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 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.