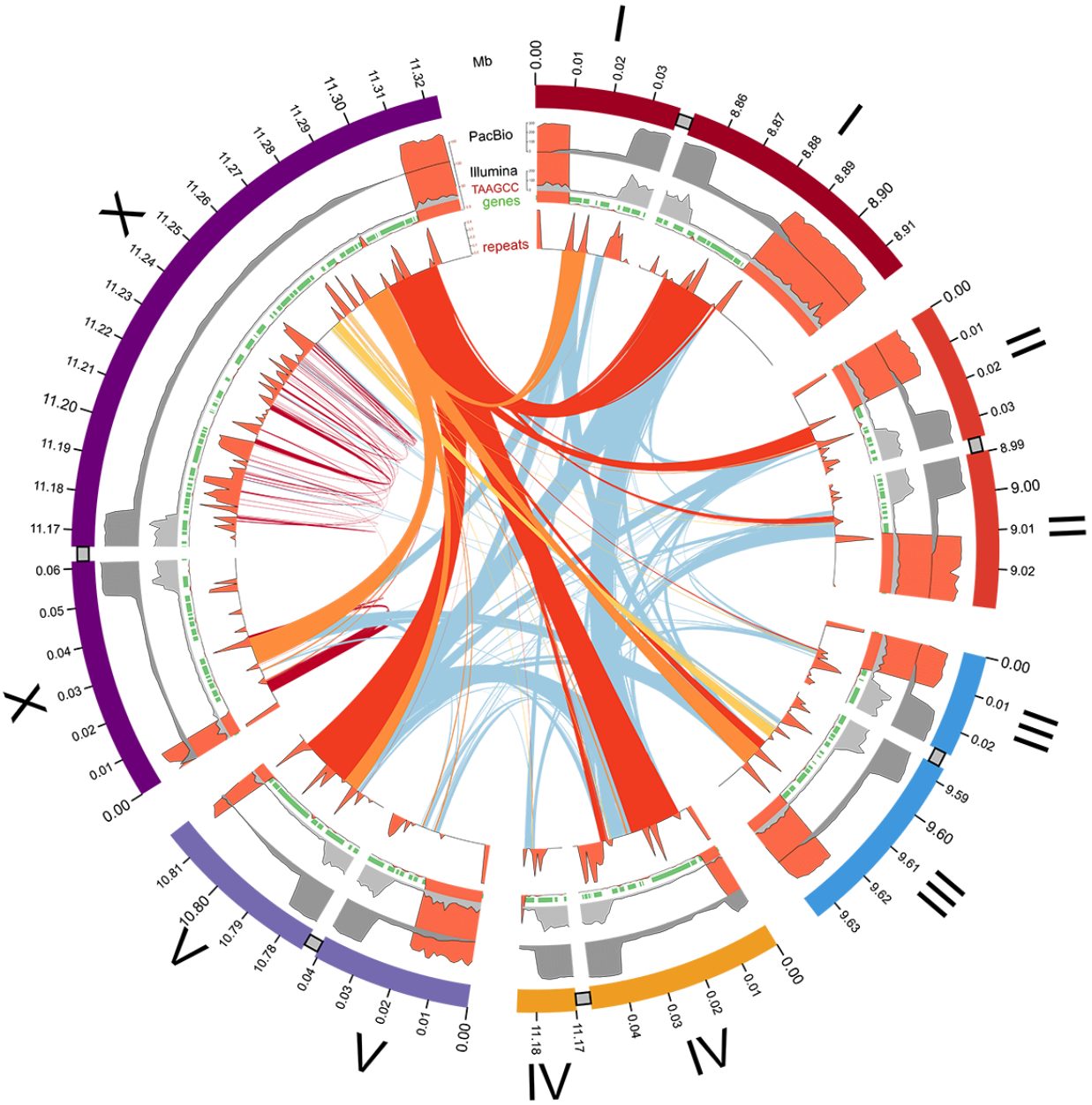
Figure 1: Sub-telomeric extensions on the chromosomes of *Oscheius tipulae* CEW1

A



bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

B



bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1: Sub-telomeric extensions on the chromosomes of *Oscheius tipulae* CEW1

A Circos plot of the *O. tipulae* CEW1 genome showing (from outside to inside) the chromosomes with a scale in Mb, coverage in PromethION reads (dark grey), coverage in Illumina short reads (light grey), count of telomeric repeats per 10 kb (red), density of repeats (red) and GC proportion in 10kb windows (line plot). The inner arcs link all significant nucleotide sequence (blastn) matches between the presumed near-complete copy of the helitron on the left end of Otip_IV (with an open reading frame of 3895 amino acids, spanning 26 kb of the telomere extension) and the rest of the genome.

B Circos plot of the low-coverage sub-telomeric extension sequences of *O. tipulae* CEW1, with 10 kb of flanking chromosome, showing (from outside to inside) the chromosomes with a scale in 100 kb, coverage in PromethION reads (dark grey), coverage in illumina short reads (light grey), count of telomeric repeats per 10 kb (red), density of repeats (red) and predicted coding genes (green). Arcs are drawn linking significant blastn matches between three sequential helitron-like components from the right telomere extension of chromosome X (yellow, orange and red links), between repeats limited to within Otip_X R (dark red), and between all other telomeric extension sequences (light blue).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 1: Telomeric extension sequences in *Oscheius tipulae*

Chromosome	*Arm*	*Extension span (bp)*	*Sub-telomeric extension start site **	*External TTAGGC telomere repeat addition site **	*External TTAGGC telomere repeat span (bp)*
Otip_I	**L**	**16822**	27184	10363	10362
	R	**24306**	8885185	8860879	31405
Otip_II	**L**	**10122**	25180	15059	15058
	R	**10413**	9010174	8999761	19801
Otip_III	**L**	**4060**	16298	12239	12238
	R	**25188**	9621518	9596330	13932
Otip_IV	**L**	**34114**	39782	5669	5668
	R	**4711**	11184555	11179844	463
Otip_V	**L**	**23599**	30663	7065	7064
	R	**29106**	10812563	10783457	4916
Otip_X	**L**	**32697**	51845	19149	19148
	R	**133498**	11309184	11175686	14155
Total		**348636**			**154210**

* Base position with reference to the full chromosome sequence length (including telomere hexamer repeat and telomere extensions). This position is the site of addition of telomeric hexamer repeat in the short-form chromosomes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Comparing chromosome structure in *O. tipulae* and other nematodes

A striking feature of the *C. elegans* genome is the strong patterning of genic and non-genic features along each chromosome (C. elegans Sequencing Consortium, 1998). Repeats are more abundant on the autosome arms, and are largely excluded from the chromosome centres, while GC proportion has higher variance in the arms. The *C. elegans* X chromosome has a similar pattern albeit less pronounced. These patterns are likely to be generated by the local recombination rate, which is high on autosome arms and lower in the autosome centres and on the X chromosome (Rockman and Kruglyak, 2009). We explored the chromosomal *O. tipulae* genome for similar patterns. The arms of *O. tipulae* chromosomes show a higher repeat and intronic sequence fraction and a lower exonic fraction compared to the centres (Figure 2). The pattern of GC proportion along *O. tipulae* chromosomes also matched expectations from *C. elegans* (and other species, Figure S6), except for two regions. Both Otip_IV and Otip_V have regions where there is a step change in GC proportion (Figure 1 A, inner circle). The boundaries of these step changes were examined and were supported by a normal (320 fold) coverage of PromethION reads. We interpret these as recent inversions where background processes have not yet restored the local GC proportion pattern.

As expected from comparisons within the genus *Caenorhabditis* (Slos et al., 2017) (Stevens et al., 2019) (Stevens et al., 2020) and between *Caenorhabditis* and the strongylomorph *H. contortus (Doyle et al., 2020)*, while neighbouring orthologous genes tended to be found on the same linkage groups in *O. tipulae* and in *C. elegans*, local gene neighbourhoods were not highly conserved (Figure 3). Interestingly, genes in the centres of the *C. elegans* autosomes were not more likely to be retained in the centres of *O. tipulae* chromosomes, and gene neighbourhood conservation was largely absent apart from conserved operonic gene sets.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 2: Metrics of the *Oscheius tipulae* CEW1 genome assembly

Species	***Oscheius tipulae* CEW1 nrOscTipu1.3**	***Oscheius tipulae* CEW1 Ot_2.0**	***C. elegans* PRJNA13758 WBPS12**
Reference	This work	(Besnard et al., 2017)	From WormBase §
Genome			
Span (Mb)	60.9	59.5	100.3
Number of Scaffolds	6 (+MT*)	191	6 (+MT)
Scaffold N50 (Mb)	10.8 (chromosomal)	1.2	17.5 (chromosomal)
Number of contigs	6 (+MT)	256	6 (+MT)
Contig N50 (Mb)	10.8 (chromosomal)	0.7	17.5 (chromosomal)
Genome BUSCO **	C:90.7%[S:88.7%,D:2.0%],F:1.8%,M:7.5%	C:91.6%[S:89.0%,D:2.6%],F:1.6%,M:6.8%	C:99.4%[S:98.9%,D:0.5%],F:0.1%,M:0.5%
Annotation			
Proportion repeat	7.93%	7.85 %	21.95 %
Proportion coding ***	22.6 (37.09%)	19.1 (32.18 %)	28.8 (28.68 %)
Proportion intron	16.6 (27.29%)	16.6 (27.9 %)	33.1 (33.02 %)
Number of protein coding genes	16367¶	14626	20184
Proteome BUSCO **	C:90.2%[S:87.5%,D:2.7%],F:1.9%,M:7.9%	C:89.1%[S:86.2%,D:2.9%],F:2.0%,M:8.9%	C:100%[S:99.6%,D:0.4%],F:0%,M:0%

§ See https://parasite.wormbase.org/Caenorhabditis_elegans_prjna13758/Info/Index/

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

* MT: mitochondrial genome

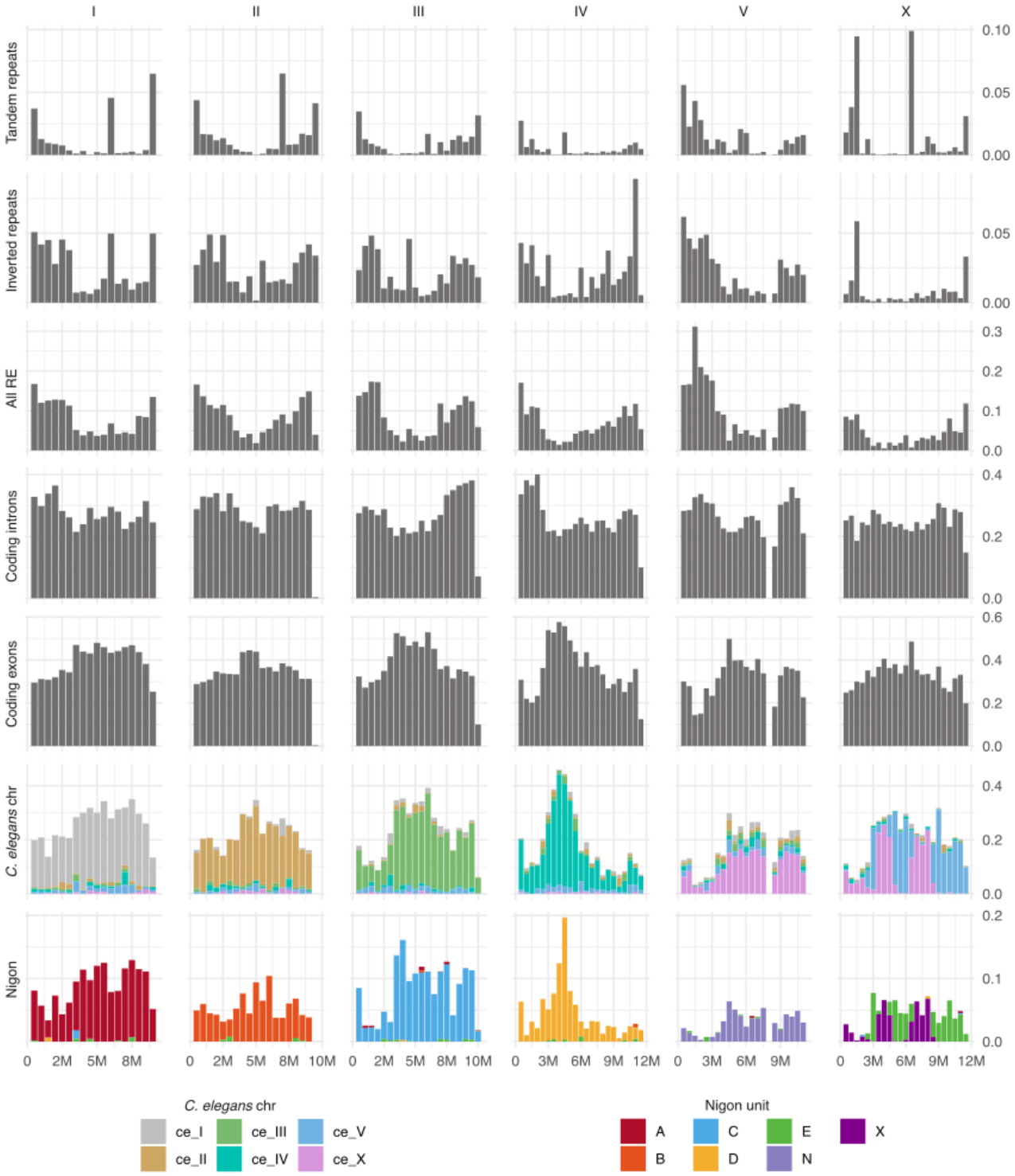
** BUSCO scores assessed against nematoda_odb10 (n=3131); reported as C = complete, S = single copy, D = duplicated, F = fragmented, and M = missing.

*** Using the longest isoform per gene.

¶ Including 698 loci with similarity to mobile elements.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2: The *Oscheius tipulae* CEW1 genome



bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2: The *Oscheius tipulae* CEW1 genome

Repeat and gene densities along *O. tipulae* chromosomes, and mapping of placement of orthologues of *C. elegans* and Nigon element assigned loci. Feature densities (repeat and gene features and orthologue mapping) were calculated in 500 kb non-overlapping windows along each chromosome. For the mapping of orthologues (rightmost two chart sets), the plots are stacked bars coloured by either *C. elegans* chromosome or Nigon element (Nigon).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3: Local gene order comparison between *Oscheius tipulae* and *Caenorhabditis elegans*

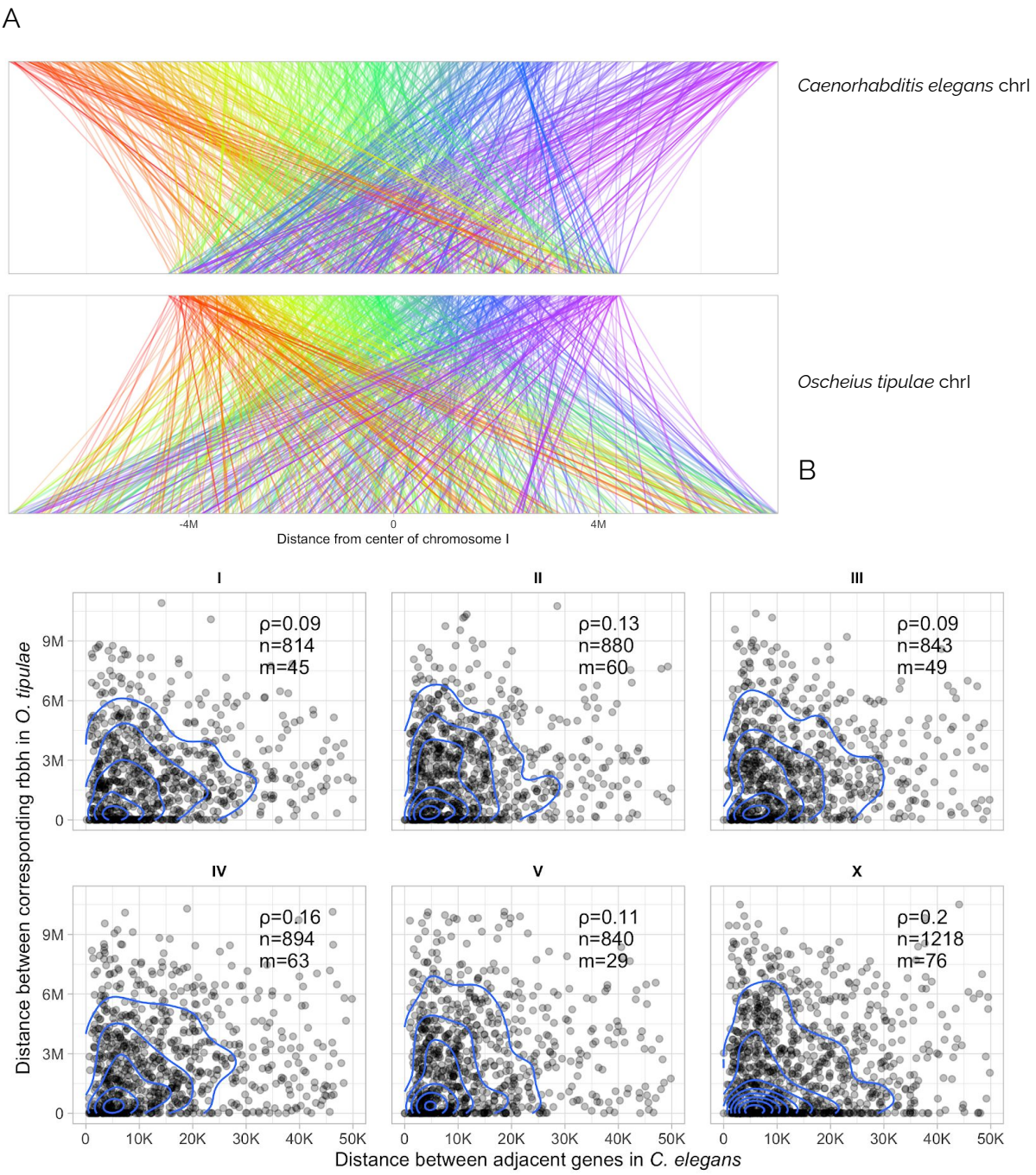


Figure 3: Local gene order comparison between *Oscheius tipulae* and *Caenorhabditis elegans*

A Whole chromosome comparison of *Oscheius tipulae* chromosome I (Otip_I) and its homologue, *Caenorhabditis elegans* chromosome I (Cele_I). Lines link loci that are

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

reciprocal best BLAST hits. In the top panel the lines are coloured by their placement on *C. elegans* Cele_I (a spectrum from red on the left arm through to purple on the right arm). In the lower panel the lines are coloured by their placement on *O. tipulae* Otip_I (a spectrum from red on the left arm through to purple on the right arm).

B Gene neighbourhoods are not conserved between *O. tipulae* and *C, elegans*. We measured the separation distance between each pair of neighbouring *C. elegans* loci for which we could identify single-copy reciprocal best-BLAST relationships between *O. tipulae* and *C. elegans*, and plotted the separation between these orthologue pairs in *C. elegans* (x axis) and *O. tipulae* (y axis) for each chromosome. Genes associated to an operon in *C. elegans* were discarded from this analysis. Correlation tests had p-values < 0.01 for all except chromosomes I and III. Ortholog pairs more than 50 kb apart in *C. elegans* were excluded. 2D kernel density (blue lines) was visualized using ggplot2::geom_density_2d. ρ : Spearman's rank correlation coefficient; n: number of orthologous pairs including pairs excluded from the plot; m: the number of pairs excluded from each panel.

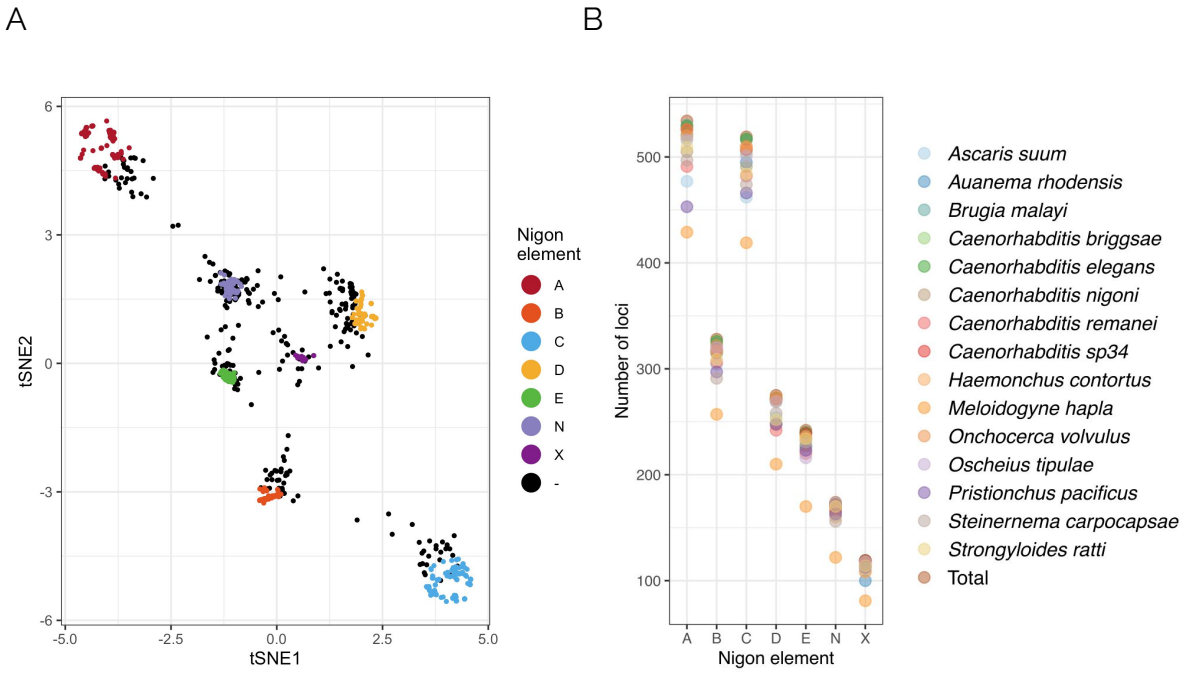
bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Chromosomal elements are conserved across the Rhabditida

Previously we defined conserved nematode chromosomal elements, called Nigon elements, through manual comparison of five genomes from species in Rhabditina (Clade V) (Tandonnet et al., 2019). Manual generation of chromosome assignments is not sustainable, and piecewise addition will fossilise initial taxonomic and data biases. We therefore developed an objective, algorithmic method to identify and group loci that define conserved elements based on shared chromosomal co-location (Figure 4 A). Using this method we were able to include all available rhabditid nematode genomes that have been reported to be chromosomally-complete or near-complete. We identified 2191 loci that had a one-to-one orthologous relationship between most species. These formed seven clusters of loci co-located on chromosomes in most species. These clusters of loci were used to paint the nematode chromosomal assemblies, and replicate our previous, manual definition of Nigon elements. Then clusters had between 534 loci (defining Nigon element A) and 119 (defining NigonX) (Figure 4 B and Table S10). The *Meloidogyne hapla* genome had low BUSCO scores and low representation of orthologues in all the Nigon element sets.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4: Loci that define Nigon elements in rhabditid nematodes



bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 4: Loci that define Nigon elements in rhabditid nematodes

A t-distributed Stochastic Neighbor Embedding (t-SNE) plot of the Gower distance between 3412 orthologous gene families mapped to the chromosomal assemblies of 9 species. The 2191 loci that are included in the Nigon element sets are coloured. The black dots represent loci not assigned to a Nigon unit. Parameters used for t-SNE are perplexity=500, max_iter=1000, initial_dims=50 and theta=0.5.

B For each Nigon-defining set of loci, we counted the number of loci found in each species. The total number of loci per set is indicated by a star. The assembly of *Meloidogyne hapla* has the fewest loci and lowest proportion of loci in all Nigon sets.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Independent chromosomes in iB. Malayi *and* O. volvulus*, and a set of* Chromosome evolution and homology in the Rhabditida

The general conservation of chromosome number (n=6) in rhabditid nematodes might suggest that these karyotypes reflect a static pattern of locus co-location and Nigon element structure. Previous analyses suggest that this is not the case (Rödelsperger et al., 2017; Tandonnet et al., 2019), and we went on to use our chromosome painting to define Nigon element structure in each species (Figure 2, Figure 5) and to build a model of rhabditid karyotype evolution (Figure 6). We recapitulated previous findings made on a limited set of species (Tandonnet et al., 2019) and extended the Nigon element schema to species across Rhabditida. In general, Nigon elements were found as independent chromosomes across Rhabditida, but many species showed patterns of Nigon element defining locus mixing that evidenced past fusions and breakages.

In Rhabditina, all seven of the Nigon elements were present as distinct chromosomes in at least one species. A relatively limited number of fusions and scissions were necessary to explain the observed patterns of Nigon assignment. For example, in *P. pacificus* Ppa-chrI was identified as a recent fusion of NigonA and NigonN (Figure 5 D). As noted previously (Rödelsperger et al., 2017; Tandonnet et al., 2019), Ppa-chrI retains structural signal of two chromosomal elements, with a dual pattern of repeat density and gene density peaks, and these features are consistent with the NigonA and NigonN portions of this chromosome. The *P. pacificus* X chromosome contained only loci from NigonX. Nigon painting of the *Caenorhabditis* species and *H. contortus* was very similar, and showed that the n=6 karyotype of these species is derived from the seven Nigon elements through fusion of NigonN with NigonX to form the X chromosome (Table 3). In contrast to the NigonA-NigonN fusion in *P. pacificus*, the NigonN and NigonX loci in the X chromosomes of *Caenorhabditis* and *H. contortus* are intermixed, suggesting that the fusion was ancestral to the split between these nematodes, and that processes of intrachromosomal rearrangement have removed evidence of distinct NigonN or NigonX domains. This NigonX-NigonN fusion was not observed in other Rhabditina species. In *O. tipulae* NigonX was found to have fused with NigonE to for the X chromosome, and the blocky pattern of locus distribution suggested that the fusion was more recent than the NigonN-NigonX fusion in

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Caenorhabditis and *H. contortus* (Figure 5 A and B). In *A. rhodensis* the loci defining NigonX were all found on Arh-lg5, which is the X chromosome (Figure 5 C). NigonA, NigonC and NigonN loci were distributed, with a blocky pattern, across three *A. rhodensis* autosomes (Arh-chr2, Arh-chr3, Arh-chr4), suggesting a relatively recent set of scission and fusion events (Figure 5 C). Thus, at the base of Rhabditina, we predict there were seven distinct linkage groups, corresponding to Nigon elements A through E, N and X (Figure 6).

In Spirurina, NigonA was an independent autosome in *B. malayi* (*Bma-chr3*) and a NigonA element was identified as recently fused with a NigonN element in *O. volvulus* to form *Ovo-OM1* (Figure 5 G and H; Supplemental Figure S7). Chromosomes consisting nearly completely of NigonA loci were found in *A. suum* (*Asu-chr1*) (Figure 5 I). NigonB is an independent autosome in both *B. malayi* and *O. volvulus*, and four *A. suum* autosomes are painted only by NigonB loci (Figure 5 G, H and I). In *A. suum*, some NigonB loci are also found on the X chromosomes. Similarly, NigonC loci paint independent chromosomes in *B. malayi* and *O. volvulus* and four autosomes in *A. suum.* NigonN appeared to have fused recently and independently in *B. malayi* (with NigonD to form the X chromosome; Figure 5 G) and *O. volvulus* (with NigonA to form *Ovo-OM1*; Figure 5 H). That these fusions are recent is supported by patterns of repeat and GC content across the chromosomes. In *A. suum* NigonN loci are largely restricted to four NigonN autosomes (Figure 5 I, Figure 6). The X chromosomes of *A. suum*, *B. malayi* and *O. volvulus* each carried evidence of a fusion between NigonD and NigonX. It was not easy to discern whether this fusion was ancestral to Spirurina or arose independently in the Ascarididomorpha and Spiruromorpha (Figure 6). Independent fusion was supported by the finding that there are distinct *A. suum* autosomes only painted by NigonD, that NigonX loci while wholly limited to the five *A. suum* X chromosomes are not always associated with NigonD loci, and that NigonD and NigonX loci form distinct blocks in the *A. suum* X chromosomes. On *A. suum* Asu_chrX1, the NigonD and NigonX domains are resolved as distinct somatic chromosomes by chromatin diminution. In *B. malayi* and *O. volvulus* the NigonD and NigonX loci were intermixed, suggesting long-term association and mixing by intrachromosomal rearrangement.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

It is notable that while all the autosomes of *A. suum* were painted by loci from a single Nigon element set, all five X chromosomes had mixed origins, involving blocks of NigonA, NigonB, NigonD and NigonX loci. The chromosomes of *A. suum* are subject to chromatin diminution in somatic cells of the embryo (Wang et al., 2020). This process generates remodelled telomeres for all chromosomes, and specific cleavage at internal sites in some chromosomes such that somatic cells have more chromosomes (but less genetic material overall) than do germline cells. It is striking that all but one of the intrachromosomal cleavage points were between blocks of chromosome with distinct Nigon identity. The one cleavage not between Nigon blocks (in *Asu-chr1*) separated two NigonA components. Not all blocks of Nigon identity were separated by cleavage during diminution (Figure 6).

In Tylenchina, retention of ancestral units, breakages and fusions also describe the present day chromosome structures observed (Figure 6). In *S. carpocapsae*, which is not fully chromosomally assembled, the twelve autosomal scaffolds, which correspond to four autosomes, each had a single Nigon identity. We predict that these scaffolds will assemble to yield four autosomes corresponding to NigonA, NigonC, NigonE and NigonN. The *S. carpocapsae* X chromosome comprised three domains corresponding to NigonB, NigonX and NigonD (Figure 5 F). *S. ratti* has two autosomes and an X chromosome. The autosomes were modeled as being a complex product of a two stage fusions-scissions-fusions process. The first fusions were between NigonA and NigonC, and between NigonE and NigonN. After some time (as evidenced by the intermixing of loci) these fusion chromosomes were split, and refused to form Sr-chr1 and Sr-chr2. The second set of scissions and fusions was relatively recent, as each autosome had distinct domains corresponding to the presumed ancestral fusions. The *S. ratti* X chromosome was painted by loci corresponding to NigonB, NigonD and NigonX (Figure 5 E). The NigonB, NigonN and NigonX loci were fully intermixed. Again it was difficult to disentangle ancient fusion versus parallel fusion of NigonD and NigonX in Tylenchina. The distinctness of the NigonX and NigonD domains in the *S. carpocapsae* argues for independent events. Nonetheless it is striking that NigonD and NigonX are associated with the X in both Tylenchina and Spirurina. *Meloidogyne hapla* has seventeen chromosomes, but Nigon painting of these did not yield definitive

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

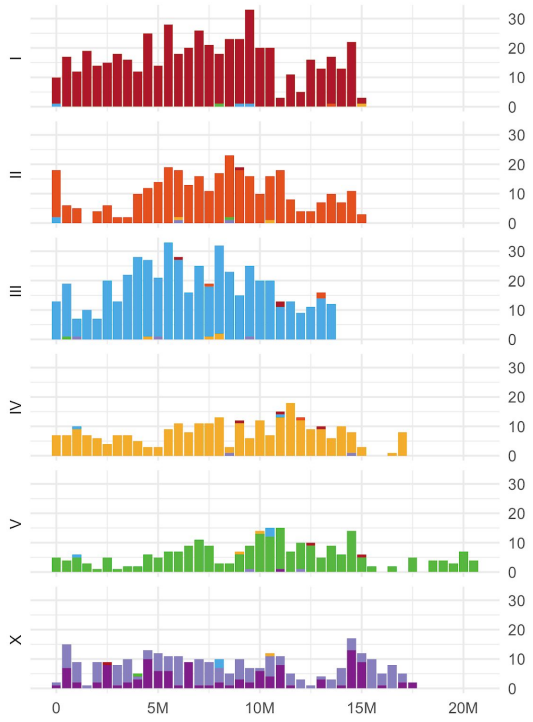
Nigon assignments (Supplementary Text 2). We noted that there was no association between NgonD and NigonX loci on the *M. hapla* chromosomes.

The independent existence in a common ancestor of Rhabditina, Tylenchina and Spirurina of elements corresponding to Nigons A, B, C, E and N was evident from the identification of chromosomes, or distinct chromosome domains, in all three suborders corresponding to these elements. NigonD and NigonX were not associated in Rhabditina, but were variably associated in Tylenchina and Spirurina. It was unclear whether this should be modeled as an ancient fusion with differential resolution in extant species, or repeated fusion events. The existence of chromosomal domains corresponding to NigonD and NigonX in *S. carpocapsae* (Tylenchina) and *A. suum* (Spirurina) and of NigonD-only autosomes in *A. suum* argued against a single, ancestral event (Figure 6).

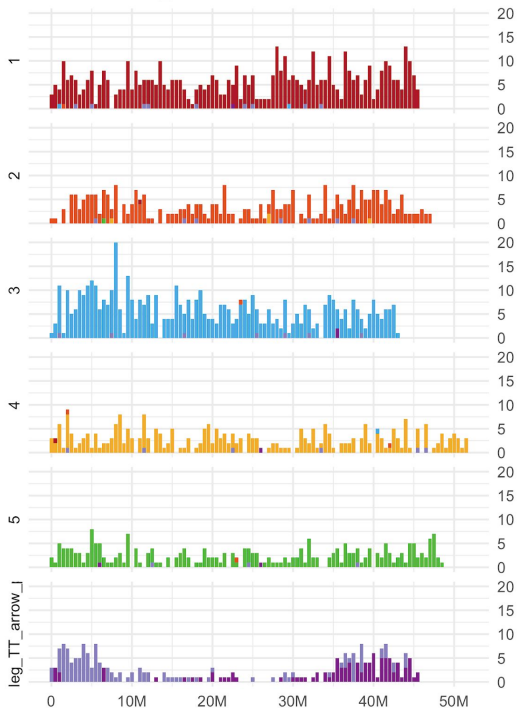
bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5: Nigon painting of Rhabditid nematode chromosomes

A *Caenorhabditis elegans*



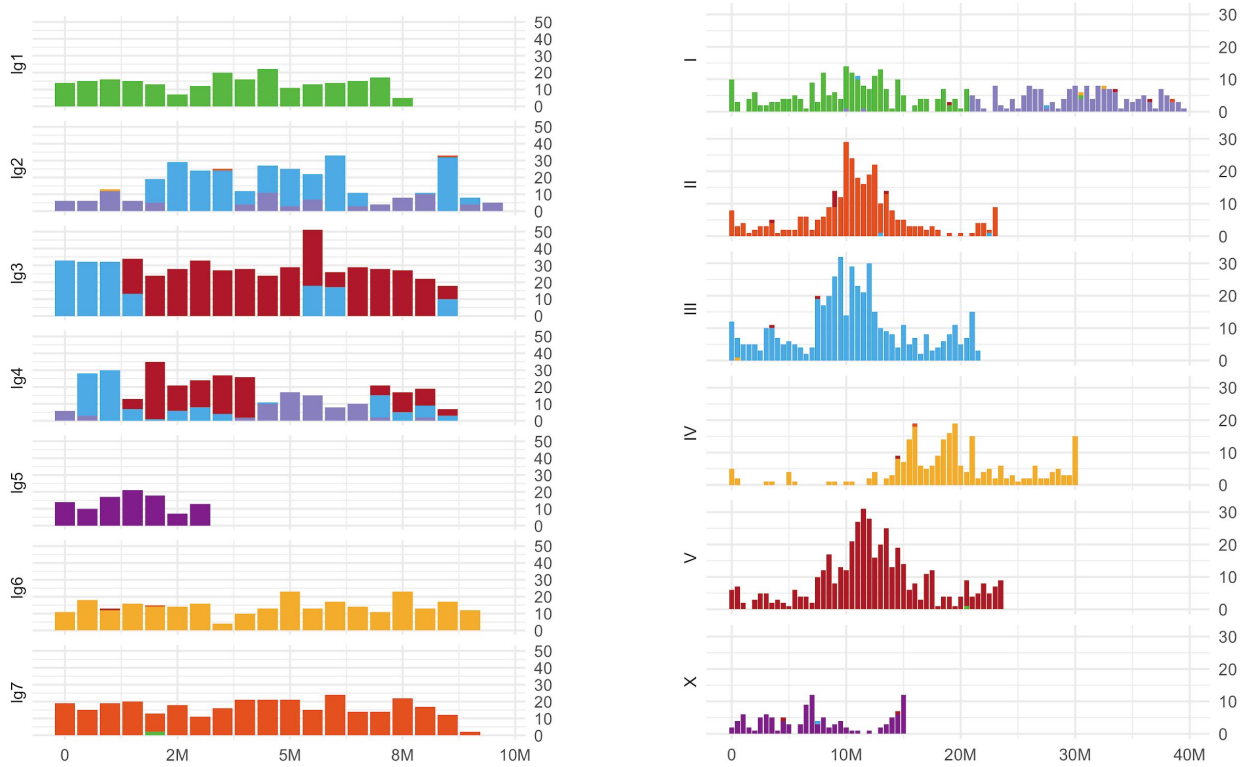
B *Haemonchus contortus*



C *Auanema rhodensis*

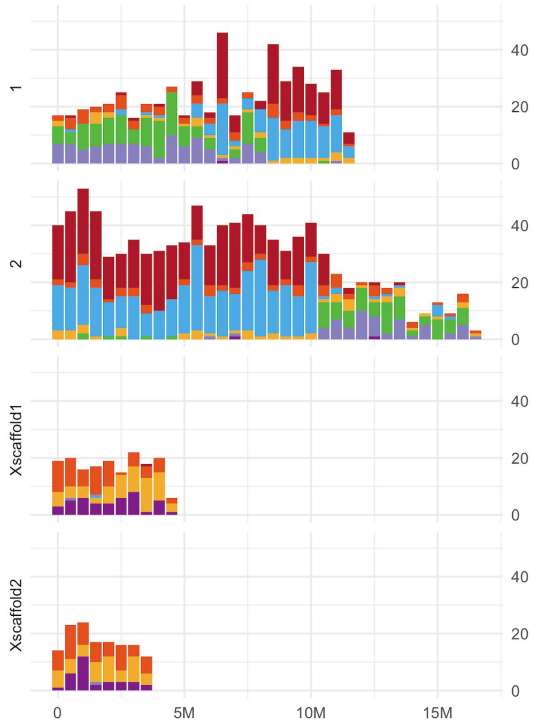
D *Pristionchus pacificus*

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

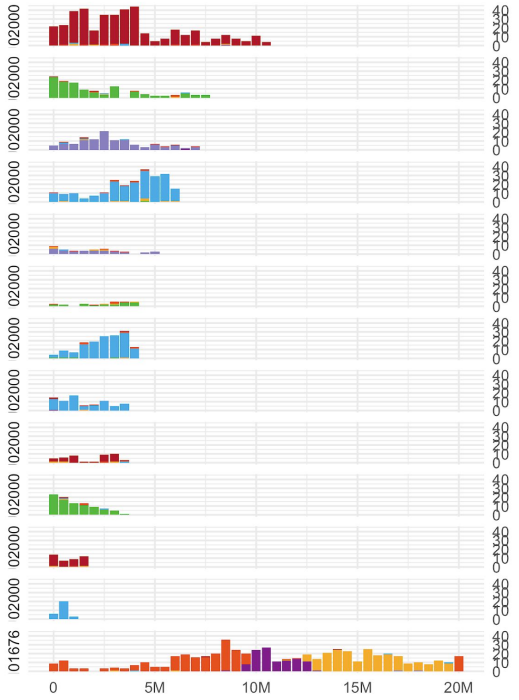


bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

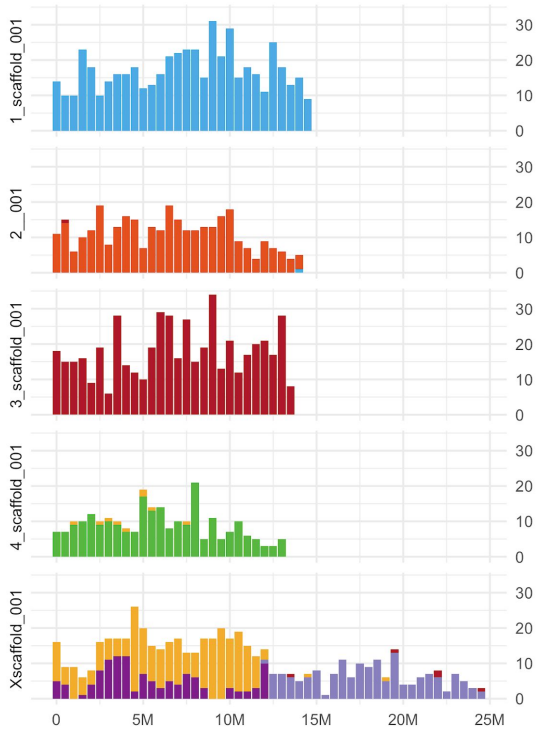
E *Strongyloides ratti*



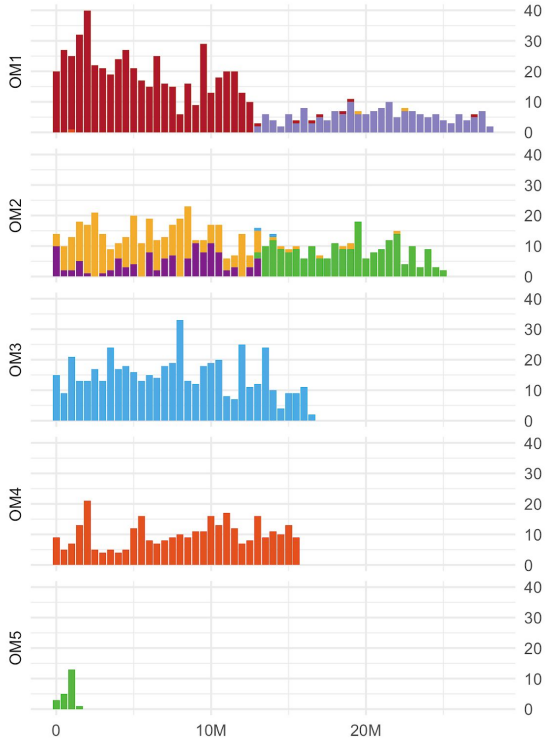
F *Steinernema carpocapsae*



G *Brugia malayi*

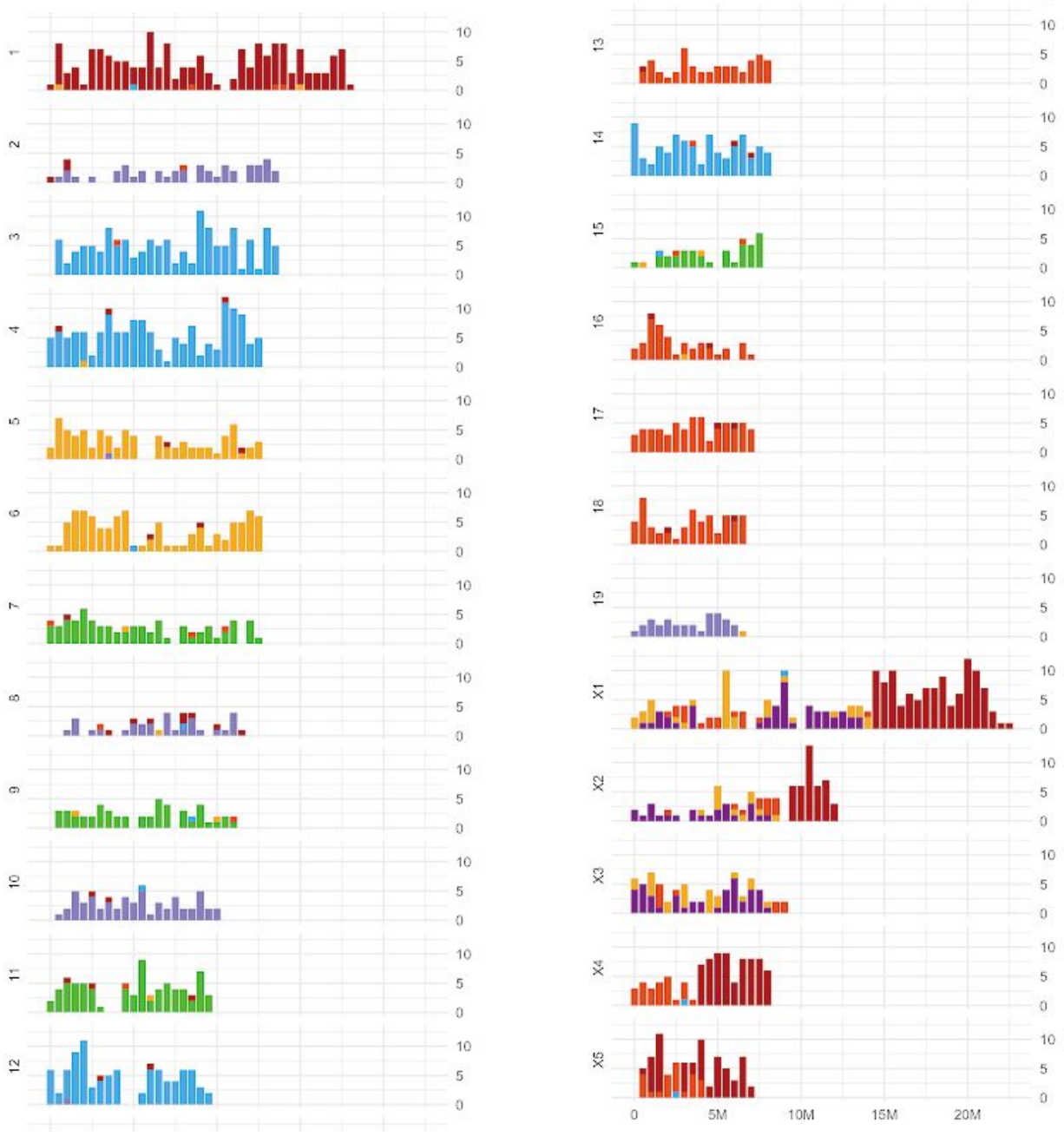


H *Onchocerca volvulus*



bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

I *Ascaris suum*



J Legend

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

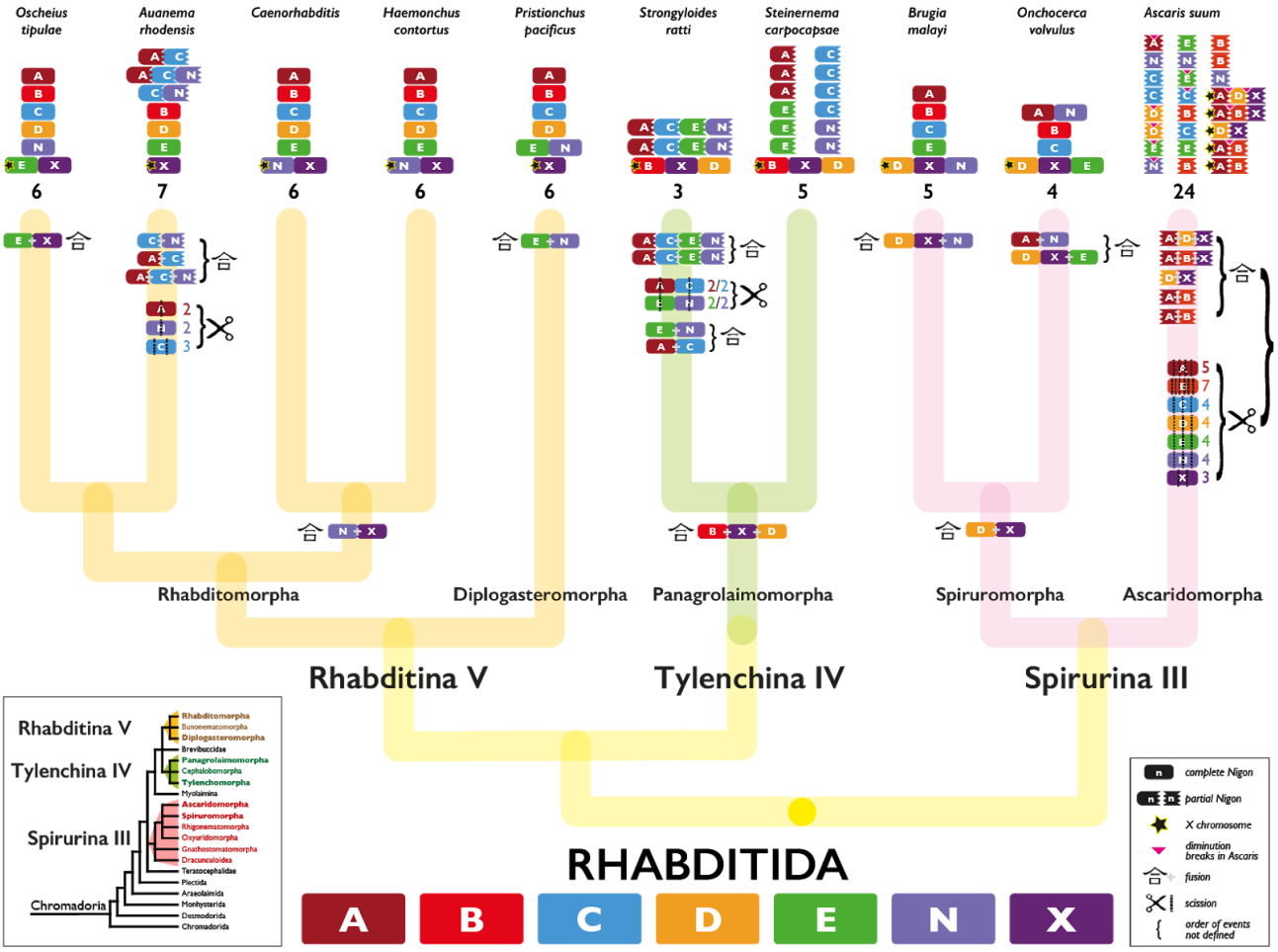
Figure 5: Nigon painting of Rhabditid nematode chromosomes

Examples of nematode chromosomal assemblies painted by their content of Nigon-defining loci. Each subgraph shows the count of loci mapped in non-overlapping 0.5 Mb windows along a chromosome as a stacked histogram coloured by Nigon origin. The X- and Y-axes are scaled to the maxima within a species in each panel (X: chromosome length, Y: Nigon-defining loci per interval) within each species. The legend in panel J applies to all nine chromosome panels.

A *Caenorhabditis elegans* (Rhabditina, Rhabditomorpha), B *Haemonchus contortus* (Rhabditina, Rhabditomorpha), C *Auanema rhodensis* (Rhabditina, Rhabditomorpha), D *Pristionchus pacificus* (Rhabditina, Diplogasteromorpha), E *Strongyloides ratti* (Tylenchina, Panagrolaimomorpha), F *Steinernema carpocapsae* (not a fully chromosomal assembly; Tylenchina, Panagrolaimomorpha), G *Brugia malayi* (Spirurina, Spiruromorpha), H *Onchocerca volvulus* (Spirurina, Spiruromorpha), I *Ascaris suum* (Spirurina, Ascaridomorpha). Panel J shows the colour key for the other panels.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 6: A model of chromosome evolution in Rhabditida



bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 6: A model of chromosome evolution in Rhabditida.

For each species the classification of chromosomes to Nigon elements is shown (see Figure 2, Figure 5 and Supplemental Figure S7), including rearranged chromosomes. X chromosomes are indicated by a star. For each lineage we have inferred the patterns of chromosome scission (the ✂ symbol; the number indicates the number of fragments resulting) and fusion (the Kanji symbol for "fusion point" 合) on the tree. Where the order in time of events is not resolved, we have bracketed them. In *A. suum* a pink triangle indicates positions of internal cleavage of germline chromosomes during diminution. Note that the "partial Nigons" in *S. carpocapsae* are assembly scaffolds rather than chromosomes. The cladogram representing the phylogeny of Chromadoria nematodes (inset, lower left; (Blaxter and Koutsovoulos, 2015)) is derived from phylogenetic analysis of shared protein coding genes. In the main figure we have resolved a previous uncertain polytomy in Rhabditomorpha to favour a closer relationship between *Haemonchus contortus* and the genus *Caenorhabditis* based on the exact conservation of linkage groups.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Dynamic evolution of rhabditid sex chromosomes

NigonX was the only element consistently associated with the sex chromosome all the nematode species analysed (Table 3). Additionally the NigonX element was much more likely to be involved in fusions with other Nigon elementsHowever, the histories of loci defining NigonD and NigonX were intertwined in Spirurina and Tylenchina. In Tylenchina, the X chromosomes were ancestral fusions of NigonB, NigonD and NigonX, and this fusion was fully intermixed in *S. ratti* (Figure 5 E) but unmixed in *S. carpocapsae* (Figure 5 F). NigonX loci were found on three of the five *A. suum* X chromosomes (As-chrX1, As-chrX2 and As-chrX3), mixed or fused with other Nigon element loci (Figure 5 I). The other two *A. suum* X chromosomes (As-chrX4 and As-chrX5) are fusions of NigonA and NigonB loci. The NigonA loci tended to form distinct domains in all the *A.suum* X chromosomes, which suggests relatively recent fusion, but the NigonD and NigonX loci were intermixed. *A. suum* also had two autosomes that contained only NigonD loci (As-chr5 and As-chr6), suggesting that this element had an independent existence and fissioned into at least two parts before fusion with NigonX. In both spiruromorph nematodes, a NigonD plus NigonX intermixed domain was found in the X chromosome, but this had fused with different, autonomous Nigon elements in *B. malayi* (with NigonN) and *O. volvulus* (with NigonE) (Figure 5 G, H).

The complete admixture of NigonD and NigonX loci in these spiruromorph X chromosome domains suggests ancient fusion, especially since the *B. malayi* and *O. volvulus* genomes are largely collinear both within non-fused chromosomes and within fused chromosome blocks (Foster et al., 2020). Because NigonX and NigonD are present, intermixed, in the X chromosomes of both Ascaridomorpha and Spiruromorpha, the NigonD-NigonX fusion could be ancestral to Spirurina. However the presence of NigonD-only autosomes in *A. suum* suggested that NigonD was present as an independent element in this lineage, and mapping of the relationships between NigonD loci in *A. suum* and spiruromorph genomes does not show any particular grouping. Thus we modeled these NigonD-NigonX fusions as independent events (Figure 6).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

NigonX and NigonD were also present in the X chromosomes of Tylenchina, as part of a B-X-D fusion. In *S. carpocapsae*, this fusion appeared to be recent, as the three sets of Nigon element-derived loci occupied distinct domains, with NigonX central (Figure 5 F). In *S. ratti* the loci from NigonB, NigonD and NigonX were intermixed on *Sra-chrX*, but some NigonD loci were found on the two autosomes (Figure 5 E). These autosomal NigonD loci were found in association with some NigonB loci, perhaps as a result of translocation from an ancestral intermixed NigonB-NigonD chromosome. The distinctness of the NigonD domain within *S. carpocapsae* Sc_X suggested that NigonD was an independent element in Tylenchina also. The B-D-X fusion is, we suggest, an association of NigonD and NigonX independent of that in Spirurina.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3: Nigon elements and the X chromosomes of rhabditid nematodes

	Nigon is distinct in			**Nigon is part of X chromosome in**									
	Rhabditina	**Tylenchina**	**Spirurina**	**Rhabditina**					**Tylenchina**		**Spirurina**		
species	9 species	2 species	3 species	*Ot*	*Ar*	*Ce*	*Hc*	*Pp*	*Sr*	*Sc*	*Bm*	*Ov*	*As*
NigonA	**yes**	**yes**	**yes**										■
NigonB	**yes**	part of X	**yes**						■	■			■
NigonC	**yes**	**yes**	**yes**										
NigonD	**yes**	part of X	yes (part)						■	■	■	■	■
NigonE	**yes**	**yes**	**yes**	■								■	
NigonN	**yes**	**yes**	**yes**			■	■				■		
NigonX	**yes**	part of X	part of X	■	■	■	■	■	■	■	■	■	■

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Functional coherence of Nigon element loci

The distinct histories of the gene sets associated with each Nigon element means that these sets of loci have been linked for a significant period of time, and may have been selected to stay together or evolved to collaborate. We explored whether such association might reflect the shared biological function of these genes. We interrogated functional enrichment of the Nigon-defining loci through the KEGG pathway annotation of *C. elegans* orthologues. We first compared KEGG pathway annotations of Nigon-defining loci to those of the full gene set of *C. elegans* (Table S11). Terms relevant to RNA transport were enriched in NigonA loci and NigonC loci, ribosome biogenesis was enriched in NigonA loci and spliceosome pathway was enriched in NigonC loci. ErbB signaling and calcium signaling pathways were enriched in the Nigon D locus set. In NigonN loci Hippo signaling was enriched. In NigonX loci, annotations relevant to axon regeneration, calcium signaling pathway and neuroactive ligand-receptor interaction were significantly enriched. No KEGG pathways were enriched among loci of NigonB or NigonE. The Nigon-defining loci were drawn from a specific, conserved subset of all *C. elegans* loci, and this conservation will have *a priori* biased the annotations being assessed. A also compared the annotations associated with each Nigon locus set to those of the set of all 2175 Nigon loci and only detected enrichment in NigonX loci, in the KEGG axon regeneration pathway (enrichment significance 4.40 e10-4).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

DISCUSSION

New technologies and complete genome sequencing of *O. tipulae*

New sequencing technologies and the development of improved assembly toolkits are generating more highly-contiguous reference genome assemblies. Here we use the Oxford Nanopore long reads to generate chromosomally-complete contigs representing all the nuclear chromosomes and the mitochondrion of the free-living nematode *Oscheius tipulae*. While the data are sufficient to generate a chromosomally contiguous assembly, not all assembly tools were able to generate this from the data, and there is evidently still development work to be done. Alternate methods of generating single molecule, long read data such as the Pacific Biosciences SEQUEL II CLR (single pass) and HiFi (circular consensus, multiple pass) have similar properties to PromethION data, with the HiFi standing out as having higher per-base accuracy. This higher per-base accuracy likely simplifies the assembly process, in particular in the resolution of repeat structures that are close to but not 100% identical. However the PromethION data, which can include very long reads, may be better at traversing recent segmental duplications and homogenised multicopy loci (Nurk et al., 2020). In our assembly of *O. tipulae* the only identified remaining collapsed repeat was the 6.8 kb repeat of the ribosomal RNA cistron (nSSU, 5.8S and nLSU loci), which we estimate is repeated about 117 times, summing to 801 kb. This repeat is homogenised and thus only very long range technologies, such as ultralong nanopore reads or BioNano mapping, could resolve it fully.

One unexpected and striking feature of the *O. tipulae* genome is the structure of the telomeres. The PromethION data robustly predicts two telomere repeat addition sites at each end of each chromosome, generating a core, high coverage chromosome with lower coverage subtelomeric extensions at each end. The extensions are supported by unique mapping of independently generated short Illumina data. Nearly 350 kb (or 0.6% of the genome) is in these extensions, which carry additional, expressed protein coding genes. We currently interpret the extensions as segments of chromosomes that have been specifically removed from the genomes of a proportion of cells in each nematode, possibly through developmentally regulated chromatin diminution. The

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

presence of helitron mobile elements specifically in the subtelomeric elements is intriguing. The helitrons contain nuclease and Pif1 DEAH-box helicase domains. In yeast and other taxa, Pif1 helicases are intimately involved in DNA metabolism, and in particular in DNA replication and telomerase function and regulation. It is possible that the *O. tipulae* telomeric helitrons are parasitic elements that have generated extended subtelomeric regions in which they reside, and that they also control the excision of these telomeric extensions and the specific addition of neo-telomeres. Alternatively the nematode may have co-opted the helitrons to regulate a chromatin diminution process that regulates expression of germline-restricted genes by eliminating them from the soma, as in *A. suum* (Wang et al., 2012). In nematodes, chromatin diminution distinguishing soma from germline has been described in Ascaridomorpha and in XX and X0 sperm made by parthenogenetic female *S. papillosus (Wang and Davis, 2014)*. Given that the diminution signal we observe is present on all chromosome ends, we currently favour an ascaridomorph-like process that distinguishes a germ-line genome from a somatic one. It must be distinct from the ascaridomorph process, as the internal breakage and addition sites in *A. suum* are associated with multi-kilobase regions while the *O. tipulae* sites are precise. It may be that similar processes are present in other nematodes, and other species, but have been overlooked because of the lack of contiguity of the previously available short-read sequence data.

Evolution of rhabditid nematode karyotypes

We have refined an approach to defining loci that define conserved linkage groups. In many taxa, it is possible to use gene neighbourhoods (gene order and synteny) to drive inference of ancestral karyotypic organisation (Kim et al., 2017). However we and others have noted that gene order is poorly conserved in rhabditid nematodes (Stein et al., 2003; Teterina et al., 2020). This is interesting, not only because the position of protein coding genes in chromosomes of *C. elegans* correlates with gene conservation, such that deeply evolutionary conserved genes tend to be found in chromosomal centres, and novel loci on the arms (C. elegans Sequencing Consortium, 1998). Despite this we observed other chromosomal features that were similar to *C. elegans*, such as the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

differential abundance of repeats on the presumed arms of *O. tipulae* autosomes. This suggests that distinct evolutionary processes may drive these patterns.

We were able to derive sets of loci that traveled together on linkage groups through rhabditid genome evolution by clustering orthologues based on a numerical representation of their chromosomal location in each species. This process is robust, and extendable to incorporate additional genomes. It is also applicable to other taxa where chromosomally-complete genomes are available. In Rhabditida, we identified seven clusters of loci that define seven chromosomal units, named Nigon elements. These elements are fully congruent with a previous manual estimate (Tandonnet et al., 2019). Painting the chromosomal genome assemblies of fourteen rhabditid species revealed that in no species were all of these elements present as distinct chromosomes, but each Nigon element was found as a distinct element in several species. In Tylenchina and Spirurina, we found that NigonX was fused with other Nigon elements to form the X chromosome. These fusions include NigonD and NigonA in Tylenchina and NigonD in Spirurina. We have modeled NigonD as a distinct unit despite this frequent association with NigonX because some NigonD loci uniquely painted two chromosomes in *A. suum* and the NigonD painting of the X chromosome in *S. carpocapsae* identified a single contiguous block. We were not able to apply the Nigon element model to *M. hapla*, where mapping to the 17 chromosomal scaffolds yielded only a few with majority assignment to one element. It will be informative to explore chromosomal evolution in the plant parasitic Heteroderidae further.

Brugia and *Onchocerca* are unusual in Spiruromorpha in having an apparent XX:XY sex determination system (Post, 2005). Within the filarial nematodes an XY system has evolved twice from an ancestral XX:X0 system, once in the ancestor of *Onchocerca* and *Dirofilaria* species and once in the ancestor of *Wuchereria* and *Brugia* species (Figure 7) (Post, 2005). It was proposed from karyotypic analyses that the neo-X chromosome in *Onchocerca* and *Brugia* arose from the fusion of an autosome with the ancestral X, and that the neo-Y chromosome in these species was just this autosomal chromosomal component (Post, 2005). Our analysis of Nigon element conservation supports this model, but additionally suggests that the enlarged X chromosomes in the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

two species are the results of two distinct fusions with an ancestral X: in *B. malayi* with NigonN and in *O. volvulus* with NigonE.

While males in *B. malayi* and *O. volvulus* have a pair of sexually dimorphic chromosomes it is not clear whether the "Y" chromosome actually carries a male determining locus, or whether the system is a modified Xo system, as is found in the tylenchine *Strongyloides papillosus* (Albertson et al., 1979). In *Strongyloides ratti*, sex determination is XX:Xo, and the X chromosome is an intermixed fusion of NigonB, NigonD and NigonX. In the related *Strongyloides papillosus*, sex determination is also XX:Xo, but in this species haploidy of X is generated by intrachromosomal chromatin diminution. While the genome assembly for *S. papillosus* is not chromosomal, mapping of genetic marker loci indicated that the part that is lost is likely the NigonB-NigonD-NigonX component, while the remainder of the *S. papillosus* X appears to be homologous to *S. ratti* chromosome I (Nemetschke et al., 2010). Thus *S. papillosus* male-determining sperm have a "reduced-X" chromosome that contains only the autosomal part, and males are still diploid for the autosomal part. So are the *B. malayi* and *O. volvulus* sex determination systems XX:XY or apparent XY systems that biologically behave as XX:Xo?

Foster *et al.* (Foster et al., 2020) have argued, based on the presence on spiruromorph X chromosomes of NigonD loci (i.e loci mapping to *C. elegans* chromosome IV) that NigonD was the ancestral sex determination element of all Rhabditida, and that sex determination function transitioned to NigonX only later in rhabditid evolution. Foster *et al.* were unable to identify NigonN. Read data from males and females identified only a very small segment of genome (2.7 Mb in many short contigs) that was unique to male *B. malayi*, and a previously identified, male-linked locus (transposon on Y, TOY) was part of this male-limited genome. Karyotypic analyses identified the *B. malayi* X chromosome as being similar in size to the X chromosomes of XX:X0 species, and similar in size to the autosomes (Post, 2005). We interpret this sequence (and TOY) as being repeat accumulation in the subtelomeric regions of the short-form X chromosome that was lost consequent to the fusion event between the ancestral X and NigonN chromosomes, and doubt it has roles in sex determination. The NigonN

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

component of the *B. malayi* X is diploid in males and females while the 12 Mb NigonD - NigonX portion is haploid in males.

Other spiruromorphs, including close relatives of *Onchocerca* and *Brugia* have an XX:X0 sex determination system (Post, 2005), suggesting this mode is ancestral. This pattern leads us to question whether sex determination in *B. malayi* and *O. volvulus* is in fact XX:XY, where, by analogy to XX:XY systems in other taxa, a sex (male) determining locus is present on the Y. In *B. malayi* and *O. volvulus*, where the non-NigonD-NigonX component of the X chromosome derives from fusion with different chromosomes, we propose that the same haploid-X mechanism operates, and that the apparent XX:XY sex determination system is in fact an XX:X0 system, where the additional component (NigonN in *B. malayi* and NigonE in *O. volvulus*) is always diploid. This must mean that the fusion partner, diploid in males and females, must be under distinct dosage compensation control compared to its sex-determining partner, as is likely the case in *S. papillosus*. Similar sex chromosome fusions where the newly fused parts have distinct dosage compensation mechanisms have been identified in another species with holocentric chromosomes, the butterfly *Danaus plexippus*. In *D. plexippus* the neo-Z chromosome has two distinct modes of compensation, spatially distributed along the fusion based on the origin of the segment (i.e. expression of the Z component, is halved in ZZ males, while expression of the autosomally derived fragment is doubled in WZ females (Gu et al., 2019).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 7: Sex chromosome evolution in the filarial nematodes (Spiruromorpha)

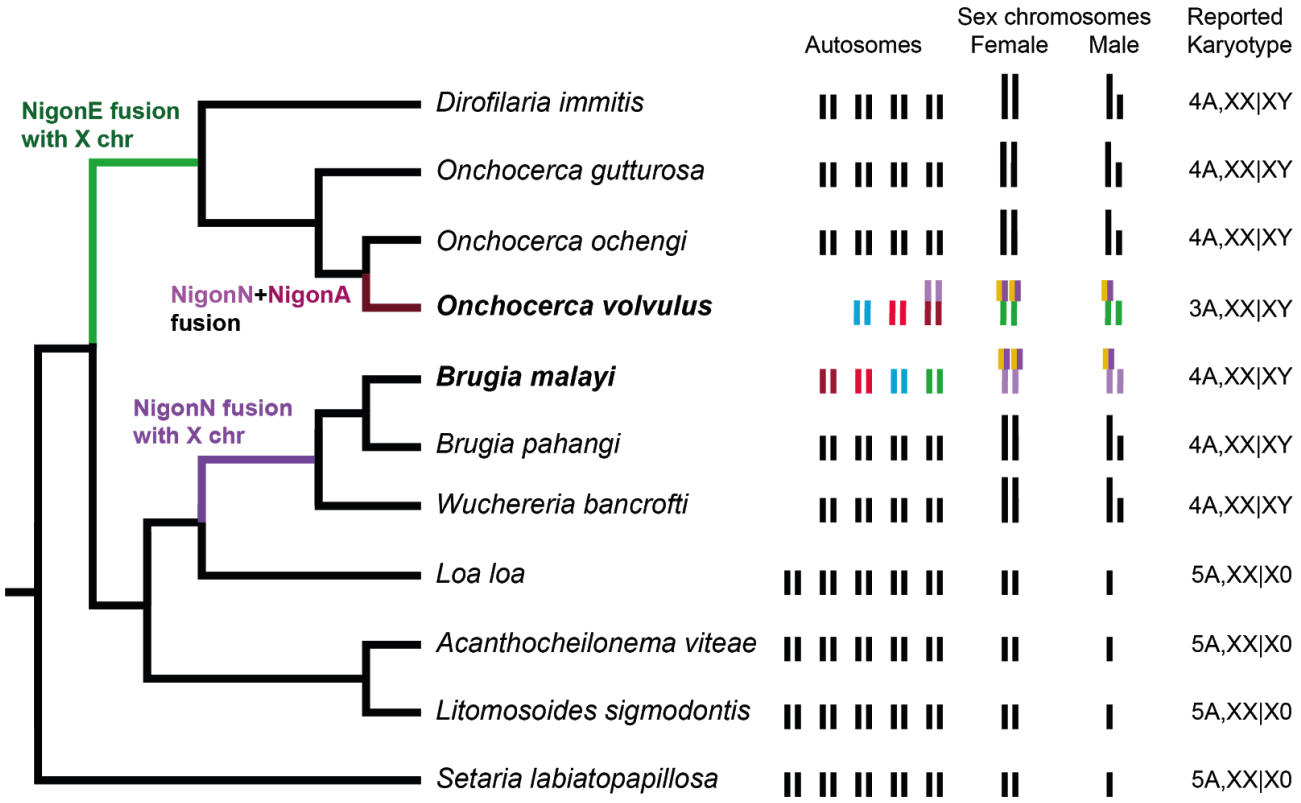


Figure 7: Sex chromosome evolution in the filarial nematodes (Spiruromorpha)

Within the filarial nematodes, karyotypes have been determined for a number of species. The phylogeny (cladogram to the left) is derived from multilocus phylogenomic analysis, and is in agreement with marker gene-based phylogeny (Lefoulon et al., 2015). The karyogram and karyotype data are from the work of Rory Post and colleagues (Post, 2005). The inferred position of fusion events within the filarial phylogeny are indicated, and the karyotypes of *O. volvulus* and *B. malayi* are coloured by their Nigon assignment (see Figure 5 G,H and Figure 6).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Functional coherence of loci that define Nigon elements

As these Nigon-defining loci have been colocated on the same karyotypic unit for much of rhabditid nematode evolution, we wondered whether each set had a functional coherence such that the colocated genes functioned together in specific pathways. We identified functional enrichment in five of the seven gene sets in *C. elegans* when the whole *C. elegans* gene set was used as comparator. As we selected these genes based on their largely one-to-one conservation across Rhabditida, it would be expected that they would be enriched in conserved function, and so this result is perhaps not surprising. When using only the set of Nigon-defining loci as a reference, however, we found functional enrichment only of NigonX-linked loci. Based on our model, the loci that define NigonX have been on the sex determination chromosome of rhabditid nematodes since the last common ancestor of Tylenchina, Spirurina and Rhabditina. These loci will thus have been exposed as haploid in males and will have had an effective population size of approximately 0.75 of the size of any autosomal locus for the entirety of rhabditid evolution. Genes with essential functions are more rarely found on the X chromosome than in the autosomes of *C. elegans*, while genes with non-lethal, post-embryonic knock-down phenotypes are enriched in the X chromosome (Kamath et al., 2003).

Outlook

The telomere-to-telomere chromosomal assembly of *Oscheius tipulae* can now stand as a platform for future work on the developmental and population genetics of this important model species. Investigation of the biological importance of the telomeric extensions, and especially of their presence or absence in other species is of particular importance. Why do some genes stay on the same chromosome, while others appear to move freely? What mechanisms drive the processes of chromosomal structure, and why do some species have distinct patterns of intra- and inter-chromosome rearrangement? In Lepidoptera there is a general conservation of karyotype (with n=31) and genes tend to be situated on homologous chromosomes in the same order (i.e. there is strong conservation of micro- and macro-synteny), but some taxa diverge

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

strongly from this pattern and have very different chromosome numbers (from n=4 to >200) (de Vos et al., 2020) that may not be simply described by fusion of whole chromosomes, or scission of chromosomes into multiple parts (Hill et al., 2019). In the Lepidoptera, karyotypic change is associated with speciation (de Vos et al., 2020), but whether this is also true in nematoda is not clear, though we note the relatively rapid karyotypic evolution in filarial nematodes (Figure 7) and the existence of genera and orders such as *Diploscapter* in Rhabditina (n=1 to 7) (Fradin et al., 2017) and the Ascaridomorpha (n=1 to 24), where chromosome counts vary greatly (Walton, 1959). We look forward to an increase of chromosomal assemblies from rhabditid and other nematodes in the near future to further explore patterns and processes in nematode chromosome evolution.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ACKNOWLEDGEMENTS

PG is funded by a PhD Scholarship from the Darwin Trust of Edinburgh. PromethION sequencing at Edinburgh Genomics was funded by a UK Natural Environment Research Council Biomolecular Analysis Facility contract. Sophie Tandonnet was funded by CAPES/CNPq (201116/2014-6) and FAPESP (2019/07285-7). We acknowledge the assistance of our colleagues at Edinburgh Genomics and of members of the University of Edinburgh Institute of Evolutionary Biology evolutionary genomics community, especially Lewis Stevens and Andrea Martinez Martinez. The *O. tipulae* CEW1 strain was supplied by Marie-Anne Félix, and was originally isolated by Carlos Winter. We also thank Tom Freeman for discussions about clustering, and Jianbin Wang and Richard Davis for early access to the chromosomal genome assembly of *Ascaris suum*.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

REFERENCES

Albertson, D.G., Nwaorgu, O.C., Sulston, J.E., 1979. Chromatin diminution and a chromosomal mechanism of sexual differentiation in Strongyloides papillosus. Chromosoma 75, 75–87. doi:10.1007/BF00330626

Albertson, D.G., Thomson, J.N., 1982. The kinetochores of Caenorhabditis elegans. Chromosoma 86, 409–428. doi:10.1007/BF00292267

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F.J., Lippman, Z.B., Schatz, M.C., 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 20, 224. doi:10.1186/s13059-019-1829-6

Besnard, F., Koutsovoulos, G., Dieudonné, S., Blaxter, M., Félix, M.-A., 2017. Toward Universal Forward Genetics: Using a Draft Genome Sequence of the Nematode Oscheius tipulae To Identify Mutations Affecting Vulva Development. Genetics 206, 1747–1761. doi:10.1534/genetics.117.203521

Bhutkar, A., Schaeffer, S.W., Russo, S.M., Xu, M., Smith, T.F., Gelbart, W.M., 2008. Chromosomal rearrangement inferred from comparisons of 12 Drosophila genomes. Genetics 179, 1657–1680. doi:10.1534/genetics.107.086108

Blaxter, M., Koutsovoulos, G., 2015. The evolution of parasitism in Nematoda. Parasitology 142 Suppl 1, S26-39. doi:10.1017/S0031182014000791

Bonfield, J.K., Whitwham, A., 2010. Gap5--editing the billion fragment sequence assembly. Bioinformatics 26, 1699–1703. doi:10.1093/bioinformatics/btq268

Bushnell, B., 2017. BBTools. sourceforge.

C. elegans Sequencing Consortium, 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282, 2012–2018. doi:10.1126/science.282.5396.2012

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. doi:10.1186/1471-2105-10-421

Coghlan, A., Coghlan, A., Tsai, I.J., Berriman, M., 2018. Creation of a comprehensive repeat library for a newly sequenced parasitic worm genome. Protoc. exch. doi:10.1038/protex.2018.054

d'Alençon, E., Sezutsu, H., Legeai, F., Permal, E., Bernard-Samain, S., Gimenez, S., Gagneur, C., Cousserans, F., Shimomura, M., Brun-Barale, A., Flutre, T., Couloux, A., East, P., Gordon, K., Mita, K., Quesneville, H., Fournier, P., Feyereisen, R., 2010. Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. Proc Natl Acad Sci USA 107, 7680–7685. doi:10.1073/pnas.0910413107

Dainat, J., 2020. NBISweden/AGAT: AGAT-v0.1.1. Zenodo. doi:10.5281/zenodo.3674778

De Ley, P., Blaxter, M.L., 2002. Systematic position and phylogeny, in: Lee, D. (Ed.), The Biology of Nematodes. Taylor & Francis, pp. 1–30.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

de Vos, J.M., Augustijnen, H., Bätscher, L., Lucek, K., 2020. Speciation through chromosomal fusion and fission in Lepidoptera. Philos. Trans. R. Soc. Lond. B. Biol. Sci 375, 20190539. doi:10.1098/rstb.2019.0539

Doyle, S.R., Illingworth, C.J.R., Laing, R., Bartley, D.J., Redman, E., Martinelli, A., Holroyd, N., Morrison, A.A., Rezansoff, A., Tracey, A., Devaney, E., Berriman, M., Sargison, N., Cotton, J.A., Gilleard, J.S., 2019. Population genomic and evolutionary modelling analyses reveal a single major QTL for ivermectin drug resistance in the pathogenic nematode, Haemonchus contortus. BMC Genomics 20, 218. doi:10.1186/s12864-019-5592-6

Doyle, S.R., Tracey, A., Laing, R., Holroyd, N., Bartley, D., Bazant, W., Beasley, H., Beech, R., Britton, C., Brooks, K., Maitland, K., Martinelli, A., Noonan, J.D., Paulini, M., Quail, M.A., Redman, E., Rodgers, F.H., Sallé, G., Sankaranarayanan, G., Howe, K.L., Cotton, J.A., 2020. Extensive genomic and transcriptomic variation defines the chromosome-scale assembly of Haemonchus contortus, a model gastrointestinal worm. BioRxiv. doi:10.1101/2020.02.18.945246

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. doi:10.1093/bioinformatics/btq461

Ellinghaus, D., Kurtz, S., Willhoeft, U., 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9, 18. doi:10.1186/1471-2105-9-18

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157. doi:10.1186/s13059-015-0721-2

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238. doi:10.1186/s13059-019-1832-y

Evans, D., Zorio, D., MacMorris, M., Winter, C.E., Lea, K., Blumenthal, T., 1997. Operons and SL2 trans-splicing exist in nematodes outside the genus Caenorhabditis. Proc Natl Acad Sci USA 94, 9751–9756. doi:10.1073/pnas.94.18.9751

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta, M., Qureshi, M., Sangrador-Vegas, A., Salazar, G.A., Tate, J., Bateman, A., 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279-85. doi:10.1093/nar/gkv1344

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA 117, 9451–9457. doi:10.1073/pnas.1921046117

Foster, J.M., Grote, A., Mattick, J., Tracey, A., Tsai, Y.-C., Chung, M., Cotton, J.A., Clark, T.A., Geber, A., Holroyd, N., Korlach, J., Li, Y., Libro, S., Lustigman, S., Michalski, M.L., Paulini, M., Rogers, M.B., Teigen, L., Twaddle, A., Welch, L., Ghedin, E., 2020. Sex chromosome evolution in parasitic nematodes of humans. Nat. Commun. 11, 1964. doi:10.1038/s41467-020-15654-6

Fradin, H., Kiontke, K., Zegar, C., Gutwein, M., Lucas, J., Kovtun, M., Corcoran, D.L., Baugh, L.R., Fitch, D.H.A., Piano, F., Gunsalus, K.C., 2017. Genome architecture and

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

evolution of a unichromosomal asexual nematode. Curr. Biol. 27, 2928-2939.e6. doi:10.1016/j.cub.2017.08.038

Gu, L., Reilly, P.F., Lewis, J.J., Reed, R.D., Andolfatto, P., Walters, J.R., 2019. Dichotomy of Dosage Compensation along the Neo Z Chromosome of the Monarch Butterfly. Curr. Biol. 29, 4071-4077.e3. doi:10.1016/j.cub.2019.09.056

Haas, B., 2007. TransposonPSI: an application of PSI-Blast to mine (retro-) transposon ORF homologies. Broad Institute, Cambridge, MA, USA.

Hammarlund, M., Davis, M.W., Nguyen, H., Dayton, D., Jorgensen, E.M., 2005. Heterozygous insertions alter crossover distribution but allow crossover interference in Caenorhabditis elegans. Genetics 171, 1047–1056. doi:10.1534/genetics.105.044834

Hill, J., Rastas, P., Hornett, E.A., Neethiraj, R., Clark, N., Morehouse, N., de la Paz Celorio-Mancera, M., Cols, J.C., Dircksen, H., Meslin, C., Keehnen, N., Pruisscher, P., Sikkink, K., Vives, M., Vogel, H., Wiklund, C., Woronik, A., Boggs, C.L., Nylin, S., Wheat, C.W., 2019. Unprecedented reorganization of holocentric chromosomes provides insights into the enigma of lepidopteran chromosome evolution. Sci. Adv. 5, eaau3648. doi:10.1126/sciadv.aau3648

Hubley, R., Finn, R.D., Clements, J., Eddy, S.R., Jones, T.A., Bao, W., Smit, A.F.A., Wheeler, T.J., 2016. The Dfam database of repetitive DNA families. Nucleic Acids Res. 44, D81-9. doi:10.1093/nar/gkv1272

Kamath, R.S., Fraser, A.G., Dong, Y., Poulin, G., Durbin, R., Gotta, M., Kanapin, A., Le Bot, N., Moreno, S., Sohrmann, M., Welchman, D.P., Zipperlen, P., Ahringer, J., 2003. Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. Nature 421, 231–237. doi:10.1038/nature01278

Kim, J., Farré, M., Auvil, L., Capitanu, B., Larkin, D.M., Ma, J., Lewin, H.A., 2017. Reconstruction and evolutionary history of eutherian chromosomes. Proc Natl Acad Sci USA 114, E5379–E5388. doi:10.1073/pnas.1702012114

Kinsella, C.M., Ruiz-Ruano, F.J., Dion-Côté, A.-M., Charles, A.J., Gossmann, T.I., Cabrero, J., Kappei, D., Hemmings, N., Simons, M.J.P., Camacho, J.P.M., Forstmeier, W., Suh, A., 2019. Programmed DNA elimination of germline development genes in songbirds. Nat. Commun. 10, 5468. doi:10.1038/s41467-019-13427-4

Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37, 540–546. doi:10.1038/s41587-019-0072-8

Krumlauf, R., 2018. Hox genes, clusters and collinearity. Int. J. Dev. Biol. 62, 659–663. doi:10.1387/ijdb.180330rr

Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., Marra, M.A., 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645. doi:10.1101/gr.092759.109

Laetsch, D.R., Blaxter, M.L., 2017. KinFin: Software for Taxon-Aware Analysis of Clustered Protein Sequences. G3 (Bethesda) 7, 3349–3357. doi:10.1534/g3.117.300233

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lefoulon, E., Bain, O., Bourret, J., Junker, K., Guerrero, R., Cañizales, I., Kuzmin, Y., Satoto, T.B.T., Cardenas-Callirgos, J.M., de Souza Lima, S., Raccurt, C., Mutafchiev, Y., Gavotte, L., Martin, C., 2015. Shaking the Tree: Multi-locus Sequence Typing Usurps Current Onchocercid (Filarial Nematode) Phylogeny. PLoS Negl. Trop. Dis. 9, e0004233. doi:10.1371/journal.pntd.0004233

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Llorens, C., Futami, R., Covelli, L., Domínguez-Escribá, L., Viu, J.M., Tamarit, D., Aguilar-Rodríguez, J., Vicente-Ripolles, M., Fuster, G., Bernet, G.P., Maumus, F., Munoz-Pomer, A., Sempere, J.M., Latorre, A., Moya, A., 2011. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. 39, D70-4. doi:10.1093/nar/gkq1061

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., Borodovsky, M., 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 33, 6494–6506. doi:10.1093/nar/gki937

Melters, D.P., Paliulis, L.V., Korf, I.F., Chan, S.W.L., 2012. Holocentric chromosomes: convergent evolution, meiotic adaptations, and genomic analysis. Chromosome Res. 20, 579–593. doi:10.1007/s10577-012-9292-1

Mongue, A.J., Nguyen, P., Volenikova, A., Walters, J.R., 2017. Neo-sex Chromosomes in the Monarch Butterfly, Danaus plexippus. G3 (Bethesda) 7, 3281–3294. doi:10.1534/g3.117.300187

Nakatani, Y., Takeda, H., Kohara, Y., Morishita, S., 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res. 17, 1254–1265. doi:10.1101/gr.6316407

Nawrocki, E.P., Burge, S.W., Bateman, A., Daub, J., Eberhardt, R.Y., Eddy, S.R., Floden, E.W., Gardner, P.P., Jones, T.A., Tate, J., Finn, R.D., 2015. Rfam 12.0: updates to the RNA families database. Nucleic Acids Res. 43, D130-7. doi:10.1093/nar/gku1063

Nemetschke, L., Eberhardt, A.G., Hertzberg, H., Streit, A., 2010. Genetics, chromatin diminution, and sex chromosome evolution in the parasitic nematode genus Strongyloides. Curr. Biol. 20, 1687–1696. doi:10.1016/j.cub.2010.08.014

Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., Miga, K.H., Eichler, E.E., Phillippy, A.M., Koren, S., 2020. HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Res. doi:10.1101/gr.263566.120

Opperman, C.H., Bird, D.M., Williamson, V.M., Rokhsar, D.S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T.D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B.R., Thomas, V.P., Windham, E., 2008. Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. Proc Natl Acad Sci USA 105, 14802–14807. doi:10.1073/pnas.0805946105

Post, R., 2005. The chromosomes of the Filariae. Filaria J. 4, 10. doi:10.1186/1475-2883-4-10

Putnam, N.H., Butts, T., Ferrier, D.E.K., Furlong, R.F., Hellsten, U., Kawashima, T.,

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Robinson-Rechavi, M., Shoguchi, E., Terry, A., Yu, J.-K., Benito-Gutiérrez, E.L., Dubchak, I., Garcia-Fernàndez, J., Gibson-Brown, J.J., Grigoriev, I.V., Horton, A.C., de Jong, P.J., Jurka, J., Kapitonov, V.V., Kohara, Y., Rokhsar, D.S., 2008. The amphioxus genome and the evolution of the chordate karyotype. Nature 453, 1064–1071. doi:10.1038/nature06967

Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., Vilo, J., 2019. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 47, W191–W198. doi:10.1093/nar/gkz369

Rockman, M.V., Kruglyak, L., 2009. Recombinational landscape and population genomics of Caenorhabditis elegans. PLoS Genet. 5, e1000419. doi:10.1371/journal.pgen.1000419

Rödelsperger, C., Meyer, J.M., Prabh, N., Lanz, C., Bemm, F., Sommer, R.J., 2017. Single-Molecule Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model Organism Pristionchus pacificus. Cell Rep. 21, 834–844. doi:10.1016/j.celrep.2017.09.077

Rzeszutek, I., Maurer-Alcalá, X.X., Nowacki, M., 2020. Programmed genome rearrangements in ciliates. Cell. Mol. Life Sci. doi:10.1007/s00018-020-03555-2

Seemann, T., 2014. snippy: fast bacterial variant calling from NGS reads. https://github.com/tseemann/snippy.

Slos, D., Sudhaus, W., Stevens, L., Bert, W., Blaxter, M., 2017. Caenorhabditis monodelphis sp. n.: defining the stem morphology and genomics of the genus Caenorhabditis. BMC Zool. 2. doi:10.1186/s40850-017-0013-2

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D.H.A., Fulton, L.A., Fulton, R.E., Griffiths-Jones, S., Harris, T.W., Hillier, L.W., Kamath, R., Kuwabara, P.E., Waterston, R.H., 2003. The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol. 1, E45. doi:10.1371/journal.pbio.0000045

Steinbiss, S., Willhoeft, U., Gremme, G., Kurtz, S., 2009. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res. 37, 7002–7013. doi:10.1093/nar/gkp759

Stevens, L., Félix, M.-A., Beltran, T., Braendle, C., Caurcel, C., Fausett, S., Fitch, D., Frézal, L., Gosse, C., Kaur, T., Kiontke, K., Newton, M.D., Noble, L.M., Richaud, A., Rockman, M.V., Sudhaus, W., Blaxter, M., 2019. Comparative genomics of 10 new Caenorhabditis species. Evol. Lett. 3, 217–236. doi:10.1002/evl3.110

Stevens, L., Rooke, S., Falzon, L.C., Machuka, E.M., Momanyi, K., Murungi, M.K., Njoroge, S.M., Odinga, C.O., Ogendo, A., Ogola, J., Fèvre, E.M., Blaxter, M., 2020. The Genome of Caenorhabditis bovis. Curr. Biol. 30, 1023-1031.e4. doi:10.1016/j.cub.2020.01.074

Stiernagle, T., 2006. Maintenance of C. elegans. WormBook 1–11. doi:10.1895/wormbook.1.101.1

Sturtevant, A.H., Dobzhansky, T., 1936. Inversions in the third chromosome of wild races of drosophila pseudoobscura, and their use in the study of the history of the

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

species. Proc Natl Acad Sci USA 22, 448–450. doi:10.1073/pnas.22.7.448

Tandonnet, S., Koutsovoulos, G.D., Adams, S., Cloarec, D., Parihar, M., Blaxter, M.L., Pires-daSilva, A., 2019. Chromosome-Wide Evolution and Sex Determination in the Three-Sexed Nematode Auanema rhodensis. G3 (Bethesda) 9, 1211–1230. doi:10.1534/g3.119.0011

Tarailo-Graovac, M., Chen, N., 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics Chapter 4, Unit 4.10. doi:10.1002/0471250953.bi0410s25

Tenenbaum, D., 2019. KEGGREST: Client-side REST access to KEGG. https://www.bioconductor.org/packages/release/bioc/html/KEGGREST.html.

Teterina, A.A., Willis, J.H., Phillips, P.C., 2020. Chromosome-Level Assembly of the Caenorhabditis remanei Genome Reveals Conserved Patterns of Nematode Genome Organization. Genetics 214, 769–780. doi:10.1534/genetics.119.303018

Thompson, J.D., Gibson, T.J., Higgins, D.G., 2002. Multiple sequence alignment using ClustalW and ClustalX. Curr. Protoc. Bioinformatics Chapter 2, Unit 2.3. doi:10.1002/0471250953.bi0203s00

Triantaphyllou, A.C., 1963. Polyploidy and parthenogenesis in the root-knot nematode meloidogyne arenaria. J. Morphol. 113, 489–499. doi:10.1002/jmor.1051130309

Vaser, R., Sović, I., Nagarajan, N., Šikić, M., 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27, 737–746. doi:10.1101/gr.214270.116

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9, e112963. doi:10.1371/journal.pone.0112963

Walton, A.C., 1959. Some parasites and their chromosomes. J. Parasitol. 45, 1–20.

Wang, J., Davis, R.E., 2014. Programmed DNA elimination in multicellular organisms. Curr. Opin. Genet. Dev. 27, 26–34. doi:10.1016/j.gde.2014.03.012

Wang, J., Mitreva, M., Berriman, M., Thorne, A., Magrini, V., Koutsovoulos, G., Kumar, S., Blaxter, M.L., Davis, R.E., 2012. Silencing of germline-expressed genes by DNA elimination in somatic cells. Dev. Cell 23, 1072–1080. doi:10.1016/j.devcel.2012.09.020

Wang, J., Veronezi, G.M.B., Kang, Y., Zagoskin, M., O'Toole, E.T., Davis, R.E., 2020. Comprehensive Chromosome End Remodeling during Programmed DNA Elimination. Curr. Biol. doi:10.1016/j.cub.2020.06.058

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. doi:10.1093/bioinformatics/btp033

Wheeler, T.J., Eddy, S.R., 2013. nhmmer: DNA homology search with profile HMMs. Bioinformatics 29, 2487–2489. doi:10.1093/bioinformatics/btt403

Xiong, W., He, L., Lai, J., Dooner, H.K., Du, C., 2014. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. Proc Natl Acad

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Sci USA 111, 10263–10268. doi:10.1073/pnas.1410068111

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure Legends

Figure 1: Dual telomeres on the chromosomes of *Oscheius tipulae* CEW1

A Circos plot of the *O. tipulae* CEW1 genome showing (from outside to inside) the chromosomes with a scale in Mb, coverage in PromethION reads (dark grey), coverage in Illumina short reads (light grey), count of telomeric repeats per 10 kb (red), density of repeats (red) and GC proportion in 10kb windows (line plot). The inner arcs link all significant nucleotide sequence (blastn) matches between the presumed near-complete copy of the helitron on the left end of Otip_IV (with an open reading frame of 3895 amino acids, spanning 26 kb of the telomere extension) and the rest of the genome.

B Circos plot of the low-coverage telomeric extensions of *O. tipulae* CEW1, with 10 kb of normal coverage flanking chromosome, showing (from outside to inside) the chromosomes with a scale in 100 kb, coverage in PromethION reads (dark grey), coverage in illumina short reads (light grey), count of telomeric repeats per 10 kb (red), density of repeats (red) and predicted coding genes (green). Arcs are drawn linking significant blastn matches between three sequential helitron-like components from the right telomere extension of chromosome X (yellow, orange and red links), between repeats limited to within Otip_X R (dark red), and between all other telomeric extension sequences (light blue).

Figure 2: The *Oscheius tipulae* CEW1 genome

Repeat and gene densities along *O. tipulae* chromosomes, and mapping of placement of orthologues of *C. elegans* and Nigon element assigned loci. Feature densities (repeat and gene features and orthologue mapping) were calculated in 500 kb non-overlapping windows along each chromosome. For the mapping of orthologues (rightmost two chart sets), the plots are stacked bars coloured by either *C. elegans* chromosome (Ce chr) or Nigon element group (Nigon).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 3: Local gene order comparison between *Oscheius tipulae* and *Caenorhabditis elegans*

A Whole chromosome comparison of *Oscheius tipulae* chromosome I (Otip_I) and its homologue, *Caenorhabditis elegans* chromosome I (Cele_I). Lines link loci that are reciprocal best BLAST hits. In the top panel the lines are coloured by their placement on *C. elegans* Cele_I (a spectrum from red on the left arm through to purple on the right arm). In the lower panel the lines are coloured by their placement on *O. tipulae* Otip_I (a spectrum from red on the left arm through to purple on the right arm).

B Gene neighbourhoods are not conserved between *O. tipulae* and *C, elegans*. We measured the separation distance between each pair of neighbouring *C. elegans* loci for which we could identify single-copy reciprocal best-BLAST relationships between *O. tipulae* and *C. elegans*, and plotted the separation between these orthologue pairs in *C. elegans* (x axis) and *O. tipulae* (y axis) for each chromosome. Ortholog pairs more than 50 kb apart in *C. elegans* were excluded. R: Spearman correlation; n: number of orthologous pairs including pairs excluded from the plot; m: the number of pairs excluded from each plot.

Figure 4: Loci that define Nigon elements in rhabditid nematodes

A t-distributed Stochastic Neighbor Embedding (t-SNE) plot of the Gower distance between 3412 orthologous gene families mapped to the chromosomal assemblies of 9 species. The 2191 loci that are included in the Nigon element sets are coloured. The black dots represent loci not assigned to a Nigon unit.

B For each Nigon-defining set of loci, we counted the number found in each species. The total number of loci per set is indicated by a star. The assembly of *Meloidogyne hapla* has the fewest loci and lowest proportion of loci in all Nigon sets.

C The proportion of Nigon-defining loci found in each species' genome correlates with the BUSCO completeness of the genome.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 5: Nigon painting of Rhabditid nematode chromosomes

Examples of nematode chromosomal assemblies painted by their content of Nigon-defining loci. Each subgraph shows the count of loci mapped in non-overlapping 0.5 Mb windows along a chromosome as a stacked histogram coloured by Nigon origin. The X- and Y-axes are scaled to the maxima within a species in each panel (X: chromosome length, Y: Nigon-defining loci per interval) within each species. The legend in panel J applies to all nine chromosome panels.

A *Caenorhabditis elegans* (Rhabditina, Rhabditomorpha), B *Haemonchus contortus* (Rhabditina, Rhabditomorpha), C *Auanema rhodensis* (Rhabditina, Rhabditomorpha), D *Pristionchus pacificus* (Rhabditina, Diplogasteromorpha), E *Strongyloides ratti* (Tylenchina, Panagrolaimomorpha), F *Steinernema carpocapsae* (not a fully chromosomal assembly; Tylenchina, Panagrolaimomorpha), G *Brugia malayi* (Spirurina, Spiruromorpha), H *Onchocerca volvulus* (Spirurina, Spiruromorpha), I *Ascaris suum* (Spirurina, Ascaridomorpha). Panel J shows the colour key for the other panels.

Figure 6: A model of chromosome evolution in Rhabditida.

For each species the classification of chromosomes to Nigon elements is shown (see Figure 2, Figure 5 and Supplemental Figure S7), including rearranged chromosomes. X chromosomes are indicated by a star. For each lineage we have inferred the patterns of chromosome scission (the ✂ symbol; the number indicates the number of fragments resulting) and fusion (the Kanji symbol for "fusion point" 合) on the tree. Where the order in time of events is not resolved, we have bracketed them. In *A. suum* a pink triangle indicates positions of internal cleavage of germline chromosomes during diminution. Note that the "partial Nigons" in *S. carpocapsae* are assembly scaffolds rather than chromosomes. The cladogram representing the phylogeny of Chromadoria nematodes (inset, lower left; (Blaxter and Koutsovoulos, 2015)) is derived from phylogenetic analysis of shared protein coding genes. In the main figure we have resolved a previous uncertain polytomy in Rhabditomorpha to favour a closer relationship between *Haemonchus contortus* and the genus *Caenorhabditis* based on the exact conservation of linkage groups.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 7: Sex chromosome evolution in the filarial nematodes (Spiruromorpha)

Within the filarial nematodes, karyotypes have been determined for a number of species. The phylogeny (cladogram to the left) is derived from multilocus phylogenomic analysis, and is in agreement with marker gene-based phylogeny (Lefoulon et al., 2015). The karyograms and karyotype data are from the work of Rory Post and colleagues (Post, 2005). The inferred position of fusion events within the filarial phylogeny are indicated, and the karyotypes of *O. volvulus* and *B. malayi* are coloured by their Nigon assignment (see Figure 5 G,H and Figure 6).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Information

Supplementary Text S1: Defining sets of orthologous loci in linkage through Rhabditid evolution

Supplementary Text S2: Nigon element analysis of *Meloidogyne hapla*

Figure S1: Raw data QC and read filtering

Figure S2: Distribution of SL and 5S RNA loci in the *Oscheius tipulae* genome

Figure S3: Clustering with different k values

Figure S4: t-SNE of orthologous loci co-locations in 9 nematode species

Figure S5: nOti 2.0 assembly aligned to nrOscTipu1.3

Figure S6: GC proportion plots of nematode chromosomes

Figure S7: Nigon painting of *Caenorhabditis* species

Figure S8: Exploring the robustness of definition of Nigon element-defining sets of orthologues

Table S1: Chromosome numbers in Nematoda

Table S2: Software used

Table S3: Raw data for genomic analysis of *Oscheius tipulae*

Table S4. Sources of nematode assemblies included in the Kraken2 database for contamination assessment

Table S5. Raw sequence allocation by Kraken grouped by Phylum

Table S6. Contiguity and BUSCO metrics of *Oscheius tipulae* assemblies

Table S7. Base fold coverage of the *Oscheius tipulae* assembly

Table S8: Sources and metrics of rhabditid nematode genome data

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S9. BUSCO scores of assemblies and proteomes of rhabditid nematodes

Table S10: Nigon element-defining loci in fifteen nematode genomes

Table S11. KEGG categories enriched among Nigon defining loci

Table S12: Nigon-defining loci in the *Meloidogyne hapla* genome assembly

Supplementary references

Supplementary Text S1: Defining sets of orthologous loci in linkage through Rhabditid evolution

We identified sets of loci that potentially define ancestral linkage groups - Nigon elements - by analysing the chromosomal colocation of sets of orthologues across nine chromosomally assembled nematode genomes. The nine genomes analysed were *Auanema rhodensis, Brugia malayi, Caenorhabditis elegans, Haemonchus contortus, Onchocerca volvulus, Oscheius tipulae, Pristionchus pacificus, Steinernema carpocapsae* and *Strongyloides ratti* (see Table S4). Proteomes of these nine species were clustered into orthogroups using orthofinder (Emms and Kelly, 2019). The orthofinder output was imported into KinFin (Dominik R Laetsch and Blaxter, 2017), and built in KinFin routines used to select sets of orthologues that conformed to particular presence-absence patterns. One to one orthologues, or single copy orthologues are orthologue sets where all the species in the analysis have exactly one copy. We defined fuzzy one to one orthologues, where an orthologue set was selected if it contained at least a given number of species with one copy in each, and the remaining species were permitted to be missing the locus (count of 0) or to have more than one copy. This allows fuzzy one to one orthologues to be defined even when one or a few genomes have experienced gene loss, where there are idiosyncratic gene duplications (or uncollapsed retained haploid duplications) or where the gene set or genome is incomplete. We explored the landscape of fuzzy one to one orthologue definition and

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

chose a conservative cutoff of requiring seven of the nine input species to have single members, and allowing two species to have 0 or >1.

For each of these orthologues we recorded which chromosome it was placed on in each species, yielding a matrix where each orthologue was described by an array of chromosomal assignments. Missing and duplicate orthologues were assigned null values. Gower distances between orthologues were calculated from these categorical data, yielding a normalised distance matrix relating all orthologues across all species. This distance matrix was examined using t-SNE with different perplexity values and Graphia Pro using different correlation cutoff values We identified seven clusters of orthologues (Figure S4). CLARA clustering was used to cluster the orthologues by Dice distance, exploring a range of values of K from 1 to 10. Clustering with K=7 had the best scores in terms of mean silhouette width (Figure S3), and these seven clusters were used to define sets of loci that are associated with seven ancestral Nigon elements.

We explored the robustness of definition of these seven Nigon-defining orthologue sets by exploring the effect of the 1-to-1 orthologous relationship in different number of species used in identifying them (Figure S8). When low numbers of species were analysed, identification of seven elements was not achieved, and elements N and X tended to be merged. However for all analyses using over three species in the analysis, the numbers of orthology groups assigned to each putative Nigon-defining set was relatively constant. When the genomes with many chromosomes (*A. suum*, *M. hapla*) or the less-well assembled genomes were used, the number of Nigon-defining loci was impacted.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary Text S2: Nigon element analysis of *Meloidogyne hapla*

Introduction

The genus *Meloidogyne* includes both diploid species and major clades of polyploid hybrids, with karyotypes of n=17 in diploids and n of up to 70 in polyploids. *Meloidogyne hapla* is diploid and has 17 chromosomes, defined by genetic mapping and karyotyping (Thomas et al., 2012). The genome assembly has been superscaffolded into 19 scaffolds using genetic cross mapping, with chromosomes 1 (mh_1A, mh_1B) and 2 (mh_2A, mh_2B) split in two. Each single superscaffold (e.g. "mh_10.12.8") corresponds to a single genetic linkage group. There are 275 short, unplaced scaffolds. The superscaffolds were analysed in detail, with the split chromosomes 1 and 2 kept as two fragments.

Methods

Nigon-unit defining loci were identified in *Meloidogyne hapla* and their mapping to superscaffolds recorded.

Results

Compared to the other nematode species analysed, the published genome sequence for *M. hapla* had a noticeably lower proportion of matches to the set of shared orthologues (Figure 6 B,C). Of 2191 Nigon-defining loci, 1688 (or 77%) had matches in the *M. hapla* genome. The mapping rate was lowest for Nigon set X (68%) and highest for Nigon set B (81%) (Table S12). Three quarters (75%) of the mapped loci from the Nigon-defining sets were present in the chromosome-sized contigs, with the highest proportion of Nigon set matches in these unplaced contigs from the X set (38%). Most of the unplaced scaffolds had a single Nigon locus mapped (range 1-17, mean 1.54).

The number of Nigon-defining loci mapped to each chromosomal molecule ranged from 8 to 157. The two chromosome 1 contigs had a total of 201 loci mapped and the two chromosome 2 contigs had a total of 144 loci mapped. Four of the shorter chromosome-mapped scaffolds had <40 Nigon-defining loci mapped, and these were not analysed further. Assessment of possible assignment to ancestral Nigon elements using simple counts or proportions of mapped Nigon-defining loci (Table S12) biased

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

assignment to those Nigon elements that contained more loci (e.g. the number of loci used to define Nigon X is less than one sixth the number used to define Nigons A and C). Thus we normalised Nigon-defining locus counts per chromosomal scaffold by expressing them as the proportion of Nigon-defining loci from each Nigon set.

Using normalised proportions, a few chromosomal superscaffolds could be assigned to ancestral Nigon elements with some confidence. Thus chromosome mh_16 contained 50% of all the NigonE set loci mapped to the chromosomes, and chromosome mh_8 contained 38% of the mapped NigonX set. Other chromosomes showed a bias towards one or a few Nigon-defining sets. Chromosome mh_5 had overrepresentation of NigonB, mh11 of NigonN, mh4 of NigonC, mh5 of NigonB and mh_5 of NigonX loci. Notably, the assignments of the two halves of chromosome 1 (mh-1A and mh_1B) were congruent (highest proportion to NigonA in both), while the mappings of mh_2A and mh_2B from chromosome 2 were discordant (highest proportions to NigonB and NigonD respectively).

Discussion

Overall, the assignment of *M. hapla* chromosomal superscaffolds to Nigon origin is complex. No scaffold was simply assigned to one Nigon, but only seven had content where one Nigon set constituted a majority of mapped loci, and these counts were biased by the count of loci in each set. In *Strongyloides ratti* and *Steinernema carpocapsae*, we identified a shared fusion of NigonB, NigonD and NigonX to form the X chromosome of each species, with additional complex rearrangements in *S. ratti*. In the *M. hapla* assignments there is no signal of this predicted fusion as NigonX-derived loci are not particularly associated with either NigonB or NigonD loci. The two scaffolds on which a majority of NigonX loci were mapped (mh6 and mh8) had very few NigonD loci (zero and 1 loci respectively). The additional fusions predicted in *S. ratti* (fusions of C+E+N and of A+C+D+E+N) were also not evident.

Thus we conclude that, if the chromosomes of *M. hapla* contain a remnant signature of their derivation from ancestral Nigon elements, the fragmentation implied in going from seven ancestral to seventeen extant chromosomes also involved a large number of translocations. It will be very informative to examine additional high-quality

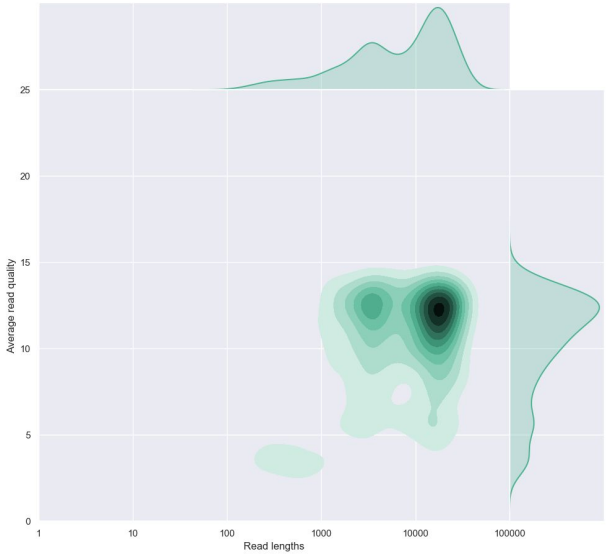
bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

genomes from additional *Meloidogyne* species, and other chromosomally-contiguous genomes from the Heteroderidae. Many *Meloidogyne* species are complex triploid and tetraploid hybrids, making assembly and interpretation difficult, but recent progress with long read data generation shows promise in generation of high quality genome estimates (Susič et al., 2020; Szitenberg et al., 2017).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S1: **Raw data QC and read filtering**

A



B

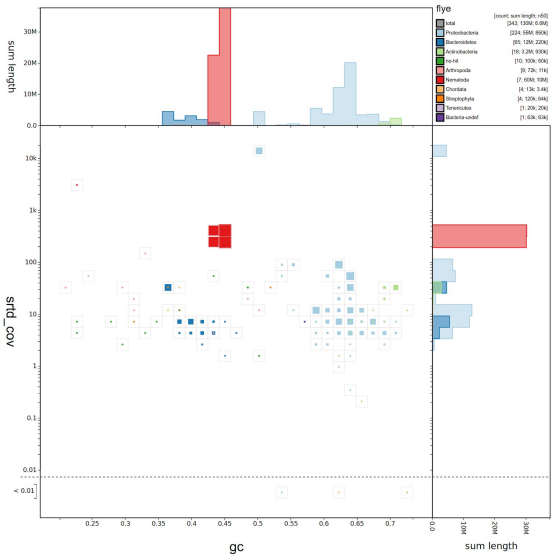


Figure S1. Raw data QC and read filtering

A. Density plot of raw PromethION nanopore read length (x axis) against average read quality (y axis) (generated by Nanoplot).

B. BlobToolKit (Challis et al., 2020) plot (GC proportion [x axis] vestus read coverage [y axis]) of the flye assembly of all the sequenced reads. Taxonomic annotation was achieved using blastx and Diamond tblastn source classification.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S2**: Distribution of SL and 5S RNA loci in the Oscheius tipulae genome**

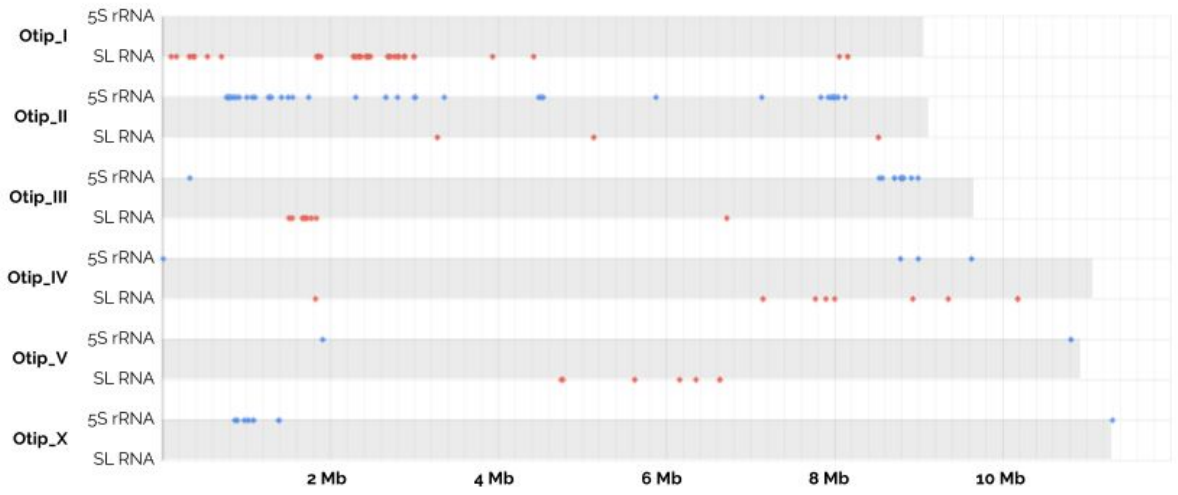


Figure S2: Distribution of SL and 5S RNA loci in the *Oscheius tipulae* genome

Spliced leader RNA and 5S ribosomal RNA loci were identified using Rfam (Kalvari et al., 2018) models and Rnammer (Lagesen et al., 2007), and are plotted by position along each of the *O. tipulae* chromosomes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S3: **Clustering with different k values**

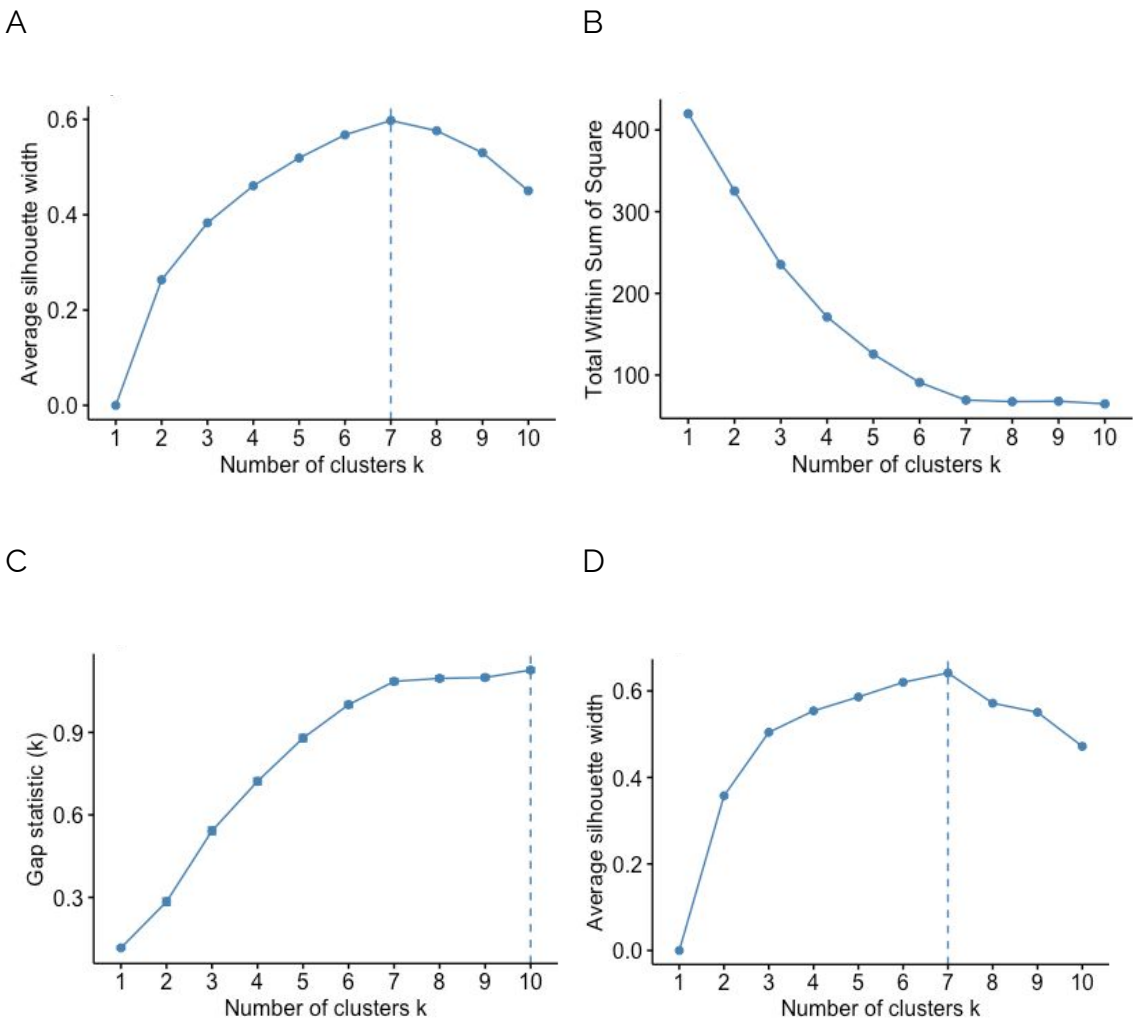


Figure S3. Clustering with different k values

Optimal number of clusters between 1 and 10 using CLARA (A, B and C) or PAM (D) on the gower distance matrix of 1172 single copy orthologous genes. Silhouette (A and D), within sum of square (B) and gap statistic (C) were used to estimate the optimal number of clusters. One hundred Monte Carlo samples were performed for the gap statistic.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S4: **t-SNE of orthologous loci co-locations in 9 nematode species**

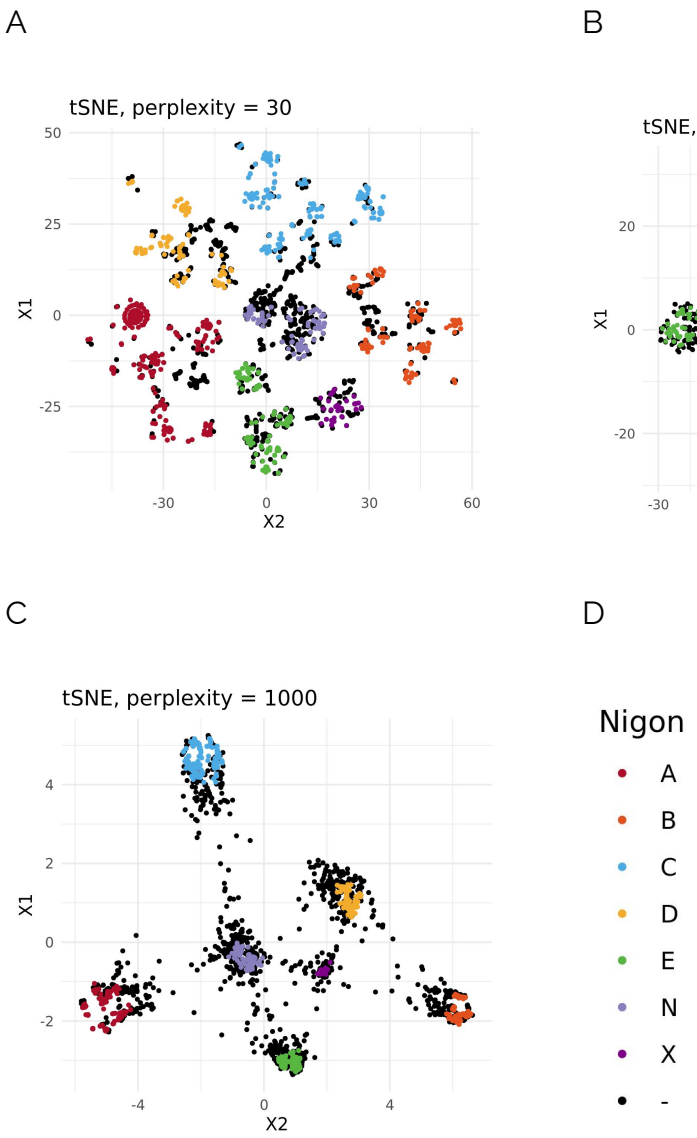


Figure S4. t-SNE of orthologous loci co-locations in 9 nematode species

Alternative t-SNE plots of 3412 orthologous gene families with different perplexity parameter values. A maximum of 1000 iterations were performed in all cases. Black dots denote orthologs not assigned to a Nigon unit. The legend in panel D applies to all tSNE panels. Parameters used for t-SNE, apart from the specified perplexity, are max_iter=1000, initial_dims=50 and theta=0.5.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S5: **nOti 2.0 assembly aligned to nrOscTipu1.3**

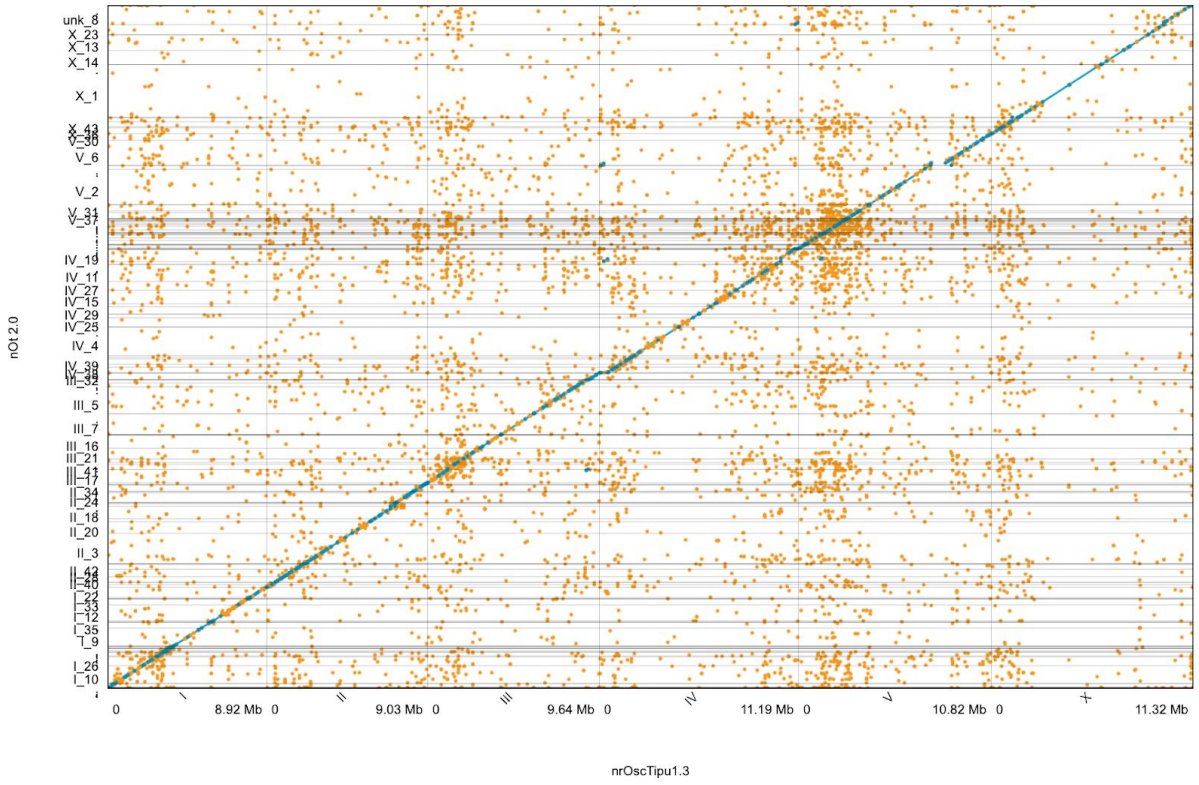


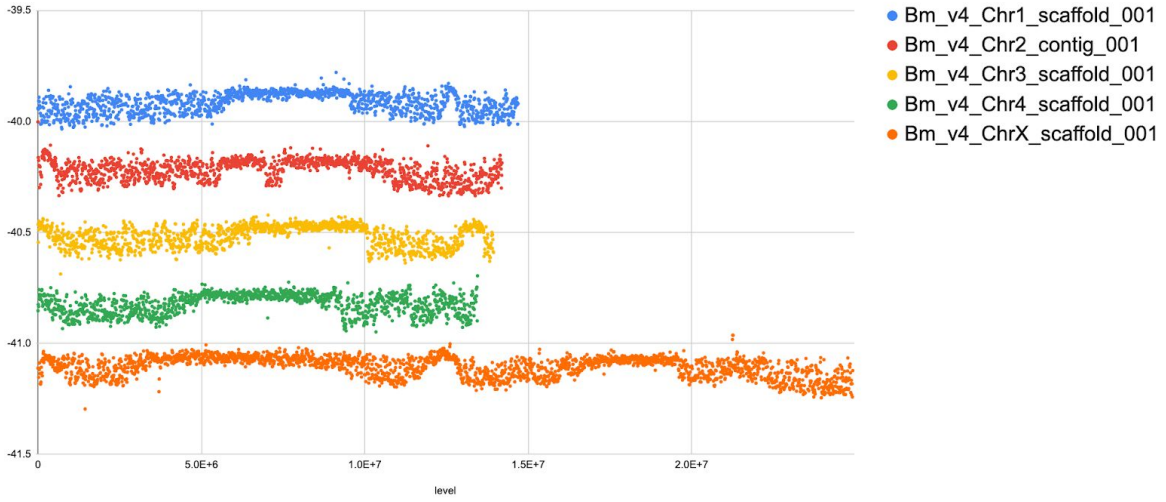
Figure S5: nOti 2.0 assembly aligned to nrOscTipu1.3

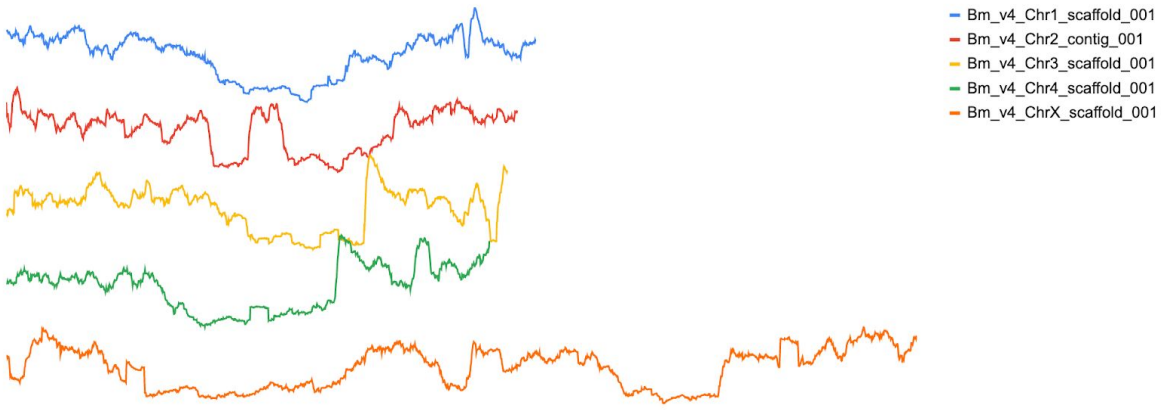
The telomere-to-telomere genome assembly (x axis) aligned to the previous assembly (Besnard et al., 2017) (y axis) in nucmer (Kurtz et al., 2004). Blue dots indicate co-oriented nucleotide matches, while orange dots indicate matches in opposite orientations.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S6: **GC proportion plots of nematode chromosomes**

A



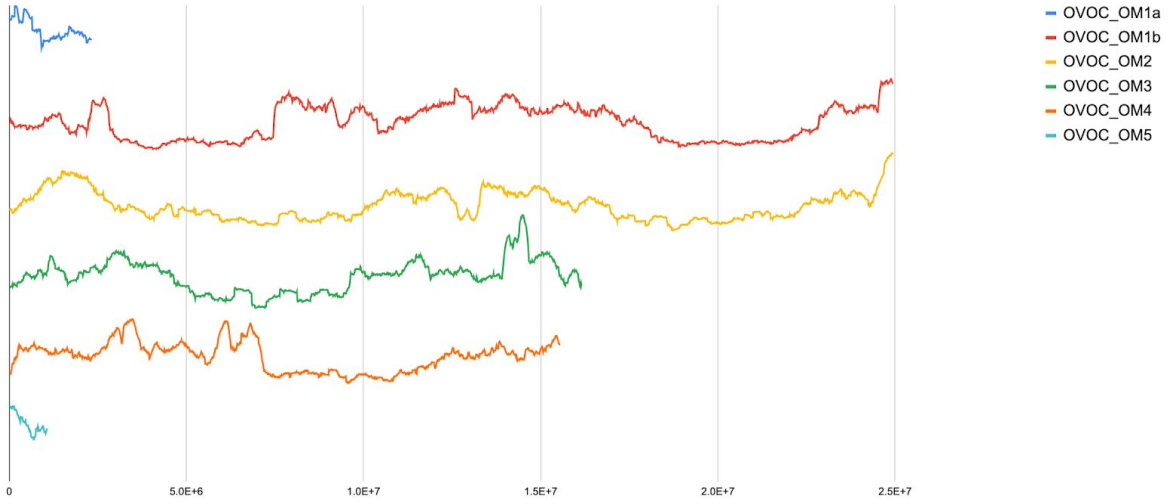


B

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

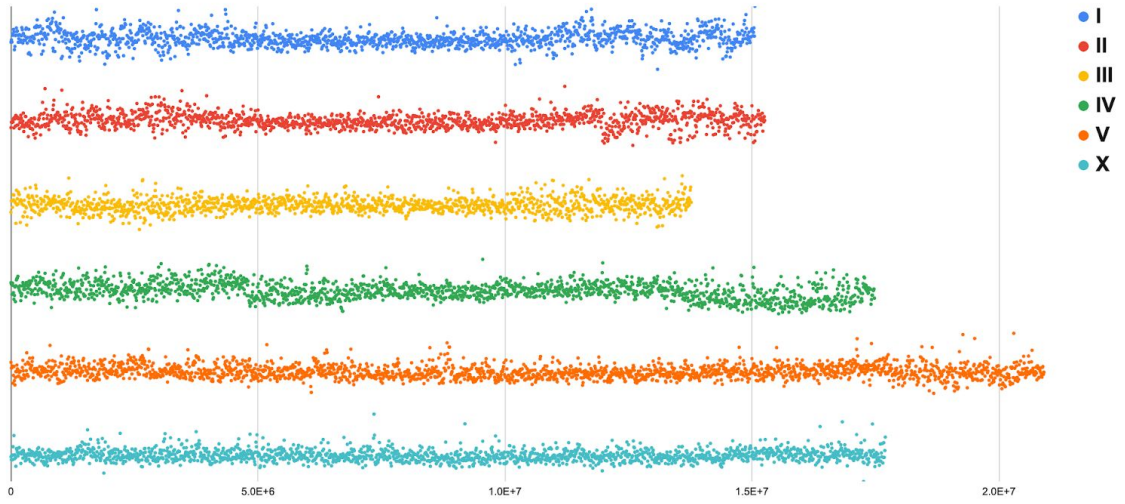
C

D



bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

E



F

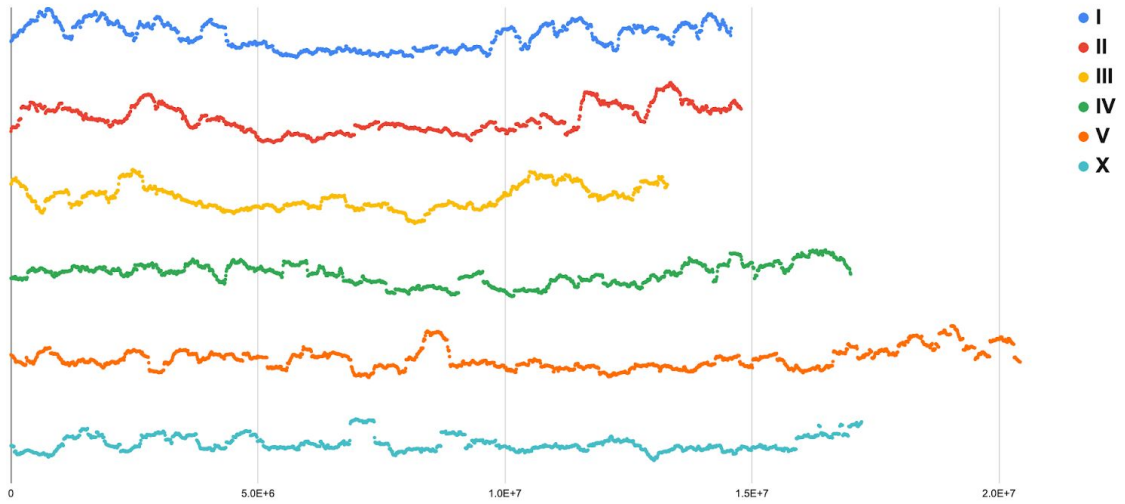


Figure S6: GC proportion plots of nematode chromosomes

A *B. malayi* GC proportion in 10kb windows across the chromosomes.

B *B. malayi* GC proportion standard deviation in 0.5 Mb sliding windows across the chromosomes.

C *O. volvulus* GC proportion in 10kb windows across the chromosomes.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

D *O. volvulus* GC proportion standard deviation in 0.5 Mb sliding windows across the chromosomes.

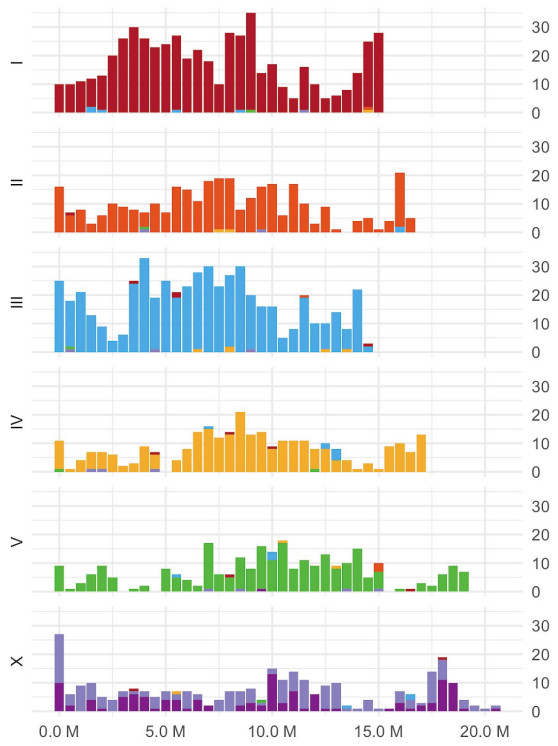
E *C. elegans* GC proportion in 10kb windows across the chromosomes.

F *C. elegans* GC proportion standard deviation in 0.5 Mb sliding windows across the chromosomes.

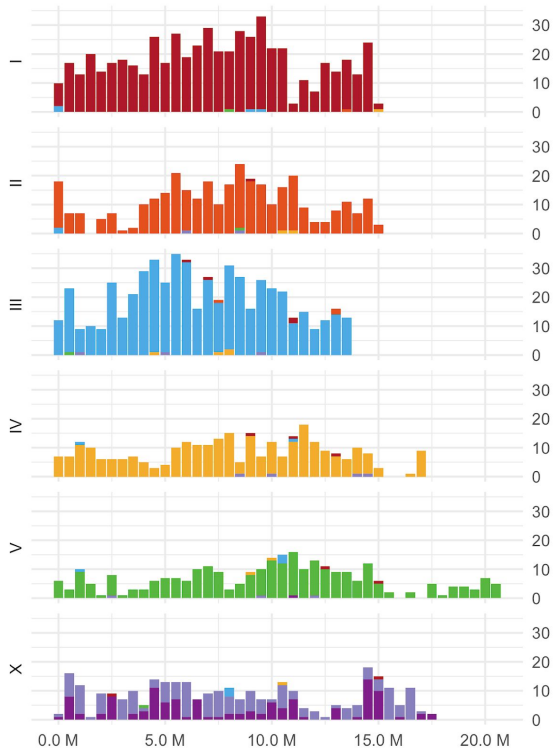
bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S7: **Nigon painting of *Caenorhabditis* species**

A



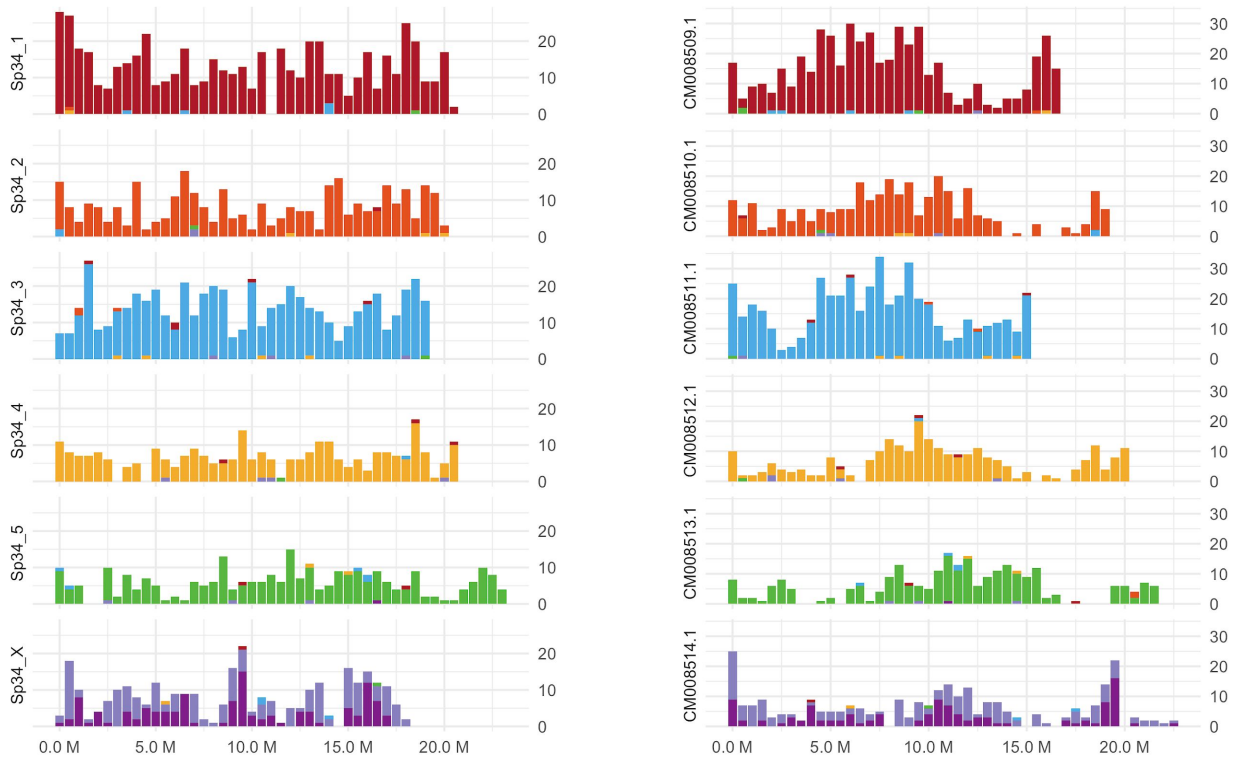
B



C

D

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

E

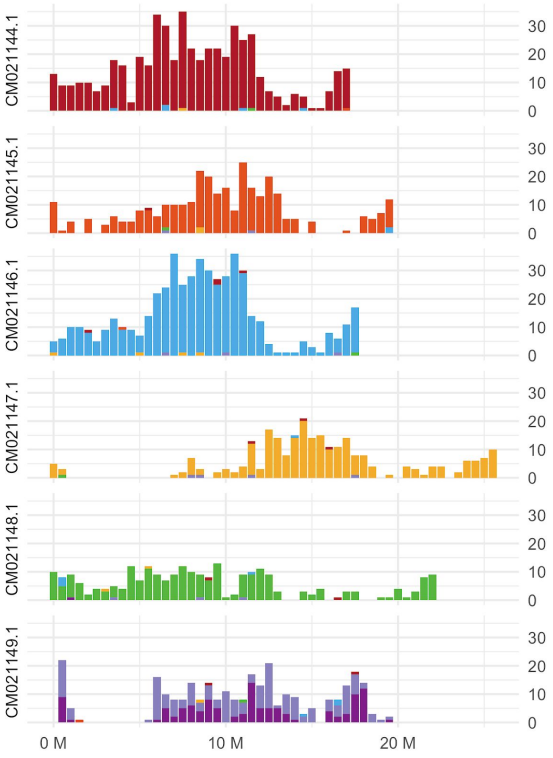


Figure S7: Nigon painting of *Caenorhabditis* species

Chromosomal assemblies of *Caenorhabditis* species painted using the Nigon-defining locus sets. See main text Figure 5 for other species. A *Caenorhabditis elegans,* B *Caenorhabditis briggsae,* C *Caenorhabditis inopinata,* D *Caenorhabditis nigoni*, E *Caenorhabditis remanei.*

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure S8: **Exploring the robustness of definition of Nigon element-defining sets of orthologues.**

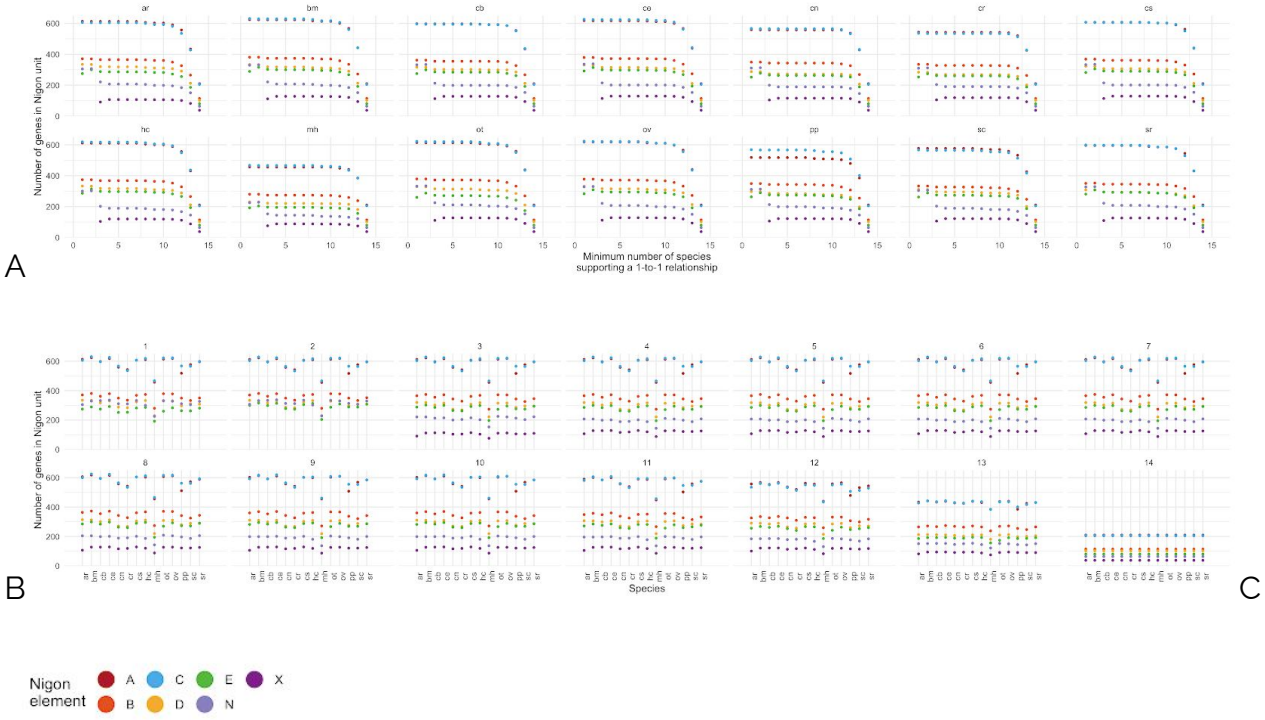


Figure S8: Exploring the robustness of definition of Nigon element-defining sets of orthologues.

A Numbers of orthologous gene sets assigned to each putative Nigon-defining set in analyses using different numbers of species (from 2-14) as the minimum number of species having a 1-to-1 orthologous relationships input to the K-means clustering. Species abbreviations are ar: *Auanema rhodensis,* bm: *Brugia malayi,* cb: *Caenorhabditis briggsae*, ce: *C. elegans*, cn: *C. nigoni*, cr: *C. remanei*, cs: *C. sp34*, hm: *Haemonchus contortus,* mh: *Meloidogyne hapla,* ot: *Oscheius tipulae,* ov: *Onchocerca volvulus,* pp: *Pristionchus pacificus,* sc: *Steinernema carpocapsae* and sr: *Strongyloides ratti*

B Same data as A, grouped by species rather than by minimum number of species having a 1 to 1 orthologous relationship. The number of orthologues assigned is stable in each species between K=3 and K=10.

C Legend corresponding to panels A and B.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S1: **Chromosome numbers in Nematoda***

Clade	***Subgroup***	***n in majority***	***others***
Clade I	Trichocephalida	3, 4	
Clade II		?	
Clade C	Desmodoridae	7	
Clade III	Oxyuromorpha	4, 8	
	Ascaridomorpha	variable	highly variable: n = 1-24
	Spiruromorpha	5, 6	rarely n = 4, 7
Clade IV	Strongyloides	3	
	Meloidogyninae	17	parthenogens variable to n > 50
	Pratylenchus	5 - 7	
	Panagrolaimus	4	triploid parthenogens n = 12 = 3x4
	Steinernema	6	also n = 8,10
Clade V	Strongylomorpha	6	
	Rhabditomorpha	6	varies n = 1 - 9 (especially variable in *Diploscapter*)
	Diplogasteromorpha	6	one with n = 5

* See (Walton, 1959) for a summary of chromosome numbers in nematode parasites.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S2: **Software used**

Software	***Version***	***Source***	***Parameters used if not default***	***Reference***
AGAT	0.2.3	https://github.com/NBISweden/AGAT		(Dainat, 2020)
bbduk	(bbtools) 38.51-0	http://sourceforge.net/projects/bbmap/	bbduk.sh ref=bbmap-38.51-0/resources/adapters.fa qtrim=w trimq=20	(Bushnell, 2017)
blast+	2.9.0	https://anaconda.org/bioconda/blast	-max_target_seqs 10 -max_hsps 1 -num_threads 8 -evalue 1e-25	(Camacho et al., 2009)
blobtools	v2.0.0alpha	https://github.com/blobtoolkit/blobtools2		(Dominik R. Laetsch and Blaxter, 2017)
BUSCO	4.0.2	https://gitlab.com/ezlab/busco	-l nematoda_odb10	(Simão et al., 2015)
bwa	0.7.17-r1188	https://github.com/lh3/bwa	mem	(Li and Durbin, 2009)
Canu	1.9	https://anaconda.org/bioconda/canu		(Koren et al., 2017)
diamond	0.9.24	https://anaconda.org/bioconda/diamond	--sensitive --max-target-seqs 1 --evalue 1e-25	(Buchfink et al., 2015)

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

flye	2.6-g0d65569	https://github.com/fenderglass/Flye	--genome-size 70m --asm-coverage 800 --plasmids --meta	(Kolmogorov et al., 2019)
GAP5	2.0.0b11	https://www.sanger.ac.uk/tool/gap5/		(Bonfield and Whitwham, 2010)
GeneMark	4.57_lic	http://topaz.gatech.edu/GeneMark/license_download.cgi		(Lomsadze et al., 2005)
Guppy	2.3.35	https://nanoporetech.com/		
KinFin	1.0.3	https://github.com/DRL/kinfin		(Dominik R Laetsch and Blaxter, 2017)
medaka	0.9.2	https://github.com/nanoporetech/medaka		(Oxford Nanopore Technologies Ltd., 2017)
minimap2	2.17	https://github.com/lh3/minimap2	-x map-ont	(Li, 2018)
MIRA	4.0.2	https://sourceforge.net/projects/mira-assembler/files/MIRA/stable/	job=est,denovo,accurate technology=454	(Chevreux et al., 1999)
Orthofinder	2.3.11	https://github.com/davidemms/OrthoFinder	-M msa -S blast	(Emms and Kelly, 2019, 2015)
PASA	2.3.3	https://github.com/PASApipeline/PASApipeline		(Haas et al., 2003)
pilon	1.22-0	https://github.com	--fix-all	(Walker et al., 2014)

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

		/broadinstitute/pilon		
R	3.6.0	https://www.r-project.org/		(Team, 2019)
racon	1.2.1	https://github.com/isovic/racon	-m 8 -x -6 -g -8 -w 500	(Vaser et al., 2017)
RaGOO	1.11	https://github.com/malonge/RaGOO	-i 0.51	(Alonge et al., 2019)
Repeatmasker	4.0.7	http://www.repeatmasker.org/RMDownload.html	-s	(Smit et al., 2015)
RepeatModeler	2.0.1	http://www.repeatmasker.org/RepeatModeler/	-LTRStruct	(Flynn et al., 2020)
TransposonPSI	08222010	http://transposonpsi.sourceforge.net/		(Haas, 2007)
LTRharvest	(GenomeTools) 1.5.9	http://genometools.org/pub/	-tabout no	(Ellinghaus et al., 2008)
LTRdigest	(GenomeTools) 1.5.9	http://genometools.org/pub/		(Steinbiss et al., 2009)
USEARCH	7.0.1090	http://drive5.com/usearch/	--id 0.8	(Edgar, 2010)
samtools	1.9	https://github.com/samtools/samtools		(Li and Durbin, 2009)
Snippy	4.6.0	https://github.com/tseemann/snippy		(Seemann, 2014)

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Freebayes	v1.3.2	https://github.com/ekg/freebayes		(Garrison and Marth, 2012)
IRF	3.05	http://tandem.bu.edu/irf/irf.download.html	Positional arguments: 3 4 12 2 1 50 30000 300000	(Warburton et al., 2004)
TRF	4.09	http://tandem.bu.edu/trf/trf.download.html	Positional arguments: 2 7 7 80 10 50 500	(Benson, 1999)
Mummer	3.1	http://mummer.sourceforge.net/		(Kurtz et al., 2004)
Dot	commit 42701ca	https://github.com/dnanexus/dot	DotPrep.py --unique-length 500 --overview 99999	(Nattestad, 2018)
Clustal W	2.1	http://www.clustal.org/		(Thompson et al., 1994)
JalView	2.10.1	https://www.jalview.org/		(Waterhouse et al., 2009)
HelitronScanner	1	https://sourceforge.net/projects/helitronscanner/		(Xiong et al., 2014)
WormMine	WS277	http://intermine.wormbase.org/tools/wormmine/begin.do		
KEGGREST	1.26.1	https://bioconductor.org/packages/r		(Tenenbaum, 2019)

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

		elease/bioc/html/KEGGREST.html		
Rtsne	0.15	https://github.com/jkrijthe/Rtsne	max_iter = 1000, dims = 2	(Krijthe, 2015)

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S3: **Raw data for genomic analysis of *Oscheius tipulae***

Accession	***Type***	***Platform***	***Read metrics***	***Use***
SRR12179520	Genomic DNA	Oxford Nanopore PromethION	108.4 Gb in 8.83 M reads with an N50 of 19.1 kb and the longest read measuring 166.0 kb	Primary genome assembly
ERX1726786	Genomic DNA	Illumina	300 bp, paired end from 400 bp insert library*	Polishing
ERX1726787	Genomic DNA	Illumina	100 bp, paired end from 3 kb mate pair library*	Polishing
ERX1941409	RNA	454 GS FLX Titanium	288.4 Mb in 610.0 K reads	Gene finding

* raw data from Besnard, Fabrice, et al. (2017).

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S4. **Sources of nematode assemblies included in the Kraken2 database for contamination assessment.**

Species	***Link***
Oscheius tipulae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/900/184/235/GCA_900184235.1_Oscheius_tipulae_assembly_v2/GCA_900184235.1_Oscheius_tipulae_assembly_v2_genomic.fna.gz
Caenorhabditis elegans	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/985/GCF_000002985.6_WBcel235/GCF_000002985.6_WBcel235_genomic.fna.gz
Caenorhabditis briggsae	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/985/GCF_000002985.6_WBcel235/GCF_000002985.6_WBcel235_genomic.fna.gz
Caenorhabditis nigoni	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/002/742/825/GCA_002742825.1_nigoni.pc_2016.07.14/GCA_002742825.1_nigoni.pc_2016.07.14_genomic.fna.gz
Caenorhabditis inopinata	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/003/052/745/GCA_003052745.1_Sp34_v7/GCA_003052745.1_Sp34_v7_genomic.fna.gz
Haemonchus contortus	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/007/637/855/GCA_007637855.2_ASM763785v2/GCA_007637855.2_ASM763785v2_genomic.fna.gz
Pristionchus pacificus	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/180/635/GCA_000180635.3_El_Paco/GCA_000180635.3_El_Paco_genomic.fna.gz
Meloidogyne hapla	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/172/435/GCA_000172435.1_Freeze_1/GCA_000172435.1_Freeze_1_genomic.fna.gz
Strongyloides ratti	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/001/040/885/GCF_001040885.1_S_ratti_ED321/GCF_001040885.1_S_ratti_ED321_genomic.fna.gz
Onchocerca volvulus	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/499/405/GCA_000499405.2_ASM49940v2/GCA_000499405.2_ASM49940v2_genomic.fna.gz

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Brugia malayi	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/002/995/GCF_000002995.3_ASM299v2/GCF_000002995.3_ASM299v2_genomic.fna.gz

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S5. **Raw sequence allocation by Kraken grouped by Phylum**

Superkingdom	***Phylum***	***Total number of bases***	***Number of sequences***
Bacteria	Proteobacteria	87,185,310,828	5,898,270
Eukaryota	Nematoda	20,728,191,131	2,835,558
Bacteria	Bacteroidetes	205,721,843	15,301
NA|Bacteria|Eukaryota	NA	109,051,955	54,269
Bacteria	Actinobacteria	89,155,134	6,327
Eukaryota	Chordata	48,258,726	10,874
Bacteria	Firmicutes	21,325,116	4,945
Eukaryota	Ascomycota	4,380,260	1,217
Bacteria	Tenericutes	3,984,814	919
Bacteria	Cyanobacteria	2,993,440	622
Bacteria	Planctomycetes	1,103,149	186
Bacteria	Fusobacteria	1,050,278	257
Bacteria	Spirochaetes	889,419	210
Bacteria	Thermotogae	207,234	49
Eukaryota	Basidiomycota	195,581	38
Bacteria	Verrucomicrobia	146,155	25
Bacteria	Chlamydiae	112,778	21
Bacteria	Acidobacteria	110,156	19
Bacteria	Chloroflexi	105,978	28
Eukaryota	Microsporidia	99,924	13
Bacteria	Chlorobi	88,431	13

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Bacteria	Aquificae	86,567	23
Bacteria	Deinococcus-Thermus	79,879	29
Bacteria	Deferribacteres	72,027	21
Bacteria	Nitrospirae	49,624	13
Bacteria	Ignavibacteriae	49,299	8
Bacteria	Synergistetes	46,203	5
Bacteria	Thermodesulfobacteria	30,584	8
Bacteria	Candidatus Gracilibacteria	23,976	10
Bacteria	Calditrichaeota	22,029	5
Bacteria	Armatimonadetes	20,877	2
Bacteria	Fibrobacteres	17,770	6
Bacteria	Gemmatimonadetes	14,963	4
Bacteria	Elusimicrobia	11,891	5
Bacteria	Kiritimatiellaeota	10,426	2
Bacteria	Dictyoglomi	8,854	4
Bacteria	Chrysiogenetes	8,744	3
Bacteria	Candidatus Cloacimonetes	8,106	3
Bacteria	Caldiserica	3,046	1
Bacteria	Balneolaeota	2,232	1

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S6. **Contiguity and BUSCO metrics of *Oscheius tipulae* assemblies**

		Genome contiguity metrics*					***BUSCO*** scores***			
Read set	***Assembler***	***# contigs***	***Largest contig***	***Total length***	***N75***	***L75***	***C****	***D****	***F****	***M****
raw	flyemeta[§]	7	11.30	60.11	9.02	5	88.4	2	2.1	9.5
	redbean	945	0.59	47.46	0.04	323	62.3	1.9	1.9	35.8
	shasta	55	5.66	59.83	1.74	12	89	2.3	2.1	8.8
de-contaminated	flyemeta	11	11.30	60.04	9.01	5	88.7	2	2.2	9.1
	flye	7	11.24	60.00	9.01	5	88.9	2	2.1	9
	redbean	19	11.17	59.63	8.82	5	88.7	2	2.2	9.1
	shasta	34	9.99	59.47	2.26	9	89.2	2.5	2	8.8
	canu	13	11.28	59.94	8.83	5	88.6	2.3	2.1	9.3
40x error corrected	flyemeta[†]	13	11.18	59.87	5.13	6	89.1	2.1	2.1	8.8
	flye	7	11.29	60.00	9.03	5	88.4	1.9	2.2	9.4
	redbean	29	11.17	59.09	2.58	9	88.5	2.1	2.2	9.3
	shasta	964	0.26	52.46	0.05	509	75.2	11.7	2	22.7

* Largest contig, Total length and N75 in megabases.

** BUSCO v4.0.2 with AUGUSTUS v3.3.3 scores of Nematoda_odb10 (3131 loci); C = complete, S = single copy; D = duplicated; F = fragmented; M = missing.

[§] Assembly used to generate nOti 3.1.

[†] Assembly used to generate nOti 3.2 through a consensus with nOti 3.1.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S7. **Base fold coverage of the *Oscheius tipulae* assembly**

Chromosome	***Mean base fold coverage***	***Standard deviation***
Ot_ChrI	332.33	35.21
Ot_ChrII	329.33	31.26
Ot_ChrIII	329.55	33.85
Ot_ChrIV	333.16	29.09
Ot_ChrV	353.59	478.29
Ot_ChrX	327.22	41.84
Ot_MT	5720.28	1,078.38

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S8: **Sources and metrics of rhabditid nematode genome data**

Species	***Clade***	***Reference***	***Genome size (Mb)***	***Karyotype***
Oscheius tipulae	V	This work	60.9	6
Auanema rhodensis	V	(Tandonnet et al., 2019)	60.6	7
Caenorhabditis elegans	V	(C. elegans Sequencing Consortium, 1998)	100.3	6
Caenorhabditis briggsae	V	(Stein et al., 2003)	108.4	6
Caenorhabditis nigoni	V	(Yin et al., 2018)	129.5	6
Caenorhabditis remanei	V	(Teterina et al., 2020)	124.8	6
Caenorhabditis inopinata	V	(Kanzaki et al., 2018)	123.0	6
Haemonchus contortus	V	(Laing et al., 2016)	283.4	6
Pristionchus pacificus	V	(Rödelsperger et al., 2017)	158.5	6
Steinernema carpocapsae	IV	(Serra et al., 2019)	84.5	5
Meloidogyne hapla	IV	(Opperman et al., 2008)	53.0	17
Strongyloides ratti	IV	(Nemetschke et al., 2010)	43.2	3
Ascaris suum	III	(Wang et al., 2020)	278.6	24
Onchocerca volvulus	III	(Cotton et al., 2016)	96.4	5

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Brugia malayi	III	(Foster et al., 2020)	88.2	6

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S9. **BUSCO scores of assemblies and proteomes of rhabditid nematodes**

	BUSCO assembly*§					***BUSCO* proteome***					***# of protein coding genes***
Species	***C***	***S***	***D***	***F***	***M***	***C***	***S***	***D***	***F***	***M***	
Oscheius tipulae (v3.0)	90.7	*88.7*	*2*	1.8	7.5	90.7	*87.9*	*2.8*	2.4	6.9	16138
Oscheius tipulae (nOt2.0)	91.6	*89*	*2.6*	1.6	6.8	89.1	*86.2*	*2.9*	2.0	8.9	14626
Ascaris suum	94.7	*91.7*	*3*	1.2	4.1	79.9	*77.4*	*2.5*	3.4	16.7	17969
Auanema rhodensis	89.9	*89.3*	*0.5*	1.4	8.7	85.4	*84.6*	*0.8*	1.2	13.4	10861
Caenorhabditis elegans	99.5	*98.9*	*0.5*	0.1	0.5	100.0	*99.6*	*0.4*	0.0	0.0	20184
Caenorhabditis briggsae	99	*98.4*	*0.6*	0.4	0.6	99.2	*98.5*	*0.7*	0.4	0.4	20829
Caenorhabditis nigoni	99.4	*98.6*	*0.8*	0.1	0.5	96.8	*95.7*	*1.1*	1.4	1.8	29167
Caenorhabditis remanei	99.3	*98.1*	*1.1*	0.3	0.5	93.1	*91.9*	*1.2*	0.9	6.0	26308
Caenorhabditis inopinata	98.8	*97.9*	*0.9*	0.4	0.8	98.7	*97.7*	*1.0*	0.2	1.1	21443
Haemonchus contortus	94.4	*93.3*	*1.1*	0.9	4.7	93.1	*91.6*	*1.5*	1.1	5.8	19473
Pristionchus pacificus	82.6	*80.9*	*1.7*	1.5	15.9	77.4	*75.2*	*2.2*	1.0	21.6	25992
Steinernema carpocapsae	78.2	*74.8*	*3.5*	3	18.7	74.7	*70.9*	*3.8*	3.8	21.5	30825
Meloidogyne hapla	59.4	*58.3*	*1.1*	2.3	38.3	52.6	*51.1*	*1.4*	2.6	44.8	14419
Strongyloides ratti	68	*66.4*	*1.6*	1.5	30.5	70.6	*67.8*	*2.8*	0.9	28.5	12464
Onchocerca volvulus	98	*97.4*	*0.6*	0.4	1.6	98.3	*97.2*	*1.1*	0.5	1.2	12110
Brugia malayi	99.1	*98.4*	*0.7*	0.2	0.8	98.9	*98.3*	*0.6*	0.2	0.9	10940

* BUSCO v4.0.4 with AUGUSTUS v3.2.3 scores of Nematoda_odb10 (3131 loci); C = complete, S = single copy; D = duplicated; F = fragmented; M = missing. For genome mode

§ Using flag for Augustus self optimization

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S10: **Nigon element-defining loci in fifteen nematode genomes**

Nigon	**A**	**B**	**C**	**D**	**E**	**N**	**X**	**total**	**Proportion**
Ascaris suum	477	304	462	254	224	162	114	1997	0.91
Auanema rhodensis	521	315	495	272	227	169	100	2099	0.96
Brugia malayi	530	324	516	271	239	171	119	2170	0.99
Caenorhabditis briggsae	518	317	504	269	235	173	119	2135	0.97
Caenorhabditis elegans	529	326	517	272	240	172	119	2175	0.99
Caenorhabditis nigoni	505	315	491	247	227	166	109	2060	0.94
Caenorhabditis remanei	491	306	482	242	220	166	115	2022	0.92
Caenorhabditis inopinata	526	318	507	271	239	172	119	2152	0.98
Haemonchus contortus	520	317	509	271	236	160	108	2121	0.97
Meloidogyne hapla	429	257	419	210	170	122	81	1688	0.77
Onchocerca volvulus	526	322	510	273	234	170	117	2152	0.98
Oscheius tipulae	516	320	501	269	216	173	118	2113	0.96
Pristionchus pacificus	453	297	466	248	223	163	113	1963	0.9
Steinernema carpocapsae	497	291	474	258	231	156	112	2019	0.92
Strongyloides ratti	509	309	484	252	234	170	114	2072	0.95
Total loci per Nigon	**534**	**328**	**519**	**275**	**242**	**174**	**119**	**2191**	1.00

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S11. **KEGG categories enriched among Nigon defining loci**

Nigon	***KEGG term***	***adjusted_p_value***
A	Ribosome biogenesis in eukaryotes	1.91×10^{-04}
A	RNA transport	1.37×10^{-02}
C	Spliceosome	1.87×10^{-02}
C	RNA transport	2.42×10^{-02}
D	Calcium signaling pathway	1.14×10^{-02}
D	ErbB signaling pathway	3.66×10^{-02}
N	Hippo signaling pathway - multiple species	2.38×10^{-03}
X	Axon regeneration	1.15×10^{-05}
X	Calcium signaling pathway	1.46×10^{-02}
X	Neuroactive ligand-receptor interaction	1.56×10^{-02}

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table S12: **Nigon-defining loci in the *Meloidogyne hapla* genome assembly**

	Nigon element loci mapped								Assignment	
scaffold*	**A**	**B**	**C**	**D**	**E**	**N**	**X**	**Total**	**>20%**	**>10%**
mh_10.12.18	17	23	12	3	4	3	1	63		
	0.27	0.37	0.19	0.05	0.06	0.05	0.02			
	0.06	0.12	0.04	0.02	0.03	0.03	0.02			b
mh_11	5	3	18	2	3	22	3	56		
	0.09	0.05	0.32	0.04	0.05	0.39	0.05			
	0.02	0.02	0.05	0.01	0.02	0.24	0.06		**N**	c
mh_13A.17	62	13	34	24	14	3	7	157		
	0.39	0.08	0.22	0.15	0.09	0.02	0.04			
	0.20	0.07	0.10	0.15	0.12	0.03	0.14		**A**	cdx
mh_13B	11	1	1	16				29		
	0.38	0.03	0.03	0.55	0.00	0.00	0.00			
	0.04	0.01	0.00	0.10	0.00	0.00	0.00			d
mh_14	1	1	2	1		3		8		
mh_15	4		3	1	1	6	1	16		
mh_16	6	5	14	5	60	2	1	93		
	0.06	0.05	0.15	0.05	0.65	0.02	0.01			
	0.02	0.03	0.04	0.03	0.50	0.02	0.02		**E**	
mh_1A	53	24	11	2	4	3		97		
	0.55	0.25	0.11	0.02	0.04	0.03	0.00			
	0.17	0.12	0.03	0.01	0.03	0.03	0.00			ab

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mh_1B	61	15	9	6	1	11	1	104		
	0.59	0.14	0.09	0.06	0.01	0.11	0.01			
	0.20	0.08	0.03	0.04	0.01	0.12	0.02		**A**	n
mh_20	4	2	2	1		1		10		
mh_21	8	2	9	8	2	6	1	36		
mh_2A	12	11	8	25	4	4		64		
	0.19	0.17	0.13	0.39	0.06	0.06	0.00			
	0.04	0.06	0.02	0.16	0.03	0.04	0.00			d
mh_2B	12	25	16	15	6	6		80		
	0.15	0.31	0.20	0.19	0.08	0.08	0.00			
	0.04	0.13	0.05	0.09	0.05	0.07	0.00			b
mh_3	6	14	31	7	11	4	1	74		
	0.08	0.19	0.42	0.09	0.15	0.05	0.01			
	0.02	0.07	0.09	0.04	0.09	0.04	0.02			
mh_4	18	3	96	15	2	4		138		
	0.13	0.02	0.70	0.11	0.01	0.03	0.00			
	0.06	0.02	0.28	0.09	0.02	0.04	0.00		**C**	
mh_5	4	40	11	21	2	6	1	85		
	0.05	0.47	0.13	0.25	0.02	0.07	0.01			
	0.01	0.21	0.03	0.13	0.02	0.07	0.02		**B**	d
mh_6	3	2	37		1	2	13	58		
	0.05	0.03	0.64	0.00	0.02	0.03	0.22			
	0.01	0.01	0.11	0.00	0.01	0.02	0.26		**X**	c
mh_7	17	6	8	6	6	6	1	50		
	0.34	0.12	0.16	0.12	0.12	0.12	0.02			
	0.06	0.03	0.02	0.04	0.05	0.07	0.02			

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mh_8	3	3	20	1			19	46		
	0.07	0.07	0.43	0.02	0.00	0.00	0.41			
	0.01	0.02	0.06	0.01	0.00	0.00	0.38		**X**	
unplaced contigs	122	64	77	51	49	30	31	424		
	0.29	0.15	0.18	0.12	0.12	0.07	0.07			
Total	**429**	**257**	**419**	**210**	**170**	**122**	**81**	**1688**		

* For each chromosomally-assigned scaffold, three sets of values are given: count of loci mapped, proportion of loci mapped to each scaffold (i.e. each row sums to 1.00) and proportion of loci mapped normalised by the size of the Nigon element defining set. Assignments to Nigon elements are given in the last two columns where proportional presence was more than 20% or 10% of the assignments. Proportions and Nigon element assignments were not made for the four scaffolds with the fewest loci mapped.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Supplementary references

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F.J., Lippman, Z.B., Schatz, M.C., 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol. 20, 224. doi:10.1186/s13059-019-1829-6

Benson, G., 1999. Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573–580. doi:10.1093/nar/27.2.573

Besnard, F., Koutsovoulos, G., Dieudonné, S., Blaxter, M., Félix, M.-A., 2017. Toward Universal Forward Genetics: Using a Draft Genome Sequence of the Nematode Oscheius tipulae To Identify Mutations Affecting Vulva Development. Genetics 206, 1747–1761. doi:10.1534/genetics.117.203521

Blaxter, M., Koutsovoulos, G., 2015. The evolution of parasitism in Nematoda. Parasitology 142 Suppl 1, S26-39. doi:10.1017/S0031182014000791

Bonfield, J.K., Whitwham, A., 2010. Gap5--editing the billion fragment sequence assembly. Bioinformatics 26, 1699–1703. doi:10.1093/bioinformatics/btq268

Buchfink, B., Xie, C., Huson, D.H., 2015. Fast and sensitive protein alignment using DIAMOND. Nat. Methods 12, 59–60. doi:10.1038/nmeth.3176

Bushnell, B., 2017. BBTools. https://sourceforge.net/projects/bbmap/.

C. elegans Sequencing Consortium, 1998. Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282, 2012–2018. doi:10.1126/science.282.5396.2012

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. BMC Bioinformatics 10, 421. doi:10.1186/1471-2105-10-421

Challis, R., Richards, E., Rajan, J., Cochrane, G., Blaxter, M., 2020. BlobToolKit - Interactive Quality Assessment of Genome Assemblies. G3 (Bethesda). doi:10.1534/g3.119.400908

Chevreux, B., Wetter, T., Suhai, S., 1999. Genome sequence assembly using trace signals and additional sequence information. Presented at the German conference on bioinformatics, Citeseer, pp. 45--56.

Cotton, J.A., Bennuru, S., Grote, A., Harsha, B., Tracey, A., Beech, R., Doyle, S.R., Dunn, M., Hotopp, J.C.D., Holroyd, N., Kikuchi, T., Lambert, O., Mhashilkar, A., Mutowo, P., Nursimulu, N., Ribeiro, J.M.C., Rogers, M.B., Stanley, E., Swapna, L.S., Tsai, I.J., Lustigman, S., 2016. The genome of Onchocerca volvulus, agent of river blindness. Nat. Microbiol. 2, 16216. doi:10.1038/nmicrobiol.2016.216

Dainat, J., 2020. NBISweden/AGAT: AGAT-v0.1.1. Zenodo. doi:10.5281/zenodo.3674778

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26, 2460–2461. doi:10.1093/bioinformatics/btq461

Ellinghaus, D., Kurtz, S., Willhoeft, U., 2008. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. BMC Bioinformatics 9, 18. doi:10.1186/1471-2105-9-18

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Emms, D.M., Kelly, S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 16, 157. doi:10.1186/s13059-015-0721-2

Emms, D.M., Kelly, S., 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 20, 238. doi:10.1186/s13059-019-1832-y

Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., Smit, A.F., 2020. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA 117, 9451–9457. doi:10.1073/pnas.1921046117

Foster, J.M., Grote, A., Mattick, J., Tracey, A., Tsai, Y.-C., Chung, M., Cotton, J.A., Clark, T.A., Geber, A., Holroyd, N., Korlach, J., Li, Y., Libro, S., Lustigman, S., Michalski, M.L., Paulini, M., Rogers, M.B., Teigen, L., Twaddle, A., Welch, L., Ghedin, E., 2020. Sex chromosome evolution in parasitic nematodes of humans. Nat. Commun. 11, 1964. doi:10.1038/s41467-020-15654-6

Garrison, E., Marth, G., 2012. Haplotype-based variant detection from short-read sequencing. arXiv.

Haas, B.J., Delcher, A.L., Mount, S.M., Wortman, J.R., Smith, R.K., Hannick, L.I., Maiti, R., Ronning, C.M., Rusch, D.B., Town, C.D., Salzberg, S.L., White, O., 2003. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 31, 5654–5666. doi:10.1093/nar/gkg770

Haas, B., 2007. TransposonPSI: an application of PSI-Blast to mine (retro-) transposon ORF homologies. http://transposonpsi.sourceforge.net/.

Kalvari, I., Argasinska, J., Quinones-Olvera, N., Nawrocki, E.P., Rivas, E., Eddy, S.R., Bateman, A., Finn, R.D., Petrov, A.I., 2018. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. Nucleic Acids Res. 46, D335–D342. doi:10.1093/nar/gkx1038

Kanzaki, N., Tsai, I.J., Tanaka, R., Hunt, V.L., Liu, D., Tsuyama, K., Maeda, Y., Namai, S., Kumagai, R., Tracey, A., Holroyd, N., Doyle, S.R., Woodruff, G.C., Murase, K., Kitazume, H., Chai, C., Akagi, A., Panda, O., Ke, H.-M., Schroeder, F.C., Kikuchi, T., 2018. Biology and genome of a newly discovered sibling species of Caenorhabditis elegans. Nat. Commun. 9, 3216. doi:10.1038/s41467-018-05712-5

Kolmogorov, M., Yuan, J., Lin, Y., Pevzner, P.A., 2019. Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. 37, 540–546. doi:10.1038/s41587-019-0072-8

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., Phillippy, A.M., 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722–736. doi:10.1101/gr.215087.116

Krijthe, J.H., 2015. Rtsne: T-distributed stochastic neighbor embedding using Barnes-Hut implementation. R package version 0.13, URL https://github.com/jkrijthe/Rtsne.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., Salzberg, S.L., 2004. Versatile and open software for comparing large genomes. Genome Biol. 5, R12. doi:10.1186/gb-2004-5-2-r12

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Laetsch, Dominik R, Blaxter, M.L., 2017. KinFin: Software for Taxon-Aware Analysis of Clustered Protein Sequences. G3 (Bethesda) 7, 3349–3357. doi:10.1534/g3.117.300233

Laetsch, Dominik R., Blaxter, M.L., 2017. BlobTools: Interrogation of genome assemblies [version 1; peer review: 2 approved with reservations]. F1000Res. 6, 1287. doi:10.12688/f1000research.12232.1

Lagesen, K., Hallin, P., Rødland, E.A., Staerfeldt, H.-H., Rognes, T., Ussery, D.W., 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100–3108. doi:10.1093/nar/gkm160

Laing, R., Martinelli, A., Tracey, A., Holroyd, N., Gilleard, J.S., Cotton, J.A., 2016. Haemonchus contortus: Genome Structure, Organization and Comparative Genomics. Adv. Parasitol. 93, 569–598. doi:10.1016/bs.apar.2016.02.016

Lefoulon, E., Bain, O., Bourret, J., Junker, K., Guerrero, R., Cañizales, I., Kuzmin, Y., Satoto, T.B.T., Cardenas-Callirgos, J.M., de Souza Lima, S., Raccurt, C., Mutafchiev, Y., Gavotte, L., Martin, C., 2015. Shaking the Tree: Multi-locus Sequence Typing Usurps Current Onchocercid (Filarial Nematode) Phylogeny. PLoS Negl. Trop. Dis. 9, e0004233. doi:10.1371/journal.pntd.0004233

Li, H., 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100. doi:10.1093/bioinformatics/bty191

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. doi:10.1093/bioinformatics/btp324

Lomsadze, A., Ter-Hovhannisyan, V., Chernoff, Y.O., Borodovsky, M., 2005. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 33, 6494–6506. doi:10.1093/nar/gki937

Nattestad, M., 2018. Dot. https://github.com/dnanexus/dot.

Nemetschke, L., Eberhardt, A.G., Viney, M.E., Streit, A., 2010. A genetic map of the animal-parasitic nematode Strongyloides ratti. Mol. Biochem. Parasitol. 169, 124–127. doi:10.1016/j.molbiopara.2009.10.008

Opperman, C.H., Bird, D.M., Williamson, V.M., Rokhsar, D.S., Burke, M., Cohn, J., Cromer, J., Diener, S., Gajan, J., Graham, S., Houfek, T.D., Liu, Q., Mitros, T., Schaff, J., Schaffer, R., Scholl, E., Sosinski, B.R., Thomas, V.P., Windham, E., 2008. Sequence and genetic map of Meloidogyne hapla: A compact nematode genome for plant parasitism. Proc Natl Acad Sci USA 105, 14802–14807. doi:10.1073/pnas.0805946105

Oxford Nanopore Technologies Ltd., 2017. Medaka. https://github.com/nanoporetech/medaka.

Post, R., 2005. The chromosomes of the Filariae. Filaria J. 4, 10. doi:10.1186/1475-2883-4-10

Rödelsperger, C., Meyer, J.M., Prabh, N., Lanz, C., Bemm, F., Sommer, R.J., 2017. Single-Molecule Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model Organism Pristionchus pacificus. Cell Rep. 21, 834–844. doi:10.1016/j.celrep.2017.09.077

Seemann, T., 2014. snippy: fast bacterial variant calling from NGS reads.

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

https://github.com/tseemann/snippy.

Serra, L., Macchietto, M., Macias-Muñoz, A., McGill, C.J., Rodriguez, I.M., Rodriguez, B., Murad, R., Mortazavi, A., 2019. Hybrid Assembly of the Genome of the Entomopathogenic Nematode Steinernema carpocapsae Identifies the X-Chromosome. G3 (Bethesda) 9, 2687–2697. doi:10.1534/g3.119.400180

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. doi:10.1093/bioinformatics/btv351

Smit, A.F., Hubley, R., Green, P., 2015. RepeatMasker. RepeatMasker website.

Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D.H.A., Fulton, L.A., Fulton, R.E., Griffiths-Jones, S., Harris, T.W., Hillier, L.W., Kamath, R., Kuwabara, P.E., Waterston, R.H., 2003. The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol. 1, E45. doi:10.1371/journal.pbio.0000045

Steinbiss, S., Willhoeft, U., Gremme, G., Kurtz, S., 2009. Fine-grained annotation and classification of de novo predicted LTR retrotransposons. Nucleic Acids Res. 37, 7002–7013. doi:10.1093/nar/gkp759

Susič, N., Koutsovoulos, G.D., Riccio, C., Danchin, E.G.J., Blaxter, M.L., Lunt, D.H., Strajnar, P., Širca, S., Urek, G., Stare, B.G., 2020. Genome sequence of the root-knot nematode Meloidogyne luci. J. Nematol. 52, 1–5. doi:10.21307/jofnem-2020-025

Szitenberg, A., Salazar-Jaramillo, L., Blok, V.C., Laetsch, D.R., Joseph, S., Williamson, V.M., Blaxter, M.L., Lunt, D.H., 2017. Comparative genomics of apomictic toot-knot nematodes: hybridization, ploidy, and dynamic genome change. Genome Biol. Evol. 9, 2844–2861. doi:10.1093/gbe/evx201

Tandonnet, S., Koutsovoulos, G.D., Adams, S., Cloarec, D., Parihar, M., Blaxter, M.L., Pires-daSilva, A., 2019. Chromosome-Wide Evolution and Sex Determination in the Three-Sexed Nematode Auanema rhodensis. G3 (Bethesda) 9, 1211–1230. doi:10.1534/g3.119.0011

Team, R.C., 2019. R: A Language and Environment for Statistical Computing.

Tenenbaum, D., 2019. KEGGREST: Client-side REST access to KEGG. https://www.bioconductor.org/packages/release/bioc/html/KEGGREST.html.

Teterina, A.A., Willis, J.H., Phillips, P.C., 2020. Chromosome-Level Assembly of the Caenorhabditis remanei Genome Reveals Conserved Patterns of Nematode Genome Organization. Genetics 214, 769–780. doi:10.1534/genetics.119.303018

Thomas, V.P., Fudali, S.L., Schaff, J.E., Liu, Q., Scholl, E.H., Opperman, C.H., Bird, D.M., Williamson, V.M., 2012. A sequence-anchored linkage map of the plant-parasitic nematode Meloidogyne hapla reveals exceptionally high genome-wide recombination. G3 (Bethesda) 2, 815–824. doi:10.1534/g3.112.002261

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting,

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680. doi:10.1093/nar/22.22.4673

Vaser, R., Sović, I., Nagarajan, N., Šikić, M., 2017. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27, 737–746. doi:10.1101/gr.214270.116

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., Earl, A.M., 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE 9, e112963. doi:10.1371/journal.pone.0112963

Walton, A.C., 1959. Some parasites and their chromosomes. J. Parasitol. 45, 1–20.

Wang, J., Veronezi, G.M.B., Kang, Y., Zagoskin, M., O'Toole, E.T., Davis, R.E., 2020. Comprehensive chromosome end remodeling during programmed DNA elimination. BioRxiv. doi:10.1101/2020.04.18.047035

Warburton, P.E., Giordano, J., Cheung, F., Gelfand, Y., Benson, G., 2004. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. Genome Res. 14, 1861–1869. doi:10.1101/gr.2542904

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. doi:10.1093/bioinformatics/btp033

Xiong, W., He, L., Lai, J., Dooner, H.K., Du, C., 2014. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. Proc Natl Acad Sci USA 111, 10263–10268. doi:10.1073/pnas.1410068111

Yin, D., Schwarz, E.M., Thomas, C.G., Felde, R.L., Korf, I.F., Cutter, A.D., Schartner, C.M., Ralston, E.J., Meyer, B.J., Haag, E.S., 2018. Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. Science 359, 55–61. doi:10.1126/science.aao0827

bioRxiv preprint doi: https://doi.org/10.1101/2020.09.04.283127; this version posted September 21, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Online data and code

A github repository containing code for the R analyses of genome features and the circos plot data and configuration files is at https://github.com/tolkit/otipu_chrom_assem .

A data sheet containing data for tables and figures is at A Chromosomal assembly of Oscheius_tipulae data .

A presentation file of the figures is at The Oscheius tipulae genome Figures for Paper .