Phytopathology Research

RESEARCH

Open Access

Complete genome sequencing of nematode *Aphelenchoides besseyi*, an economically important pest causing rice white-tip disease

Hongli Ji[1], Jialian Xie[1], Ziduan Han[2], Fang Yang[1], Wenjuan Yu[1], Yunliang Peng[1] and Xue Qing[3*]

Abstract

Aphelenchoides besseyi is a seed-borne plant-parasitic nematode that causes severe rice yield losses worldwide. In the present study, the *A. besseyi* Anhui-1 strain isolated from rice in China was sequenced with a hybrid method combining PacBio long reads and Illumina short reads, and subsequently annotated using available transcriptome references. The genome assembly consists of 166 scaffolds totaling 50.3 Mb, with an N50 of 1.262 Mb and a maximum scaffold length of 9.17 Mb. A total of 16,343 genes were annotated in the genome, with 94 gene families expanded while 70 families contracted specifically in *A. besseyi*. Furthermore, gene function analysis demonstrated that the genes related to drought tolerance were enriched, and cellulase genes were horizontally acquired from eukaryotic origin. Our findings provide resources to interpret the biology, evolution, ecology, and functional diversities of *Aphelenchoides* spp. in the light of genomics.

Keywords Aphelenchoidea, Nematoda, PacBio, Horizontal gene transfer, *Aphelenchoides oryzae*, *Oryza sativa*

Background

Aphelenchoides besseyi is a seed-borne plant-parasitic nematode (PPN) that parasitizes rice (*Oryza sativa*), strawberry (*Fragaria grandiflora*), as well as other plants belonging to 35 genera (Duncan and Moens 2013). *A. besseyi* could cause severe rice yield losses of up to 70% in some cases (Lin et al. 2005; Tulek and Cobanoglu 2010) and is considered as one of the major PPNs in world crop production (Jones et al. 2013).

A. besseyi was first isolated from strawberries in the USA (Christie 1929). Later, Yokoo (1948) described *A. oryzae* from rice, but it was considered as a junior synonym of *A. besseyi* due to the overlapping of many morphological characters (Allen 1952). Recently, molecular and phylogenetic analyses suggested that *A. besseyi* may be a species complex consisting of several cryptic species that are not well morphologically delimited (Oliveira et al. 2019; Xu et al. 2020). The species complex consists of three described species: *A. pseudobesseyi* parasitizing ornamental plants, *A. oryzae* parasitizing rice, and *A. besseyi* parasitizing strawberries (Subbotin et al. 1942). Since the identification of these species primarily relies on molecular tools and their morphology is nearly identical, we retain the name of species complex *A. besseyi* in the present study.

Unlike most PPNs that infect root tissues in the soil, *A. besseyi* feeds on growing points of stems and leaves of seedlings, causing the disease called 'white-tip' (Perry and Moens 2013). *A. besseyi* bears a series of interesting

*Correspondence:
Xue Qing
qingxue@njau.edu.cn
[1] MOA Key Laboratory of Integrated Management of Pests on Crops in Southwest China, Institute of Plant Protection, Sichuan Academy of Agricultural Sciences, Chengdu 610066, China
[2] Department of Plant Pathology, Northwest A&F University, Yangling 712100, China
[3] Department of Plant Pathology, Nanjing Agricultural University, Nanjing 210095, China

characters that can be a model to study nematode evolution and adaptation. For example, it can survive in stored rice grains for several years through anhydrobiosis (Tiwari and Khare 2003), thus can be used to study the adaption of desiccation; it is a facultative parasite propagating on both fungi and plants (Perry and Moens 2013), which could be an excellent object to study the evolution of plant parasitism. *A. besseyi* is predominantly amphimictic, and males are usually abundant (Huang et al. 1979), but parthenogenetic reproduction has also been found in some populations (Nandini et al. 2001); thus, it can be an example to study the mechanisms of reproduction mode.

Regardless of their importance in agriculture, the genome of *A. besseyi* has not yet been sequenced. This is also the case in a broader of aphelenchs nematodes. Among 453 valid aphelenchs nematodes listed by Hunt (2008), genomic information was only available for *Bursaphelenchus xylophilus, B. okinawaensis*, *B. mucronatus*, and *Aphelenchus avenae.* More high-quality genomic data would provide valuable insights into the evolution of aphelenchs (Kikuchi et al. 2011; Wan et al. 2021).

In this study, both PacBio long reads and Illumina short reads sequences were used to study the genome of *A. besseyi*. Protein-coding, non-coding genes, and transposable elements (TE) were predicted using newly sequenced data together with the available RNA-seq transcriptome. Cellulase is an iconic gene that PPNs acquired through horizontal gene transfer (HGT), and we also investigated the possible origin of cellulase genes from *A. besseyi*. So far, this new genome is the most contiguous and most complete annotated one for aphelenchs species, which could provide a robust reference for further analyses with important evolutionary and agro-economic implications.

Results

Genome features of *A. besseyi*

Our assembly of the *A. besseyi* Anhui-1 strain (NCBI BioProject: PRJNA901680) consists of 166 scaffolds totaling 50.3 Mb, with an N50 of 1.262 Mb and a maximum scaffold length of 9.17 Mb (Table 1). A total of 143 ncRNAs were identified, including 17 miRNA (62–148 bp), 148 rRNA (113–7556 bp), 61 snRNA (72–217 bp), and 276 tRNA (71–127 bp) (Additional file 1: Tables S1–S4). The genome completeness was assessed by mapping BUSCOs onto the genome assembly. The assembled genome represents 78.2% of the Anhui-1 genome as it carries 744 single-copy (75.8%), 24 duplicated (2.4%), 161 missing BUSCOs (16.4%). In addition, 53 fragmented BUSCOs (5.4%) were aligned to the genome. The assembled genome is about half of the model nematode *Caenorhabditis elegans* (100 Mb), and it is the smallest genome known in aphelenchs (Table 1). GC content of the genome assembled is 42.2%, which is similar to *A. avenae* (42.1%) and *B. xylophilus* (40.4%) but higher than *B. okinawaensis* (36.2%).

Analyses of repetitive elements suggested a total of 303 kb tandem repeats, occupying 0.6% of the genome. The size of TE varies depending on different methods, with the largest when using RepeatModeler and LTR-FINDER database (de novo methods, 9810 kb, 19.5% of genome) and smallest when using Repbase database (Repbase TEs, 988 kb, 1.97% of genome). After combining these methods and removing redundancy, significantly longer TEs were recovered (Combined TEs, 10.4 Mb), accounting for 21.2% of the genome (Table 2 and Additional file 2).

Gene annotation and comparison with other nematodes

The *A. besseyi* genome is predicted to encode 16,343 protein-coding genes, whose number is similar to *B. xylophilus* (15,860) but much less than *A. avenae* (43,724) (Table 1). Among these annotated genes, 93.9% of protein-coding genes can be assigned to orthogroups (15,348). A total of 452 species-specific orthogroups containing 2334 protein-coding genes were found in *A. besseyi*. Within 44 examined species, the root-knot nematode *Meloidogyne graminicola*s has the smallest gene number (10,895), while animal parasitic species tend to have more genes; for example, the largest gene number was found in the insect parasite nematode *Romanomermis culicivorax* (48,376).

A total of 7495 annotated orthogroups in *A. besseyi* contain a single copy (5299 orthogroups), followed by

Table 1 The genome statistics of newly sequenced *Aphelenchoides besseyi* and other sequenced aphelenchs nematodes

Sample	Scaffolds	Genome size (Mb)	N50 (Mb)	GC%	Gene	References
A. besseyi	166	50.30	1.26	42.2	16,343	This study
B. xylophilus	11	78.30	12.79	40.4	15,860	Dayi et al. (2020)
B. okinawaensis	6	69.96	11.60	36.2	14,593	Sun et al. (2020)
B. mucronatus	72	73.05	1.60	39.3	13,696	Wu et al. (2020)
A. avenae	18,660	264.01	0.14	42.1	43,724	Wan et al. (2021)

Table 2 Summary of *Aphelenchoides besseyi* transposable elements (TE) annotation statistics based on different reference databases and methods

TE types	Repbase TEs		TE proteins		De novo		Combined TEs	
	Length (kb)	% genome	Length (kb)	% genome	Length (kb)	% genome	Length (kb)	% genome
DNA	599	1.19	1500	2.98	2386	4.74	2902	5.77
LINE	130	0.26	378	0.75	558	1.11	714	1.42
SINE	2	0.004	0	0	0	0	2	0.004
LTR	255	0.51	483	0.96	1931	3.84	2126	4.23
Unknown	2	0.004	0	0	4935	9.81	4936	9.81
Total	988	1.97	2361	4.69	9810	19.5	10,680	21.2

two copies (1114 orthogroups), while multiple copies are relatively rare (a total of 1082 orthogroups). This is in line with *B. xylophilus* as well as most PPNs, but different from polyploid root-knot species like *M. arenaria* and *M. enterolobii* in which two copies were more abundant (Fig. 1a). Besides, the genome of *A. besseyi* shows a longer average gene length, with 3 kb to be the most frequently recovered length (Additional file 3: Figure S1).

Identifying homologous relationships among the sequences of different species plays a pivotal role in

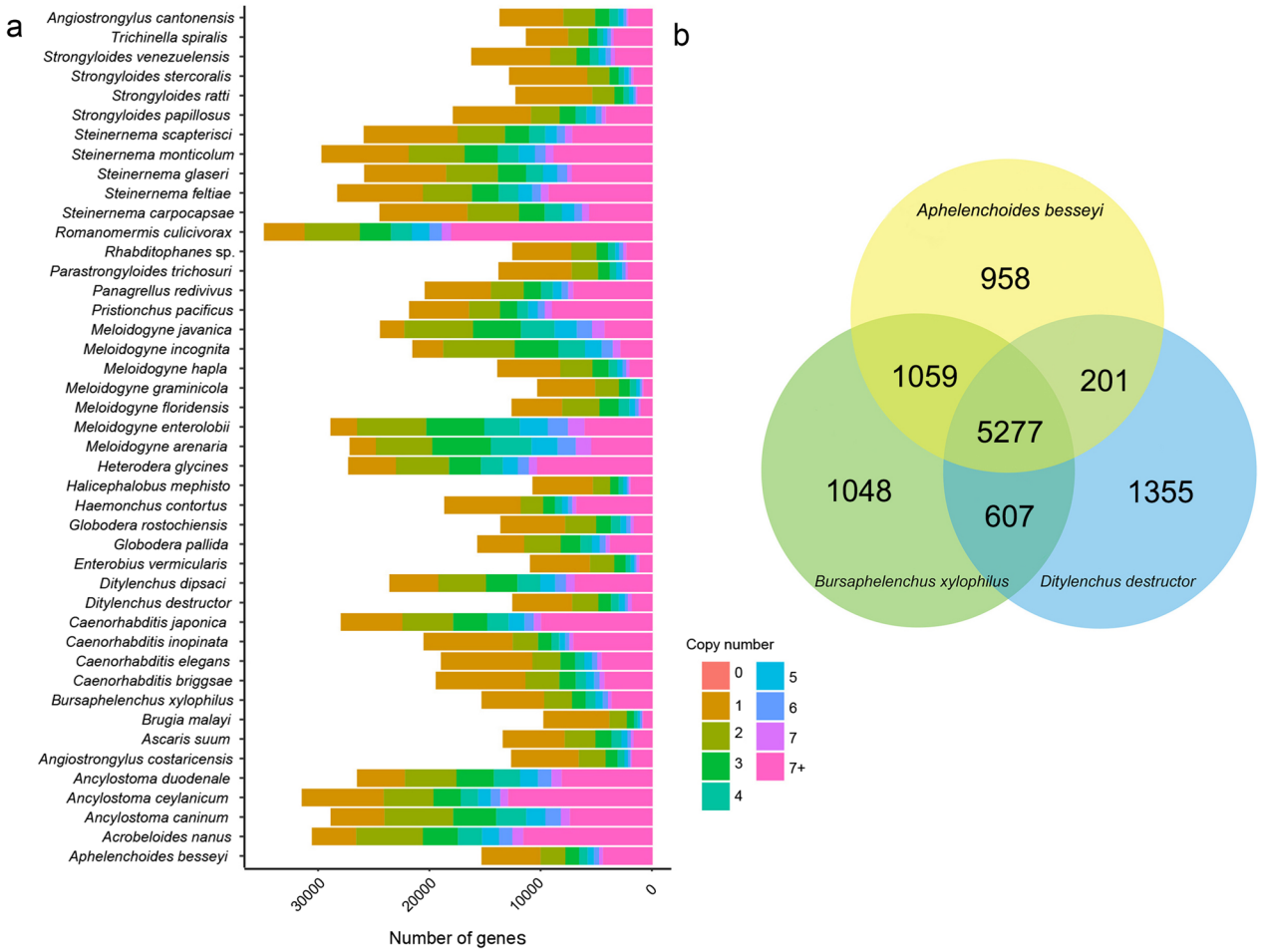

Fig. 1 The comparison of orthologous genes of *Aphelenchoides besseyi* and other closely related species. **a** The number of annotated genes in orthogroups and their copy numbers. **b** Orthologous genes shared among different species

enhancing our understanding of evolution and diversity. Therefore, we compared the protein-coding gene families shared by *A. besseyi*, *B. xylophilus*, and *Ditylenchus destructor*. A total of 5277 orthogroups were resolved among the three species. The two aphelenchs nematodes *A. besseyi* and *B. xylophilus* share the most abundant unique orthogroups (1059), followed by the stem nematodes *D. destructor* and *B. xylophilus* (607), and there are only 201 orthogroups shared by *D. destructor* and *A. besseyi*. In respect to unique genes, *A. besseyi* has a reduced number of unique gene families (958) compared to *B. xylophilus* (1048) and *D. destructor* (1355) (Fig. 1b).

Phylogenetic placement and molecular dating

We identified 242 orthogroups that have single-copy genes for a minimum of 50% of species (44 nematodes), and these genes were subsequently used for phylogeny reconstruction. As expected, *A. besseyi* is fully supported as a sister to pine wood nematode *B. xylophilus* (BS = 100) and forms a basal clade within Tylenchomorpha (Fig. 2a). Further molecular dating was performed using 1126 orthogroups that have single-copy genes for 12 out of 17 species. The results suggests that *A. besseyi* splits with *B. xylophilus* in an average of 163.8 million years ago, similar to the splitting of sedentary endoparasite cyst and root-knot nematodes (160.3 million years ago) (Fig. 2b).

The gene family expansion and function prediction

The analysis for gene family expansion and contraction reveals that 94 and 70 gene families are respectively expanded and contracted in *A. besseyi* (Additional file 4: Tables S1, S2), similar to *B. xylophilus*, which has 88 expanded and 69 contracted gene families (Fig. 3).

To better evaluate the gene ontology and functional classification of annotated genes, we performed functional analysis using gene ontology (GO), eukaryotic orthologous groups (KOG), Kyoto encyclopedia of genes and genomes (KEGG) (Figs. 4, 5 and Additional file 4: Tables S3–S5), NCBI NR, and SwissProt databases. The NR search annotated 12,031 genes and the SwissProt resulted in 7646 annotations; details for these two annotations are given in Additional file 4: Tables S6, S7.

A total of 8238 protein-coding genes were functionally annotated using the GO database (Additional file 4: Table S3). GO terms include biological process (BP), cellular component (CC), and molecular function (MF), comprising 22, 16, and 10 elements, respectively. The top three annotated BPs were the cellular process (GO:0009987), metabolic process (GO:0008152),

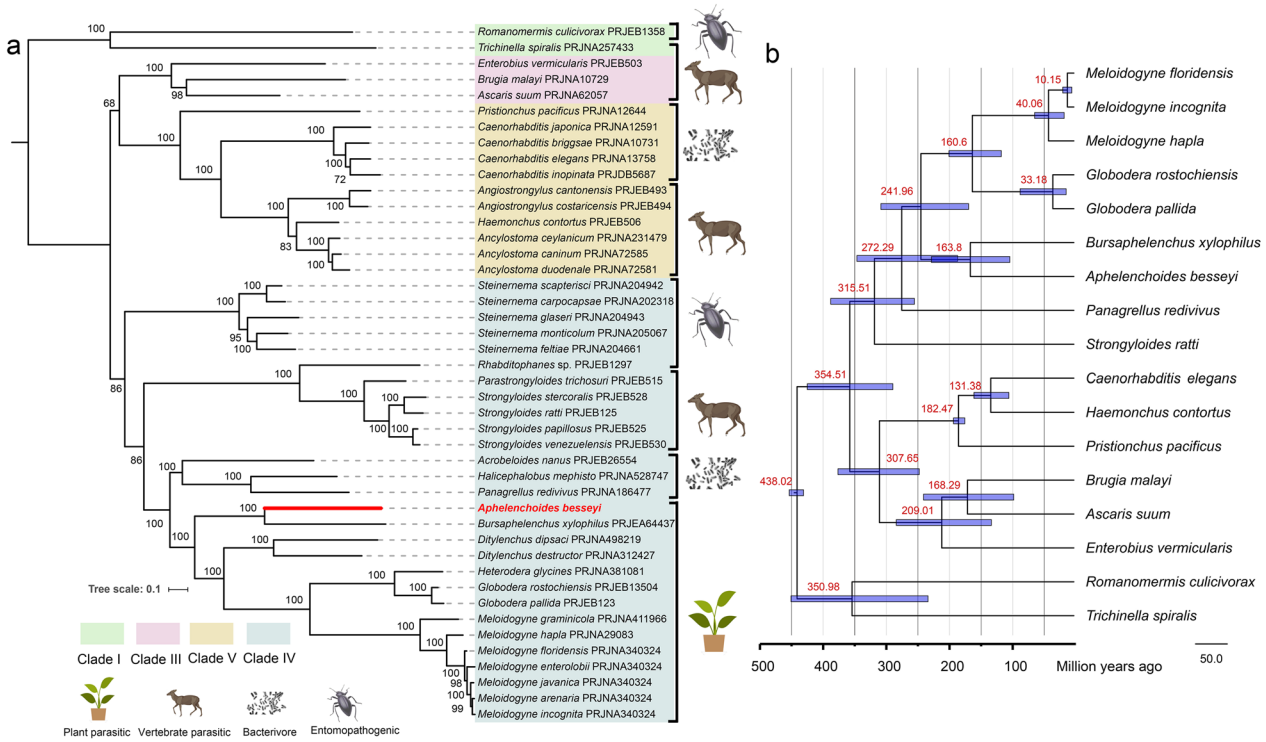

Fig. 2 Phylogenetic placement and molecular dating of *Aphelenchoides besseyi* within Nematoda. **a** The tree was constructed by most single-copy conserved orthologous groups acquired by comparison of protein sequences in 44 species. The bootstrap values are shown at nodes. **b** The molecular dating of *A. besseyi*. The estimated species divergence times (millions of years ago) are indicated as ranges at each branch node

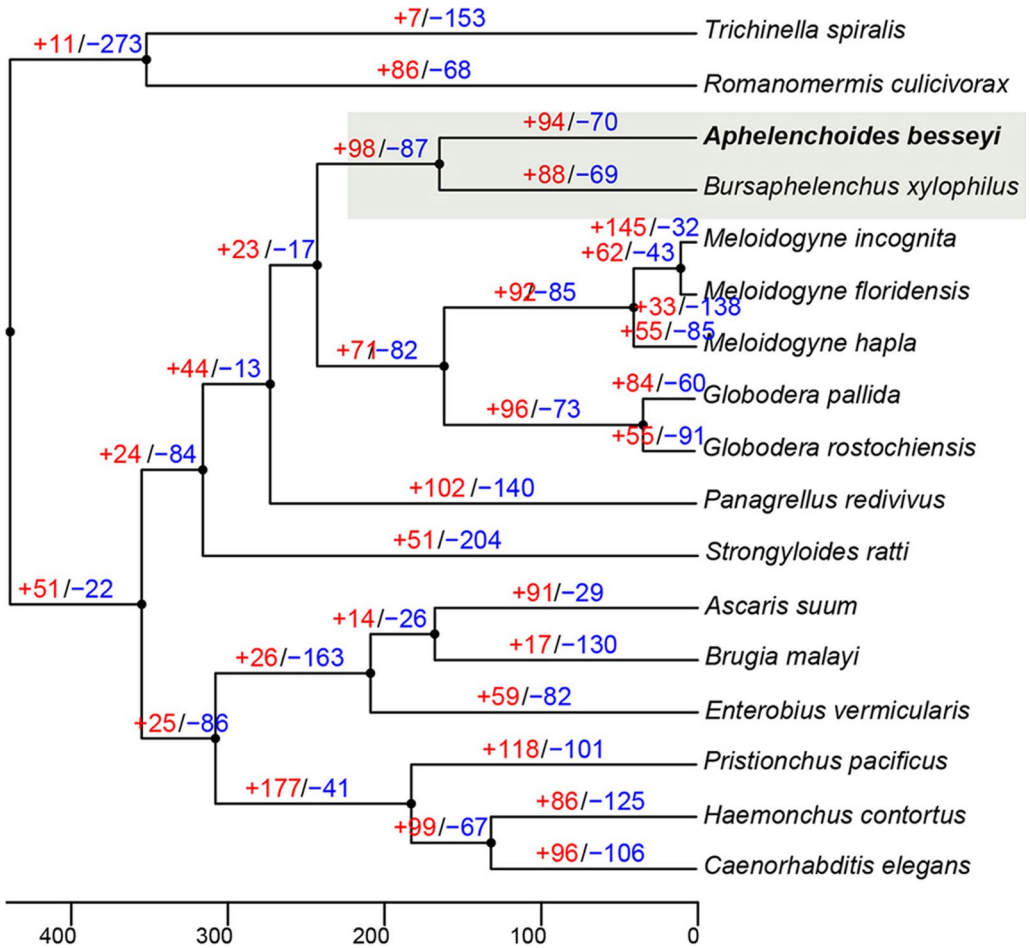

Fig. 3 The analysis for the gene family expansion and contraction. The red and blue colors indicate the number of expanded and contracted gene numbers, respectively

and single-organism process (GO:0044699), in which 947, 942, and 733 genes were included, respectively. There were 495, 495, and 368 genes included in the top three CCs, including the cell (GO:0005623), cell part (GO:0044464), and membrane (GO:0016020), respectively. A total of 1025, 734, and 108 genes were included in the most annotated MFs: catalytic activity (GO:0003824), binding (GO:0005488), and transporter activity (GO:0005215), respectively (Fig. 4a).

For KOG, a total of 10,832 protein-coding genes involving 25 categories were annotated (Additional file 4: Table S4). Among them, 1802 (16.64%) genes were annotated as the general function, which was the most abundant category, followed by 1639 (15.13%) genes assigned in signal transduction mechanisms (Fig. 4b).

KEGG pathway analyses annotated 2834 protein-coding genes (Additional file 4: Table S5), and the main pathways were 'global and overview maps' for metabolism, 'translation' for genetic information processing, 'signal transduction' for environmental information processing, 'transport and catabolism' for cellular processes, and 'aging' for organismal systems (Fig. 5). Further analysis suggested that *A. besseyi* has several different metabolism pathways compared to other related species (Additional file 3: Figures S2–S4). For instance, for vitamin B6 metabolism, *A. besseyi* is similar to *B. xylophilus* in lack of aldehyde oxidase (K00157) but bears threonine synthase (K01733), which is absent in *Pristionchus pacificus*, *C. elegans* and *B. xylophilus*. The potato rot nematode *Ditylenchus destructor* has similar life cycles being both mycophagous and plant parasitic species. In comparison to *D. destructor*, *A. besseyi* is similar to free-living *P. pacificus* and *C. elegans* in missing pyridoxal 5'-phosphate synthase pdxS subunit (K06215) and pyridoxine 5-phosphate synthase (K03474). With respect to biotin, *A. besseyi* is similar to *B. xylophilus* in having 3-oxoacyl-[acyl-carrier protein] reductase (K00059), which is absent in free-living *P. pacificus* and *C. elegans*. In comparison to *D. destructor*, *A. besseyi* lacks 8-amino-7-oxononanoate synthase (K00652), biotin synthase (K01012),

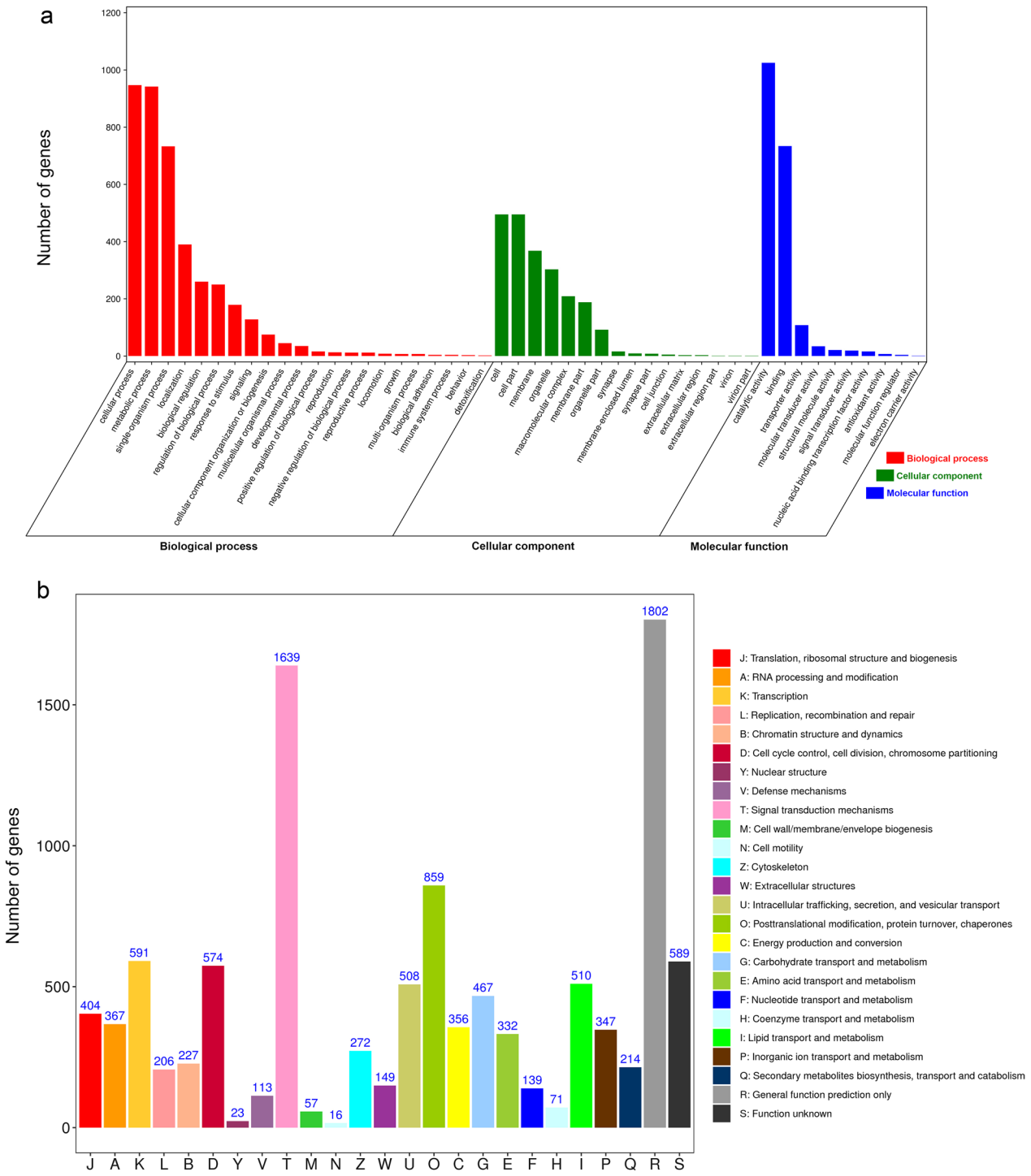

Fig. 4 Functional annotations of the *Aphelenchoides besseyi* genes. **a** GO functional annotation. **b** KOG functional annotation

and biotin-protein ligase (K01942). The riboflavin metabolism is generally similar to *B. xylophilus,* except riboflavin kinase (K00861) is absent. However, flavin prenyltransferase (K03186), ectonucleotide pyrophosphatase/phosphodiesterase family member 1/3 (K01513), and FAD synthetase (K00953) are present in *A. besseyi*. A total of seven proteins are missing in comparison to *D. destructor*; they are GTP cyclohydrolase II (K01497), diaminohydroxyphosphoribosylaminopyrimidine deaminase (K01498), 5-amino-6-(5-phosphoribosylamino)

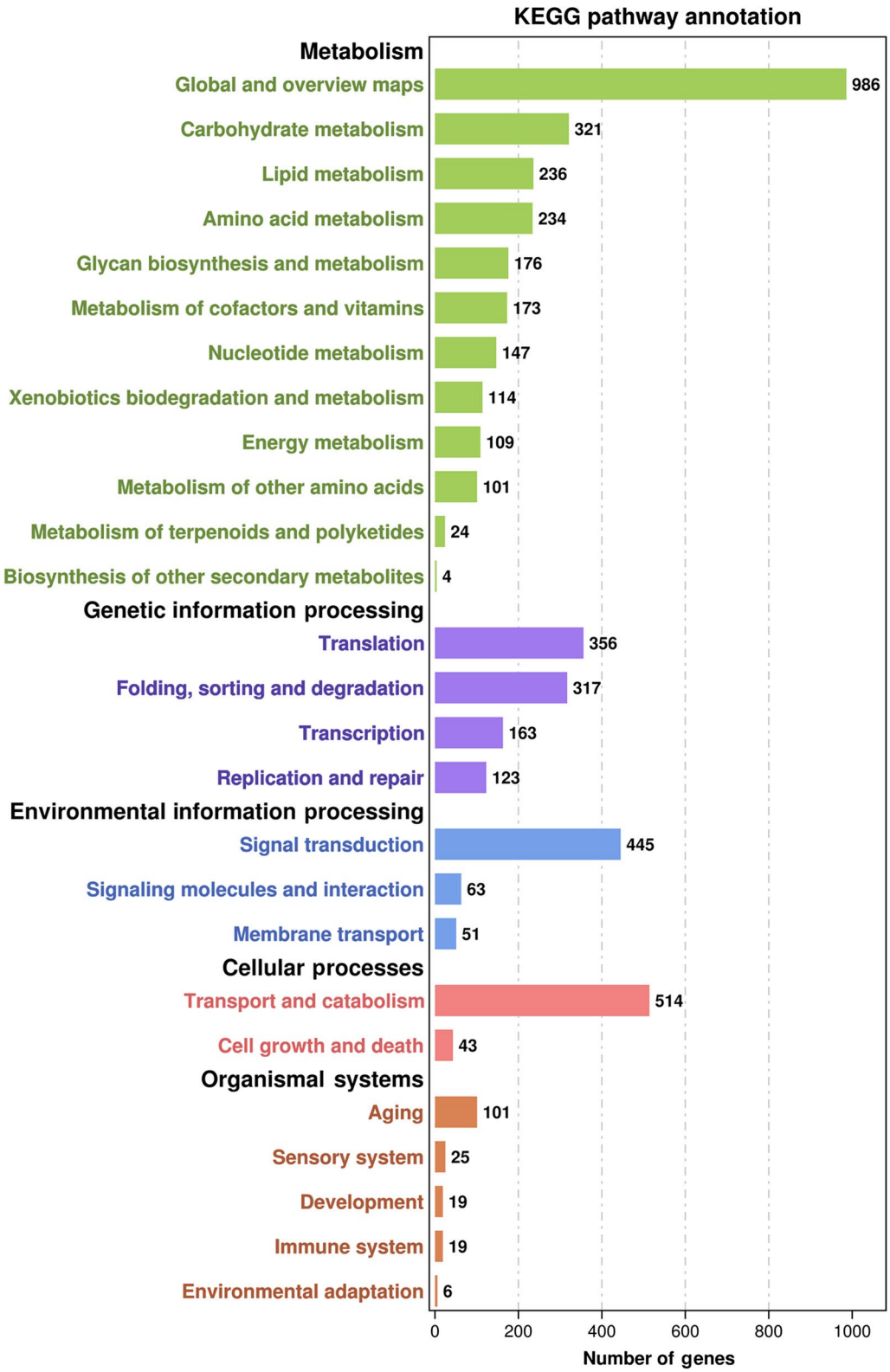

Fig. 5 Functional annotation of the *Aphelenchoides besseyi* genes by KEGG

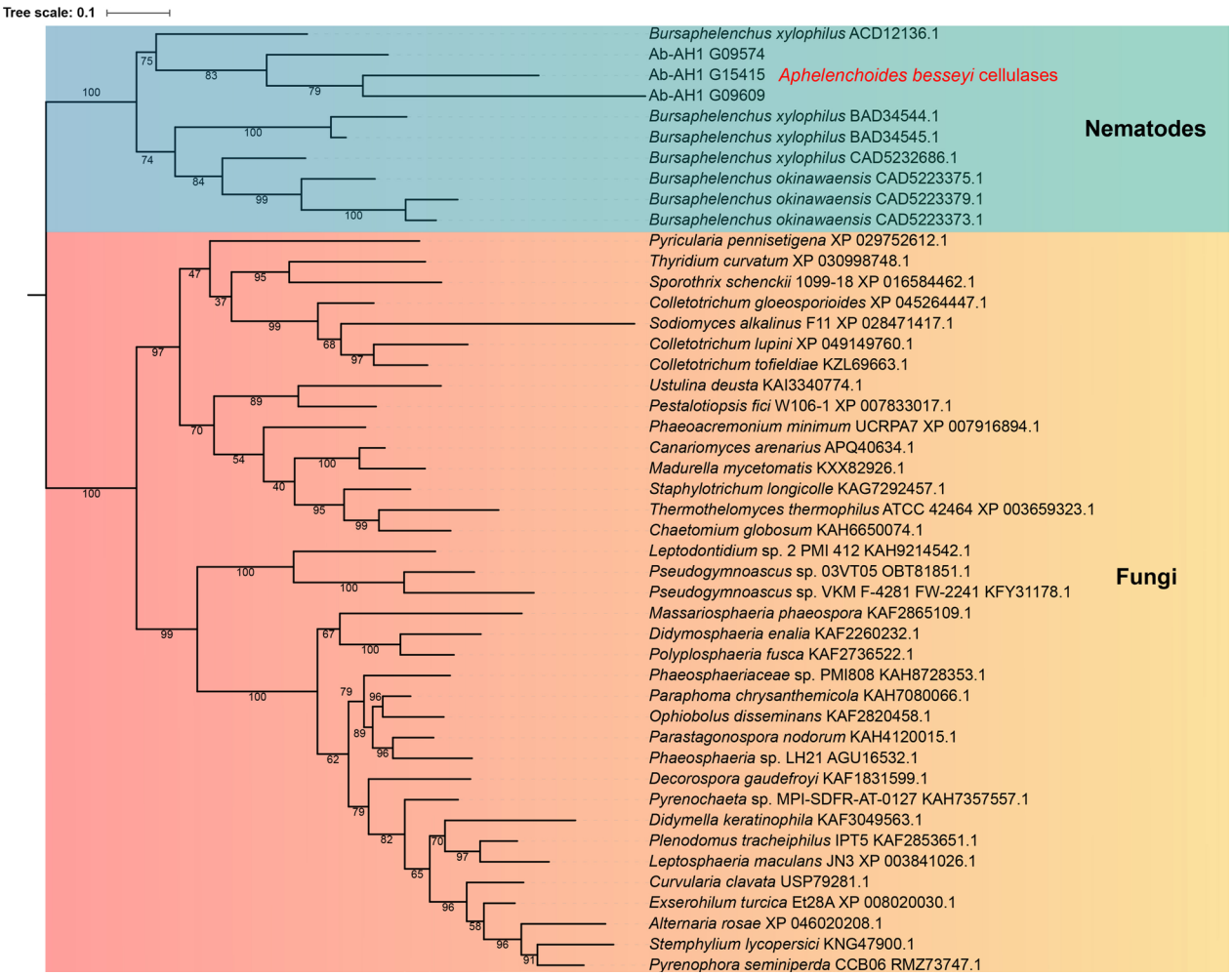

Fig. 6 A phylogenetic tree of cellulases from *Aphelenchoides besseyi* and other organisms. Cellulases in *A. besseyi* and their homologs were used to construct a phylogenetic tree using the model WAG + I + G4

uracil reductase (K00082), 5-amino-6-(5-phospho-D-ribitylamino) uracil phosphatase (K22912), 3,4-dihydroxy 2-butanone 4-phosphate synthase (K02858), riboflavin kinase (K00861), and FMN Hydrolase/5-amino-6-(5-phospho-D-ribitylamino) uracil phosphatase (K20860). Besides, ectonucleotide pyrophosphatase/Phosphodiesterase family member 1/3 (K01513) and flavin prenyltransferase (K03186) were found in *A. besseyi* but not in *D. destructor.*

Gene related to drought tolerance

The survival of the *A. besseyi* is to remain anhydrobiotic in the seed until planting; thus, we suspected a series of drought tolerance/resistance genes were possibly involved. Indeed, we recovered significantly more transcription factors in *A. besseyi* in comparison to other studied species. In particular, there are 83 proteins similar to the LysR family transcriptional regulator (LTTRs) in the bacterium *Bradyrhizobium japonicum*, and 7 proteins are similar to the 12-oxophytodienoate reductase (OxyR) presents in *Oryza sativa* subsp. Japonica. Interestingly, both LTTRs and OxyR are related to transcriptional regulation during the expression of drought tolerance/resistance genes (Additional file 4: Table S8).

Aphelenchoides besseyi horizontally acquired cellulase genes from eukaryotic origin

Cellulose is one of the major components in plant tissues. In this study, we found three cellulase genes in the *A. besseyi* genome, and they are endo-glucanases that belong to the glycosyl hydrolase family 45 (GHF45). When blasted against the NCBI database, the best-hit homologs of *A. besseyi* cellulases match to the pinewood nematode *Bursaphelenchus* species and fungi (Fig. 6). The phylogenetic tree showed that *A. besseyi* and *Bursaphelenchus* cellulases are clustered in one clade, while all fungal cellulases

are in a separate clade. *Aphelenchoides* and *Bursaphelenchus* are closely related genera. Based on limited data, we could not determine if nematodes from the two genera acquired cellulases from the same origin, but it is likely that *A. besseyi* also gains cellulase from the fungal origin as *Bursaphelenchus* species (Kikuchi et al. 2004).

Discussion

Use of long-read sequence technologies to generate genomes in the plant-parasitic nematode

The first PPN was sequenced in 2008 based on the Sanger method using BAC libraries (Abad et al. 2008). Later, with the development of high throughput sequencing technologies and decreasing cost, a growing number of PPNs has been sequenced. Currently, a total of 27 PPN species are genomes available in GeneBank (accessed on 01 July 2022). Among these, approximately half of them were assembled based on short reads generated through the Illumina platform, resulting in highly fragmented contigs, e.g., 17,125 in *Subanguina moxae*, 31,341–34,316 in *Meloidogyne javanica*, 129,028 contigs in *Rotylenchulus reniformis*, and 5944 in *Hoplolaimus columbus* (Takeuchi et al. 2015; Szitenberg et al. 2017; Ma et al. 2021). The poor quality of these draft genomes reduces the reliability of downstream gene annotation and limits further sensitive studies, such as comparative genomics or population genomics at the species level. A typical example is the *A. avenae.* This species is related to *Aphelenchoides* and *Bursaphelenchus* but has nearly three times more annotated genes (43,724 vs. 16,343 in *A. besseyi* and 15,860 in *B. xylophilus*). The assembly of *A. avenae* has 28,772 contigs, which are highly fragmented, including a considerable number of duplications or even contaminations. Therefore, it is difficult to draw any solid conclusion based on the quality of the given dataset (Wan et al. 2021).

The utilization of long-read sequencing technologies, such as PacBio and Nanopore, has greatly advanced our ability to assemble high-quality genomes in animals. With these technologies, obtained nematode genomes can reach a few hundred scaffolds, with an N50 at a level of several Mb, greater consensus accuracy, and a lower degree of sequencing bias (Amarasinghe et al. 2020). In the present study, we demonstrated a hybrid genome sequencing strategy, combining long reads (PacBio) with high-accuracy and low-cost Illumina short reads, which can be used to correct long reads assemblies, and finally obtain a more complete and contiguous genome assembly. This resource will also pave the way for comparative genomics towards pinpointing the evolution of plant parasitism, the genome bases of anhydrobiosis, and the mechanism of reproduction model switch in this plant parasite.

More recently, *A. besseyi* complex was sequenced in an independent study (Lai et al. 2022) during the revision of this manuscript. In that study, they used the hybrid strategy of Illumina HiSeq 2500 to produce 150 bp paired-end reads, PacBio and Nanopore sequencing system to produce long-read, and, more importantly, Hi-C was used to generate chromosome level assembly. The acquired populations of *A. besseyi* have genome sizes ranging from 44.7 to 47.4 Mb, slightly smaller than our sequenced population, and are amongst the smallest in the clade IV. This method can be further used for genome sequencing for other PPNs.

Horizontally acquired cellulases in *A. besseyi*

The acquisition of plant cell-wall degrading genes through HGT is a symbolic event for the evolution of PPNs (John et al. 2005; Haegeman et al. 2011; Kikuchi et al. 2017). Among those enzymes, cellulase exists the most in all known plant parasites and some free-living nematodes. However, cellulases from different GHFs were found in nematodes, and those genes were gained from different origins independently. As cellulases from most plant-parasites are of bacterial origins (Danchin et al. 2010), it has also been shown that fungi are potential donors of cellulases in nematodes, although it is less frequent than bacteria (Haegeman et al. 2011). Here using the complete genome, we showed that *A. besseyi* along with the pine wood nematodes *Bursaphelenchus* species had acquired cellulases from fungal origins, which belong to the GHF 45. This is in agreement with earlier studies in *B. xylophilus* and *A. besseyi* from the pre-sequencing era (Kikuchi et al. 2004, 2014; Palmoares-Ruis et al. 2014). However, due to the lack of data, we are not able to confirm whether *Aphelenchoides* and *Bursaphelenchus* gained cellulases through one HGT event. Along with recent HGT studies in nematodes (Han et al. 2022) and insects (Xia et al. 2021), our data provide new insights into the adaption of animals through HGT.

Conclusion

In this study, we sequenced *A. besseyi* isolated from rice in China using PacBio long reads and Illumina short reads. The assembly consists of 166 scaffolds totaling 50.3 Mb, with an N50 of 1.262 Mb and a maximum scaffold length of 9.17 Mb. A total 16,343 genes were annotated in the genome, with 94 expanded and 70 contracted gene families. Further gene function analysis demonstrated that the transcription factors related to drought tolerance were enriched, and cellulase genes were horizontally acquired from eukaryotic origin.

Methods

Nematode culture and DNA isolation

A. besseyi was isolated from the infected seeds of *O. sativa* subsp. japonica, cv. AnHui-1 as described in Xie et al. (2019). Nematodes were subsequently cultured on the fungus *Botrytis cinerea* at 25°C using one male and one female for 10 generations. The nematodes were collected in double-distilled water for 24 h before further application. Genomic DNA was extracted from mix stages.

DNA extraction and sequencing

High molecular weight DNA of *A. besseyi* for PacBio sequencing was extracted from *c.a.* 50,000 individuals. DNA quantity and quality were assessed using a Qubit Fluorometer (ThermoFisher, Waltham, MA, USA) and 2100 bioanalyzer (Agilent Technology, Santa Barbara). DNA molecules were ruptured into smaller fragments; BluePippin (Saga Science, Beverly, MA, USA) was used to size select DNA fragments of > 20 kb. Libraries were prepared using the SMRTbell Template Prep Kit-SPv3 following the manufacturer's recommendations. Sequencing was performed on a PacBio Sequel platform at Gendenovo (Guangzhou, China). Illumina libraries were prepared using the Paired-End Sample Prep Kit (Illumina Inc., San Diego, CA) with an insert size of 500 bp. Sequencing was performed on Illumina Novaseq 6000 platform.

De novo assembly

De novo assembly was performed with PacBio long reads using MECAT (Xiao et al. 2017). The parameters '-n 50' for mecat2pw and 'Overlapper = mecat2asmpw' for mecat2canu were used for assembly. Illumina reads were used to correct the PacBio long reads using Pilon (Walker et al. 2014). To evaluate the accuracy of genome assembly and sequencing, Illumina short reads were realigned to genome assembly to obtain statistical indicators, including mapping rate, genome coverage, depth distribution, and homozygous and heterozygous SNP number. Expressed Sequence Tags (ESTs) from *A. besseyi* were aligned to the genome assembly by BLAT software to evaluate the integrity of genome assembly. BUSCO (Simão et al. 2015) pipelines were also performed to evaluate the completeness of genome assembly using nematoda_odb9 (with the number of species = 8).

Gene annotation

A hybrid strategy using transcriptome, homologous, and de novo annotation was adopted to predict gene structure. The de novo prediction was conducted using Augustus (Stanke et al. 2005) and GeneMark (Lukashin and Borodovsky 1998) based on the Hidden Markov Model. References were then used to search for and annotate homologous genes in MAKER (Cantarel et al. 2008). RNA-Seq data were used for prediction by combining hisat2 alignment and StringTie (Pertea et al. 2016) assembly results to obtain predicted gene sets. Finally, MAKER (Cantarel et al. 2008) software was used to integrate the prediction resulted from the three above-mentioned methods to obtain the final gene set.

We annotated genes using NCBI NR, GO, SwissProt, KEGG, and KOG databases. The predicted protein-coding gene sequences are aligned with different databases through BLAST 2.2.29+ (McGinnis et al. 2004) with a threshold of e-value less than 10^{-5} for filtering, and the top 20 hits with the highest score value are selected.

For Non-coding RNAs, rRNAs were predicted using RNAmmer (Lagesen et al. 2007), tRNA were predicted by tRNAscan-SE (Lowe and Eddy 1997), and both sRNA and miRNA were predicted by comparing the Rfam database (V13, http://rfam.xfam.org/).

Tandem repeats finder (Benson 1999) was used to predict tandem repeats. Three prediction methods were used to extract interspersed repeats: (1) based on signature using Repbase (Jurka et al. 2005) through LTR_FINDER (Xu and Wang 2007), Helitroscanner (Xiong et al. 2014), MITE-Hunter (Han and Wessler 2010), and MGEscan-nonLTR (Rho and Tang 2009). (2) construction in de novo method using programs PILER (Edgar et al. 2005), RepeatScout (Price et al. 2005), and RepeatModeler (Flynn et al. 2020); (3) homology construction. RepeatMasker (Chen 2004) software was used to predict the repeat sequences based on the constructed repeat sequence database in structure prediction (signature) and de novo prediction.

Phylogeny, molecular dating, and gene function analysis

For phylogenetic analysis, the ortholog gene identification and clustering were performed using OrthoFinder (Emms and Kelly 2019). MAFFT (Katoh and Standley 2013) was used to align amino acid sequences in each orthogroup. Aligned sequences were concatenated, and a maximum-likelihood Species tree was constructed using IQ-TREE (Nguyen et al. 2015) using 1000 bootstrap replications.

The divergence time between *A. besseyi* and 16 other nematodes was estimated using the MCMCtree program implemented in PAML (Yang 2007). Calibration time was obtained from the TimeTree database (http://www.timetree.org/). Gene family expansion and contraction were determined using CAFÉ (De Bie et al. 2006) based on gene family changes in the inferred phylogenetic history. Two methods were employed for function prediction. For the gene families, the GO

terms were obtained through BLAST2GO (Conesa et al. 2005) searching against NCBI non-redundant database and using the Gene Set Enrichment Analysis tool in WormBase7 (https://wormbase.org).

To further analyze drought-related genes, the reference database was built using all available drought-resistant and drought-tolerant genes in UniProt (https://www.uniprot.org/). The annotated *A. besseyi* gene, together with other species, was used as a query. The genes were extracted using the BLASTP search for those showing > 50% similarity with > 30% identity.

Analysis of cellulase genes

To search for potential cellulase genes, multiple known nematode cellulases (Han et al. 2022) were used to BLAST against the *A. besseyi* genome. DIAMOND blastp with the '–more-sensitive' option was used and resulted in five hits from the *A. besseyi* genome (Buchfink et al. 2021). These genes were manually examined in SMART (http://smart.embl-heidelberg.de/), and only three of them contain a cellulase domain, which belongs to the glycoside hydrolase (GH) family 45.

To investigate the potential origin of cellulase genes in *A. besseyi*, we first searched for homologs of the *A. besseyi* cellulase with the domain amino sequences in the NCBI non-redundant database using the BLASTp algorithm. Matching sequences with *e* values less than 1.25×10^{-87} were collected, and pre-existing homologs from *Aphelenchoides* were manually removed. These sequences were clustered using a 90% identity threshold through cd-hit (Li and Godzik 2006), and the remaining 43 and three *A.besseyi* sequences from this study were aligned using MAFFT (Katoh and Standley 2013). IQ-TREE was used to construct phylogenetic trees (Nguyen et al. 2015). A total of 541 substitution models were tested with 1000 ultrafast bootstraps (Kalyaanamoorthy et al. 2017; Hoang et al. 2018). Based on Bayesian Information Criterion, WAG + I + G4 was identified as the best-fit model for the given data.

Abbreviations

BP	Biological process
CC	Cellular component
ESTs	Expressed sequence tags
GH	Glycoside hydrolase
GHF45	Glycosyl hydrolase family 45
GO	Gene ontology
HGT	Horizontal gene transfer
KEGG	Kyoto encyclopedia of genes and genomes
KOG	Eukaryotic orthologous groups
LTTRs	LysR family transcriptional regulator
MF	Molecular function
PPN	Plant-parasitic nematode
TE	Transposable elements

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42483-023-00158-0.

Additional file 1: Table S1. The statistics of predicted miRNA in *Aphelenchoides besseyi*. **Table S2.** The statistics of predicted rRNA in *Aphelenchoides besseyi*. **Table S3.** The statistics of predicted snRNA in *Aphelenchoides besseyi*. **Table S4.** The statistics of predicted tRNA in *Aphelenchoides besseyi*.

Additional file 2. The tandem repeats and transposable elements in the genome of *Aphelenchoides besseyi*.

Additional file 3: Figure S1. Gene length distribution of *Aphelenchoides besseyi*. **Figure S2.** The metabolism pathway for vitamin B6. **Figure S3.** The metabolism pathway for biotin. **Figure S4.** The metabolism pathway for riboflavin.

Additional file 4: Table S1. The expanded gene family and corresponding gene family. **Table S2.** The annotation for the expanded genes. **Table S3.** Gene function prediction: GO analysis-Cellular component/molecular function/biological process. **Table S4.** Gene function prediction: KOG analysis. **Table S5.** Annotated KEGG pathway. **Table S6.** The annotated genes by NR search. **Table S7.** The annotated genes by SwissProt. **Table S8.** The comparison of genes related to tolerance/resistance in different nematode species.

Acknowledgements

Not applicable.

Authors' contributions

YLP and HLJ designed the research; JLX, FY, and WJY prepared the materials, XQ, ZDH, HLJ, and JLX analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (32001876) and China Agriculture Research System (Grant No. CARS-01-41).

Availability of data and materials

The datasets generated and/or analysed during the current study are available in the NCBI repository, https://www.ncbi.nlm.nih.gov/nuccore/?term=PRJNA901680.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 20 October 2022 Accepted: 16 January 2023
Published online: 31 January 2023

References

Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EG, Deleury E, et al. Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. Nat Biotechnol. 2008;26(8):909–15. https://doi.org/10.1038/nbt.1482.

Allen MW. Taxonomic status of the bud and leaf nematodes related to *Aphelenchoides fragariae* (Ritzema Bos, 1891). Proc Helminthol Soc Wash. 1952;19:108–20.

Amarasinghe SL, Su S, Dong X, Zappia L, Ritchie ME, Gouil Q. Opportunities and challenges in long-read sequencing data analysis. Genome Biol. 2020;21:30. https://doi.org/10.1186/s13059-020-1935-5.

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

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

Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. 2008;18:188–96. https://doi.org/10.1101/gr.6743907.

Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr Protoc Bioinform. 2004;5:4–10. https://doi.org/10.1002/0471250953.bi0410s25.

Christie JR. A description of *Aphelenchoides besseyi* n. sp., the summer dwarf nematode of strawberries, with comments on the identity of *Aphelenchoides subtenuis* (Cobb, 1929) and *Aphelenchoides hodsoni* Goodey, 1935. Proc Helminthol Soc Wash. 1942;9:82–4.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6. https://doi.org/10.1093/bioinformatics/bti610.

Danchin EGJ, Rosso MN, Vieira P, Almeida-Engler J, Coutinho P, Henrissat B, et al. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. PNAS. 2010;107:17651–6. https://doi.org/10.1073/pnas.1008486107.

Dayi M, Sun S, Maeda Y, Tanaka R, Yoshida A, Tsai IJ, et al. Nearly complete genome assembly of the pinewood nematode *Bursaphelenchus xylophilus* strain Ka4C1. Microbiol Resour Announc. 2020;9:e01002-e1020. https://doi.org/10.1128/MRA.01002-20.

De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. Bioinformatics. 2006;22:1269–71. https://doi.org/10.1093/bioinformatics/btl097.

Duncan LW, Moens M. Migratory endoparasitic nematodes. In: Perry RN, Moens M, editors. Plant nematology. 2nd ed. Wallingford: CAB International; 2013. p. 144–78.

Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. Bioinformatics. 2005;21:152–8. https://doi.org/10.1093/bioinformatics/bti1003.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:1–14. https://doi.org/10.1186/s13059-019-1832-y.

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

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

Han Y, Wessler SR. MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. Nucleic Acids Res. 2010;38:e199. https://doi.org/10.1093/nar/gkq862.

Han Z, Sieriebriennikov B, Susoy V, Lo WS, Igreja C, Dong C, et al. Horizontally acquired cellulases assist the expansion of dietary range in *Pristionchus* nematodes. Mol Biol Evol. 2022;39:msab370. https://doi.org/10.1093/molbev/msab370.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018;35:518–22. https://doi.org/10.1093/molbev/msx281.

Huang CS, Huang SP, Chiang YC. Mode of reproduction and sex ratio of rice white-tip nematode *Aphelenchoides besseyi*. Nematologica. 1979;25:255–60. https://doi.org/10.1163/187529279x00271.

Hunt DJ. A checklist of the Aphelenchoidea (Nematoda:Tylenchina). J Nematode Morphol Syst. 2008;10:99–135.

John JT, Furlanetto C, Kikuchi T. Horizontal gene transfer from bacteria and fungi as a driving force in the evolution of plant parasitism in nematodes. Nematology. 2005;7:641–6. https://doi.org/10.1163/156854105775142919.

Jones JT, Haegeman A, Danchin EG, Gaur HS, Helder J, Jones MG, et al. Top 10 plant-parasitic nematodes in molecular plant pathology. Mol Plant Pathol. 2013;14:946–61. https://doi.org/10.1111/mpp.12057.

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. Repbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 2005;110:462–7. https://doi.org/10.1186/s13100-015-0041-9.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017;14:587–9. https://doi.org/10.1038/nmeth.4285.

Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80. https://doi.org/10.1093/molbev/mst010.

Kikuchi T, Jones JT, Aikawa T, Kosaka H, Ogura N. A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode *Bursaphelenchus xylophilus*. FEBS Lett. 2004;572:201–5. https://doi.org/10.1016/j.febslet.2004.07.039.

Kikuchi T, Cotton JA, Dalzell JJ, Hasegawa K, Kanzaki N, McVeigh P, et al. Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. PLoS Pathog. 2011;7:e1002219. https://doi.org/10.1371/journal.ppat.1002219.

Kikuchi T, Cock PJA, Helder J, Jones JT. Characterisation of the transcriptome of *Aphelenchoides besseyi* and identification of a GHF 45 cellulase. Nematology. 2014;16:99–107. https://doi.org/10.1163/15685411-00002748.

Kikuchi T, Eves-van den Akker S, Jones JT. Genome evolution of plant-parasitic nematodes. Annu Rev Phytopathol. 2017;55:333–54. https://doi.org/10.1146/annurev-phyto-080516-035434.

Lagesen K, Hallin P, Rødland EA, Stærfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 2007;35:3100–8. https://doi.org/10.1093/nar/gkm160.

Lai CK, Lee YC, Ke HM, Lu MR, Liu WA, Lee HH, et al. The *Aphelenchoides* genomes reveal major events of horizontal gene transfers in clade IV nematodes. Biorxiv. 2022. https://doi.org/10.1101/2022.10.13.512134.

Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics. 2006;22:1658–9. https://doi.org/10.1093/bioinformatics/btl158.

Lin M, Ding X, Wang Z, Zhou F, Lin N. Description of *Aphelenchoides besseyi* from abnormal rice with 'small grains and erect panicles' symptom in China. Rice Sci. 2005;12:289–94.

Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64. https://doi.org/10.1093/nar/25.5.955.

Lukashin AV, Borodovsky M. GeneMark hmm: new solutions for gene finding. Nucleic Acids Res. 1998;26:1107–15. https://doi.org/10.1093/nar/26.4.1107.

Ma X, Baeza JA, Richards VP, Agudelo P. First genomic resource of the Columbia lance nematode *Hoplolaimus columbus*. Phytopathology. 2021;111:2396–8. https://doi.org/10.1094/PHYTO-12-20-0536-A.

McGinnis S, Madden TL. BLAST: at the core of a powerful and diverse set of sequence analysis tools. Nucleic Acids Res. 2004;32:20–5. https://doi.org/10.1093/nar/gkh435.

Nandini GN, Mathur VK, Ramasundaram P, Sabesh M. Reproductive variations in *Aphelenchoides besseyi* populations. Indian J Nematol. 2001;31:115–9.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74. https://doi.org/10.1093/molbev/msu300.

Oliveira CJ, Subbotin SA, Álvarez-Ortega S, Desaeger J, Brito JA, Xavier KV, 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. 2019;103:2825–42. https://doi.org/10.1094/PDIS-04-19-0752-RE.

Palomares-Rius JE, Hirooka Y, Tsai IJ, Masuya H, Hino A, Kanzaki N, et al. Distribution and evolution of glycoside hydrolase family 45 cellulases in nematodes and fungi. BMC Evol Biol. 2014;14:69. https://doi.org/10.1186/1471-2148-14-69.

Perry RN, Moens M. Plant nematology. Cham: Centre Agriculture Bioscience International Publishing; 2013.

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. Transcript-level expression analysis of RNA-seq experiments with HISAT. StringTie and Ballgown Nat Protoc. 2016;11:1650–67. https://doi.org/10.1038/nprot.2016.095.

Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. Bioinformatics. 2005;21:351–8. https://doi.org/10.1093/bioinformatics/bti1018.

Rho M, Tang H. MGEScan-non-LTR: computational identification and classification of autonomous non-LTR retrotransposons in eukaryotic genomes. Nucleic Acids Res. 2009;37:e143. https://doi.org/10.1093/nar/gkp752.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210–2. https://doi.org/10.1093/bioinformatics/btv351.

Stanke M, Morgenstern B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. 2005;33:465–7. https://doi.org/10.1093/nar/gki458.

Subbotin S, De Oliveira C, Álvarez-Ortega S, Desaeger J, Crow W, Overstreet C, 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;23:1–33. https://doi.org/10.1163/15685411-bja10048.

Sun S, Shinya R, Dayi M, Yoshida A, Sternberg PW, Kikuchi T. Telomere-to-telomere genome assembly of *Bursaphelenchus okinawaensis* strain SH1. Microbiol Resour Announc. 2020;9(43):e01000-e1020. https://doi.org/10.1128/MRA.01000-20.

Szitenberg A, Salazar-Jaramillo L, Blok VC, Laetsch DR, Joseph S, Williamson VM, et al. Comparative genomics of apomictic root-knot nematodes: hybridization, ploidy, and dynamic genome change. Genome Biol Evol. 2017;9:2844–61. https://doi.org/10.1093/gbe/evx201.

Takeuchi T, Yamaguchi M, Tanaka R, Dayi M, Ogura N, Kikuchi T. Development and validation of SSR markers for the plant-parasitic nematode *Subanguina moxae* using genome assembly of Illumina pair-end reads. Nematology. 2015;17:515–22. https://doi.org/10.1163/15685411-00002885.

Tiwari SP, Khare MN. White tip caused by *Aphelenchoides besseyi*, an important seed-borne disease of rice. In: Trivedi PC, editor. Advances in Nematology. India: Scientific Publishers; 2003. p. 103–14.

Tulek A, Cobanoglu S. Distibution of the rice white tip nematode, *Aphelenchoides besseyi*, in rice growing areas in the Thrace region of Turkey. Nematol Mediterr. 2010;38:215–7.

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

Wan X, Saito JA, Hou S, Geib SM, Yuryev A, Higa LM, et al. The *Aphelenchus avenae* genome highlights evolutionary adaptation to desiccation. Commun Biol. 2021;4:1–8. https://doi.org/10.1038/s42003-021-02778-8.

Wu S, Gao S, Wang S, Meng J, Wickham J, Luo S, et al. A reference genome of *Bursaphelenchus mucronatus* provides new resources for revealing its displacement by pinewood nematode. Genes (basel). 2020;11:570. https://doi.org/10.3390/genes11050570.

Xia J, Guo Z, Yang Z, Han H, Wang S, Xu H, et al. Whitefly hijacks a plant detoxification gene that neutralizes plant toxins. Cell. 2021;184:3588. https://doi.org/10.1016/j.cell.2021.02.014.

Xiao CL, Chen Y, Xie SQ, Chen KN, Wang Y, Han Y, et al. MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. Nat Methods. 2017;14:1072–4. https://doi.org/10.1038/nmeth.4432.

Xie J, Yang F, Wang Y, Peng Y, Ji H. Studies on the efficiency of different inoculation methods of rice white-tip nematode *Aphelenchoides besseyi*. Nematology. 2019;21:673–8. https://doi.org/10.1163/15685411-00003244.

Xiong W, He L, Lai J, Dooner HK, Du C. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. Proc Natl Acad Sci USA. 2014;111:10263–8. https://doi.org/10.1073/pnas.1410068111.

Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. 2007;35:265–8. https://doi.org/10.1093/nar/gkm286.

Xu X, Qing X, Xie JL, Yang F, Peng YL, Ji HL. Population structure and species delimitation of rice white tip nematode, *Aphelenchoides besseyi* (Nematoda: Aphelenchoididae), in China. Plant Pathol. 2020;69:159–67. https://doi.org/10.1111/ppa.13113.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24:1586–91. https://doi.org/10.1093/molbev/msm088.

Yokoo T. *Aphelenchoides oryzae* Yokoo n. sp., a nematode parasitic to rice plant. Jpn J Phytopathol. 1948;13:40–3. https://doi.org/10.3186/JJPHYTOPATH.13.40.