The *Aphelenchoides* genomes reveal major events of horizontal gene transfers in clade IV nematodes

Cheng-Kuo Lai[1,2], YiChien Lee[1,3,4], Huei-Mien Ke[1], Min R Lu[1], Wei-An Liu[1], Hsin-Han Lee[1], Yu-Ching Liu[1], Toyoshi Yoshiga[5], Taisei Kikuchi[6], Peichen J. Chen[7]* and Isheng Jason Tsai[1,2]*

Correspondence: janetchen@nchu.edu.tw and ijtsai@gate.sinica.edu.tw

[1]Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan.
[2]Genome and Systems Biology Degree Program, National Taiwan University and Academia Sinica, Taipei, Taiwan.
[3]Biodiversity Program, Taiwan International Graduate Program, Academia Sinica and National Taiwan Normal University, Taipei, Taiwan
[4]Department of Life Science, National Taiwan Normal University, 116 Wenshan, Taipei, Taiwan
[5]Faculty of Agriculture, Saga University, Saga 840-8502 Japan
[6]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, 277-8562, Japan
[7]Department of Plant Pathology, National Chung Hsing University, Taichung, Taiwan

Abstract

Aphelenchoides besseyi is a plant-parasitic nematode (PPN) in the Aphelenchoididae family capable of infecting more than 200 plants. *A. besseyi* is also a species complex with strains exhibiting varying pathogenicity to plants. We present the genome and annotations of six *Aphelenchoides* species, four of which belonged to the *A. besseyi* species complex. Most *Aphelenchoides* have a genome size of 44.7-47.4 Mb and are amongst the smallest in the clade IV, with the exception of *A. fujianensis*, which has a size of 143.8 Mb and is the largest. Phylogenomic analysis successfully delimited the species complex into *A. oryzae* and *A. pseudobesseyi* and revealed a reduction of transposon elements in the last common ancestor of *Aphelenchoides*. Synteny analyses between reference genomes indicated that three chromosomes in *A. besseyi* were derived from fission and fusion events. A systematic identification of horizontal gene transfer (HGT) genes across 27 representative nematodes allowed us to identify two major episodes of acquisition corresponding to the last common ancestor of clade IV or major PPNs, respectively. These genes were mostly lost and differentially retained between clades or strains. Most HGT events were acquired from bacteria, followed by fungi, and also from plants which was especially prevalent in *Bursaphelenchus mucronatus*. Our results establish a comprehensive understanding on new origins of horizontal gene transfer in nematodes.

Introduction

The ability to parasitise plants has evolved in the phylum Nematoda on at least four occasions[1,2]. The major plant parasites belonged to the Aphelenchodidae and Parasitaphelenchidae families making up the Aphelenchoidea superfamily and the Tylenchida order of clade IV nematodes[3]; these plant parasitic nematodes (PPNs) collectively cause worldwide agriculture damages of over US$80 billion each year[4]. Root-knot nematodes in *Meloidogyne* genus cause the majority of these losses and were the first of PPNs to have their genomes sequenced[5], followed by pinewood nematode *Bursaphelenchus xylophilus*[6,7], potato cyst nematode *Globodera pallida*[8], soybean cyst nematode *Heterodera glycines*[9] and others[10,11]. Comparing these genomes yield insight into several adaptions that allow PPNs to parasitize plants. Examples include effectors such as carbohydrate active enzymes (CAZyme), which are known to be secreted by PPNs and are hypothesized to be involved in degrading or modifying the composition of different plant structural tissues[12,13]. Some of these PPN-specific genes are known to be acquired from bacteria or fungi through horizontal gene transfer (HGT)[14], giving nematodes the ability to adapt to different environments[14]. Although numerous HGT genes have been identified and documented in different nematodes, research on the timing and subsequent maintenance of these genes, and why their copy numbers differ, has been restricted to a few PPN clades[15].

Currently, the only major groups containing plant parasitic nematodes that lack a reference genome are Trichodoridae and Aphelenchoididae. Of particular interest is the *Aphelenchoides besseyi***,** which is a foliar nematode that infects almost 200 plants in 35 genera[16]. This nematode is 10mm in body size, has a life cycle of this nematode is around 10 to 12 days and it can reproduce in extreme environments, making it hard to eliminate. Better known as the rice white tip, *A. besseyi* infects important agronomic crops such as rice, soybeans and strawberries[17,18], causing necrosis and distortion of its host's leaves[16,18,19]. The nematode has reportedly been responsible for up to a 60% crop loss in some cases[20,21] and was listed among the top ten plant parasitic nematodes in a recent review[22]. Despite the economic damage these parasitic

nematodes inflict, particularly in the Asian region, little is known about the basic biology, genetic diversity or evolution of *A. besseyi* and other Aphelenchoididae members. It has been reported that *A. besseyi* isolated from different hosts have different levels of pathogenicity. For instance, the populations of *A. besseyi* isolated from strawberries were unable to parasitise rice[21]. However, the populations of this species from bird's-nest fern can reproduce in both rice and strawberries[19]. Despite their almost overlapping morphological features, we previously identified copy number variations of genes encoding cell-wall-degrading enzymes including glycosyl hydrolase family 5 (GH5) and GH45 cellulases between *A. besseyi* of different host origins[23]. An 18S phylogeny separated the strains isolated from rice and fern unambiguously, suggesting that *A. besseyi* might be a species complex, with literatures also identifying variations in different molecular markers in different hosts that are original to *A. besseyi*[18,24]. Subbotin *et al*. recently used a combination of molecule makers (28S, ITS and mitochondria *COI* gene)[17] to reclassify foliar nematodes into three separated clades: the *A. besseyi* isolated mainly from strawberries, the *A. oryzae* mainly isolated from rice and *A. pseudobesseyi* from wood fern suggesting this species complex may be well differentiated at the genome level. From an evolutionary perspective, *A. besseyi* is also interesting because its primitive plant parasitism was a relatively recent evolutionary adaptation[25].

In this study, we sequenced and annotated the genomes of four *A. besseyi* species complex strains isolated from different plants which we later designated as *A. pseudobesseyi* and *A. oryzae,* and another two species in the Aphelenchoididae family (*Aphelenchoides bicaudatus* and *Aphelenchoides fujianensis*). We compared the proteomes of six Aphelenchoididae members with 21 other representative nematodes to delimit species relationships and investigated their gene family dynamics. We identified synteny with representative nematodes and inferred rearrangement events to determine how the three chromosomes of *A. besseyi* was evolved. The availability of the *Aphelenchoides* assemblies allowed us to systematically determine the horizontal gene transfer-acquired genes in nematode genomes. By inferring the evolutionary origins of these HGT genes we found historical HGT events that shaped nematode evolution.

125 The major event occurred in the last common ancestor of clade IV nematodes and may
126 have contributed to the early adaptation of these nematodes.
127
128
129 ## Results
130
131 **Genome assemblies and annotations of six *Aphelenchoides* species**
132 We sequenced and assembled the genomes of six nematodes in the
133 *Aphelenchoides* genus (four *A. besseyi*, one *A. bicaudatus* and one *A. fujianensis*).
134 These species were chosen to represent the Aphelenchoididae family and *A. besseyi*
135 strains isolated from three plant hosts (**supplementary table S1**) to delimit their
136 relationship within the species complex. For each species, an initial assembly was
137 produced from either 70-148X Oxford Nanopore or 113-422X Pacbio reads using Flye
138 assembler[26] and further polished using Illumina reads (**supplementary table S2**).
139 Among *A. besseyi* assemblies, the VT strain isolated from tape grass *Vallisneria spiralis*
140 had the highest genome quality with N50 5.4 Mb (hereafter denoted as APVT). The
141 contigs of this strain were further scaffolded with 150X Hi-C reads using the Juicer
142 program[27] (**supplementary fig. S1**) yielding a final assembly of 44.7 Mb (N50 = 16.9
143 Mb). More than 99% of this assembly was in three scaffolds, presumably corresponding
144 to three chromosomes[28] (2n=6). Five *Aphelenchoides* assemblies ranged from 44.7 to
145 47.4 Mb (N50 = 12.2-17.8 Mb; **supplementary table S3**), and a sixth assembly (*A.*
146 *fujianensis*) of 143.8 Mb (N50 = 553 kb; **supplementary table S3**) was estimated to be
147 triploid[29] (**and supplementary fig. S2**). Although not present in the assemblies, the
148 telomere motif TTAGGC was identified in the reads of *A. peseudobesseyi* at low
149 coverage (**Supplementary Info**), which is consistent to the sister group species of *B.*
150 *xylophilus* (**supplementary table S4**) indicating the presence of telomeres in these
151 species.
152 Using the proteomes of *Bursaphelenchus xylophilus* and *Caenorhabditis*
153 *elegans*, and the transcriptomes of pooled worms in each species as evidences, we
154 predicted 11,701 to 12,948 protein coding genes in six *Aphelenchoides* species with
155 Maker2 pipeline[30] (**supplementary table S3**). With the exception of *A. fujianensis*,
156 these were fewer protein coding genes in these species than in Tylenchida nematodes

157 (12,762 to 19,212) and free-living nematodes (20,184 to 20,992). The completeness of
158 annotated genes was estimated to be 76.4–81.3% based on a BUSCO assessment,
159 lower than that of *Bursaphelenchus* species (83.0–89.4%), but are higher than that of
160 Tylenchida (59.8–73.8%) nematodes. The lower BUSCO completeness in
161 *Aphelenchoides* was likely clade specificm as re-annotation of the APVT strain with
162 trained models based on manual curation of 975 genes also gave a similar score of
163 78.2%. Among them, 66.5% to 71.0% of genes in *Aphelenchoides* could be assigned at
164 least a domain from the Protein family (Pfam) database[31]. In addition, orthologous
165 groups were inferred with the proteomes of six *Aphelenchoides* with 21 other
166 nematodes using Orthofinder[32]. With the exception of *A. fujianensis*, 78.5–85.4% (*A.*
167 *fujianensis* = 48.7%), 69.4–76.9% (*A. fujianensis* = 42.8%) and 87.5–98.7% of
168 *Aphelenchoides* genes were orthologous to *B. xylophilus*, *C. elegan*s and at least one
169 other nematode species, respectively, suggesting that the reduced proteome in most
170 *Aphelenchoides* was mainly comprised of conserved genes among nematodes.
171
172 **Phylogenomics delimit species complex of *Aphelenchoides besseyi***
173 To investigate the evolution of plant-parasitic nematodes and the relationships
174 among members in the *A. besseyi* species complex, a maximum-likelihood phylogenetic
175 tree was constructed based on 74 low-copy orthologues. The phylogeny is consistent
176 with a the previous study[33]: the major plant parasitic nematodes were divided into
177 Aphelenchoidea and Tylenchida, and six *Aphelenchoides* species were grouped as
178 sister to *Bursaphelenchus* (**fig. 1a**). The *A. besseyi* strains were clustered into two
179 groups based on their hosts, suggesting that relationships in these species within the *A.*
180 *besseyi* species complex can be unambiguously resolved based on their different
181 lifestyles and host preferences. Combined with the previous 28S phylogeny of the *A.*
182 *besseyi* species complex[17] (**supplementary fig. S3**), we further designated these two
183 groups as *A. oryzae* and *A. pseudobesseyi* groups isolated from rice or other plants
184 (land grass and bird's-nest fern). The median nucleotide and amino acid identity was
185 86.6% and 90% between these two groups, respectively (**supplementary fig. S4**).
186 Strains in each group also differed in heterozygosity (0.017-0.019% in *A. oryzae* vs
187 0.071-0.075% in *A. pseudobesseyi*) and changes in recent effective population sizes

188 inferred using pairwise sequentially Markovian coalescent (PSMC) analysis[34]
189 (**supplementary fig. 5**). Together these results emphasised that relationships among
190 species in the *A. besseyi* species complex were highly diversified at the genome level
191 despite being challenging to differentiate based solely on morphology[17].
192
193 **Genome reduction as a result of transposable element loss**
194 The *Aphelenchoides* genomes were smaller than those of other plant-parasitic
195 nematodes (**fig. 1b** and **supplementary table S3**), indicating that genome reduction
196 took place in the last common ancestor of the *Aphelenchoides* genus. Much of the
197 reduction can be explained by the reduced markup of repeat content compared to other
198 nematodes (**fig. 1b**). The dominant transposable elements of *Aphelenchoides* were
199 DNA transposons—which were reduced in content (0.14–1.36 Mb vs. 4.2–22.1 Mb)—
200 and number of families (1–7 in *Aphelenchoides* compared to 9 and 26 in *B. xylophilus*
201 and *H. glycines*, respectively) compared to other nematodes. Fewer LTR (0.07–0.8 Mb
202 vs. 0.24–9.3 Mb) and LINE (0.0006–0.66 Mb vs 0.02–4.5 Mb) retrotransposons were
203 also observed in this genus. These results suggest that the reduced genome sizes in
204 *Aphelenchoides* might have been caused by the rapid loss of transposable elements
205 and led to the eventual loss of entire families in some cases (**fig. 2a**). Within the *A.*
206 *besseyi* species complex, *A. peseudobesseyi* contained significantly fewer DNA
207 transposons, LTR and LINE retrotransposons than *A. oryzae* (**fig. 2a and**
208 **supplementary fig. 6**).
209
210
211
212
213
214
215

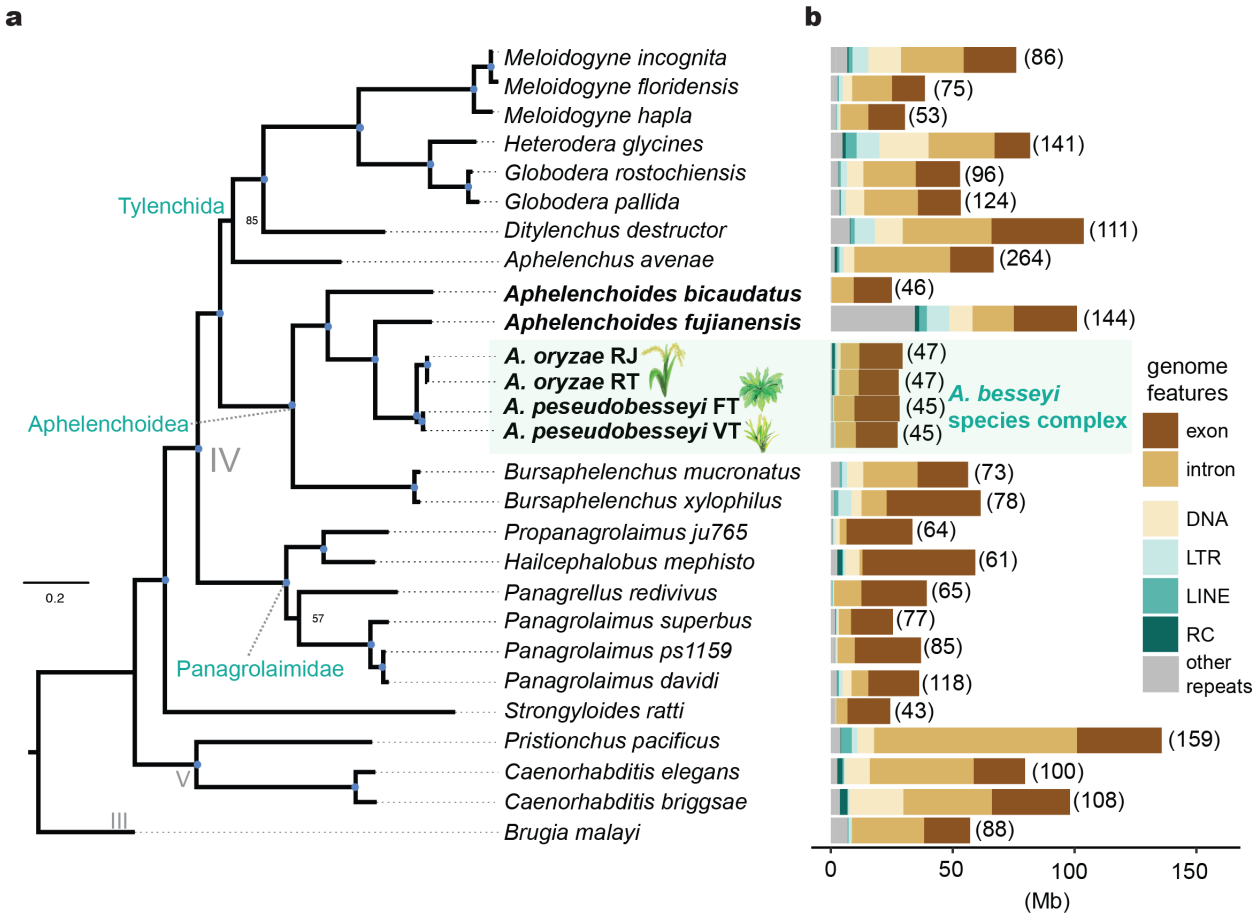

Figure 1 Phylogenomic analysis of plant-parasitic nematodes. a. The phylogeny of 27 representative nematodes was inferred based on the concatenation of protein sequences from 74 low-copy orthologues. The blue dots in branches denote a bootstrap value of 100. **b.** The size of the genome features in nematodes, repeats containing DNA transposons (DNA), long interspersed nuclear elements (LINE), long terminal repeats (LTR), rolling-circular (RC) and the unclassified repeats are labelled as “other repeats”. Numbers in brackets denote genome assembly sizes in megabases.

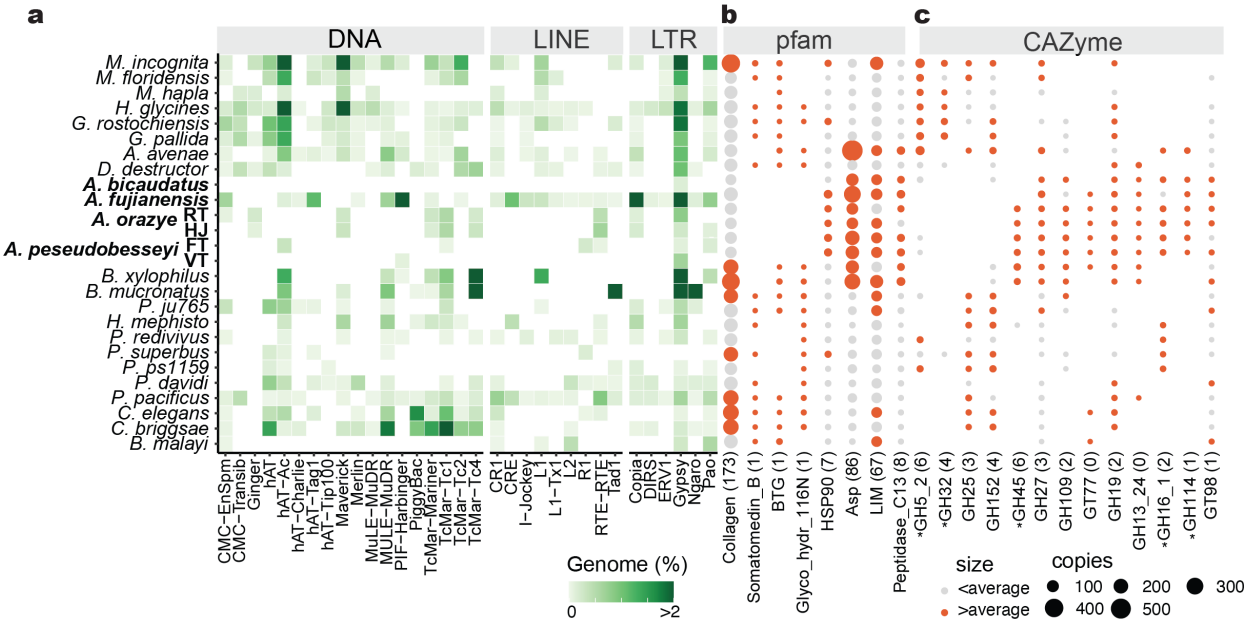

Figure 2 Repeat and proteome contents in nematodes. a. The genome proportions of DNA, LINE and LTR transposable elements in nematodes shown by genome percentage **b-c.** Protein families and CAZyme gene copy numbers vary significantly among nematodes. The dot size represents the copy number of each domain and the different colour represents the copy number of domains larger or lower than average copies shown in brackets.

Gene family specialization in the *Aphelenchoides* species

We observed 66 enriched and 31 reduced protein domains in the four member of the *A. besseyi* species complex compared to 21 other nematodes. (**fig. 2b** and **supplementary table S5**). Domain reduction included collagen (90–109 copies in the *A. besseyi* species complex vs. 72–407 in others), Somatomedin B and BTG. Genes containing collagen domains were reportedly associated with capsule formation; the reduced copy of collagen domains in *Trichinella spiralis* were thought to contribute its lower host-specificity than other nematodes[35], and may be related to the wide host range of *A. besseyi*. In contrast, Aphelenchoidea members possess on average four-fold (91–314 vs. 4–555 copies) more aspartic proteases (ASP) than other nematodes (**supplementary table S5**). ASPs were reported to associate with the digestion of host haemoglobin in *Haemonchus controtus*[36], and also skin penetration in hookworms[37],

248 and may play an important role in the *Aphelenchoides*'s parasitism process. Other
249 expansions included LIM and peptidase C13 domains, which participate in participating
250 in the regulation of cell motility and cell growth[38] or degradation of protein tissues in a
251 host[39], emphasizing that these domain dynamics are associated with adaptations to
252 plant parasitism.
253 The plant cell wall acts as a primary defensive barrier and the production of
254 carbohydrate-active enzyme (CAZyme) families are important for PPNs to infect plants.
255 A total of 132 CAZyme families were identified in the representative the 27 nematodes.
256 Of these, 59–67% of the CAZyme families were observed in Aphelenchoidea which is
257 similar to the 55–66% and 58-68% of the families in Tylenchida and free-living
258 nematodes (**supplementary table S6**), respectively. A total of 13 families were
259 significantly expanded or lost in the *Aphelenchoides* genus **(fig. 2c**), including GH16,
260 GH27 and GH45. GH16 serves as the putative β-glycanases involved in the
261 degradation or remodelling of cell wall polysaccharides[40], GH16 had one to six copies in
262 Aphelenchoididae nematodes and was not identified outside this clade except in *D.*
263 *destructor*, in which there were three copies. There are three to 11 copies of GH27—
264 which are reportedly involved in the function of hemicellulose and associated with α-
265 galactosidase activity in both bacteria and fungi[41]— in Aphelenchoidea, but fewer in the
266 Tylenchida nematodes. The previously identified GH45 present in Aphelenchoidea
267 nematodes[23]—involved in the degradation of beta-1,4-glucans in the plant cell wall[19]—
268 possess different copy numbers between *A. pseudobesseyi* and *A. oryzae* and were
269 absent in *A. fujianensis* and *A. bicaudatus,* suggesting differential maintenance of these
270 genes in the same genus may have contributed to variations of pathogenicity to plants.
271
272 **Chromosome evolution of PPNs**
273 To investigate the extent of the karyotype rearrangements in *Aphelenchoides*, we
274 inferred the synteny relationships among *A. peseudobesseyi* (chromosome n=3)[28], *B.*
275 *xylophilus* (n=6) and *C. elegans* (n=6) using single copy orthologs. Within the three *A.*
276 *peseudobesseyi* chromosomes, orthologs belonging to all *C. elegans* chromosomes
277 were clustered into distinctive blocks (**fig. 3a**) suggesting a fusion of ancestral
278 chromosomes. These regions remained contiguous and contained 148-801 orthologous

279 genes that could be assigned from individual chromosomes presumably not yet broken
280 down yet by recombination, allowing us to pinpoint the fusion points and infer the order
281 of rearrangement events based the constitution of chromosomes (**fig. 3b**). We
282 encountered instances of where an ancestral chromosome was found in different parts
283 of the *A. peseudobesseyi* chromosomes, suggesting fission also took place. In the case
284 of chr IV—which remained homologous in *C. elegans* and *B. xylophilus*—corresponding
285 synteny blocks in *A. peseudobesseyi* were identified in the arm of chr 2 and chr 1
286 separated by regions of chr III origin (**fig. 3b**). The majority of the ancestral sex
287 chromosomes were unambiguously assigned to chr 2, and remapping of male
288 sequences showed equal coverage along the chromosomes (**supplementary figure**
289 **S7**), suggesting that the Aphelenchoidea superfamily including *A. besseyi* exhibited a
290 stochastic sex determination system that was recently characterized in *B. xylophilus*[42].
291 Within the *A. besseyi* species complex, a total of 91% and 88% of genomes were in
292 synteny between APVT and AORT, respectively. Intra-chromosomal inversions were
293 common at chromosome arms. In addition, we identified a major inversion of length 3.4
294 Mb long located in the centre of chr 2 (**fig. 3c**) suggesting rearrangement is still
295 ongoing. Both the LTR and LINE retrotransposons were enriched in the chromosome
296 arms of the *A. oryzae* strain (AORT) (**fig. 3c and supplementary fig. S6**), which is
297 consistent with the hallmark of nematode chromosome evolution[43]. In contrast, only the
298 LTR retrotransposons were found in the two chromosome arms of *A. peseudobesseyi*,
299 suggesting that these repeats were differentially maintained after speciation.
300

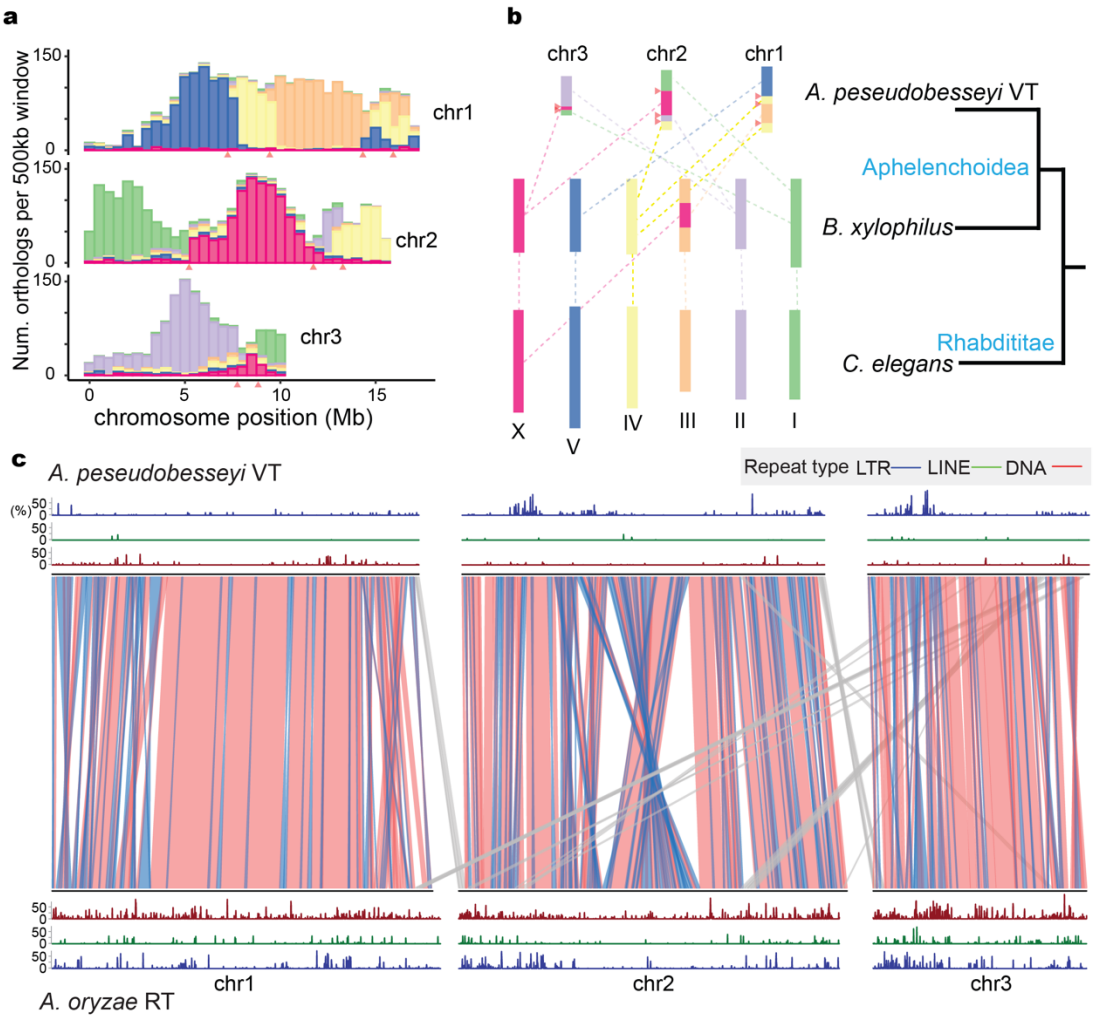

Figure 3. Chromosome evolution of plant-parasitic nematodes. a. The density of pairwise single-copy orthologs between *A. pseudobesseyi* VT and *B. xylophilus*. Colours denote *B. xylophilus* chromosomes and the putative chromosome fusion sites in APVT are labelled with red triangles. **b.** Colours denote corresponding *C. elegans* chromosomes, and dashed lines indicate linkage groups corresponding to ancestral chromosomes. **c.** The synteny relationship and the distribution of transposable between *A. pseudobesseyi* VT and *A. oryzae* RT**.** Blocks indicate synteny links between the two strains, and the line colours correspond to inversion (blue) and inter-chromosomal rearrangement (grey). Distribution of DNA transposons (red), long interspersed elements (green) and long terminal repeats (blue) between two stains are shown.

316 **Major episode of HGT in clade IV nematodes**

317 In plant parasitic nematodes, the GH5 cellulase was found present in Tylenchida
318 and only *A. peseudobesseyi* and *A. bicaudatus* within the Aphelenchoidea clade[23,44],
319 raising the possibility that many of the horizontal gene transferred genes were acquired
320 in the last common ancestor of major PPNs but were differentially lost. To identify such
321 events, a total of 27 proteomes from representative nematodes including the
322 *Aphelenchoides* genomes were searched for evidence of HGT by calculating the Alien
323 Index (AI) score using Alienness[45]. We identified a total of 1,675 HGT orthogroups in 21
324 nematodes. Placing these orthologs designated as events onto the species phylogeny
325 assuming a parsimonious scenario[46], indicated that HGT started in the last common
326 ancestor of clade IV nematodes (**fig 4a**). Examples include GH16, GH32, GH43 and the
327 aforementioned GH5 cellulases. We inferred a total of 161 orthogroups were acquired in
328 this episode, and most of their origins were inferred to be bacteria (78.3%)
329 (**supplementary table S7**) belonging to different genera, suggesting multiple
330 acquisitions took place. Of these, we found 36 Pfam terms such as ABC transporter that
331 were identified in multiple orthogroups suggesting some convergence in the acquired
332 functions (**supplementary table S8**).

333 The revised GH5 cellulase phylogeny indicated an ancient duplication took place
334 before the divergence of PPNs (**fig 5a**). One clade contains orthologs of the three
335 *Panagrolaimid* (*P.* sp. PS1159, *P. superbus* and *P. davidi*), and Tylenchida, and the
336 other clade contains members of Aphelenchoidea and Tylenchida nematodes, which
337 emphasises that the fate of the HGT genes was governed by duplications and loss.
338 Interestingly, the closest GH5 bacterial orthologs were *Salinimicrobium xinjiangense*
339 and *Leeuwenhoekiella sp*., which belonged to Flavobacteriaceae family and were from
340 marine environments. We observed two GH16 subfamilies in nematodes. GH16_3 in
341 Tylenchida and *Bursaphelenchida* nematodes were clustered with bacterial origin
342 sequences, whereas GH16_1 of *Aphelenchoides* and *Panagrolaimus* nematodes were
343 clustered with fungal origin (**fig 5b**), suggesting that the two GH16 groups arose
344 independently. GH32 in *G. pallida*[13] is believed to play a role in the function of fructose
345 hydrolysis and was found in one *Panagrolaimus* in addition to several Tylenchida
346 nematodes (**supplementary fig. S8**). GH43 was identified at two distinct clusters of

347 bacterial origin in Tylenchida and *Panagrolaimid* nematodes which have been proposed
348 to be involved in degradation of the hemicellulose in plants[47] (**supplementary fig. S9**).
349 The next major episode of acquisition took place in the common ancestor of
350 PPNs, with 47 orthogroups (**fig. 4a)**. These families included pectate lyases 3 (PL3)
351 which is associated with cell wall degradation[48]. The orthologs of PL3 in *Aphelenchus*
352 *avenae* and two *Bursaphelenchus* nematodes were grouped together with distinct
353 clusters of *Meloidogyne* species (**fig. 5c**) is consistent with previous phylogeny finding
354 in PPNs[44]. The closest bacterial ortholog in the *Meloidogyne* clade was from
355 *Curtobacterium flaccmfaciens* which is also known to cause bacterial wilt in the
356 Fabaceae family[49]. Together, these results suggested some genes that were thought to
357 play important roles in plant parasitism were in fact acquired earlier than the common
358 ancestor of plant parasitic nematodes.
359 The majority of HGT gene families were of bacterial followed by fungal origin (**fig**
360 **4b**). We also identified genes that were acquired from non-bacterial donors in the last
361 common ancestors of clade IV, as well as in more recent, different PPN lineages (**fig**
362 **4a**). This included the previously characterised fungal origin of GH45[23,50], This cellulase
363 family is present in most Aphelenchoidea nematodes except *A. fujinensis* and *A.*
364 *bicaudatus*. The GH16 family was independently acquired from a bacterial and fungal
365 donor in the last common ancestor of clade IV nematodes and the *Aphelenchoides*
366 genus, respectively (**fig 5b**). Notably, we identified 40 orthogroups among PPNs that
367 were transferred from the plant phylum Streptophyta, which is consistent with the finding
368 of several sequences that are highly similar to plants in *H. glycine*[51] (**fig. 4b**). The
369 closest plant orthologs included rice, maple and oak (**fig. 5d**) which are common hosts
370 to many PPNs. Strikingly, of these orthogroups, 27 were present in *B. mucronatus* and
371 enriched in the detoxification of cadmium and copper ion function (**supplementary**
372 **table S9**), suggesting these genes may help *Bursaphelenchus* nematodes to degrade
373 the toxin in pine wood hosts.
374

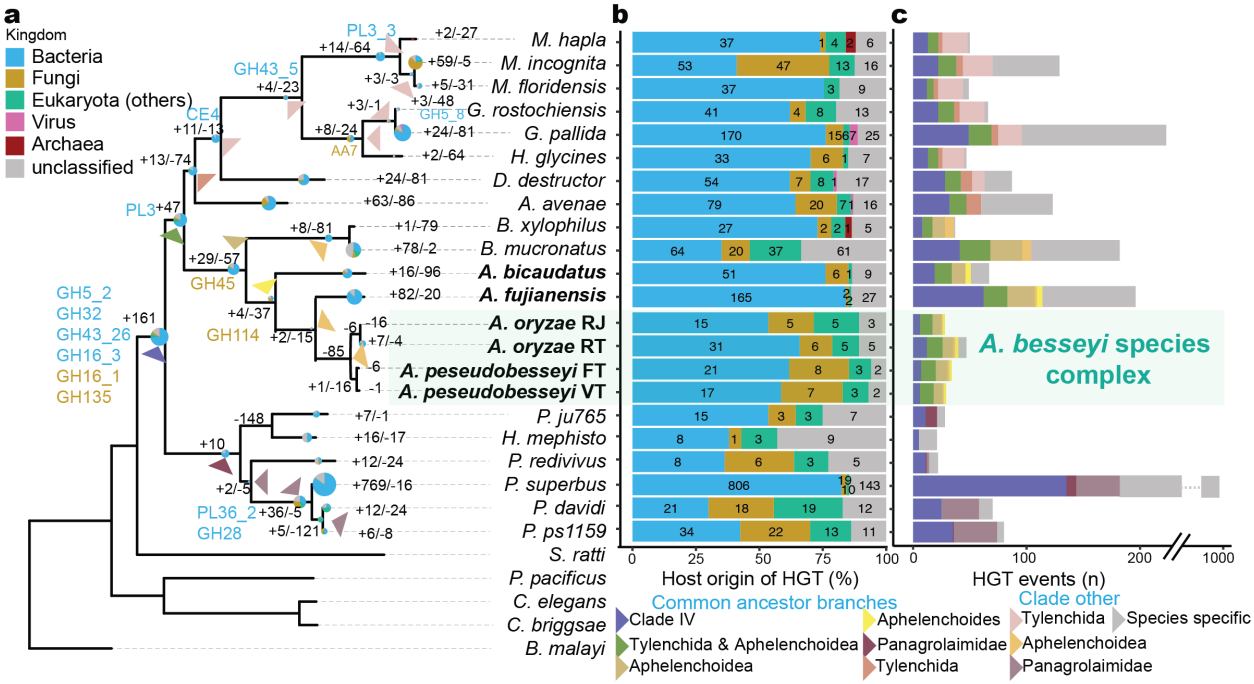

Figure 4 The evolution of genes acquired from horizontal gene transfer (HGT) in clade IV nematodes. a. HGT orthogroups were inferred by the AI score[52] > 0 across 27 representative nematodes; the HGT families gained or lost are shown in the branches. Horizontally acquired CAZymes are annotated. The proportions of donor origins in each HGT orthogroup belonging to different kingdoms of donors are shown as pie charts. The size of the pie chart corresponds to the total number of HGT orthologues in branches; the chart was normalized using: $\log_5$(total number). **b.** The distribution of HGT families transferred from different kingdoms, with the same denoted color scheme as same as figure a. **c.** Number of HGT events among different inferred time points corresponding with the branches which are triangle marked in the phylogeny.

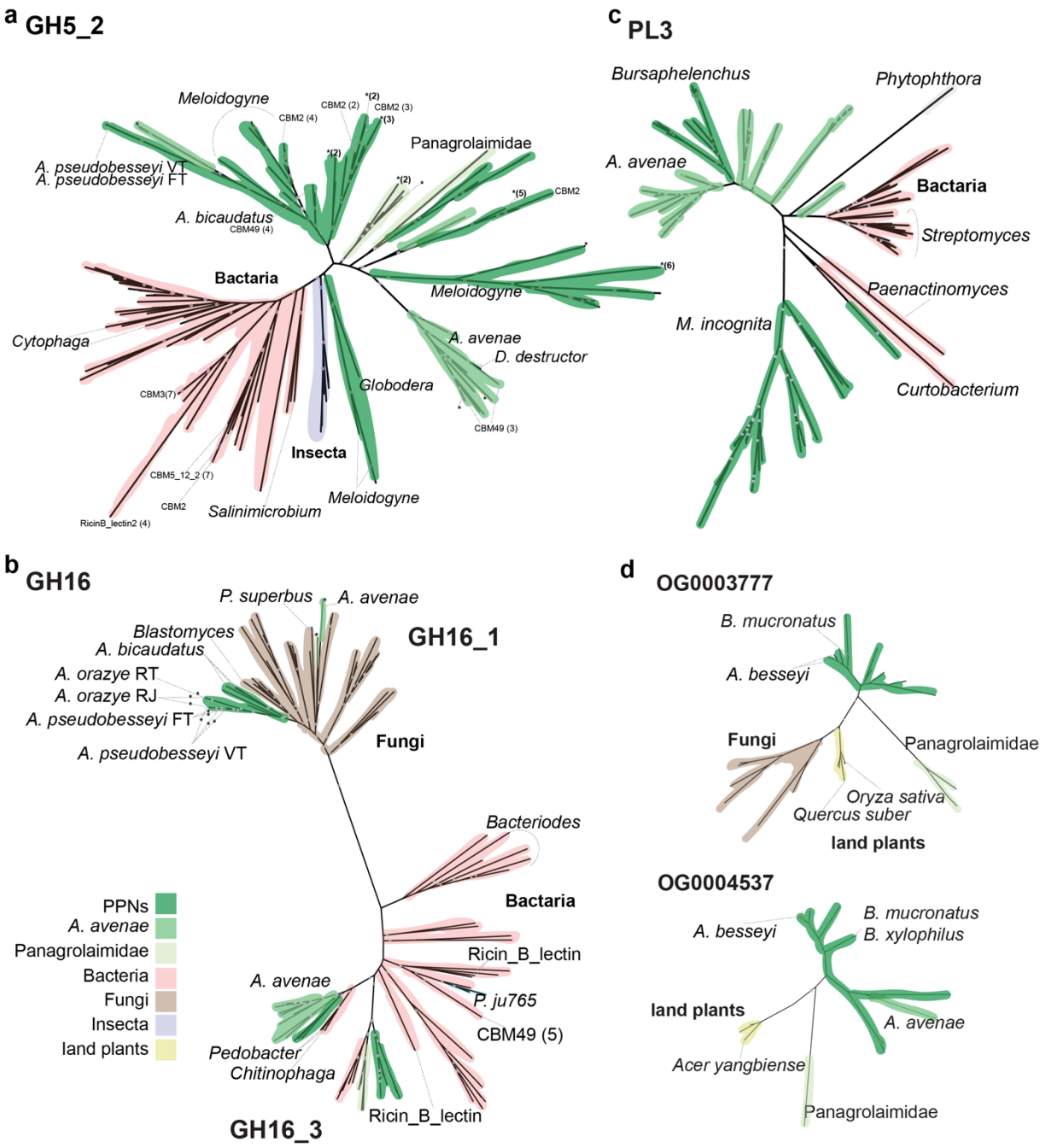

Figure 5 Phylogenies of nematode HGT orthogroups and their donor origin. a. GH5_2 **b.** GH16 **c.** PL3 **d.** OG0003777 and OG0004537. Different colours denote different kingdoms and species as shown in legend. Nematode gene copies with negative AI values were marked with an asterisk. Additional Pfam domains are labelled when available. Nodes with iqtree UFBoot and SH-aLRT bootstrap support > 80% are labelled as grey circles.

392 We identified 0.3-2.4%, 0.6-2.1% and 0.1-5.4% proteomes among
393 Aphelenchoidea, Tylenchida and *Panagrolaimomorpha* nematodes that were predicted
394 to be HGT (**fig. 4c**). The majority of these differences were the result of clade-specific
395 evolution after speciation. The high copy number of HGT genes observed in *M.*
396 *incognita* was a result of duplication[53], indicated by the fact that the number of HGT
397 orthologs of bacteria origin were over two times higher than any other species
398 (**supplementary fig. S10**). The high number of HGT genes in *P. superbus* was
399 consistent with a previous study[54] and likely to be species specific.
400 To independently assess the accuracy of our approach and interrogate the fate
401 of HGT genes, we constructed a phylogeny for every orthogroup containing identified
402 HGT candidates. Members of Aphelenchoididae and Tylenchida orthologs in the
403 majority of these orthogroups were predicted to be all HGT genes (with AI > 0; 54.6-
404 76.5% vs. 77.3-89.4%). Genes from a species were typically grouped together in the
405 orthogroup phylogeny regardless of being identified as HGT candidates, suggesting the
406 genes that were not detected using our threshold shared common ancestries with those
407 that were. Presumably, this was a result of accumulating substitutions over time.
408 Consistent with this observation, the more ancient acquired HGT orthogroups in PPNs
409 contained higher copy number of these genes compared to recently acquired families
410 (**supplementary fig. S11**). The instances included GH5 families with 12.5-70.6% of
411 copies in Tylenchida were failed to identify as HGT candidates, suggesting duplication
412 and possibly neo-functionalisation of the GH family in PPNs after being acquired from
413 bacterial origin (**fig. 5a**). The differentiation was ongoing and observed in the *A. besseyi*
414 species complex, which included the GH45 orthogroup with negative AI in two *A. oryzae*
415 strains (**supplementary fig. S12**).
416

417 ## Discussion

418 Characterising the diversity and comparing the genomes of plant parasitic
419 nematodes has been of fundamental importance in understanding how such lifestyles
420 arise and of practical importance in identifying candidate effectors and control methods.
421 The latter has been addressed in several studies, focusing mainly on *Meloidogyne*[55].
422 The *Aphelenchoides* genome assemblies presented in this study allowed us to gain a

423 holistic view of the evolution of clade IV nematodes, which appeared to gain and lose
424 many adaptations, including plant parasitism[56]. In their evolution, HGT genes have
425 played important roles in functions related to these adaptations. The most recent
426 comprehensive analyses of HGT in nematodes focused on plant parasitic nematodes[15]
427 and found many of these genes were PPN specific. Of these, donors of gene families
428 involved in plant cell-wall modifications were previously found to be associated with
429 plants which was appealing that were sympatric with plant parasitic nematodes which
430 made HGT possible[44]. Additional HGT events were identified in other clade IV
431 nematodes[13,54,57–59] but were part of analysing new *de novo* genomes.
432 Our systematic investigation of HGT has instead shown that many of the
433 aforementioned families were acquired much earlier in the last common ancestor of
434 clade IV. Many clade IV nematodes are known to survive extreme desiccation[54,59,60] and
435 the acquired HGT genes may be central to their resistance in harsh environment and
436 subsequently catalysis to successful plant parasitism[61]. Sources of these donors may
437 be the symbionts like the case of insects[62], but currently nematode endosymbionts are
438 restricted to *Wolbachia* and *Cardinium*[63] and were not identified in our analyses.
439 Interestingly, many of the closest bacterial donors were from marine environments,
440 raising the possibility that the last common ancestor of clade IV may have lived in a
441 marine environment that underwent habitat transition[64]. However, we also identified
442 donors of non-bacterial origin that were usually found in the environments that fit
443 nematodes' present day lifestyle. Now that more genome sequences are available,
444 historical HGT events were detected in the most recent common ancestor of major
445 organism groups such as land plants[65] , of moths and butterflies[66], which contributed
446 the hosts' developmental roles and adaptations. These acquisitions were found to be
447 episodic and likely took place in a time when either the host development or genome
448 defence was vulnerable. We speculate that the gain and absence of gene families in
449 clade IV nematodes may have played a role in retention of HGT genes.
450 The successful delimitation of the *A. besseyi* species complex unambiguously
451 into *A. oryzae* and the recently proposed *A. pseudobesseyi* has important implications
452 in nematode management. Congruent delimitation was observed between genomes
453 and 28S phylogenies confirming the utility of species identification with existing

454 molecular markers[18]. *A. besseyi* is generally believed to have limited mobility in natural
455 habitats, so its lack of population structure in China[24] was suggested as a consequence
456 of human-mediated dispersal. Our results also supported that *A. oryzae* appears to be
457 more rice plant specific compared to *A. pseudobesseyi* which was isolated more
458 frequently in ornamental plants and other agronomic crops[18]. A comprehensive
459 collection across a wider geographical range and resequencing of strains previously
460 designated as *A. besseyi* could confirm whether *A. oryzae* was responsible for all the
461 white tip disease in rice plants and may lead to better characterisations of the
462 biogeography and evolution of different cryptic species.

463 The reduction of genome size and reduced chromosome numbers of *A. besseyi*
464 represent an interesting outcome for the typical six nematode ancestral chromosomes
465 around hundred megabases in length. Genome rearrangement and reduction are
466 common across the tree of life including plants[67], butterflies[68] and nematodes[69]. We
467 show that *A. besseyi* underwent multiple chromosome fission and fusion events, and a
468 possible explanation together with genome reduction may be the missing of meiosis
469 genes and the telomeric repeat maintenance genes, which resulted in truncated meiosis
470 (**supplementary table S10**); this was observed in an extreme case of *Diploscapter*
471 *pachys*[70] possessing a single chromosome. Alterations in meiosis may lead to genome
472 shrinkage due to a loss of transposable elements as a result of imbalanced chromatin
473 as observed in *Caenorhabditis nigoni*[71]. It is likely that *Aphelenchoides* underwent a
474 similar scenario. However, members of *Bursaphelenchus* with six chromosomes also
475 failed to identify these aforementioned orthologs (**supplementary table S10**),
476 suggesting their divergence has taken place since the last common ancestor of
477 Aphelenchoidea. Further cell and molecular evidence are needed to confirm the
478 integrity of meiosis in *A. besseyi*.

479 To conclude, the availability of the *Aphelenchoides* genome and our comparative
480 analyses allowed us to pinpoint the major events of horizontal gene transfer in clade IV
481 nematodes. The results have reinforced the importance of horizontal gene transfers
482 contributing to multiple adaptations of these nematodes including plant parasitism. In
483 addition, the various *A. besseyi* genomes will assist in developing molecular diagnostic
484 tools to distinguish the specific diseases caused by the species complex.

485 ## Methods

486 **DNA, RNA extraction and sequencing**

487 Nematodes were cultured with *Alternaria citri* on PDA (potato dextrose agar)
488 medium. All stages of nematodes were collected from the medium, washed with sterile
489 distilled water, and purified by sucrose gradients. Genomic DNA was extracted using
490 Qiagen Genomic-tip 100/G according to the manufacturer's instructions, RNA extraction
491 was conducted using Trizol, and then purified using a lithium chloride purification
492 method. The DNA paired-end libraries were constructed using either a Nextera DNA
493 Flex or KAPA hyper library prep kit (Illumina, San Diego, USA); the RNA paired-end
494 libraries were constructed using a TruSeq Stranded mRNA library prep kit (Illumina, San
495 Diego, USA). Both DNA and RNA pair-end followed with standard protocol and were
496 sequenced by Illumina HiSeq 2500 (Illumina, USA) to produce 150-bp paired-end reads.
497 The HiC library preparation was performed by Phase Genomics (Seattle, WA, USA)
498 proximo HiC animal protocol with some modification in tissue processing. The enriched
499 worms were finely chopped by microtube pellet pestle rods for about 2 minutes. The
500 tissues were crosslinked by adding 1 ml crosslinking solution and incubate for 25
501 minutes with occasional mixing by rotation. 100 ul quenching solution was added to the
502 crosslinked tissue and mixed for 20 minutes by rotation. The rest of the preparation
503 steps follow the protocol. The library was sequenced by Illumina HiSeq 2500 (Illumina,
504 USA) to produce 150-bp paired-end reads. APFT and AORT were using Pacbio
505 sequencing system to produce long-read, and the rest of 4 *Aphelenchoides* strains
506 (APVT, AORJ, *A. bicaudatus*, *A. fujianensis*) were sequenced using the Oxford
507 Nanopore sequencing platform. The raw nanopore signals were basecalled by Guppy[72]
508 (ver 0.5.1) producing a total of 5.0-28.4 Gb sequences at least 1 kb in length.
509

510 **Assemblies of six *Aphelenchoides* spp.**

511 Raw reads of each species were assembled using Flye (ver 2.8.2)[26] assembler.
512 The assemblies from Nanopore reads were corrected using Nanopore reads using
513 Racon[73] (ver 1.4.6) and Medaka[74] (ver 0.10.0). All assemblies were further corrected
514 using Illumina reads using Pilon[75] (ver 1.22) with five iterations. The *A. pesudobesseyi*
515 VT assembly was scaffolded using HiC reads and subsequently curated in Juice-box[27]

516 tools. The other five *Aphelenchoides* genomes were reference scaffolded based on this
517 assembly using Ragtag[76] (ver 1.1).
518
519 **Gene prediction and functional annotation**
520 The identification of repetitive elements were computed by RepeatModeler[77] (ver
521 1.0.8), TransposonPSI (ver 1.0.0; https://github.com/NBISweden/TransposonPSI) and
522 USEARCH[78] (ver 8.1) based on the protocol by Berriman *et al*.[79]. Repeat locations were
523 then identified by Repeatmasking[80] (ver 4.0.9). RNA-seq reads of six *Aphelenchoides*
524 strains were trimmed by Trimmomatic[81] (ver 0.36), and aligned to corresponding
525 assemblies using STAR[82] (ver 2.7.1a). From these mappings, transcripts were inferred
526 using three approaches: i) assembled based on the mappings as guides using Trinity[83]
527 (ver 2.84; option: default setting), reconstructed using ii) Stringtie[84] (ver 1.3.4; option:
528 default setting) and iii) CLASS2[85] (ver 2.17; option: default setting). Transcripts
529 generated from Trinity were realigned to the reference using GMAP[86] (ver 2017-11-15).
530 The RNAseq mappings were also used in BRAKER[87] to train species parameter and
531 generate an initial set of annotations. Proteomes of *Bursaphelenchus xylophilus* and
532 *Caenorhabditis elegans* were downloaded from Wormbase (WBPS14;
533 https://wormbase.org) and used as homology guides to pick the best transcripts for
534 each putative locus using MIKADO[88] (ver 1.2.4; option: three Mikado steps, containing
535 "prepare", "serialize" and "pick" procedures), and were also used to train MAKER2.
536 Finally, MAKER2 was invoked to generate a final set of gene annotations using picked
537 EST evidence and protein evidence from MIKADO transcript and proteomes from
538 closely related species (*Bursaphelenchus xylophilus* and *Caenorhabditis elegans*), and
539 used gene models (BUSCO[89], BRAKER, SNAP[90] and Augustus[91]) as EST hints to train
540 predicted data with three iterations.
541
542 **Comparative analyses**
543 Proteomes of five plant-parasitic nematodes (*Bursaphelenchus xylophilus,*
544 *Meloidogyne hapla, Meloidogyne incognita, Globodera pallida, Ditylenchus destructor*),
545 two free-living nematodes (*Caenorhabditis elegans, Caenorhabditis briggsae*), six
546 *Panagrolaimomorpha* (*Propanagrolaimus* sp. JU765, *Panagrellus revidius,*

547 *Panagrolaimus superbus, Panagrolaimus* sp. PS1159*, Panagrolaimus davidi* and
548 *Halicephalobus mephisto*) and one animal parasitic nematode (*Brugia malayi*) were
549 downloaded from Wormbase[92,93](ver 14). Orthogroups were determined by
550 Orthofinder[32] (ver 2.2.7; options: -S diamond). Sequence alignments of each of the
551 single-copy orthogroups were generated by MAFFT (ver 7.310; options: --maxterate
552 1000). Then, the concatenated alignment of all single-copy orthogroups was used to
553 compute a maximum likelihood phylogeny using RAxML[94] (ver 8.2.3; options: -s -T 32 -
554 N 100 -f a -m PROTGAMMILGF) with 100 bootstrap replicates. Pfam copy numbers of
555 all 27 nematodes were identified from the results of nematode proteomes blast against
556 the database of Pfam website (ver 31; https://pfam.xfam.org/) using HMMER engine
557 with e-value smaller than 0.001. Effector enzymes were identified by searching the
558 nematode proteomes against the CAZyme[95] database (http://www.cazy.org) using
559 HMMER engine with sequence length larger than 80bp. The identified sequence was at
560 least larger than 0.35 proportion of conserved domain from database and had an e-
561 value smaller than 1e-15.
562
563 **Identification of the HGT genes**
564 The probability of genes having been acquired via HGT was estimated by using
565 Alienness Index (AI)[45]. Our donor group were generated by non-Metazoans from NCBI
566 nr database, and the recipient were Metazoans excluding the following species to
567 prevent self-alignment: Aphelenchoidea, Tylenchida, Rhabditina, Spirurina and
568 Cephaloboidea. The Alien Index (AI) was estimated by calculating the e-value of
569 diamond[96] (ver 2.0.14; option: blastx --evalue 0.001) best hits between the donor and
570 recipient database. Orthogroups having at least one gene with an AI value over 30 were
571 selected for further analysis. Gains and losses at each node were inferred using Phylip-
572 Dollop[46] (ver 3.69; options: fdollop -method d -ancseq). Some of the HGT family
573 acquired branches were manually curated by their evolutionary place of gene phylogeny
574 due to the fact that nematode genes with AI < 0 were clustered with other HGT genes.
575 The highest AI value of nematode genes with classified taxonomy hit were chosen to
576 represent the HGT origin in each orthogroup. Orthogroups with the same CAZyme
577 annotated and nematode orthology gene AI higher than -50 in those Orthogroups were

578 selected. AI < 0 genes were labelled with "*". The orthologs were further combined with
579 the HGT identified donor sequence from nr database and the specific cellulase
580 sequence from CAZyme database. To reduce contamination, orthologs of Pfam domain
581 were annotated and filtered by having at least one major domain (cellulase or pectate
582 lyase). Sequences of each HGT orthogroup were aligned using MAFFT (options: --
583 maxiterate 1000 --genafpair) and trimmed by using trimAl[97] (ver 1.4; options: -
584 gappyout). The ortholog phylogeny were computed by using IQtree[98] (ver 1.6.6; options:
585 -bb 1000 -alrt 1000). For the CAZyme unclassified HGT orthogroups, the top 2 blast hits
586 sequences from separated Uniprot (bacteria, fungi, land plants and insect) were used to
587 confirm the HGT origin.
588
589 **AUTHORS CONTRIBUTION**
590 IJT and PJC conceived the study. IJT led the study. YCL, TY and PJC sampled the
591 *Aphelenchoides* nematodes. YiCL, HMK, WAL conducted the experiments. CKL
592 analysed the data with input from HHL, YuCL and MRL. IJT and CKL wrote the
593 manuscript with input from TY, TK, PJC
594
595 **DATA AVAILABILITY**
596 The sequencing data and annotation of six *Aphelenchoides* nematodes are publicly
597 available in NCBI with the BioProject accession PRJNA834627 and scheduled in the
598 next release WBPS18 of WormBase Parasite. The accession numbers of individual
599 samples are listed in **supplementary table S2**. The information of Clade IV acquired
600 HGT orthogroups could be found in github
601 (https://github.com/lihowfun/CladeIV_HGT.git).
602
603 **Funding**
604 This work was supported by Academia Sinica grant (AS-CDA-107-L01) to IJT.
605
606 ## References
607 1. Blaxter, M. L. *et al.* A molecular evolutionary framework for the phylum Nematoda. *Nat.*
608 *1998 3926671* **392**, 71–75 (1998).
609 2. Van Megen, H. *et al.* A phylogenetic tree of nematodes based on about 1200 full-length

610 small subunit ribosomal DNA sequences. *Nematology* **11**, 927–950 (2009).
611 3. Bird, D. M. K., Jones, J. T., Opperman, C. H., Kikuchi, T. & Danchin, E. G. J. Signatures
612 of adaptation to plant parasitism in nematode genomes. *Parasitology* **142**, S71 (2015).
613 4. Nicol, J. M. *et al.* Current Nematode Threats to World Agriculture. *Genomics Mol. Genet.*
614 *Plant-Nematode Interact.* 21–43 (2011). doi:10.1007/978-94-007-0434-3_2
615 5. Abad, P. *et al.* Genome sequence of the metazoan plant-parasitic nematode Meloidogyne
616 incognita. *Nat. Biotechnol. 2008 268* **26**, 909–915 (2008).
617 6. Kikuchi, T., Cotton, J. A., Dalzell, J. J., Hasegawa, K. & Kanzaki, N. Genomic Insights
618 into the Origin of Parasitism in the Emerging Plant Pathogen Bursaphelenchus xylophilus
619 Genomic Insights into the Origin of Parasitism in the Emerging Plant Pathogen. (2011).
620 doi:10.1371/journal.ppat.1002219
621 7. Dayi, M. *et al.* Nearly Complete Genome Assembly of the Pinewood Nematode
622 Bursaphelenchus xylophilus Strain Ka4C1 . *Microbiol. Resour. Announc.* **9**, (2020).
623 8. Cotton, J. A. *et al.* The genome and life-stage specific transcriptomes of Globodera
624 pallidaelucidate key aspects of plant parasitism by a cyst nematode. *Genome Biol. 2014*
625 *153* **15**, 1–17 (2014).
626 9. Masonbrink, R. *et al.* The genome of the soybean cyst nematode (Heterodera glycines)
627 reveals complex patterns of duplications involved in the evolution of parasitism genes.
628 *BMC Genomics* **20**, 119 (2019).
629 10. Wu, S. *et al.* A reference genome of bursaphelenchus mucronatus provides new
630 resources for revealing its displacement by pinewood nematode. *Genes (Basel).* **11**,
631 (2020).
632 11. Koutsovoulos, G. D. *et al.* Genome assembly and annotation of Meloidogyne enterolobii,
633 an emerging parthenogenetic root-knot nematode. *Sci. Data 2020 71* **7**, 1–13 (2020).
634 12. Ali, M. A., Azeem, F., Li, H. & Bohlmann, H. Smart parasitic nematodes use multifaceted
635 strategies to parasitize plants. *Frontiers in Plant Science* (2017).
636 doi:10.3389/fpls.2017.01699
637 13. Danchin, E. G. J., Guzeeva, E. A., Mantelin, S., Berepiki, A. & Jones, J. T. Horizontal
638 Gene Transfer from Bacteria Has Enabled the Plant-Parasitic Nematode Globodera
639 pallida to Feed on Host-Derived Sucrose. *Mol. Biol. Evol.* **33**, 1571–1579 (2016).
640 14. Jones, J. T., Furlanetto, C. & Kikuchi, T. Horizontal gene transfer from bacteria and fungi
641 as a driving force in the evolution of plant parasitism in nematodes. *Nematology* (2005).
642 doi:10.1163/156854105775142919
643 15. Grynberg, P. *et al.* Comparative genomics reveals novel target genes towards specific

control of plant-parasitic nematodes. (2020). doi:10.20944/preprints202010.0449.v1

16. Jen, F. Y., Tsay, T. T. & Chen, P. Aphelenchoides bicaudatus from ornamental nurseries in Taiwan and its relationship with some agricultural crops. *Plant Dis.* **96**, 1763–1766 (2012).

17. Subbotin, S. A. *et al.* The taxonomic status of Aphelenchoides besseyi Christie, 1942 (Nematoda: Aphelenchoididae) populations from the southeastern USA, and description of Aphelenchoides pseudobesseyi sp. N. *Nematology* (2020). doi:10.1163/15685411-bja10048

18. Oliveira, C. J. *et al.* Morphological and Molecular Identification of Two Florida Populations of Foliar Nematodes (Aphelenchoides spp.) Isolated from Strawberry with the Description of Aphelenchoides pseudogoodeyi sp. N. (Nematoda: Aphelenchoididae) and notes on their bionomics. *Plant Dis.* **103**, 2825–2842 (2019).

19. Wang, F. *et al.* Transcriptomic analysis of the rice white tip nematode, phelenchoides besseyi (Nematoda: Aphelenchoididae). *PLoS One* **9**, (2014).

20. Lilley, C. J., Kyndt, T. & Gheysen, G. Nematode Resistant GM Crops in Industrialised and Developing Countries. in *Genomics and Molecular Genetics of Plant-Nematode Interactions* 517–541 (Springer Netherlands, 2011). doi:10.1007/978-94-007-0434-3_24

21. Koenning, S. R. *et al.* Survey of crop losses in response to phytoparasitic nematodes in the United States for 1994. *Journal of Nematology* **31**, 587–618 (1999).

22. Jones, J. T. *et al.* Top 10 plant-parasitic nematodes in molecular plant pathology. (2013). doi:10.1111/mpp.12057

23. Wu, G. L., Kuo, T. H., Tsay, T. T., Tsai, I. J. & Chen, P. J. Glycoside hydrolase (gh) 45 and 5 candidate cellulases in aphelenchoides besseyi isolated from bird's-nest fern. *PLoS One* **11**, (2016).

24. Xu, X. *et al.* Population structure and species delimitation of rice white tip nematode, Aphelenchoides besseyi (Nematoda: Aphelenchoididae), in China. *Plant Pathol.* **69**, 159–167 (2020).

25. Rybarczyk-Mydłowska, K. *et al.* Small subunit ribosomal DNA-based phylogenetic analysis of foliar nematodes (Aphelenchoides spp.) and their quantitative detection in complex DNA backgrounds. *Phytopathology* (2012). doi:10.1094/PHYTO-05-12-0114-R

26. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* (2019). doi:10.1038/s41587-019-0072-8

27. Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Syst.* (2016). doi:10.1016/j.cels.2015.07.012

678 28. Yoshida, K., Hasegawa, K., Mochiji, N. & Miwa, J. Early Embryogenesis and Anterior-
679 Posterior Axis Formation in the White-Tip Nematode Aphelenchoides besseyi
680 (Nematoda: Aphelenchoididae). *J. Nematol.* **41**, 17–22 (2009).
681 29. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and
682 Smudgeplots: Reference-free profiling of polyploid genomes. *bioRxiv* 747568 (2019).
683 doi:10.1101/747568
684 30. Cantarel, B. L. *et al.* MAKER: An easy-to-use annotation pipeline designed for emerging
685 model organism genomes. *Genome Res.* (2008). doi:10.1101/gr.6743907
686 31. Finn, R. D. *et al.* Pfam: The protein families database. *Nucleic Acids Research* (2014).
687 doi:10.1093/nar/gkt1223
688 32. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative
689 genomics. *Genome Biol.* **20**, 238 (2019).
690 33. Kikuchi, T., Eves-Van Den Akker, S. & Jones, J. T. Genome Evolution of Plant-Parasitic
691 Nematodes. (2017). doi:10.1146/annurev-phyto-080516
692 34. Nadachowska-Brzyska, K., Burri, R., Smeds, L. & Ellegren, H. PSMC analysis of effective
693 population sizes in molecular ecology and its application to black-and-white Ficedula
694 flycatchers. *Mol. Ecol.* **25**, 1058–1072 (2016).
695 35. Mitreva, M. *et al.* The draft genome of the parasitic nematode Trichinella spiralis. *Nat.*
696 *Genet.* (2011). doi:10.1038/ng.769
697 36. Tcherepanova, I., Bhattacharyya, L., Rubin, C. S. & Freedman, J. H. Aspartic proteases
698 from the nematode Caenorhabditis elegans: Structural organization and developmental
699 and cell-specific expression of asp-1. *J. Biol. Chem.* **275**, 26359–26369 (2000).
700 37. McKerrow, J. H. *et al.* Strongyloides stercoralis: Identification of a protease that facilitates
701 penetration of skin by the infective larvae. *Exp. Parasitol.* **70**, 134–143 (1990).
702 38. Koch, B. J., Ryan, J. F. & Baxevanis, A. D. The Diversification of the LIM Superclass at
703 the Base of the Metazoa Increased Subcellular Complexity and Promoted Multicellular
704 Specialization. doi:10.1371/journal.pone.0033261
705 39. Dall, E. & Brandstetter, H. Structure and function of legumain in health and disease.
706 *Biochimie* **122**, 126–150 (2016).
707 40. Holm Viborg, A. *et al.* A subfamily roadmap for functional glycogenomics of the
708 evolutionarily diverse Glycoside Hydrolase Family 16 (GH16).
709 doi:10.1074/jbc.RA119.010619
710 41. Coconi Linares, N., Dilokpimol, A., Stålbrand, H., Mäkelä, M. R. & de Vries, R. P.
711 Recombinant production and characterization of six novel GH27 and GH36 α-

galactosidases from Penicillium subrubescens and their synergism with a commercial mannanase during the hydrolysis of lignocellulosic biomass. *Bioresour. Technol.* (2020). doi:10.1016/j.biortech.2019.122258

42. Shinya, R. *et al.* Possible stochastic sex determination in Bursaphelenchus nematodes. *Nat. Commun. 2022 131* **13**, 1–14 (2022).
43. Woodruff, G. C. Patterns of putative gene loss suggest rampant developmental system drift in nematodes. *bioRxiv* 627620 (2019). doi:10.1101/627620
44. Danchin, E. G. J. *et al.* Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 17651–17656 (2010).
45. Rancurel, C., Legrand, L. & Danchin, E. G. J. Alienness: Rapid detection of candidate horizontal gene transfers across the tree of life. *Genes (Basel).* **8**, (2017).
46. Campoy, E. & González-Martín, A. The Geography as a Regulator of Genetic Flow and Genetic Structure in Andorra. *Adv. Anthropol.* (2017). doi:10.4236/aa.2017.72008
47. Morais, M. A. B. *et al.* Two distinct catalytic pathways for GH43 xylanolytic enzymes unveiled by X-ray and QM/MM simulations. *Nat. Commun. 2021 121* **12**, 1–13 (2021).
48. Atanasova, L. *et al.* Evolution and functional characterization of pectate lyase PEL12, a member of a highly expanded Clonostachys rosea polysaccharide lyase 1 family. *BMC Microbiol.* **18**, 1–19 (2018).
49. PATHOGENICITY OF CURTOBACTERIUM FLACCUMFACIENS pv. FLACCUMFACIENS TO SEVERAL PLANT SPECIES on JSTOR. Available at: https://www.jstor.org/stable/45156052#metadata_info_tab_contents. (Accessed: 26th August 2022)
50. Kikuchi, T., Jones, J. T., Aikawa, T., Kosaka, H. & Ogura, N. A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode Bursaphelenchus xylophilus. *FEBS Lett.* **572**, 201–205 (2004).
51. Elling, A. A. *et al.* Sequence mining and transcript profiling to explore cyst nematode parasitism. *BMC Genomics* **10**, (2009).
52. Rancurel, C., Legrand, L. & Danchin, E. G. J. Alienness: Rapid detection of candidate horizontal gene transfers across the tree of life. *Genes (Basel).* **8**, (2017).
53. Szitenberg, A. *et al.* Comparative genomics of apomictic root-knot nematodes: Hybridization, ploidy, and dynamic genome change. *Genome Biol. Evol.* **9**, 2844–2861 (2017).
54. Schiffer, P. H. *et al.* Signatures of the Evolution of Parthenogenesis and Cryptobiosis in

746 the Genomes of Panagrolaimid Nematodes. *iScience* **21**, 587–602 (2019).
747 55. Grynberg, P. *et al.* Comparative genomics reveals novel target genes towards specific
748 control of plant-parasitic nematodes. (2020). doi:10.20944/preprints202010.0449.v1
749 56. Holterman, M. *et al.* Disparate gain and loss of parasitic abilities among nematode
750 lineages. *PLoS One* (2017). doi:10.1371/journal.pone.0185445
751 57. Han, Z. *et al.* Horizontally Acquired Cellulases Assist the Expansion of Dietary Range in
752 Pristionchus Nematodes. *Mol. Biol. Evol.* **39**, (2022).
753 58. Zheng, J. *et al.* The ditylenchus destructor genome provides new insights into the
754 evolution of plant parasitic nematodes. *Proc. R. Soc. B Biol. Sci.* (2016).
755 doi:10.1098/rspb.2016.0942
756 59. Wan, X. *et al.* The Aphelenchus avenae genome highlights evolutionary adaptation to
757 desiccation. *Commun. Biol. 2021 41* **4**, 1–8 (2021).
758 60. Jagdale, G. B. & Grewal, P. S. Infection behavior and overwintering survival of foliar
759 nematodes, Aphelenchoides fragariae, on Hosta. *J. Nematol.* (2006).
760 61. Haegeman, A., Jones, J. T. & Danchin, E. G. J. Horizontal Gene Transfer in Nematodes:
761 A Catalyst for Plant Parasitism? *http://dx.doi.org/10.1094/MPMI-03-11-0055* **24**, 879–887
762 (2011).
763 62. Xia, J. *et al.* Whitefly hijacks a plant detoxification gene that neutralizes plant toxins. *Cell*
764 (2021). doi:10.1016/j.cell.2021.02.014
765 63. Brown, A. M. V. *et al.* Comparative genomics of wolbachia-cardinium dual endosymbiosis
766 in a plant-parasitic nematode. *Front. Microbiol.* **9**, 2482 (2018).
767 64. Holterman, M., Schratzberger, M. & Helder, J. Nematodes as evolutionary commuters
768 between marine, freshwater and terrestrial habitats. *Biol. J. Linn. Soc.* **128**, 756–767
769 (2019).
770 65. Ma, J. *et al.* Major episodes of horizontal gene transfer drove the evolution of land plants.
771 *Mol. Plant* **15**, 857–871 (2022).
772 66. Li, Y. *et al.* HGT is widespread in insects and contributes to male courtship in
773 lepidopterans. *Cell* **185**, 2975-2987.e10 (2022).
774 67. Ren, L., Huang, W., Cannon, E. K. S., Bertioli, D. J. & Cannon, S. B. A mechanism for
775 genome size reduction following genomic rearrangements. *Front. Genet.* (2018).
776 doi:10.3389/fgene.2018.00454
777 68. Cicconardi, F. *et al.* Chromosome Fusion Affects Genetic Diversity and Evolutionary
778 Turnover of Functional Loci but Consistently Depends on Chromosome Size. *Mol. Biol.*
779 *Evol.* **38**, 4449–4462 (2021).

780 69. Wang, J. *et al.* Comparative genome analysis of programmed DNA elimination in
781 nematodes. *Genome Res.* **27**, 2001–2014 (2017).
782 70. Fradin, H. *et al.* Genome Architecture and Evolution of a Unichromosomal Asexual
783 Nematode. *Curr. Biol.* **27**, 2928-2939.e6 (2017).
784 71. Yin, D. *et al.* Rapid genome shrinkage in a self-fertile nematode reveals sperm
785 competition proteins. *Science (80-.).* **359**, 55–61 (2018).
786 72. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for
787 Oxford Nanopore sequencing. *Genome Biol.* (2019). doi:10.1186/s13059-019-1727-y
788 73. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome
789 assembly from long uncorrected reads. *Genome Res.* (2017). doi:10.1101/gr.214270.116
790 74. Medaka: Sequence correction by using ONT research.
791 https://github.com/nanoporetech/medaka
792 75. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection
793 and genome assembly improvement. *PLoS One* (2014).
794 doi:10.1371/journal.pone.0112963
795 76. Alonge, M. *et al.* RaGOO: Fast and accurate reference-guided scaffolding of draft
796 genomes. *Genome Biol.* **20**, 224 (2019).
797 77. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable
798 element families. *Proc. Natl. Acad. Sci. U. S. A.* (2020). doi:10.1073/pnas.1921046117
799 78. Edgar, R. C. & Bateman, A. Search and clustering orders of magnitude faster than
800 BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
801 79. Coghlan, A., Coghlan, A., Tsai, I. J. & Berriman, M. Creation of a comprehensive repeat
802 library for a newly sequenced parasitic worm genome. *Protoc. Exch.* (2018).
803 doi:10.1038/protex.2018.054
804 80. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in
805 genomic sequences. *Current Protocols in Bioinformatics* (2009).
806 doi:10.1002/0471250953.bi0410s25
807 81. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina
808 sequence data. *Bioinformatics* (2014). doi:10.1093/bioinformatics/btu170
809 82. Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Curr. Protoc.*
810 *Bioinforma.* (2015). doi:10.1002/0471250953.bi1114s51
811 83. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a
812 reference genome. *Nat. Biotechnol.* (2011). doi:10.1038/nbt.1883
813 84. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-

seq reads. *Nat. Biotechnol.* (2015). doi:10.1038/nbt.3122

85. Song, L., Sabunciyan, S. & Florea, L. CLASS2: Accurate and efficient splice variant annotation from RNA-seq reads. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw158

86. Wu, T. D. & Watanabe, C. K. GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* (2005). doi:10.1093/bioinformatics/bti310

87. Hoff, K. J., Lange, S., Lomsadze, A., Borodovsky, M. & Stanke, M. BRAKER1: Unsupervised RNA-Seq-based genome annotation with GeneMark-ET and AUGUSTUS. *Bioinformatics* (2016). doi:10.1093/bioinformatics/btv661

88. Venturini, L., Caim, S., Kaithakottil, G. G., Mapleson, D. L. & Swarbreck, D. Leveraging multiple transcriptome assembly methods for improved gene structure annotation. *Gigascience* **7**, (2018).

89. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* (2015). doi:10.1093/bioinformatics/btv351

90. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 1–9 (2004).

91. Stanke, M. *et al.* AUGUSTUS: A b initio prediction of alternative transcripts. *Nucleic Acids Res.* (2006). doi:10.1093/nar/gkl200

92. Howe, K. L., Bolt, B. J., Shafie, M., Kersey, P. & Berriman, M. WormBase ParaSite − a comprehensive resource for helminth genomics. *Mol. Biochem. Parasitol.* **215**, 2 (2017).

93. Davis, P. *et al.* WormBase in 2022—data, processes, and tools for analyzing Caenorhabditis elegans. *Genetics* **220**, (2022).

94. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* (2014). doi:10.1093/bioinformatics/btu033

95. Xu, J. -H. Carbohydrate Active Enzyme database. in *Catalysis from A to Z* (2020). doi:10.1002/9783527809080.cataz02801

96. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* (2014). doi:10.1038/nmeth.3176

97. Capella-Gutiérrez, S., Silla-Martínez, J. M. & Gabaldón, T. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* (2009). doi:10.1093/bioinformatics/btp348

98. Nguyen, L. T., Schmidt, H. A., Von Haeseler, A. & Minh, B. Q. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* (2015). doi:10.1093/molbev/msu300