

國立臺灣大學生命科學院基因體與系統生物學學位學程



博士論文

Genome and Systems Biology Degree Program

College of Life Science

National Taiwan University

Doctoral Dissertation

比較基因體學區分葉芽線蟲複合種並探討線蟲水平
轉移基因

Comparative genomics delimit *Aphelenchoides*
besseyi species complex and pinpoint horizontal gene transfer
events in clade IV nematodes

賴政國

Cheng-Kuo Lai

指導教授：蔡怡陞 博士

Advisor: Isheng Jason Tsai, Ph.D.

中華民國111年 11 月

November 2022

國立臺灣大學博士學位論文

口試委員會審定書

PHD DISSERTATION ACCEPTANCE CERTIFICATE
NATIONAL TAIWAN UNIVERSITY

比較基因體學區分葉芽線蟲複合種並探討線蟲水平轉移基因
Comparative genomics delimit *Aphelenchoides besseyi* species
complex and pinpoint horizontal gene transfer events in clade IV
nematodes

本論文係賴政國D05B48005在國立臺灣大學基因體與系統生物學
學位學程完成之博士學位論文，於民國111年11月14日承下列考
試委員審查通過及口試及格，特此證明。

The undersigned, appointed by the Genome and Systems Biology Degree Program
on 14 (date) 11 (month) 2022 (year) have examined a PhD dissertation entitled above
presented by Cheng-Kuo Lai D05B48005 candidate and hereby certify that it is worthy of
acceptance.

口試委員 Oral examination committee:

蔡怡陞 莊樹諤

(指導教授 Advisor)

王忠信 薛雁冰 陳佩臻

系主任/所長 Director: 鄭石通

致謝

關於這份論文，首先要感謝我的指導教授 Jason，他給予精闢的研究方向指引，並且修正了我許多文字及文法上的錯誤。我也要感謝我的論文口試委員們對於論文的指教，在口委們的幫助下我修正了許多不合理以及錯誤的地方，特別要感謝王忠信博士，他找出了我初稿許多的錯誤。另外，我要感謝口委之一的陳珮臻博士，他不僅提供了葉芽線蟲，還修正論文裡葉芽線蟲不精確的地方。此外，我要感謝我的口委之一薛雁冰博士，不僅給予了我論文寶貴意見，也在博士班 Rotation 過程中讓我學到許多。我也要感謝實驗室夥伴宜謙，從線蟲培養到線蟲的定序資料大多都是仰賴於她。我也感謝阿棉幫忙定序葉芽 Pacbio 以及 Hi-C，謝謝育菁及 Rubie 對於葉芽定序資料上傳以及葉芽註解校正的貢獻。也謝謝 Caroline 植物及線蟲手繪圖片的支援。

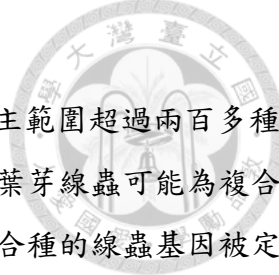
在博士班求學過程中，我要特別感謝我的指導教授 Jason，在研究卡關時您總是能用最快的速度帶我走出困境，當我覺得這份研究資料差不多就這樣時您也總是能從不一樣的角度再為其重新注入生命，同時您也給予我極大的包容，不管是研究進度卡關或是我各種粗心，您都適時的指正並且還保有我自由發揮的空間，謝謝您不管是研究上以及為人處世的各種寶貴教誨。另外，我也要感謝實驗室的大掌櫃阿棉，在研究上，你幫忙了我許多並且給予我很多寶貴的意見，你也總是偷塞許多食物給我，讓我有時懶得出門買東西時至少有食物裹腹。

我還要感謝實驗室其他的夥伴。率先去業界發大財的昕翰，擁有教科書級的 python 編寫能力，研究上的交流真的使我受益良多，也許哪一天我們真的會一起用虛幻引擎打造邊玩遊戲邊基因註解的軟體（誤）。也感謝套牢小夥伴 Tom，昕翰去業界後股市及幣圈的大小消息只有你懂，於是你也離開實驗室後我幫實驗室夥伴們科普了幾堂課。也感謝股市神童宜謙，你的正 500% 未實現損益讓我們歎為觀止，沒有你那些寶貴的葉芽資料支援，我真的無法畢業。我還感謝育菁各種行政上的支援，這些繁雜的事情總是能很放心的交給你去處理，也感謝你各種活動邀約，你與 Daphne 的爽朗為我們宅宅實驗室灑入了那們一點陽光氣息。我還感謝 Rubie，雖然你與宜謙一直在挑食界中爭輸贏，但你們之間的爭霸賽娛樂了我們好幾個中午，也感謝你三不五時路過我座位時來跟我話家常，沒有那些五四三可能當天我連一句話都沒開口。也謝謝培峰，你是實驗室禮貌冠軍，只有你會

提醒我居然還是個學長，於走廊擦肩而過時總是能獲得你恭敬地招呼，三不五時健身房的偶遇，我都怕我突然地打招呼會讓你神經繃緊而發生意外。也謝謝 Jason 為實驗室舉辦的各種活動，雖然台啤形狀聞森笛曾獲選最爛禮物，但我還是覺得那其實是一個實用的禮物。謝謝實驗室這些年的人、事、物，小到日常不知道吃什麼就選國王便當或久賀，大到為研究上山下海，這些點點滴滴會是我珍貴的回憶。謝謝你們和我一起走過的這段旅程。

最後，我要感謝我的父母，他們一直以來都給予我很多的愛和支持。他們讓我無後顧之憂地在學習和生活上自由地發展，謝謝你們的所有付出。

中文摘要



葉芽線蟲為一種植物寄生線蟲屬於 Aphelenchoidea 總科，宿主範圍超過兩百多種植物，研究顯示不同植物分離的葉芽對於宿主的範圍有所分別，證實葉芽線蟲可能為複合種。論文第一章裡，我們研究葉芽線蟲複合種的不同，四隻屬於此複合種的線蟲基因被定序組裝成基因體，44-47 Mb 大小的基因體幾乎為 clade IV 線蟲中最小的存在，我們發現葉芽線蟲基因體變小主要是因為轉座子(Transposable elements)在近期演化過程中快速消失有關，隨著葉芽基因體的解序，複合種可以成功的被分成 *A. oryzae* 以及 *A. pseudobesseyi*，並且能在 28S 分子序列得到驗證，僅存的三條染色體顯示葉芽線蟲基因體經歷過染色體融合。

論文第二章，我們探討葉芽線蟲染色體的演化，同線性的基因體研究證實葉芽的三個染色體可能源自於多種染色體融合及裂解，此外我們找出葉芽線蟲的染色體同源片段但雄性葉芽基因序列回貼結果顯示葉芽性別決定機制可能不由單一性染色體決定。我們發現葉芽複合種染色體上重複序列的密集程度也有很大的分別，顯示此複合種正持續分化中。

論文第三章中，植物線蟲生活型態的轉變可能由水平轉移基因造成，因此我們研究線蟲水平轉移基因的演化過程，我比較了 Clade IV 線蟲的水平轉移基因發現目前植物線蟲的水平轉移基因主要是從兩個演化分支點而來。這些基因大多從細菌轉移而來並且隨著時間推進演化出不同基因數量。水平轉移事件結合物種演化樹使我們能夠了解線蟲水平轉移基因的演化。

ABSTRACT

Aphelenchoides besseyi is a plant-parasitic nematode (PPN) in the Aphelenchoidea lineage that can infect nearly 200 plant species. Research has shown that nematode strains isolated from different plants exhibit varying host range, suggesting that they may be species complex.

In chapter 1, I investigated the differences within *A. besseyi* species complex. I generated the assemblies belonging to *A. besseyi* species complex. The assemblies of *Aphelenchoides* ranged from 44.7-47.4 Mb which is amongst the smallest in the clade IV nematodes. This genome reduction was mainly due to the rapid reduction in transposable elements. Phylogenomic analysis successfully delimited the species complex strains into *A. oryzae* and *A. pseudobesseyi* which was consistent with the 28S phylogeny.

In chapter 2, we investigated the chromosome evolution in *A. besseyi*. Synteny analyses between nematodes suggested that the three chromosomes in *A. besseyi* may be a result of multiple fission and fusion events. In addition, features enriched in nematode sex chromosome were identified in *A. besseyi* despite the male reads coverage of *A. besseyi* were even amongst the three chromosomes suggesting it might possess stochastic sex determination system similar to *Bursaphelenchus* species. To investigate the differences within *A. besseyi* species complex, I identified their differential repeat abundance along chromosomes, indicating ongoing genetic differentiation.

In chapter 3, the acquirement of horizontal gene transfer (HGT) genes was proposed to lead the lifestyle changes of PPNs from free-living. We sought to pinpoint the HGT evolution in PPNs. I compared the inferred HGT families across clade IV nematodes, revealing that HGT genes retained in PPNs were mostly acquired from two major episodes. These genes were mainly originated from bacteria but differentially lost between clades. The combination of HGT events and species phylogeny allowed us to pinpoint the HGT evolution in nematodes.

TABLE OF CONTENTs



致謝.....	ii
中文摘要.....	iv
ABSTRACT.....	v
TABLE OF CONTENTs	vi
LIST OF FIGUREs.....	vii
LIST OF TABLEs.....	ix
CHAPTER 1.	1
1.1 Introduction.....	1
1.2 Materials and Methods.....	2
1.3 Results	6
1.4 Discussion.....	11
CHAPTER 2.	13
2.1 Introduction.....	13
2.2 Materials and Methods.....	13
2.3 Results	14
2.3.6 Discussion.....	16
CHAPTER 3.	18
3.1 Introduction.....	18
3.2 Materials and Methods.....	19
3.3 Results	20
3.4 Discussion.....	24
Figures.....	26
Tables	58
References	69

LIST OF FIGURES

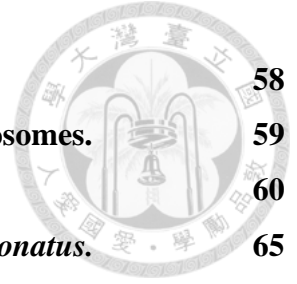
Figure 1	The signal of HiC reads in <i>A. pseudobesseyi</i> VT were curated using Juicerbox.	26
Figure 2	The polyploid prediction of <i>A. fujianensis</i> Dali generated using Smudge	27
Figure 3	Transposable elements in nematodes.	28
Figure 4	Repeat distribution between <i>A. pseudobesseyi</i> VT and <i>A. oryzae</i> RT. The	29
Figure 5	Species phylogeny and genome features of 27 representative nematodes.	30
Figure 6	Separation of <i>Aphelenchoides</i> nematodes by using DA2 and DA3 of 28S.	31
Figure 7	The median similarity of nucleotide coding sequence and protein within the <i>A. besseyi</i> species complex.	32
Figure 8	Prediction of history effective population size of <i>A. oryzae</i> RT and <i>A. pseudobesseyi</i> VT.	33
Figure 9	Heterozygosity of <i>A. oryzae</i> RT and <i>A. pseudobesseyi</i> VT strains.	34
Figure 10	Copy number distribution of Pfam and CAZyme.	35
Figure 11	CAZyme abundance in nematodes.	36
Figure 12	Chromosome evolution of <i>A. besseyi</i> .	37
Figure 13	Nigon elements of rhabditid nematodes in <i>C. elegans</i> .	38
Figure 14	Nigon elements of rhabditid nematodes in <i>B. xylophilus</i> .	39
Figure 15	Nigon elements of rhabditid nematodes in <i>A. pseudobesseyi</i> VT.	40
Figure 16	The male Illumina reads of APVT were normalized.	41
Figure 17	The synteny relationship within the <i>A. besseyi</i> species complex.	42
Figure 18	Meiosis and telomeric maintenance genes in PPNs.	43
Figure 19	HGT families in PPNs.	44
Figure 20	Phylogenetic relationship between GH5 copies in Clade IV nematodes and their candidate bacterial origin.	45
Figure 21	Total number of HGT events and genes in Clade IV nematodes.	46
Figure 22	Two HGT families identified to the genes transferred from plants.	47
Figure 23	The representative terms of HGT genes originated from plants in <i>B. mucronatus</i> .	48
Figure 24	The representative terms of 597 families were gained in Clade IV.	49
Figure 25	The representative terms of 60 families were lost in Clade IV.	50
Figure 26	HGT families occurred time points among Clade IV nematodes.	51

Figure 27 Phylogeny of GH16 in Clade IV nematodes.	52
Figure 28 Phylogenetic relationship between GH32 copies in PPNs and <i>Panagrolaimomorpha</i> nematodes.	53
Figure 29 Phylogenetic relationship between GH43 copies in PPNs and <i>Panagrolaimomorpha</i> nematodes.	54
Figure 30 Phylogenetic relationship between PL3 copies in PPNs and their donor origin copies.	55
Figure 31 GH45 relationship between Aphelenchoidea nematodes and their candidate	56
Figure 32 Different levels of HGT families having HGT genes.	57



LIST OF TABLES

Table 1	Reads information of <i>Aphelenchoides</i> used in this study.	58
Table 2	<i>Bursaphelenchus xylophilus</i> telomere identified in four chromosomes.	59
Table 3	Genome statistic of 27 representative nematodes.	60
Table 4	GO enrichment of 27 families acquired from plants in <i>B. mucronatus</i>.	65
Table 5	HGT enriched Pfam in two major nematodes HGT episodes.	67



CHAPTER 1.



Comparative genomics of *Aphelenchoides besseyi* species complex

Note: the information contained in this thesis may be found, in an alternative form, in a recently submitted study¹

1.1 Introduction

Plant parasitic nematodes (PPNs) have arisen at least on four occasions^{2,3}, which belong to the Tylenchida, Aphelenchoidea, Triplonchida and Dorylamide orders. The PPNs in Triplonchida and Dorylamide have arisen independently, however, Tylenchida and Aphelenchoidea were proposed to share a common PPN ancestor⁴. Tylenchida causes the most serious agriculture losses, hence several studies have been focused on this order especially *Meloidogyne*⁵. Currently, two major orders Trichodoridae and Aphelenchoididae containing PPNs still lack reference genomes.

Here, I focused on *Aphelenchoides besseyi* (*A. besseyi*), a foliar nematode that can infect 200 plant species belonging to 35 genera⁶. This nematode is known for its resistant to extreme environments, making it hard to be controlled. It can enter an anoxybiosis state for seven days, after exposed to anaerobic environment; when the rice kernel dries, it slowly desiccates and can remain viable in the kernel for up to three years. Besides the serious damage in the rice fields, the necrosis syndrome also could be found in other crops, such as soybeans and strawberries^{7,8}. However, the basic genetic and biological knowledge of *A. besseyi* is still unclear.

The limited genetic information and ambiguous morphological features of *A. besseyi* for separation make this species hard to study. It has been reported that *A. besseyi* isolated from different host plants exhibited varying host ranges, suggesting it may be a species complex. Recently, Subbotin *et al.* employed 28S, ITS and mitochondria makers to reclassify this species complex into three different groups: *A. besseyi* from strawberry,

A. oryzae from rice and *A. pseudobesseyi* from fern, revealing genetic differentiation within the complex.

In this chapter, I aimed to address the knowledge gap regarding the evolution of PPNs and identify the genetic differences within *A. besseyi* species complex. Hence, we isolated different *A. besseyi* strains, which were later designated as *A. oryzae* and *A. pseudobesseyi*, and species, and compared them to species in the same genus (*Aphelenchoides bicaudatus* and *Aphelenchoides fujianensis*) from different plants. Next, I sequenced and annotated the genomes of these six strains, and compared them to the genomes of 21 other representative nematodes to identify their diversity at the genome level. Genome reduction in *A. besseyi* species complex may be due to the loss of transposon elements. I observed the dynamic protein families among PPNs.

1.2 Materials and Methods

1.2.1 Species collection

To investigate the genetic differences between members of the *A. besseyi* species complex, two *A. oryzae* strains (RT and RJ) were collected from the rice of Linnei, Nantou and Hiroshima, Japan; two *A. pseudobesseyi* strains (FT and VT) were isolated from bird's-nest fern in Mingjian, Nantou; and *Vallisneria spiralis* was collected from Taichung aquarium. The other two same genus (*A. bicaudatus* Fsh and *A. fujianensis* Dali) strains were isolated from ornamental nurseries of bird's-nest fern in Taiwan and a strawberry field in Dali, Taichung. To identify the species type among these DNA samples, a PCR with a different expected product size between *A. oryzae* (926 bp) and *A. pseudobesseyi* (1,386 bp) was designed (5' TATGTCCGGAGTAAGTATTG and 3' TTAAACGAAAAGAATAAGCG) according to one-to-one ortholog information between APFT and AORT.

1.2.2 28S phylogeny in *Aphelenchoides* genus

To confirm the relationship among members of the *A. besseyi* species complex, 28S sequences of *A. pseudobesseyi* (NCBI accession ID: KY510841.1, MH187564.1, KX356765.1, MT271870.1, MT271869.1 and MT271868.1), *A. oryzae* (NCBI accession ID: KX356775.1, MK880169.1, KX356764.1, KY123694.1, KT692690.1, KX622689.1,

KT692703.1, KX356776.1, KX356774.1, KY123697.1, MT271867.1 and KX356773.1), *Aphelenchoides gorganensis* (NCBI accession ID:KX357652.1) and *Aphelenchoides ritzemabosi* (NCBI accession ID: KX119133.1, KX356837.1 and KT692713.1) were collected according to 28S phylogeny tree from the study of Subbotin *et al.*⁷; then the download 28S sequences were aligned to our six *Aphelenchoides* genomes using BLAT⁹ (options: -t=dna -q=dna -out=blast9), the 28S sequences among *Aphelenchoides* assemblies were collected using those BLAT hit information. Then 28S phylogeny of *Aphelenchoides* genus was produced by using RAxML with 100 bootstrap replicates.

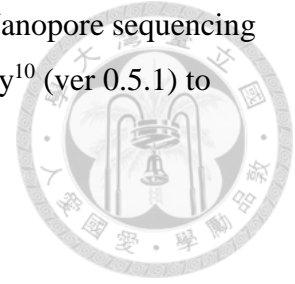
1.2.3 DNA, RNA extraction and sequencing

Six *Aphelenchoides* strains were grown on PDA (potato dextrose agar) medium and were fed with *Alternaria citri*. Nematodes regardless of stage were pooled together and collected. We washed the plates with sterile distilled water, and then purified the samples using sucrose gradients. Next, we purified the Genomic DNA of each nematode strains using Qiagen Genomic-tip 100/G following the manufacturer's instructions; RNA extraction was performed using TRIzol, followed by lithium chloride purification. DNA paired-end Illumina libraries were generated using the KAPA hyper library prep kit (Illumina, San Diego, USA) or Nextera DNA Flex; RNA libraries were constructed using the TruSeq Stranded mRNA library prep kit (Illumina, San Diego, USA).

Both DNA and RNA preparation followed with the standard protocols, and the samples were subsequently sequenced using Illumina HiSeq 2500 (Illumina, USA) to generate 150-bp paired-end reads. The HiC library preparation was performed by Phase Genomics (Seattle, WA, USA) using the proximo HiC animal protocol with some modification in tissue processing. The worms were finely chopped using a microtube pellet pestle for 2 minutes. The tissues were then crosslinked by adding 1 ml of crosslinking solution and incubate for 25 minutes with occasional mixing by rotation. 100 ul quenching solution was added to the crosslinked tissue and mixed for 20 minutes by rotation. The remaining steps in the preparation were carried out according to the protocol.

For a better resolution of genomes, long-reads of APFT and AORT strains were produced using PacBio sequencing system, and the other four *Aphelenchoides* strains

(APVT, AORJ, *A. bicaudatus*, *A. fujianensis*) applied the Oxford Nanopore sequencing platform. Base calling of the nanopore raw signals were used Guppy¹⁰ (ver 0.5.1) to produce a range of 5.0-28.4 Gb total sequences.



1.2.4 Assemblies of *Aphelenchoides* species

Raw reads of four *A. besseyi* species complex and two same genus were assembled using Flye (ver 2.8.2)¹¹ assembler. The draft assemblies from Nanopore (*A. pseudobesseyi* VT, *A. oryzae* RJ, *A. fujianensis* Dali and *A. bicaudatus* Fsh) and Pacbio (*A. pseudobesseyi* FT, *A. oryzae* RT) were corrected using Racon¹² (ver 1.4.6) and Medaka¹³ (ver 0.10.0) with their long reads. All assemblies were furtherly corrected using Pilon¹⁴ (ver 1.22) with Illumina reads for five iterations. In order to achieve a higher genome quality, all the strains have produced a HiC reads but only the *A. pseudobesseyi* VT assembly can be scaffolded into chromosome level using Juice-box¹⁵ tools. The other three genomes of *A. besseyi* species complex were reference scaffolded based on the chromosome level of APVT genome using Ragtag¹⁶ (ver 1.1).

1.2.5 Gene prediction and functional annotation

For comparing the genome structures, the identification of repeat elements between six *Aphelenchoides* genus and 21 representative nematodes were computed using RepeatModeler¹⁷ (ver 1.0.8), TransposonPSI (ver 1.0.0; <https://github.com/NBISweden/TransposonPSI>) and USEARCH¹⁸ (ver 8.1) based on the pipeline of Berriman *et al.*¹⁹. Repeat regions among 27 nematode genomes were predicted by Repeatmasking²⁰ (ver 4.0.9). For the gene annotation of six *Aphelenchoides* strains, repeats regions, evidence alignments and gene model from other gene prediction tools were gathered to predict the best possible gene model among strains using Maker pipeline⁴.

For the expressed sequence tags (ESTs), raw RNA-seq reads were trimmed using Trimmomatic²¹ (ver 0.36), and were aligned to their respective genomes using STAR²² (ver 2.7.1a). RNA-seq transcripts in genomes were inferred using the MIKADO pipeline²³ with a combination tools: i) assembled transcripts based on the mapping results

of STAR using Trinity²⁴ (ver 2.84; option: default setting), and then realigned to genome using GMAP²⁵ (ver 2017-11-15), reconstructed the mappings using ii) Stringtie²⁶ (ver 1.3.4; option: default setting) and iii) CLASS2²⁷ (ver 2.17; option: default setting). The combination of these transcripts and the proteomes of *Bursaphelenchus xylophilus* and *Caenorhabditis elegans* from Wormbase (WBPS17; <https://wormbase.org>) were used as evidence guides to predict the best transcripts using MIKADO²³ (ver 1.2.4; option: three Mikado steps, containing “prepare”, “serialize” and “pick” procedures), and this expression hints were used to train MAKER2. Finally, a step of gene predictions were computed using MAKER2 with MIKADO predicted EST hints, proteomes from closely related species (*Bursaphelenchus xylophilus* and *Caenorhabditis elegans*) as evidence hints, and the genes from other prediction tools (BUSCO²⁸, BRAKER, SNAP²⁹ and Augustus³⁰) supported by evidence hints were chosen as putative genes.

1.2.6 Comparative analyses

For comparing nematodes, proteomes of eight Tylenchida nematodes (*Meloidogyne hapla*, *Meloidogyne incognita*, *Meloidogyne floridensis*, *Heterodera glycines*, *Globodera rosotochiensis*, *Globodera pallida*, *Ditylenchus destructor* and *Aphelenchus avenae*), two Aphelenchoidea nematodes (*Bursaphelenchus xylophilus* and *Bursaphelenchus mucronatus*), three free-living nematodes (*Caenorhabditis elegans*, *Caenorhabditis briggsae* and *Pristionchus pacificus*), six *Panagrolaimomorpha* (*Propanagrolaimus* sp. JU765, *Panagrellus revidius*, *Panagrolaimus superbus*, *Panagrolaimus* sp. PS1159, *Panagrolaimus davidi* and *Halicephalobus mephisto*) and two animal parasitic nematode (*Brugia malayi* and *Strongyloides ratti*) were downloaded from Wormbase^{31,32}(ver 14). Orthogroups relationship across nematodes were predicted by Orthofinder³³ (ver 2.2.7; options: -S diamond). For the species phylogeny, low-copy orthologues were applied due to plenty duplication events in *M. incognita*, *A. avenae* and two *Panagrolaimus* species. The orthogroups present single-copy in other nematodes but have many duplicated copies in these four nematodes were only kept the longest copies. I applied MAFFT (ver 7.310; options: --maxterate 1000) to compute the alignments of each low-copy orthogroups. The concatenated alignment of all low-copy orthogroups was then used to estimate a maximum likelihood phylogeny using RAxML³⁴ (ver 8.2.3;

options: -s -T 32 -N 100 -f a -m PROTGAMMILGF) with 100 bootstrap replicates. Pfam domains of nematodes were identified by using the nematode proteomes to search the Pfam database (ver 31; <https://pfam.xfam.org/>) using HMMER engine with an e-value threshold of 0.001. Effector enzymes were identified by searching the nematode proteomes against the CAZyme³⁵ database (<http://www.cazy.org>) using the HMMER engine with a sequence length threshold of 80bp and an e-value threshold of 1e-15. The predicted enzymes were required to have at least 0.35 proportion of their length covered by conserved domain from database.

1.2.7 Heterozygosity and PSMC

The Illumina reads of APVT and AORT were remapped to their genome using BWA³⁶ (ver 0.7.17-r1188), then the SNP calling were estimated using BCFtool³⁷ mpileup (option: default setting) and BCFtool call (option: default setting) function. Then the VCF format file with SNP information can be visualized by the ggplot2 in R.

The historical population size between APVT and AORT were inferred using pairwise sequentially Markovian coalescent (PSMC) pipeline³⁸ (ver 0.6.5-r67). SNP calling were applied Samtools mpileup (option: -C50 -uf) and Bcftool call (default setting). Number generation ago were applied due to the lacking of molecular time in *A. besseyi* and its closely related species.

1.3 Results

1.3.1 Genome assemblies of *Aphelenchoides* spp.

We sequenced and assembled the genomes of six nematodes in the *Aphelenchoides* genus (two *A. oryzae*, two *A. pseudobesseyi*, one *A. bicaududatus* and one *A. fujianensis*) with the initial aim of delimiting their relationship within the species complex. We sequenced using Pacbio platform to produce 70X and 148X reads (Reads information of ***Aphelenchoides***; **Table 1**) of *A. oryzae* RT and *A. pseudobesseyi* FT strains, respectively. We produced 113-422X Oxford Nanopore reads in another four nematodes (*A. oryzae* RJ, *A. pseudobesseyi* VT, *A. bicaududatus* and *A. fujianensis*). The draft assemblies of these strains were generated by using Flye assembler¹¹ and then

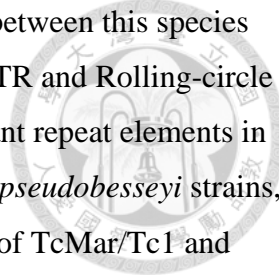
applied Racon¹², Medaka¹³ and Pilon¹⁴ to polish these genomes. Among these assemblies, the *A. pseudobesseyi* VT strain had the best genome quality (N50 = 5.4Mb) and was selected for further scaffolding with 150X Hi-C reads with the manual scaffolding tool—Juicer¹⁵. The final assembly of VT strain which is at chromosome level (N50 = 16.9 Mb), more than 99% contigs was in three large scaffolds (**Figure 1**), which is consistent with the karyotyping number (2n=6,) in the previous *A. besseyi* study³⁹. The genome size of five *Aphelenchoides* is around 44.7 to 47.4 but the *A. fujianesis* Dali is 143.8 Mb which was estimated to be an aneuploidy using Smudge⁴⁰ with Illumina reads (**Figure 2**).

Telomere motif TTAGGC was not found in assemblies, however, we have identified this consensus motif in nanopore reads of APVT strains with low coverage, indicating it might be the limitation of Flye assembler. I searched the same motif in the sister group species of *Bursaphelenchus xylophilus* which have identified in four chromosomes (**Table 2**), suggesting that *A. besseyi* telomeres were present. The genome size of *Aphelenchoides* strains were smaller than most of plant-parasitic nematodes (**Table 3**), indicating genome reduction is likely occurred at the ancestor of the *Aphelenchoides* genus.

1.3.2 Repeat contents

The reduced size of *Aphelenchoides* genomes can be explained by the losses of repeat contents compared to other nematodes (**Figure 3**). DNA transposable elements were the major repeats in *Aphelenchoides* nematodes which is comparably less than other plant-parasites (0.14–1.36 Mb vs. 4.2–22.1 Mb in other nematodes), as well as fewer DNA families were present (1–7 in *Aphelenchoides* versus 9 and 26 in *B. xylophilus* and *H. glycines*, respectively). In addition, *Aphelenchoides* genus contained fewer long terminal repeats (LTR) (0.07–0.8 Mb vs. 0.24–9.3 Mb) and long interspersed nuclear elements (LINE) (0.0006–0.66 Mb vs 0.02–4.5 Mb) retrotransposons. These results show that the genome reduction in *A. besseyi* is likely due to the losses of transposable elements since the common ancestor of this genus.

This phenomenon of genome reduction could be observed within *A. besseyi* species complex, two *A. pseudobesseyi* strains (APVT and APFT) have similar repeat contents in genome, but approximately 2.7 Mb less transposable elements compared to



two *A. oryzae* strains (**Figure 4**). The differences in repeat content between this species complex were primarily contributed by DNA transposons, LINE, LTR and Rolling-circle (RC) (**Figure 5**). DNA transposable elements were the most abundant repeat elements in *A. oryzae* strains, which is almost 10 time longer as compared to *A. pseudobesseyi* strains, the major DNA transposons in *A. oryzae* belonged to superfamilies of TcMar/Tc1 and TcMar/Mariner, respectively (**Figure 3**). It has been reported that TcMar repeat elements affect the genome size and structure during evolution⁴¹. In addition, 0.7-1 Mb RC were only identified in *A. oryzae* strains, which has been proposed associated with gene expression and genome rearrangement in plants and animals and RC were enriched nearby chromosome arms in *C. elegans* due to purifying selection⁴². Due to the higher diversity and abundance of repeats were found in other nematodes, we suggested *Aphelenchoides* lost their repeat contents during evolution, and the *A. pseudobesseyi* strains lost more repeat elements than *A. oryzae* strains during this process.

1.3.3 Annotation

To annotate the genomes of six *Aphelenchoides* strains, I used the Maker2 pipeline⁴³. The evidence hints were generated using the proteomes of *B. xylophilus* and *C. elegans*, which is the most closely related species with annotation available and the model organism in nematode, respectively, and the transcriptome sequencing of each strain. I predicted a total 11,701 to 12,948 protein-coding genes in six strains (**Table 3**). Except for *A. fujianensis* Dali, the number of predicted genes are comparably fewer compared to those of Tylenchida nematodes (12,762 to 19,212) and free-living nematodes (20,184 to 20,992). The lower intergenic region may lead to a smaller genome size for the complex. Based on BUSCO assessment (**Table 1**), the completeness of annotated genes were predicted to be 76.4–81.3%, which is lower when compared to 83.0% and 89.4% in *B. xylophilus* and *B. mucronatus*, respectively, but are higher when comparing to Tylenchida nematodes (59.8%, 50.9% and 73.8% in *M. incognita*, *G. pallida* and *D. destructor*). I therefore re-annotated the APVT strain with manually curated 975 genes but also show a similar score of 78.2%, suggesting the lower completeness of the genomes is likely clade-specific in the *A. besseyi* species complex.

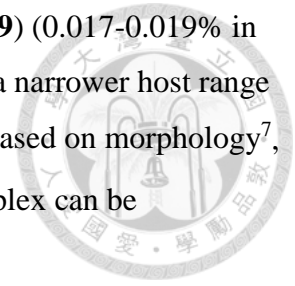
I further assigned the functions of annotated genes using protein family (Pfam)⁴⁴. We found that 66.5% to 71.0% of *Aphelenchoides* genes could be inferred containing at least one domain from Pfam database. We inferred orthologous groups of *Aphelenchoides* genes with other 21 PPNs and free-living nematodes using Orthofinder³³, 78.5–85.4% (*A. fujianensis* = 48.7%) of *Aphelenchoides* genes were orthologous to the sister specie of *B. xylophilus*, and 69.4–76.9% (*A. fujianensis* = 42.8%) were grouped with *C. elegans*. In addition, 87.5–98.7% of *Aphelenchoides* genes were identified as sharing at least one copy with other 21 nematodes, suggesting the genes of *Aphelenchoides* was mainly comprised of those conserved genes among nematodes. Together, the genome reduction of these nematodes is mainly caused by the shortage of intergenic contents and a less gene copy number as compared to other nematodes.

1.3.4 Phylogeny

To investigate the evolution of PPNs and to delimit the relationship among *A. besseyi* species complex, I inferred the species phylogeny based on 74 low-copy orthologues (**Figure 5**). Consistent with a previous study⁴⁵, the same genus of *Aphelenchoides* strains were grouped sister to *Bursaphelenchus* and the plant parasitic nematodes were separated into Aphelenchoidea and Tylenchida. The phylogeny shows the two PPN families originated from the last common ancestor of Clade IV nematodes, and is sister group to the Panagrolaimidae clade, suggesting the lifestyle of nematodes from free-living to two groups of PPN took place at this moment. Within the *A. besseyi* species complex, the strains were divided into two groups, two *A. oryzae* strains (AROT and AROJ) and two *A. pseudobesseyi* strains (APVT and APVF) were clustered together according to their hosts.

The phylogenetic relationships among this species complex in the species tree was consistent with the 28S phylogeny (**Figure 6**), *A. oryzae* is majorly isolated from rice (12 of 14 isolates from rice) but the *A. pseudobesseyi* shows a broader host range (7 plants in 8 isolates). Strains also differed in genome level, 86.6% and 90% nucleotide and amino acid identity (**Figure 7**) was estimated between these two groups. The historical population size³⁸ between these two groups were also different as *A. oryzae* exhibit a more convergent effective population size (**Figure 8**) as compared to *A. pseudobesseyi*.

In addition, *A. oryzae* strains show a lower heterozygosity (**Figure 9**) (0.017-0.019% in *A. oryzae* vs 0.071-0.075% in *A. pseudobesseyi*) is consistent with a narrower host range of *A. oryzae*. Although it is challenging to differentiate them only based on morphology⁷, these results suggested that members of the *A. besseyi* species complex can be distinguished at the genome level.



1.3.5 Gene family specialization in *Aphelenchoides* species

I observed 31 reduced and 66 expanded protein domains in the six members of *A. besseyi* species complex compared to 21 other nematodes included in study. The reduced domains included collagen (90–109 copies in the *A. besseyi* species complex vs. 72–407 in the others), BTG and Somatomedin B. The collagen domains of were thought to be involved in the capsule development in nematodes. A reduction in the number of collagen domain copies may contributed to the lower host specificity of *Trichinella spiralis*⁴⁶ which could explain the wide host range of *A. besseyi*. In contrast, Aphelenchoidea members have on average 8 times (91–314 vs. 4–555 copies) more aspartic protease (ASP) than other nematodes (**Figure 10**). The ASP domain was suggested to involve in the process of host haemoglobin digestion in *Haemonchus controtus*⁴⁷, and in larval skin penetration and migration in hookworms⁴⁸, which could significantly influence the parasitic process of *Aphelenchoide* genus. Interestingly, the ASP difference also identified within the members of *A. besseyi* species complex, *A. oryzae* strains contains only half copy number of ASP as compared to *A. pseudobesseyi* (91–96 vs. 160-194) and it might lead to differential pathogenicity to plants. Other expanded domains included LIM and peptidase C13 domains, which are involved in the regulation of cell motility and growth⁴⁹ or degradation of protein tissues in host⁵⁰. This suggests that the domain dynamics may be associated with the invasion process of PPNs to plants.

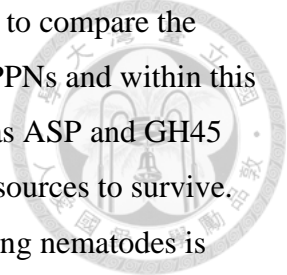
The plant cell wall acts as a primary defensive barrier against most the infection from PPNs, the production of enzymes that can degrade the cell wall or remodel its conjugates is important for PPNs to infect plants. Therefore, we focused on the copy number variation of carbohydrate-active enzyme (CAZyme) families between the analyzed nematodes. A total of 132 CAZyme families were identified in 27 nematodes. Of these, 59-67% of the families were found in the six members of *Aphelenchoides*

genus, which is similar to the 55-66% and 58-68% families found in Tylenchida and free-living nematodes (**Figure 11**). The *Aphelenchoides* genus showed significant expansion in 13 families, including GH16_1, GH27 and GH45 (**Figure 10**). GH16 is a putative β -glycanases that has been reported to be involved in PPNs degradation or remodeling of cell wall polysaccharides⁵¹ to plants, the copy number of GH16_1 in the *Aphelenchoides* genus range from 1 to 6 copies with the orthologs in Panagrolaimidae and *A. avenae*. The GH27 family is found in 3 to 11 copies in Aphelenchoidea nematodes, but is reduced in the Tylenchida nematodes. This family has been suggested to play a role in the activation function of hemicellulose and is associated with α -galactosidase activity in bacteria and fungi⁵². Interestingly, the previously identified GH45 present in clade 10b nematodes⁵³ involved in the degradation of plant cell wall⁵⁴ were absent in *A. fujianensis* and *A. bicaudatus*, suggesting differential maintenance of these genes even within the same genus. Such observation was found at the strain level, GH5 was found to be present in every Tylenchida nematodes but only *A. pseudobesseyi* strains and two *Panagrolaimus* genus have been detected. GH5 is an endoxylanase and have been reported transferred from bacteria via horizontal gene transfer event (HGT)⁵³.

1.4 Discussion

Characterizing the diversity and comparing the genomes of plant parasitic nematodes has been of fundamental in understanding the origins of PPN lifestyles. Whether those changes could be linked to the differential effectors across nematodes warrants further investigation. Applying these results to control the damage caused by PPNs is important. *Meloidogyne*⁵ causes the most serious agricultural damage around the world, hence many efforts have been focused on it.

The analysis of the *Aphelenchoide* genome presented in this chapter provided a comprehensive view of the evolution of two major PPNs clades, and confirmed the delimitation of the *A. besseyi* species complex into *A. oryzae* and the recently proposed *A. pseudobesseyi*. This finding has important implications for nematode management, and is supported by the 28S phylogeny of *Aphelenchoide* genus, which confirm the usefulness of existing molecular markers⁸ for species identification.



The annotated genome of *A. besseyi* in this study allowed us to compare the dynamics of gene family copy number between different clades of PPNs and within this species complex. The significant expansions of gene families such as ASP and GH45 might help *A. besseyi* to infect plants and have acquired essential resources to survive. PPNs generally contain a smaller genome size compared to free-living nematodes is consistent with the finding in *Caenorhabditis bovis*⁵⁵, revealing they acquired resources from their host and those families were lost because they were no longer required. In addition, *A. besseyi* has amongst the smallest genome size and gene number in PPNs, implying those genome reduction or gene losses might contribute to the adaption process for their hosts.

It has been generally believed that *A. besseyi* has limited mobility in natural habitats, so its lack of population structure in China⁵⁶ is likely a result of human-mediated dispersal. Our results suggested that *A. oryzae* is primarily isolated from rice, whereas *A. pseudobesseyi*⁸ has a wider host range. To confirm whether all cases of white tip disease in rice fields are caused by *A. oryzae*, a wider geographical range collections and sample resequencing of this species complex are required. This may provide insight into the population structure and the evolution of different cryptic species within the complex.

CHAPTER 2.

Chromosome evolution of *A. besseyi*



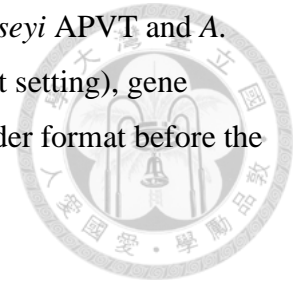
2.1 Introduction

In chapter 1, I observed three chromosomes in the *A. besseyi* species complex which is only half the number compared to closely related *B. xylophilus* (n=6) and *C. elegans* (n=6), initially suggesting a chromosome fusion took place during the evolution process of *A. besseyi*. In addition, a reduction of genome size and repeat content has been observed in this species. In this chapter, I sought to investigate the extent of karyotype rearrangements and how the *A. besseyi* chromosomes were fused into three in this chapter. I inferred the distribution of repeat and heterozygosity along three chromosomes to compare the differential within this species complex. The clade V nematode *Diploscapter pachys*⁵⁷ has only one chromosome due to the missing telomere structure and the alteration of telomeric maintenance and meiosis genes. Hence, I also compared these families across nematodes.

2.2 Materials and Methods

Synteny relationships between *A. pseudobesseyi*, *B. xylophilus* and *C. elegans* were inferred based on pairwise single copy orthologs in Orthofinder³³. Those families containing pairwise one-to-one orthologs were selected to present a link corresponding to their genome loci using custom R script. The synteny information between two pairwise species was converted into a density plot using ggplot, then the macro-synteny information across these three species could be observed as well as fusion and fission events in APVT chromosomes. Nigon elements of Rhabditida in *A. pseudobesseyi*, *B. xylophilus* and *C. elegans* were inferred using the common orthologs of Rhabditida nematodes identified in *Oscheius tipulae* study⁵⁸, we generated those Nigon elements following their pipeline.

Synteny blocks within the species complex of *A. pseudobesseyi* APVT and *A. oryzae* AROT were estimated using DAGchainer⁵⁹ (options: default setting), gene locations of nematodes GFF format was converted into the gene order format before the running of DAGchainer.



The male APVT Illumina pair-end reads were aligned onto the genome using BWA³⁶ (ver 0.7.17-r1188), and the bam file read depths were estimated applying Mosdepth⁶⁰. The depth were normalized by the median depth of the largest scaffold, and then present using the dotplot function in R ggplot. The distribution of transposable elements in three chromosomes of APVT and AROT were estimated using Bedtools⁶¹ in 10kb non-overlapping window.

The telomere repeats in APVT and *B. xylophilus* genome were scanned by using Tandem Repeat Finder (ver 4.0.9). I then identified the typical nematode telomere (TTAGGC) from the result. Neither this motif nor other motifs have been identified in the assembly of APVT, but the consensus motif has been found in nanopore raw reads. The alignment results showed it is likely to be trimmed due to the low read coverage in the assembler of Flye (ver 2.8.2)¹¹. However, the telomere motif TTAGGC was identified in four chromosomes of *B. xylophilus* assembly.

2.3 Results

2.3.1 Macro-synteny

The synteny relationships among *A. pseudobesseyi* APVT strain, *B. xylophilus* and *C. elegans* were inferred using single copy orthologs (**Figure 12a**). I identified macro-synteny blocks in three chromosomes of APVT corresponding to many distinctive blocks in the six chromosomes of *B. xylophilus* (**Figure 12b**). However, most of macro-synteny is conserved between *B. xylophilus* and *C. elegans*, revealing the chromosome fusions and fissions occurred in *Aphlenechoides* genus.

Macro-synteny blocks between APVT and *B. xylophilus* contained 148-801 orthologous genes, those synteny blocks in APVT could be traced back to co-linear

regions of *B. xylophilus* chromosomes (**Figure 12a**), allowing us to infer the fusion sites in the chromosomes of APVT. In addition, these contiguous blocks were presumably still not broken yet by recombination, revealing these fusion and fission events in *A. besseyi* might have been relatively recent. We also observed fission events, many ancestral chromosomes were identified in different blocks of APVT chromosomes. The instance case of chr I which identified in the same chromosome between *B. xylophilus* and *C. elegans*, whereas separated in different APVT chromosome arms of chr 2 and chr 3 (**Figure 12b**).

2.3.2 Sex chromosome

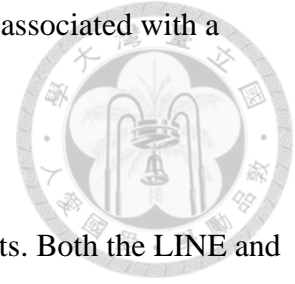
C. elegans X chromosome was identified in two chromosomes of *B. xylophilus*, the majority X chromosomes of *C. elegans* were unambiguously assigned to chr 2 with reduction size (**Figure 12b**). However, most of *C. elegans* chromosome X originated from ancestral autosomes⁵⁸ (**Figure 13**), as well as *B. xylophilus* (**Figure 14**). We then explored the Nigon elements of rhabditid⁵⁸ in *A. besseyi* and found that the ancestral nematode sex chromosome were mainly located in *A. besseyi* chr1 (**Figure 15**).

The sequencing coverage could help to identify the sex chromosomes of *A. besseyi*, the remapping of APVT male Illumina reads showed even read depth along three chromosomes (**Figure 16**), which is consistent with the recently published model of stochastic sex determination system of *B. xylophilus*⁶², indicating this system might have occurred at the last common ancestor of Aphelenchoidea nematode and remained in all Aphelenchoidea members.

2.3.3 Synteny within *A. besseyi* species complex

In the *A. besseyi* species complex, a total of 91% and 88% of genomes were in synteny between APVT and AORT, respectively. Of the synteny break regions, 49.8% and 74.2% were intergenic, suggesting the breaks were more likely to take place here. In addition, I identified a major inversion of 3.4 Mb located in the centromere of chr 2 (**Figure 17**), with a possible inversion could be identified in the arms of three chromosomes. The rearrangement frequency around the chromosome arms tend to be higher compared to chromosome center; synteny blocks were generally smaller in the

arms and have a higher frequency of translocation, which might be associated with a higher repeat density in these regions.



2.3.4 Repeat distribution

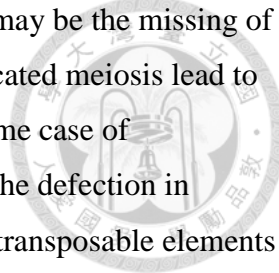
I sought to identify regions enriched in transposable elements. Both the LINE and LTR retrotransposons were abundant in the chromosome arms of the *A. oryzae* strain (AORT) (**Figure 4** and **Figure 17**), which is consistent with the hallmark of nematode chromosome evolution⁶³. In contrast, only the LTR retrotransposons were enriched in the two chromosome arms of *A. pseudobesseyi* (APVT), and the expanded DNA transposons strain were evenly located at chromosomes in *A. oryzae*, suggesting that these transposable elements were differentially maintained within this species complex.

2.3.5 Chromosome maintenance genes

In this chapter, we inferred the chromosome fusion and fission in *A. pseudobesseyi*. Those events might be derived from either telomere alteration or the defects in telomere maintenance and meiosis. I searched for a total of 99 conserved telomere and meiosis associated gene families in *C. elegans*⁵⁷ across other nematodes. Only 52 out of 91 and 1 of 10 telomere and meiosis families were identified in *Aphelenchoides* and *Bursaphelenchus* genus (**Figure 18**), respectively. MRT-1 is essential for the telomere addition in *C. elegans* and the domains of POT-1, POT-2 and POT-3 is associated with telomere maintenance^{57,64,65}. The members of ZIM-1, ZIM-2, ZIM-3 and HIM-8 are related to the central chromosome pairing⁵⁷, and the absence of these orthologs might lead to chromosome fusion in *A. besseyi*. However, the absence or divergence of these families were likely not the primary reason for the karyotype rearrangement observed in *A. pseudobesseyi* as *Bursaphelenchus* species still contains the classic six chromosomes. Nevertheless, we still cannot exclude the possibility that these genes might lead to the fusions or fissions in *A. besseyi* chromosomes.

2.3.6 Discussion

Genome rearrangement and reduction are common in plants⁶⁶, insects⁶⁷ and nematodes⁶⁸. I found that *A. besseyi* exhibited multiple chromosome fission and fusion



events, and a possible explanation together with genome reduction may be the missing of the telomeric repeat maintenance genes and meiosis genes, the truncated meiosis lead to the alteration of chromosome pairing; this was observed in an extreme case of *Diploscapter pachys*⁵⁷ which fused all the chromosomes into one. The defection in meiosis genes may lead to genome shrinkage due to a loss of DNA transposable elements derived from chromatin imbalanced in *Caenorhabditis nigoni*⁶⁹, suggesting *Aphelenchoides* may have followed a similar scenario. However, members of *Bursaphelenchus* with six chromosomes also failed to identify these aforementioned orthologs, suggesting their divergence may have occurred since the last common ancestor of Aphelenchoidea. The ancestral sex chromosomes are likely to present in *Aphelenchoides* but whether the reduced size affected the chromosome pairing or those lost orthologs have been replaced is still unclear. Further RNAseq with different stages of meiosis and the identification of *A. besseyi* mating system are needed to confirm the integrity of meiosis in *A. besseyi*.

CHAPTER 3.



Horizontal gene transfers in clade IV nematodes

3.1 Introduction

In chapter 1, I compared the CAZyme gene number across 27 representative nematodes. The dynamic copy number of families present in Tylenchida and Aphelenchoidea were identified. Some effector enzyme were proposed that have involved in the degrading or modifying the composition of different plant tissues^{70,71}. The instances included GH5, GH32, GH16 and GH45, these genes have been reported to be horizontal gene transferred (HGT) from bacteria or fungi^{53,71,72}, giving nematodes the ability to adapt to different environments⁷³. Most of GH5 were lost in Aphelenchoidea but many copies were present in Tylenchida nematodes, raising the possibility that many of HGT genes were acquired in the last common ancestor of these two clades but were differentially retained. Despite the presence of numerous HGT genes in various nematode species, research on the timing and subsequent maintenance of HGT genes, as well as the factors influencing differences in their copy numbers, has been limited to only a few clades within the nematode group⁷⁴. Further investigation into the dynamics of HGT genes in nematodes is necessary to gain a better understanding of these processes.

In this chapter, I aim to provide a holistic view of the evolution history in plant parasitic nematodes. The *Panagrolaimid* nematodes contain plenty HGT genes in their genome⁷⁵, suggesting HGT events are not PPN specific. The HGT candidates in PPNs were likely originated from the last common ancestor of Clade IV nematodes if those genes shared homology with *Panagrolaimid* species. I then included *Panagrolaimid* nematodes in our study to infer the evolutionary origin of these genes. The HGT genes in different clade of nematodes were identified. I traced the origin of the donors and inferred the time points they were acquired. I investigated the fate of these HGT genes after having acquired. By inferring the evolutionary origins of these HGT genes, we were able to identify historical HGT events that have influenced the evolution of nematodes. Our finding suggests that major HGT events

that occurred in the common ancestor of clade IV nematodes may have contributed to their adaptations to diverse environments.



3.2 Materials and Methods

3.2.1 Identification of the HGT genes

The probability of genes have been acquired via HGT was estimated by applying Alieness Index (AI)⁷⁶. The donor groups were generated by excluding all the Metazoans sequences from the NCBI nr database, and the recipient were Metazoans removing the following clade of species to prevent the bias of self-alignment: Aphelenchoidea, Tylenchida, Rhabditina, Spirurina and Cephaloidea. AI score of every nematode genes was estimated by using the e-value of diamond⁷⁷ (ver 2.0.14; option: blastx --evalue 0.001) best hits between the donor and recipient database. The genes lacking the hits in two databases were replaced with the e-value to 1. To reduce contamination, orthogroups containing at least one gene with AI value larger than 30 were selected for further HGT study.

HGT gains or losses at each branches of the phylogeny were estimated using Phylip-Dollop⁷⁸ (ver 3.69; options: fdollop -method d -ancseq). Some of the HGT family acquired branches were manually curated by their evolutionary place due to those nematode copies in HGT orthogroups were estimated to be $AI < 0$ but were clustered with other HGT genes. The highest AI value of nematode genes with classified taxonomy hits were chosen to represent the HGT origin in each orthogroups but unclassified hits were abandoned until the second or third highest AI value of orthologs with classified hits were found.

To identify HGT genes in nematodes, we first selected orthogroups with CAZyme annotations and nematode orthology genes with an AI value higher than -50. Orthologs with AI values less than 0 were labeled with a "*". We then combined these orthologs with HGT donor sequences from the NR database and specific cellulase sequences from the CAZyme database. To reduce contamination from HGT genes, we also annotated the Pfam domains of the orthologs and removed those that did not have at least one major domain (cellulase or pectate lyase). We aligned the sequences of each HGT orthogroup using MAFFT (options: --maxiterate 1000 --genafpair) and trimmed them using trimAl⁷⁹

(ver 1.4; options: -gappyout). We then used IQtree⁸⁰ (ver 1.6.6; options: -bb 1000 -alrt 1000) to compute the phylogeny of each HGT family gene trees. For CAZyme unclassified HGT orthogroups, we used the top 2 blast hits sequences from separate Uniprot databases (bacteria, fungi, land plants, and insects) to confirm the HGT origin. To infer the fate of HGT families after acquisition from the donor, we calculated the proportion of species orthologs that were detected as HGT within the same HGT families.

3.2.2 Gene ontology (GO) enrichment analysis

Gene enrichment was predicted by topGO⁸¹ program in R (ver 3.6.1). The proteomes of nematodes with GO annotation were computed by eggno-mapper⁸². For exploring the representative terms, we applied GO-Figure⁸³ to present concise functions which have considered the relationship between GO terms in DAG structure and removed the redundant terms.

3.3 Results

3.3.1 HGT families among nematodes

The availability of the *Aphelenchoides* genomes allowed us to investigate the evolutionary dynamics of HGT across the plant parasitic nematodes. A total of 27 proteomes from representative nematodes including the *Aphelenchoides* genomes were searched for evidence of HGT using the Alieness (AI) score⁷⁶. A total of 1,675 HGT orthogroups were identified in 21 nematodes, 0.3-2.4% and 0.6-2.1% proteomes among Aphelenchodea and Tylenchida nematodes were predicted to be HGT (**Figure 19**) such differences were likely the result of genome evolution after speciation. The presence of HGT genes in *Panagrolaimomorpha* (0.1-5.4% in six proteomes) is consistent with the identification of currently published *Panagrolaimid* study⁷⁵, the high HGT genes in *P. superbus* is consistent with the report⁷⁵ and most of them were estimated to be species specific. The shared HGT included a GH5 family (**Figure 20**) which is present in all clade 12 nematodes, three *Panagrolaimid* (*P. ps1159*, *P. superbus* and *P. davidi*) and *Aphelenchoide* strains (APVT, APFT and *A. bicaudatus*) and were previously inferred to

have been acquired from bacterial origin⁵³, revealing a new clade with evidence of HGT in multiple species that is sister to the two clades comprising PPNs.

The majority of HGT gene families were acquired from bacteria, and followed by fungal origin (**Figure 19**). The high copy number of HGT genes identified in *M. incognita* was a result of duplication after speciation⁸⁴, indicated by 2 times differences against HGT orthologues number (**Figure 21**). In addition, I have identified 40 HGT families that were transferred from the Streptophyta phylum in PPNs, which is consistent with the previous finding of plant-transferred genes in *H. glycine*⁸⁵. One instance shows this HGT family of nematode sequences is closely related to *Oryza sativa* (rice) and *Quercus suber* (oak) (**Figure 22**) which are the common hosts to many PPNs. Interestingly, of these orthogroups, 27 were identified in *B. mucronatus* and the GO term enrichment analysis suggested these families of plant origin were highly enriched in the detoxification of cadmium and copper ion function (**Table 4** and **Figure 23**), suggesting these genes might help *Bursaphelenchus* nematodes against the toxin present in their pinewood hosts. In addition, we identified 597 that were gained and 60 families that were lost in the last common ancestor clade IV nematode. For the inferred gain families, the GO enrichment analysis suggests these families were enriched in the ion transport functions including manganese, calcium and potassium (**Figure 24**). In contrast, the lost families show an absence or divergence functions in clade IV such as ubiquitin recycling, histamine response and insulin metabolism (**Figure 25**). The HGT events have arisen since the losses or gains of these families, suggesting which might be related to the early development of PPNs.

3.3.2 Major episode of HGT in the clade IV nematodes

One of the outstanding questions surrounding the evolution of HGT is why different PPNs possess differential origin and copies of HGT genes. By placing the identified HGT orthogroups onto the species phylogeny using dollop⁷⁸, we were able to infer the HGT origins in the species tree. I found HGT started in the last common ancestor of clade IV nematodes, comprising *Panagrolaimomorpha* and two major PPN clades (Tylenchida and Aphelenchoidea). A total of 161 HGT families were identified

that occurred in this episode (**Figure 26**), and most of these HGT were acquired from bacteria (78.3%).

Examples of HGT families included GH16, GH32, GH43 and GH5. The revised GH5 cellulase gene tree suggests an ancient duplication occurred due to many GH5 copies identified in *Panagrolaimid* (*P. sp.* PS1159, *P. superbus* and *P. davidi*) (**Figure 20**). In addition, many GH5 copies could be found in Tylenchida but only one copy was identified in *A. pseudobesseyi* (APVT and APFT) and *A. bicaudatus*, suggesting differential maintenance and duplication of GH5 in these two PPN clades. Interestingly, the closest GH5 bacterial copies were *Salinimicrobium xinjiangense* and *Leeuwenhoekiella sp.*, which belonged to Flavobacteriaceae family from marine, revealing the possibility of last common ancestor of clade IV nematode may be residing in the marine environment.

I observed two GH16 subfamilies in nematodes (**Figure 19**). Both families containing nematode orthologs were identified to be HGT, which were acquired from bacteria and fungi. The GH16_3 HGT orthologs in Tylenchida and *Bursaphelenchida* nematodes were grouped together with bacterial origin sequences. However, GH16_1 of *Aphelenchoides* and *Panagrolaimus* nematodes were clustered with fungal origin sequences (**Figure 27**), suggesting that these two GH16 groups arose independently. GH32 in *G. pallida*⁷¹ was proposed to have been involved in hydrolyzing fructose function and was commonly found in Tylenchida nematodes but also present in one *Panagrolaimus* (**Figure 28**). GH43 were observed in two separate bacterial clusters of Tylenchida and *Panagrolaimid* clades which have been proposed to be related to the degradation of their host cell wall⁸⁶ (**Figure 29**). In addition, five HGT orthogroups with a total 27 pfam domains of ATP-binding cassette (ABC) transporters (**Table 5**) was also identified in this episode, which has been proposed to associate with PPN infection⁸⁷, suggesting the acquirement of this detoxification motifs from bacteria might help the adaption process of clade IV nematodes from free-living to PPNs. Together, these results suggested some HGT genes that were thought to play important roles in plant parasitism were in fact acquired earlier than the common ancestor of plant parasitic nematodes.

3.3.3 Other HGT episodes in PPN nematodes

I inferred 47 HGT orthogroups occurred in the episode of the last common ancestor of PPNs (**Figure 26**). The instance families included pectate lyases 3 (PL3) which were reported to associated with the cell wall degradation of host plants⁸⁸. PL3 orthologs identified in *Aphelenchus avenae* and two *Bursaphelenchus* nematodes were clustered with distinct clade of *Meloidogyne* species (**Figure 30**) is consistent with previous finding in PPNs⁷². The closest bacterial ortholog in the *Meloidogyne* clade was from *Curtobacterium flaccmfaciens* which is also known to cause bacterial wilt in the Fabaceae family⁸⁹.

Moreover, 29 gained HGT families were identified in the common ancestor of Aphelenchoidea, the instance families included the fungal origin of GH45 which is present in all of clade 10b nematode except *A. fujianensis* and *A. bicaudatus*, the GH45 families were primarily clustered into *Bursaphelenchida* and *Aphelenchoides* genus, the latter have identified an ortholog group with negative AI in two *A. orazyae* strains (**Figure 31**), suggesting this mutated homolog group might have acquired after speciation.

3.3.4 The fate of HGT genes

I inferred many cases of HGT genes clustered together with those genes with negative AI in the same orthology groups, suggesting those differential parts of genes might accumulate after transferring from their donor species. We sought to estimate the proportion of every family having HGT candidates in each representative nematode. Genes in the majority of HGT families in Aphelenchoidea and Tylenchida were mainly predicted to be all HGT genes (with AI > 0; 54.6-76.5% vs. 77.3-89.4%). Although members in some orthogroups did not pass such criteria, they were grouped with members that were inferred as HGT and followed the species phylogeny, suggesting this might be a result of accumulating substitutions over time. Consistent with this finding, the more ancient the HGT took place, the higher proportions of these members were present (**Figure 32**). The instances included GH5 families with 12.5-70.6% of copies in Tylenchida failed to identify as putative HGT, suggesting duplication and possibly neo-functionalization of GH family in PPNs after being acquired from bacterial donor species. Those findings were also observed in the *A. besseyi* species complex, which

present a GH45 genes with failed HGT identified score in *A. oryzae* but are all HGT candidates in *A. pseudobesseyi* (**Figure 31**).



3.4 Discussion

Our systematic investigation of HGT showed that many of the HGT events were acquired from the last common ancestor of clade IV. Many clade IV nematodes are known to have the ability to survive in extreme desiccation^{75,90,91} environment. The acquired HGT genes may help against the harsh environment and subsequently adapt to successful plant parasitism⁹². Further biological experiments are needed to confirm the existence and function of those HGT events we identified.

Interestingly, many of the closest bacterial donors for HGT genes in clade IV nematodes were found to inhabit marine environments, suggesting that the last common ancestor of these may have lived in a marine environment, and subsequently underwent a habitat transition⁹³. However, we also identified non-bacterial donors that were typically found in environments that align with the lifestyle of nematodes, revealing that symbionts may have contributed to the acquirement of HGT genes in nematodes. However, currently only *Wolbachia* and *Cardinium*⁹⁴ are known to be endosymbionts of nematodes, and these were not observed in our analyses. Our result shows that most of the acquired HGT genes were lost over time, suggesting that most of HGT genes may be neutral to the genomes of the host nematodes.

As more and more genome sequence become available, historical HGT events have been detected in the primary episode of common ancestor groups such as land plants⁹⁵, moths and butterflies⁹⁶, which have contributed to the developmental roles and adaptations in their hosts. These acquisitions are thought to have occurred at a time when the host's development or defense system was weak, which might related to expanded or reduced families. Hence, I suggest that the presence or absence of gene families in clade IV nematodes might have affected the acquisition and retention of HGT genes.

In summary, the availability of *Aphelenchoides* genomes and our comparative analyses allowed us to infer the major events of HGT in clade IV nematodes. The fate of HGT differentially lost in different clades or *Aphelenchoides* genus may have contributed to

multiple adaptations. Additionally, the availability of various *A. besseyi* genomes will help in the development of molecular diagnostic tools to distinguish the specific diseases caused by this species complex.



Figures

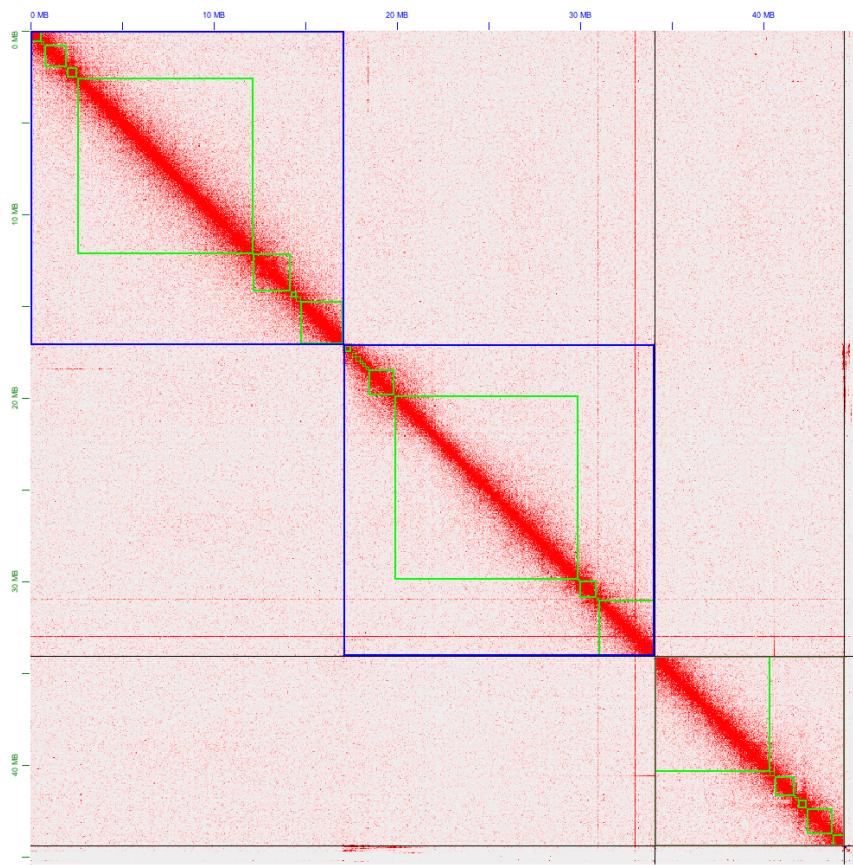


Figure 1 The signal of HiC reads in *A. pseudobesseyi* VT were curated using Juicerbox. The self-blast against results is denoted with the red dots; the more red dots indicates the more closer the genetic distance. The three blue blocks correspond to the three chromosomes in this strain.

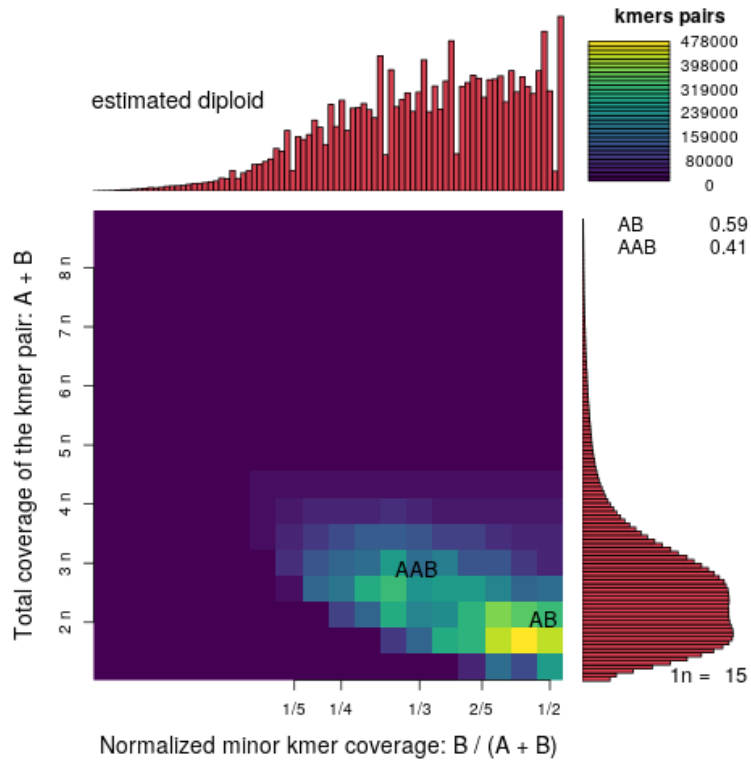


Figure 2 The polyploid prediction of *A. fujianensis* Dali generated using Smudge⁴⁰ Illumina paired-end reads. The brighter colours denote a higher number of kmer pairs. The Y-axis represents the total coverage of the kmer pairs, and the X-axis shows the normalized minor kmer coverage.

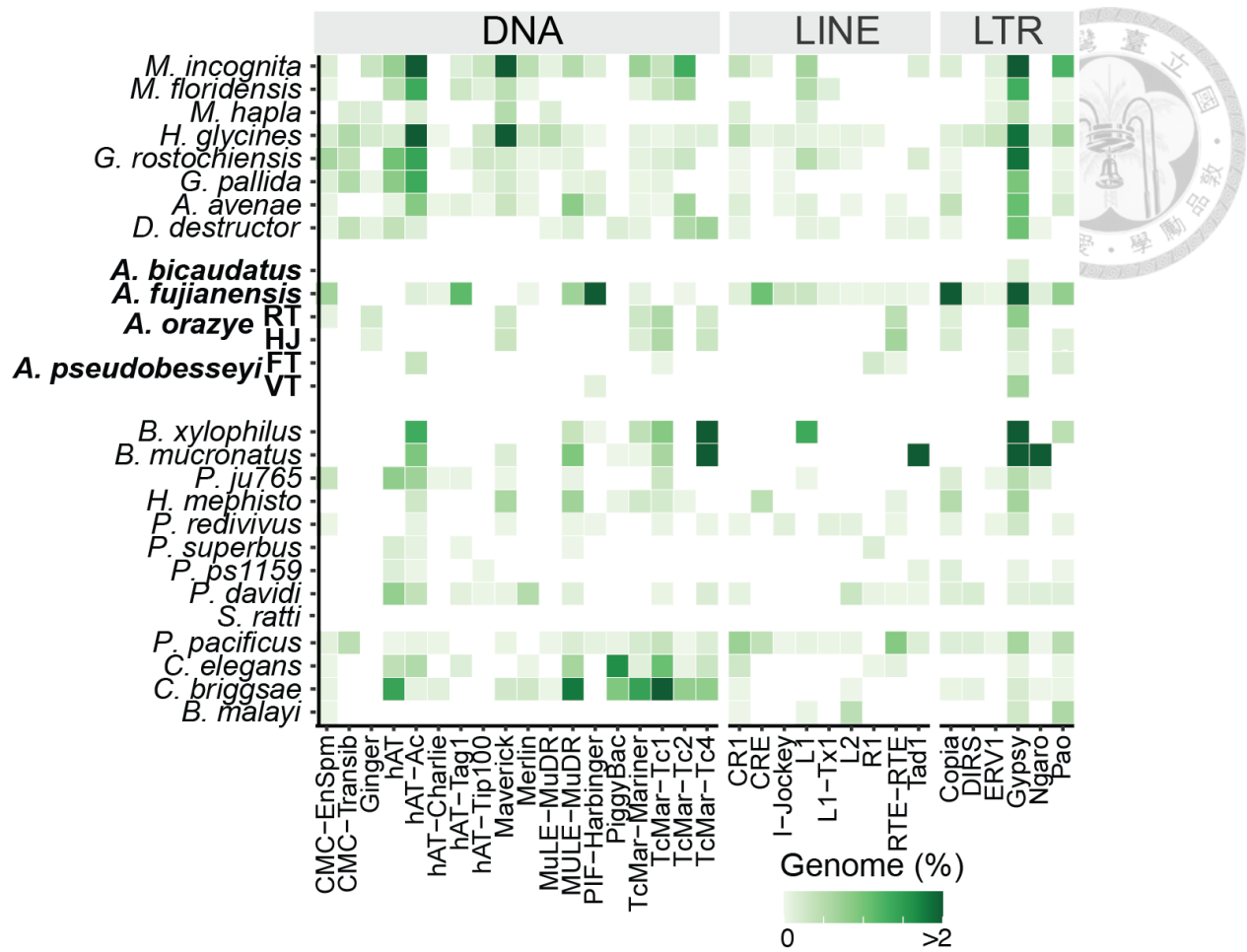


Figure 3 Transposable elements in nematodes.

The genome proportions of each DNA, LINE and LTR transposable elements families in nematodes. The highest proportions ($\geq 2\%$) denote with same dark green.

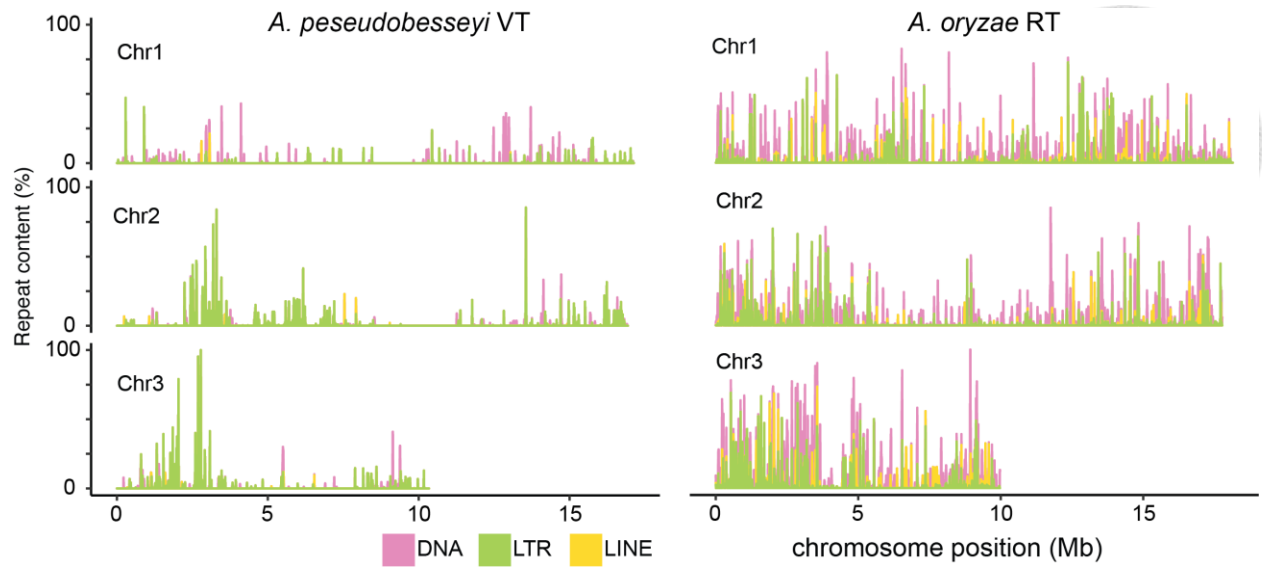


Figure 4 Repeat distribution between *A. pseudobesseyi* VT and *A. oryzae* RT. The DNA, LTR and LINE transposable contents in non-overlapping 10kb windows along the three chromosomes of the strains are denoted by pink, green and yellow colours, respectively.

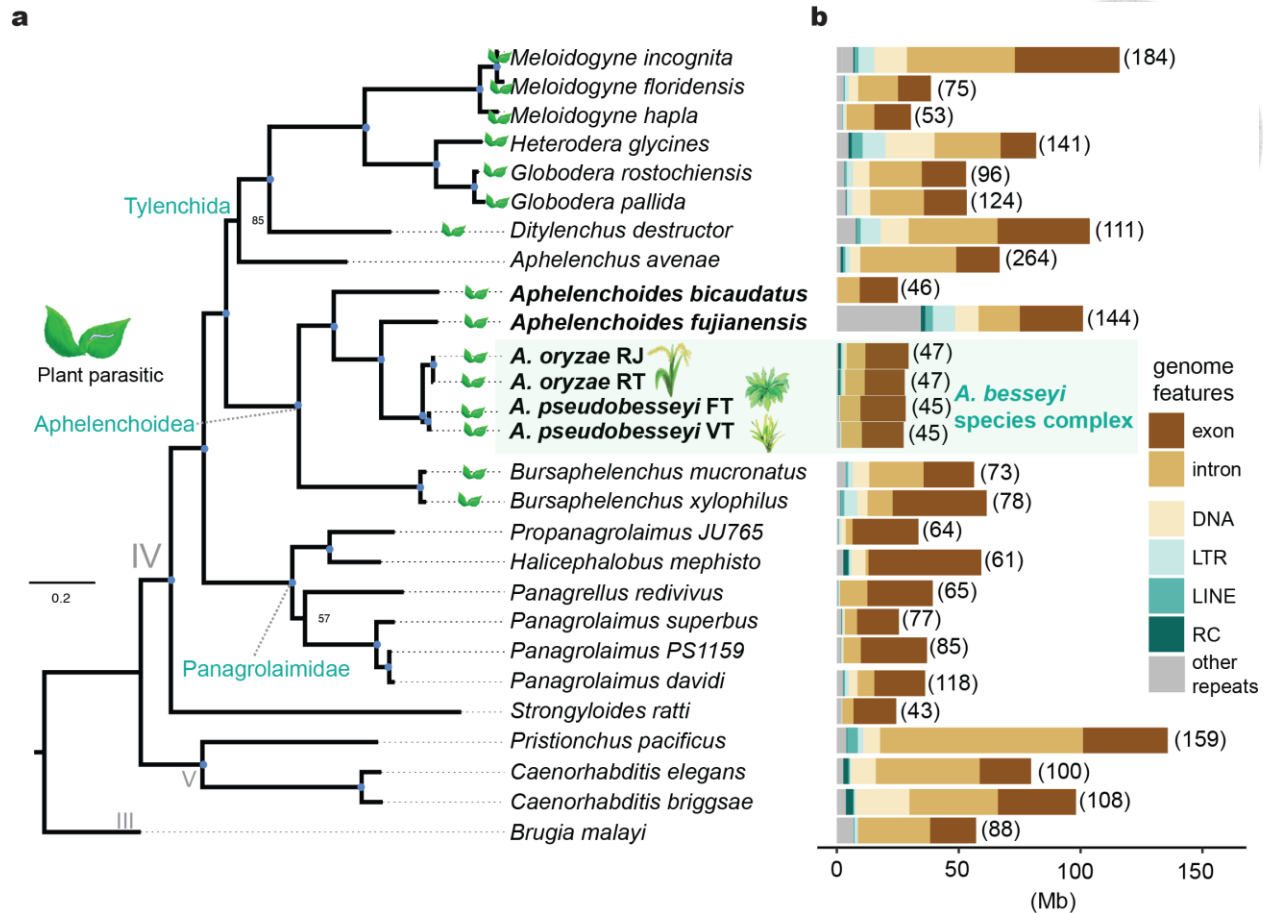


Figure 5 Species phylogeny and genome features of 27 representative nematodes.

a. The phylogeny of nematodes was inferred based on 74 low-copy orthologues. Blue dots in branches indicate a bootstrap value of 100. The grey box labelled with the position of *A. besseyi* species complex. **b.** The size of gene contents and repeat elements (including DNA transposons, long interspersed nuclear elements, long terminal repeats, rolling-circle, and other repeats) are shown, but non-repeat intergenic regions are not displayed. The number in brackets indicates the genome size of each nematodes.

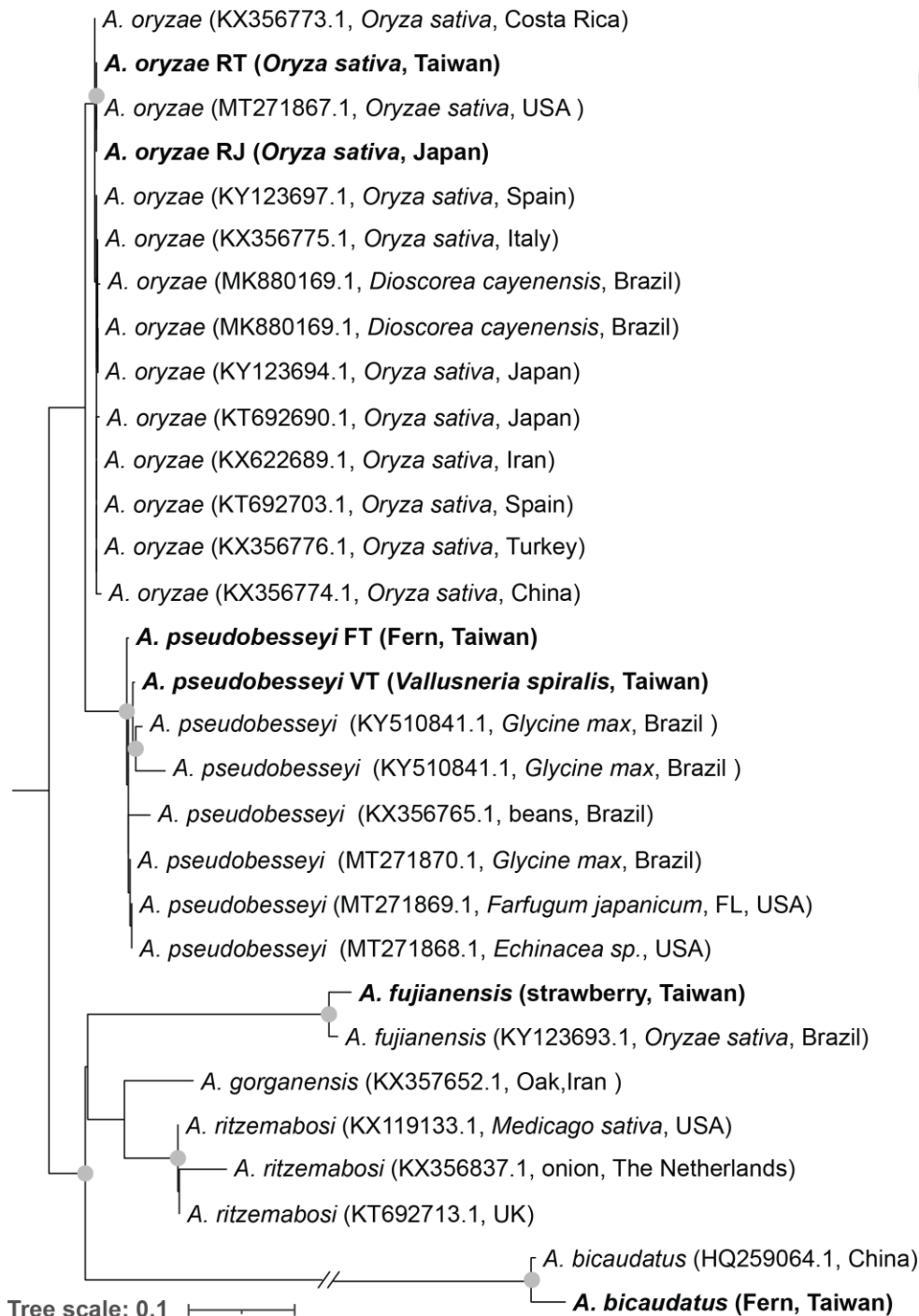


Figure 6 Separation of *Aphelenchoides* nematodes by using DA2 and DA3 of 28S.

The sequences that were sequenced in this study are in bold. The grey dot in the branches indicate a bootstrap support > 80%.

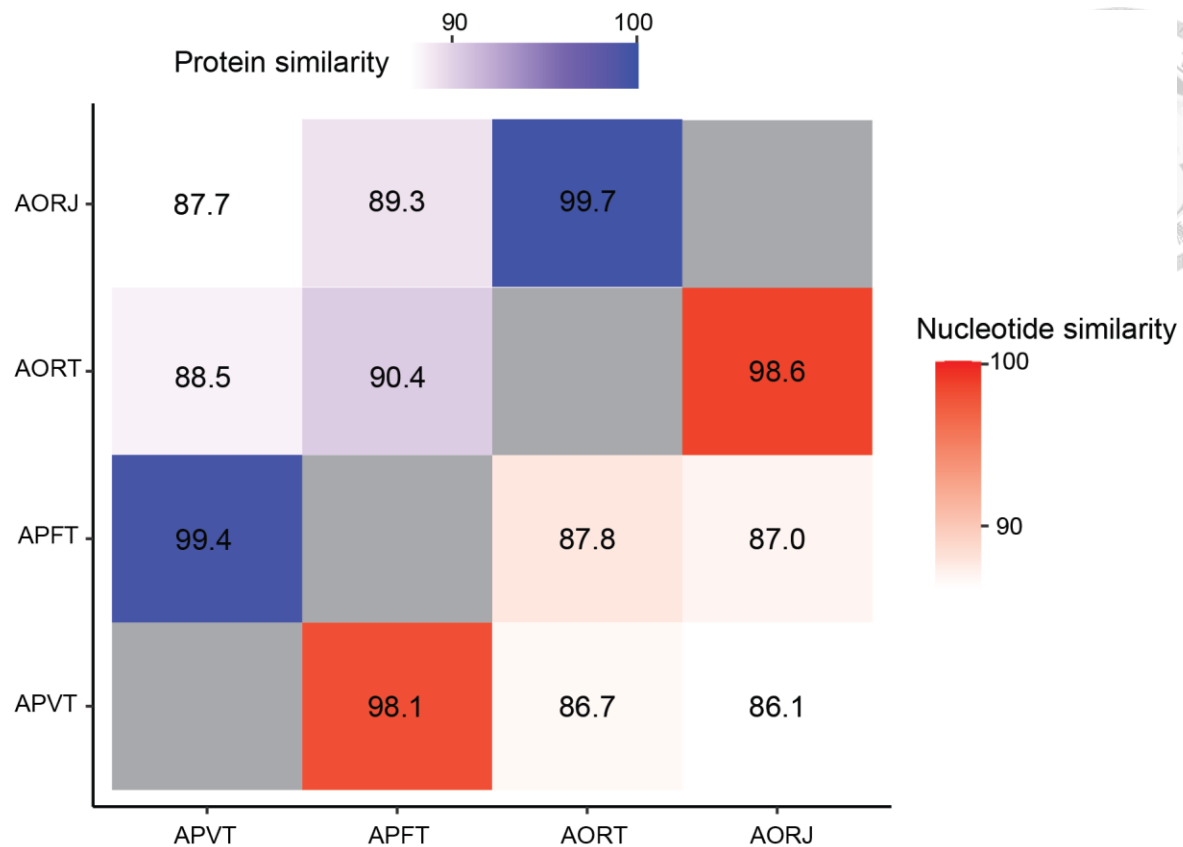


Figure 7 The median similarity of nucleotide coding sequence and protein within the *A. besseyi* species complex.

Pairwise protein similarity and nucleotide identity among four species in the *A. besseyi* complex are labelled and indicated by purple and red gradients, respectively.

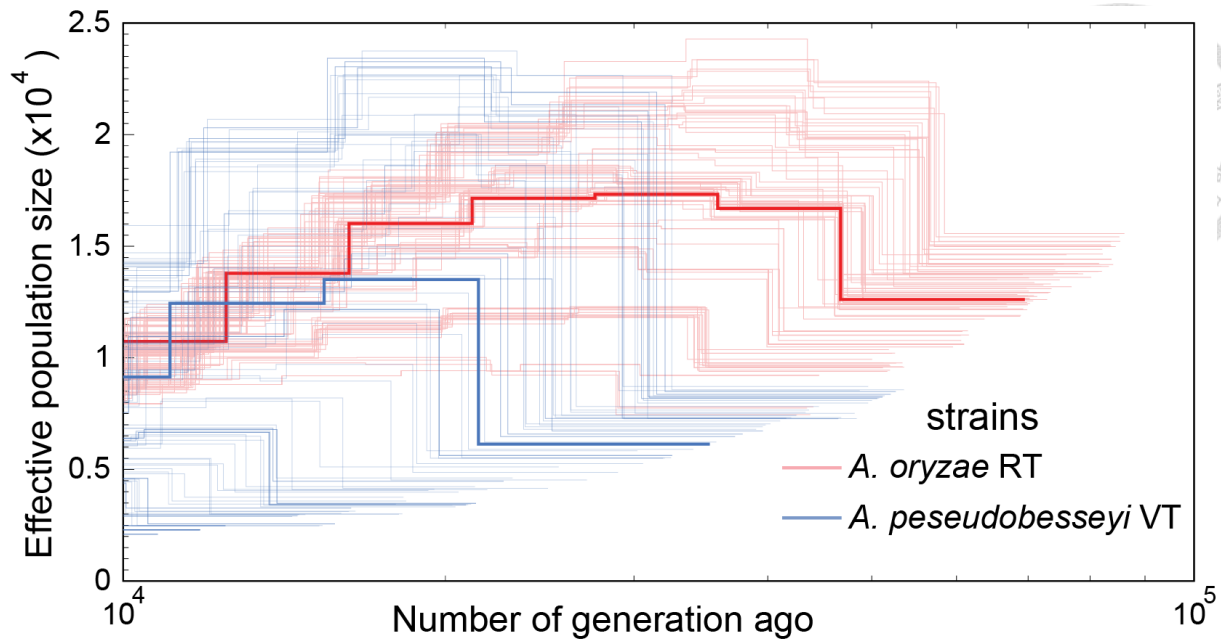


Figure 8 Prediction of history effective population size of *A. oryzae* RT (red) and *A. pseudobesseyi* VT (blue).

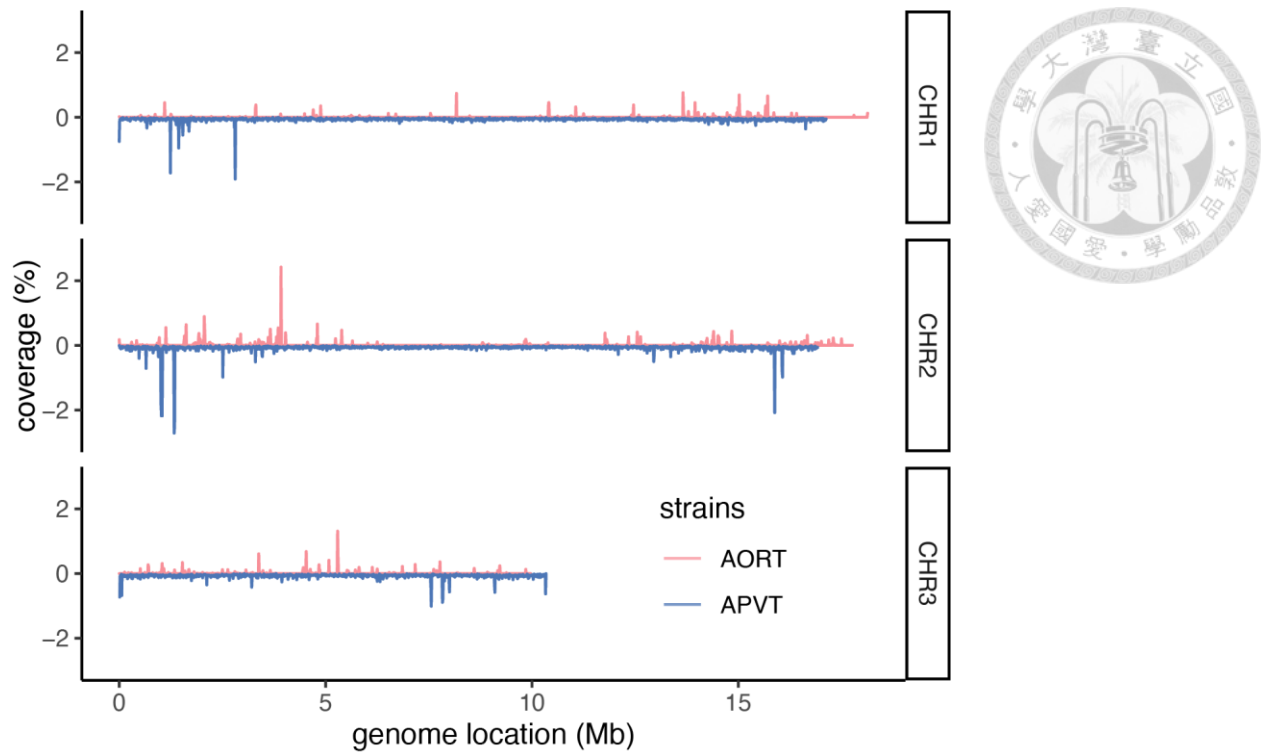


Figure 9 Heterozygosity of *A. oryzae* RT and *A. pseudobesseyi* VT strains.

Heterozygosity was inferred using remapping Illumina reads along non-overlapping 10kb windows, the positive (red) or negative (blue) value denote with the levels of heterozygosity in AORT and APVT, respectively.

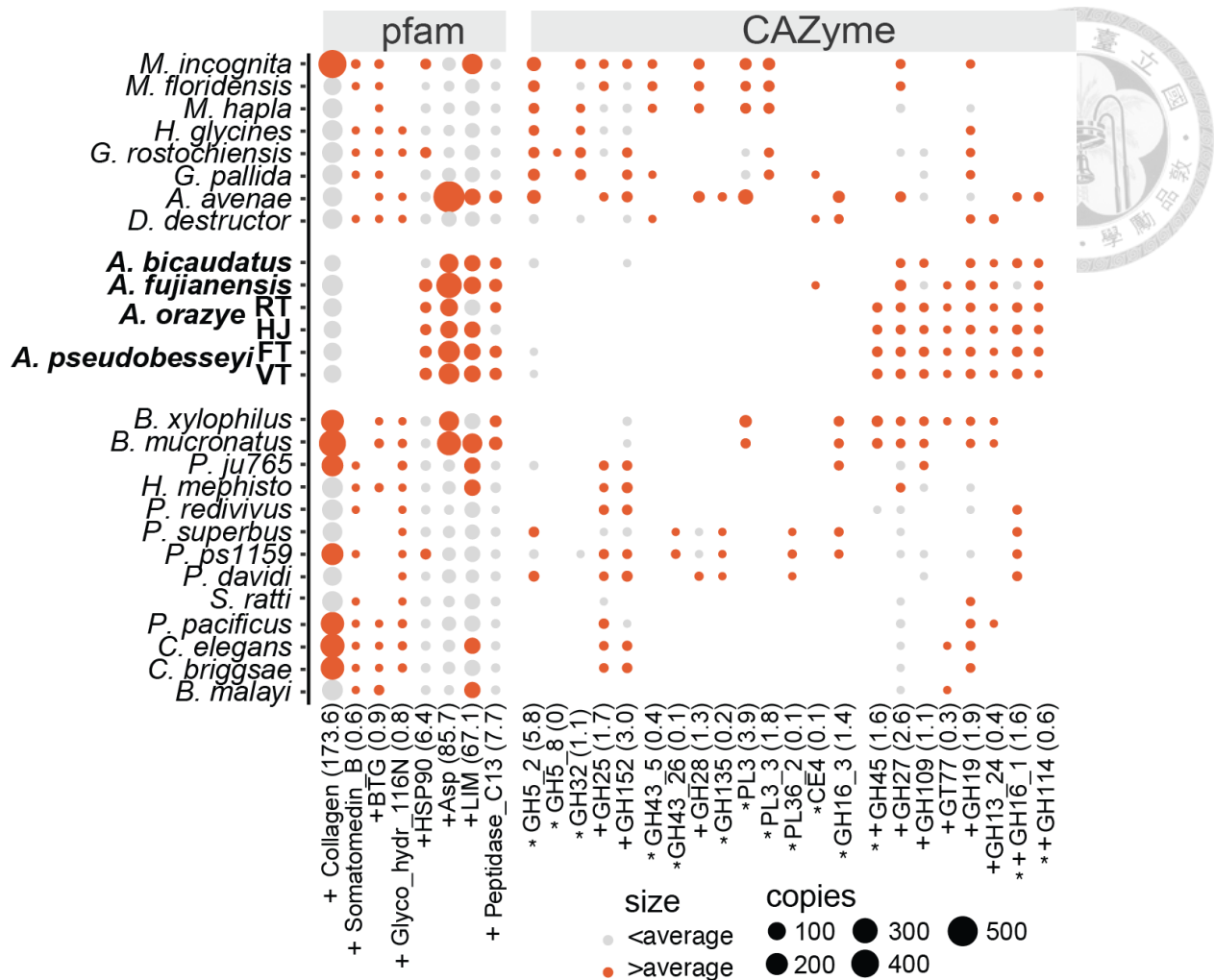


Figure 10 Copy number distribution of Pfam and CAZyme.

The dot size represents the copy number of each domains and the colour denotes the domain copy number with more or less than the average number shown in brackets. The "+" symbol indicates Pfam and CAZyme families that are significantly different between Aphelenchoides species and other nematodes. The "*" symbol denotes CAZyme families that have been identified as having been acquired horizontally.

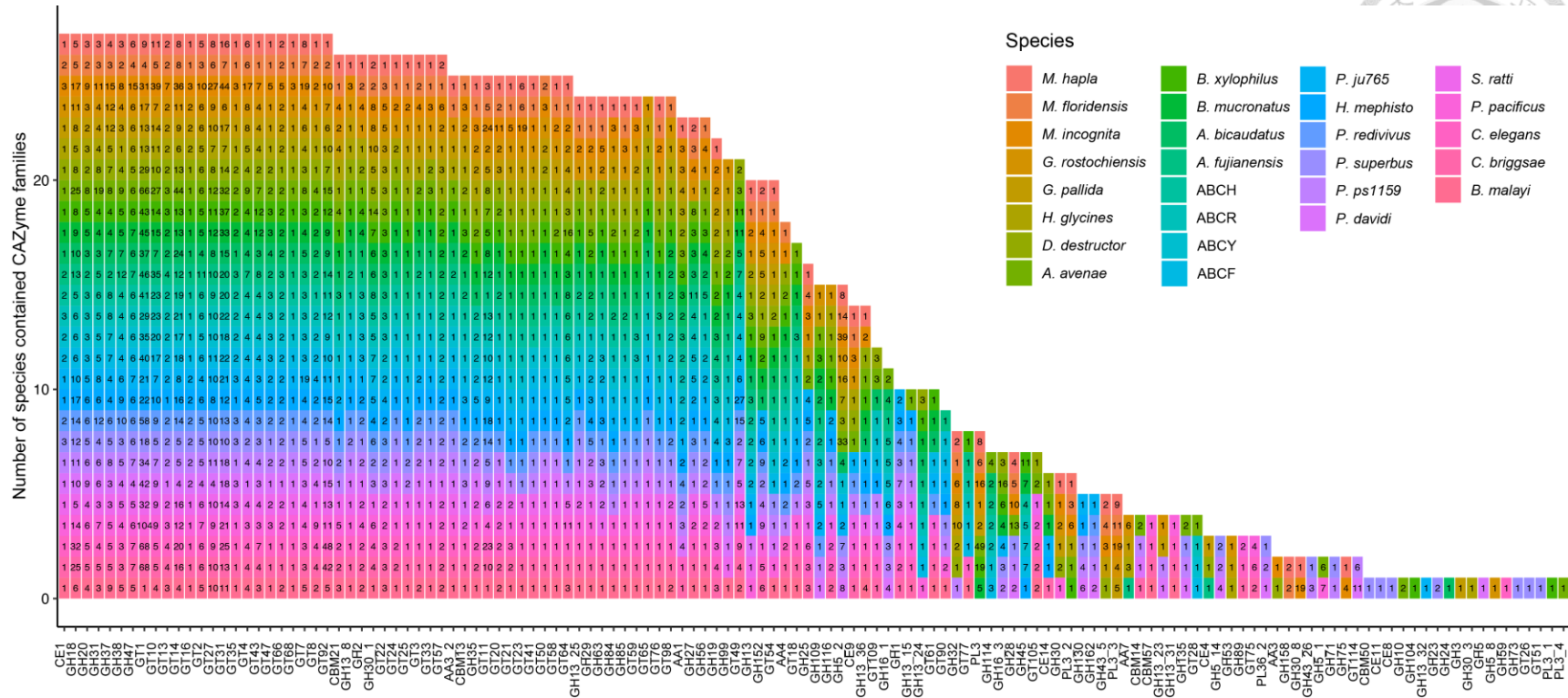
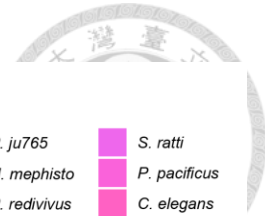


Figure 11 CAZyme abundance in nematodes.

Y-axis present the total number of nematodes containing domains, and different type of CAZyme were shown in X-axis. Different colours denote different nematodes with the dynamic copy number labelled in figures.

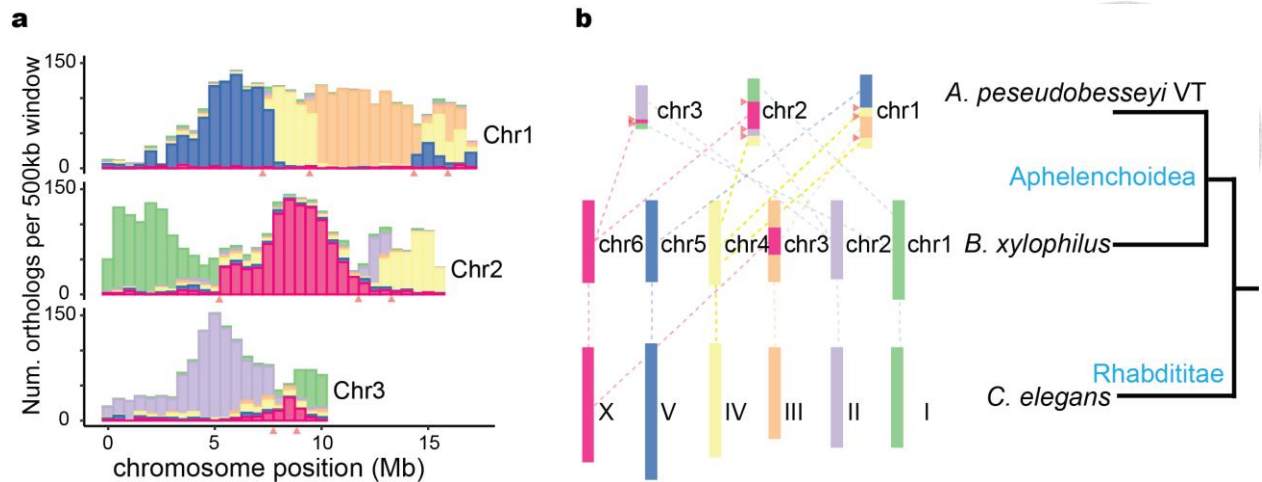


Figure 12 Chromosome evolution of *A. besseyi*.

a. The density of single-copy orthology sets between three chromosomes of *A. pseudobesseyi* VT and their corresponding macro-synteny from six chromosomes of *B. xylophilus*, the putative chromosome fusion sites are labelled with red triangles. **b.** The chromosome fragment colours in nematodes were assigned based on their major corresponding *C. elegans* chromosome using single-copy orthology pairs between given nematodes and *C. elegans*. The size of the chromosome strips in each nematode were adjusted according their chromosome size.

C. elegans

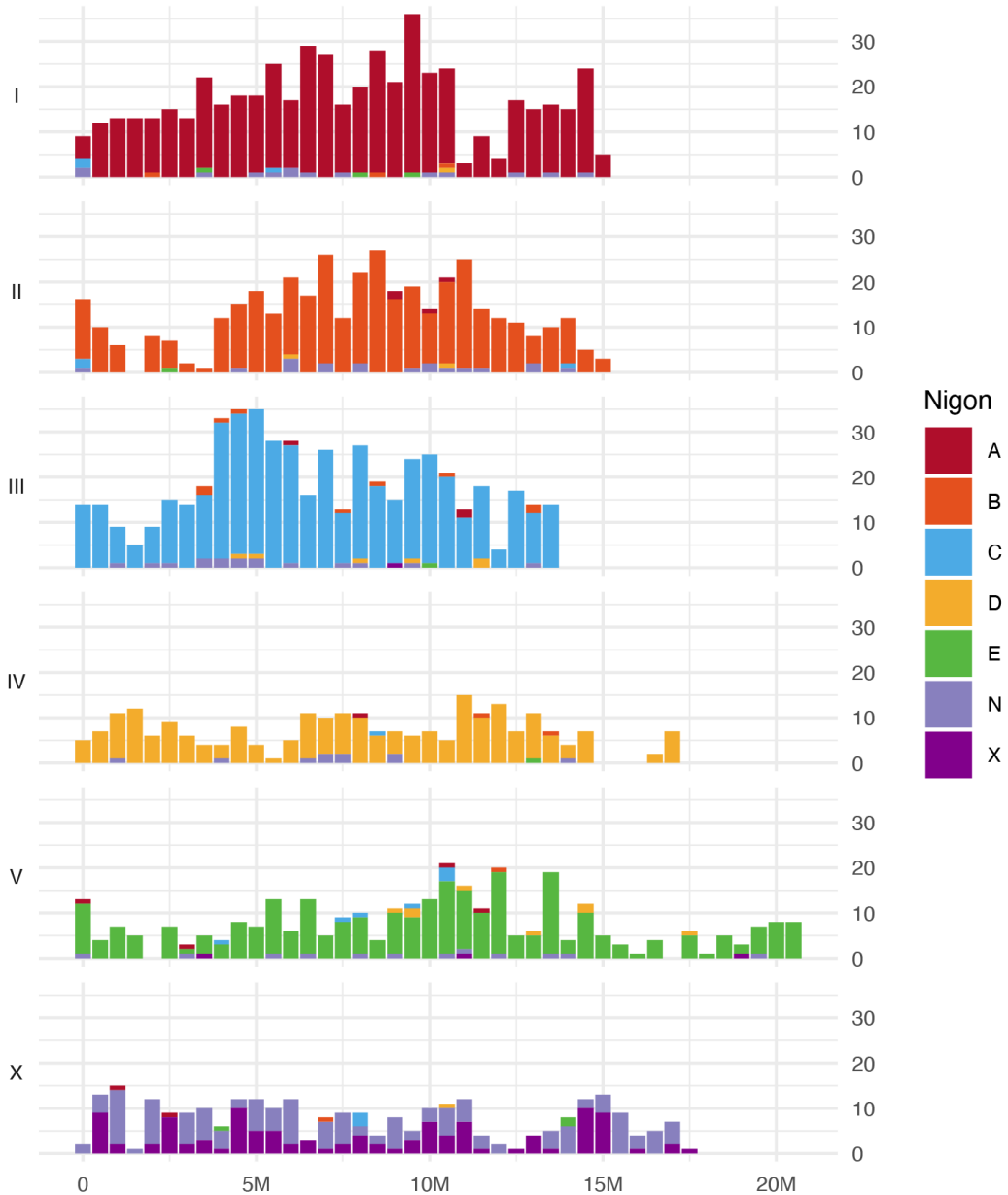


Figure 13 Nigon elements of rhabditid nematodes in *C. elegans*.

The ancestral chromosomes units of 15 rhabditid nematodes were assigned into six chromosomes of *B. xylophilus*, the six elements of A-N denoted the ancestral autosomes and the X element indicated ancestral sex chromosome.

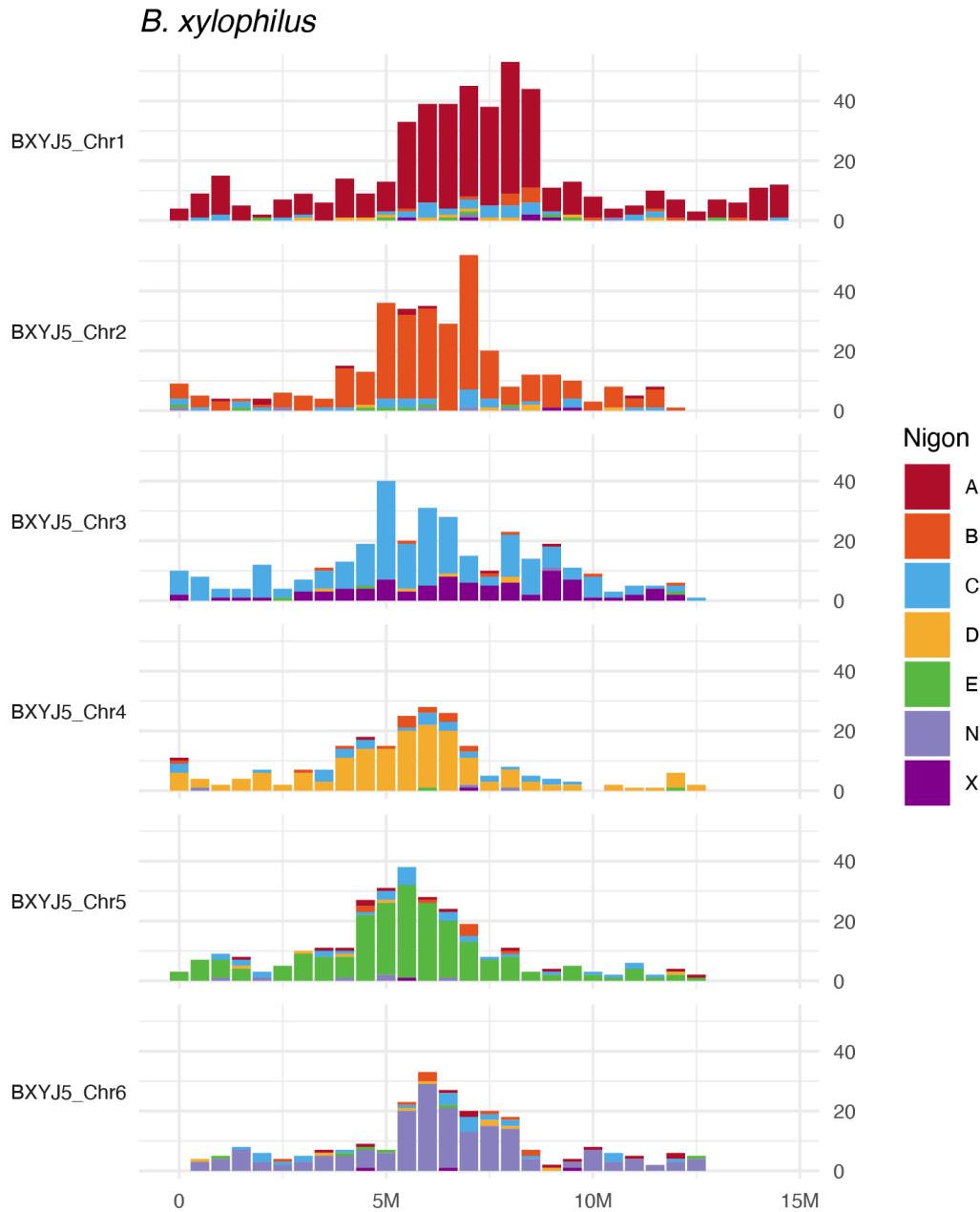


Figure 14 Nigon elements of rhabditid nematodes in *B. xylophilus*.

The ancestral chromosomes of 15 rhabditid nematodes were assigned into six *B. xylophilus* chromosomes, the elements of A-N denote ancestral autosomes and the X element indicated ancestral sex chromosome.

A. pseudobesseyi VT

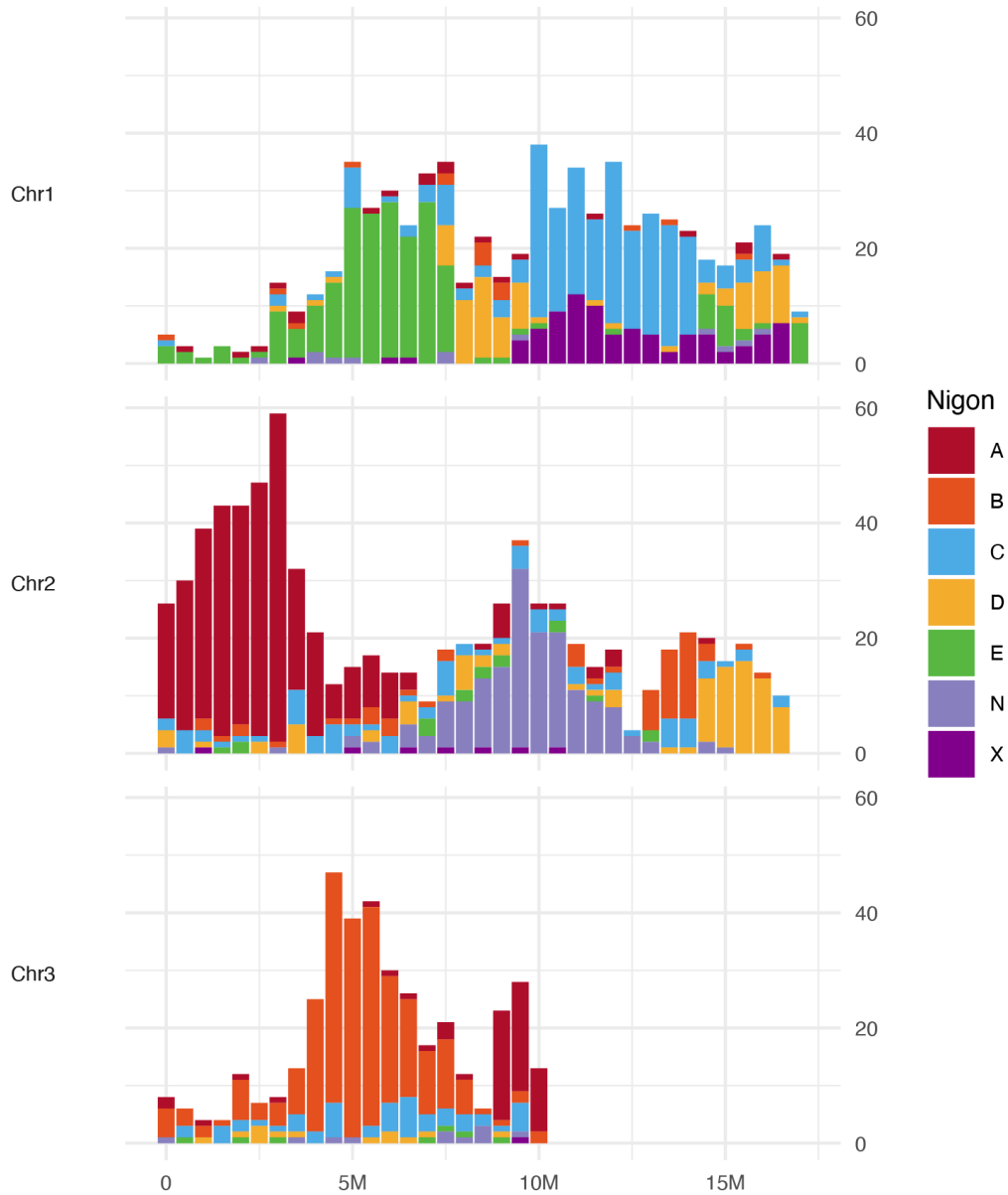


Figure 15 Nigon elements of rhabditid nematodes in *A. pseudobesseyi* VT.

The ancestral chromosomes of 15 rhabditid nematodes were assigned into APVT, the elements of A-N denote ancestral autosomes and the X denote ancestral sex chromosome.

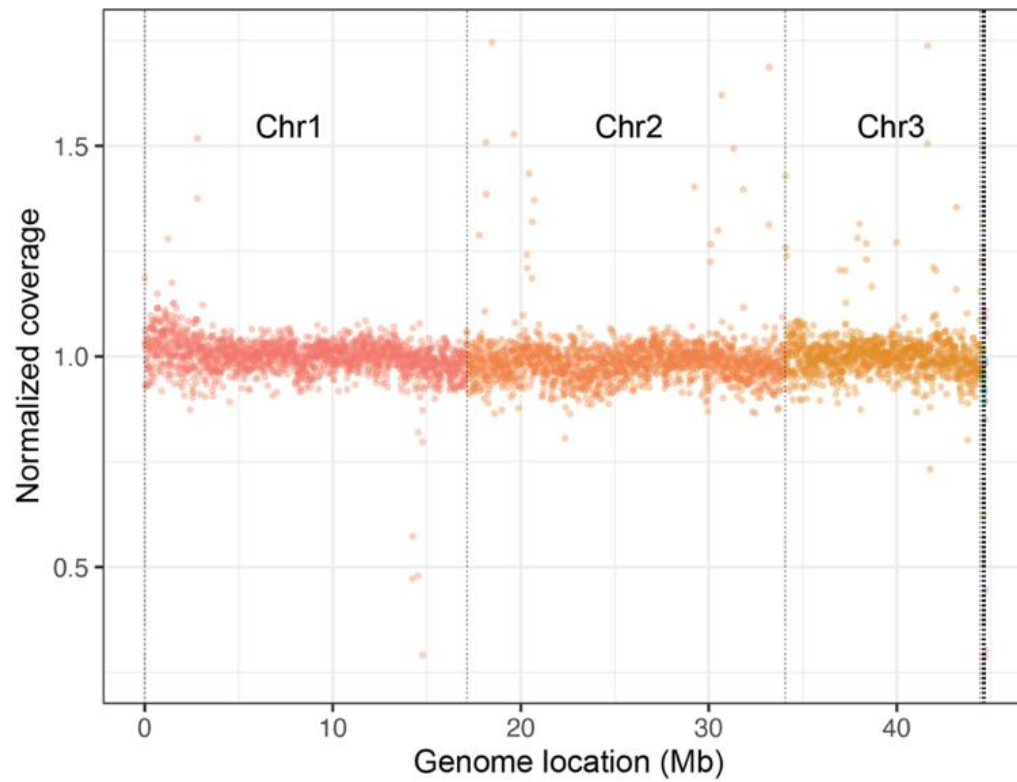


Figure 16 The male Illumina reads of APVT were normalized.

The reads depth of coverage of non-overlapping 10kb windows were normalized by the median coverage of the longest chromosome. Different dot colours denote reads aligned onto different chromosomes.

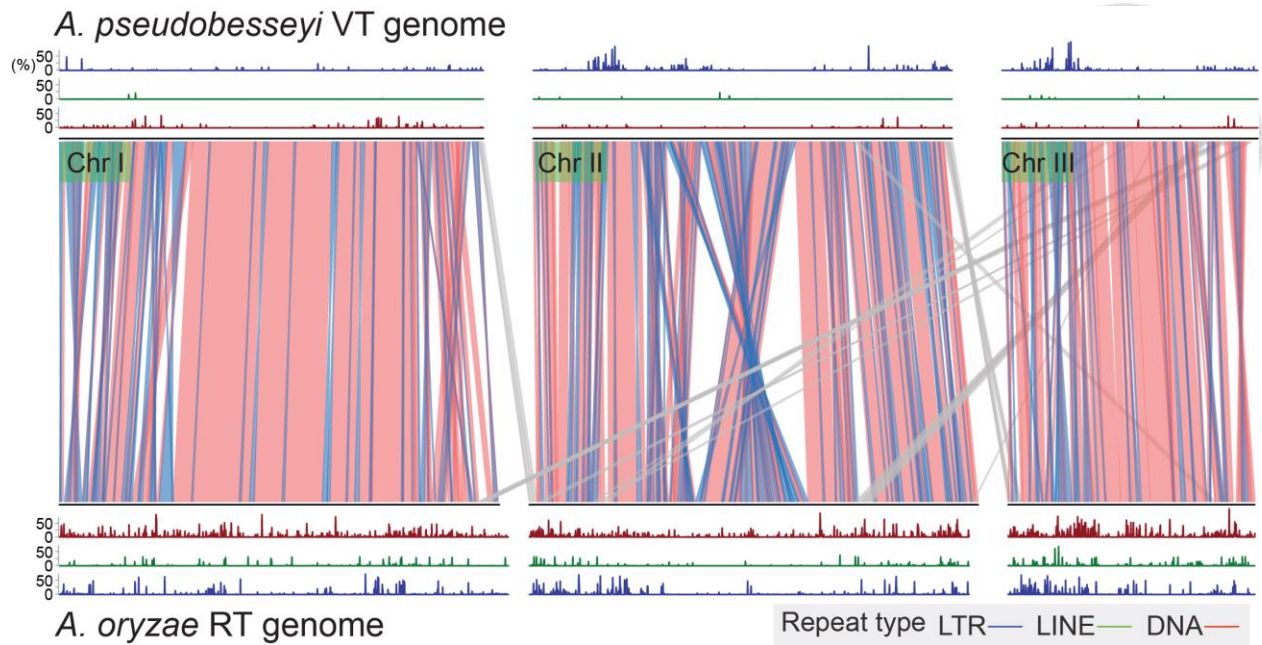


Figure 17 The synteny relationship within the *A. besseyi* species complex.

The synteny blocks are shown between the two strains, the intra-inversion chromosomes denote with blue colours and the gray denote inter-chromosomal rearrangement. correspond to the in distribution of transposon element denote between two. The distribution of DNA transposon (red), long LINE (green) and LTR (blue) between two strains were shown.

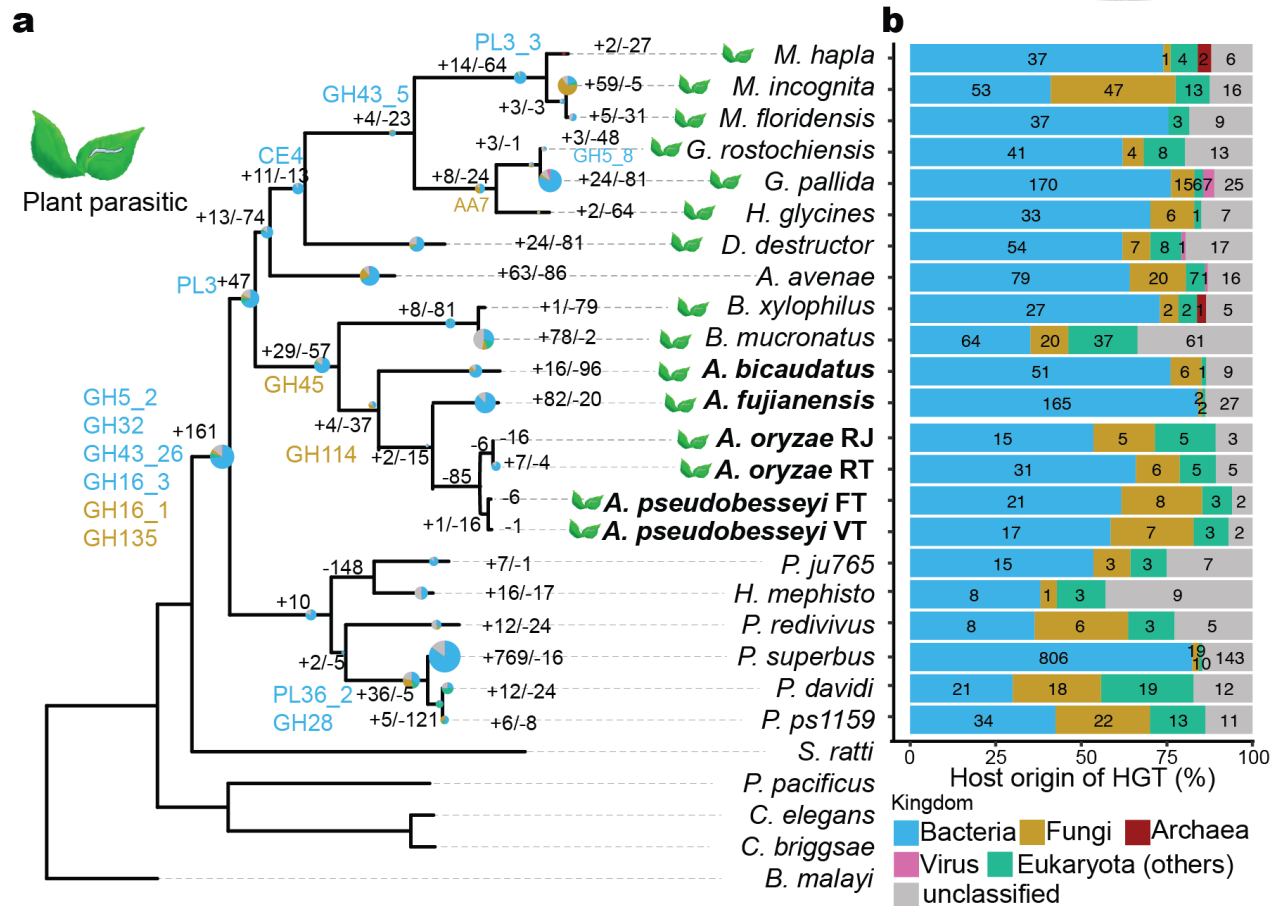


Figure 19 HGT families in PPNs.

a. The original HGT genes were inferred by AI score⁷⁶ across 27 representative nematodes. Candidate original CAZyme are labelled in the branches of the phylogeny and the proportions of HGT orthologues belonging to different potential kingdoms of donor are shown as pie charts. The size of the pie chart corresponds to the total number of HGT orthologues in branches **b**. Proportion of nematode HGT orthogroups from different kingdoms.

GH5_2

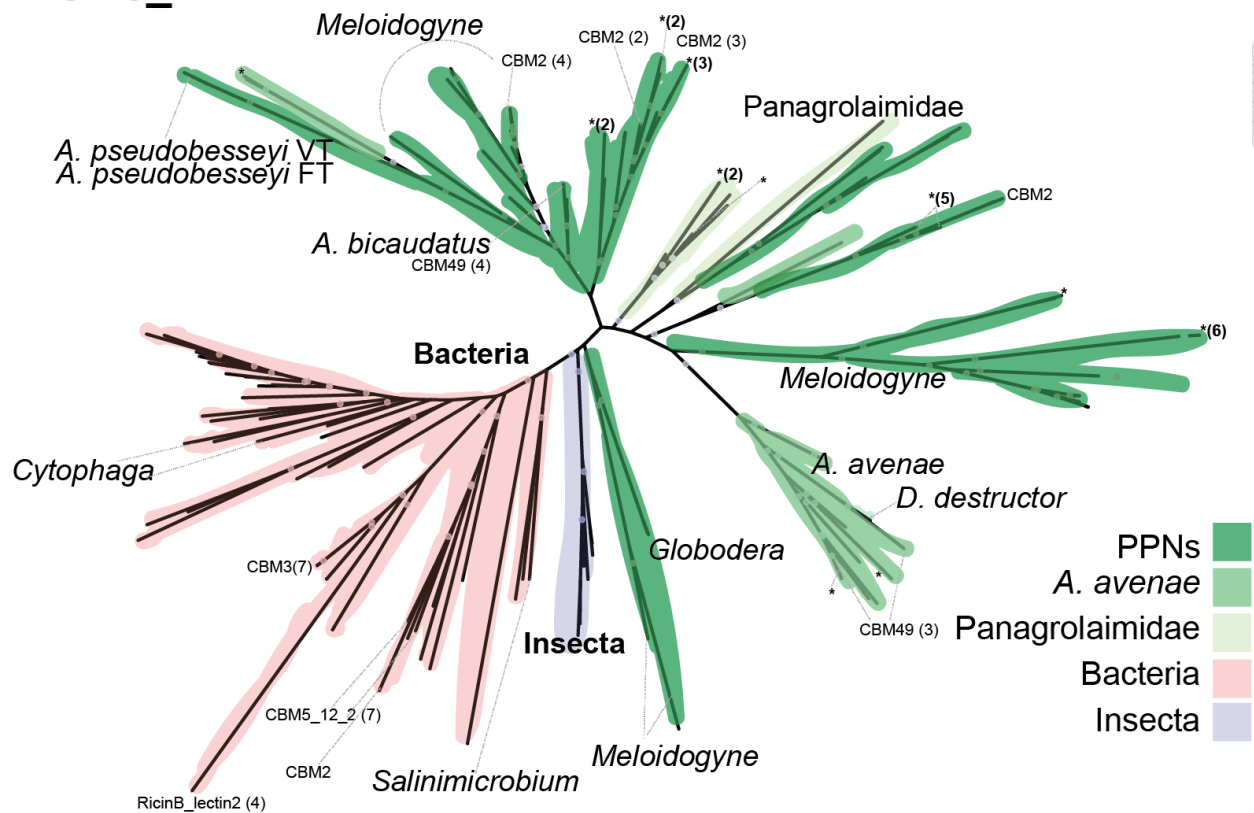


Figure 20 Phylogenetic relationship between GH5 copies in Clade IV nematodes and their candidate bacterial origin.

GH5 copies contains pfam domain were labelled with the domain name and the GH5 copies with $AI < 0$ were labelled with “*”. **Different colours denote different kingdoms and species as shown in legend.**

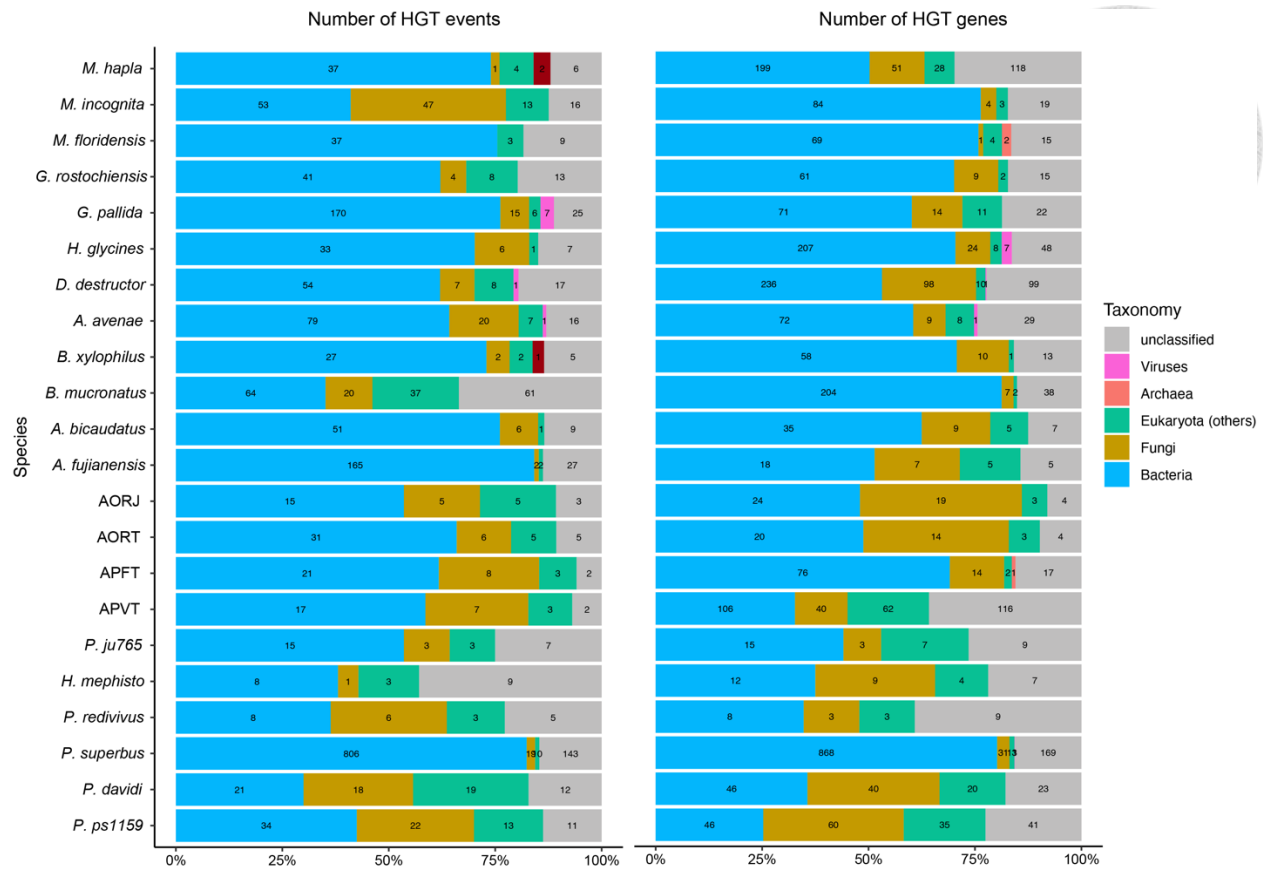
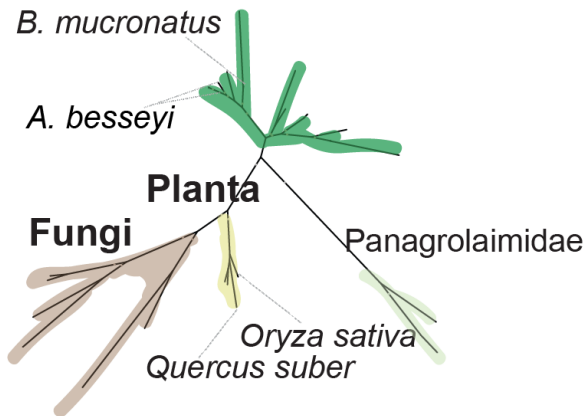


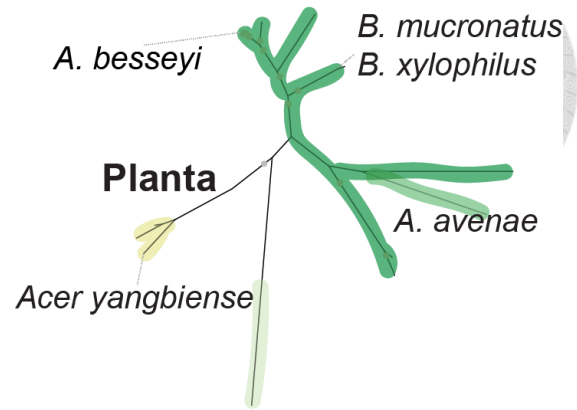
Figure 21 Total number of HGT events and genes in Clade IV nematodes.

Different kingdoms of host origin are denoted with different colours.

OG0003777



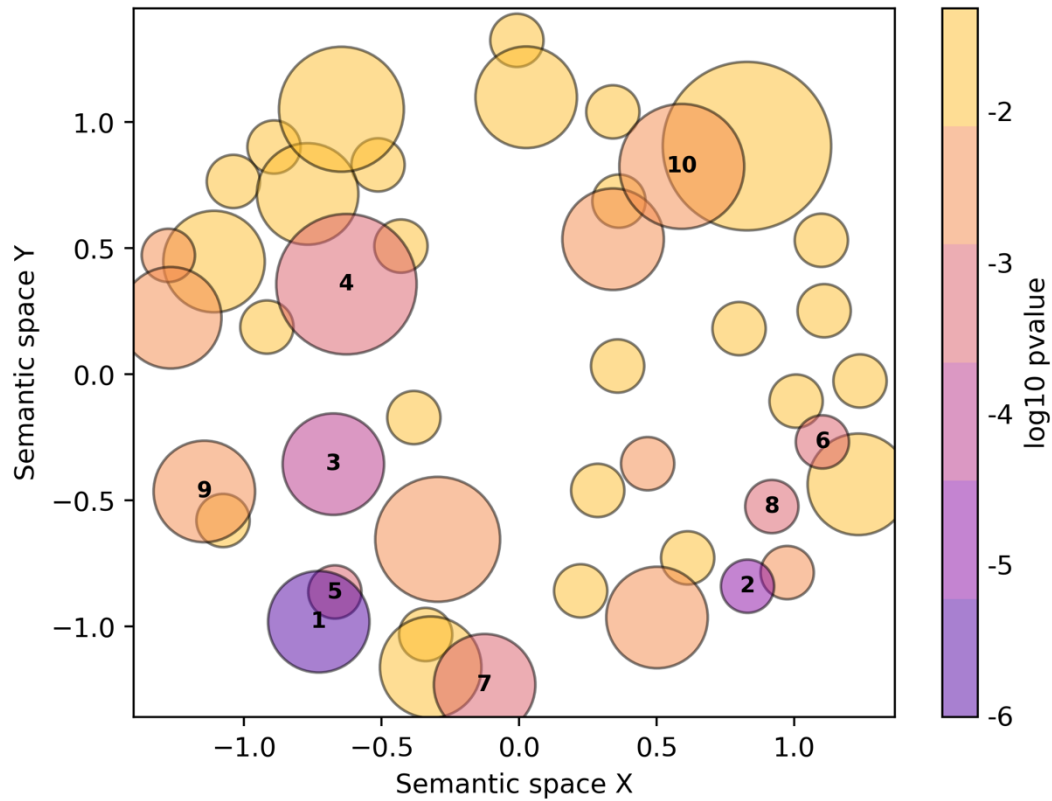
OG0004537



PPNs ■ *A. avenae* ■ Panagrolaimidae ■ Fungi ■ Plantia ■

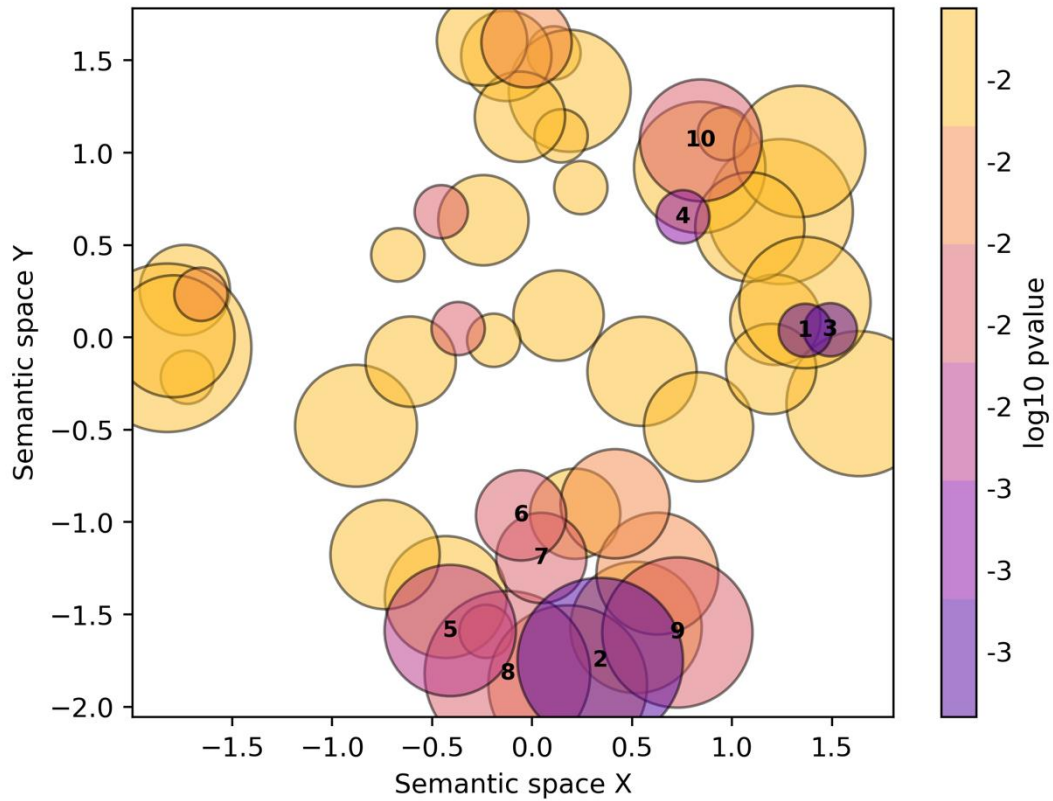
Figure 22 Two HGT families identified to the genes transferred from plants.

Different colours denote different donor kingdoms and different clades of nematodes as labelled in legend.



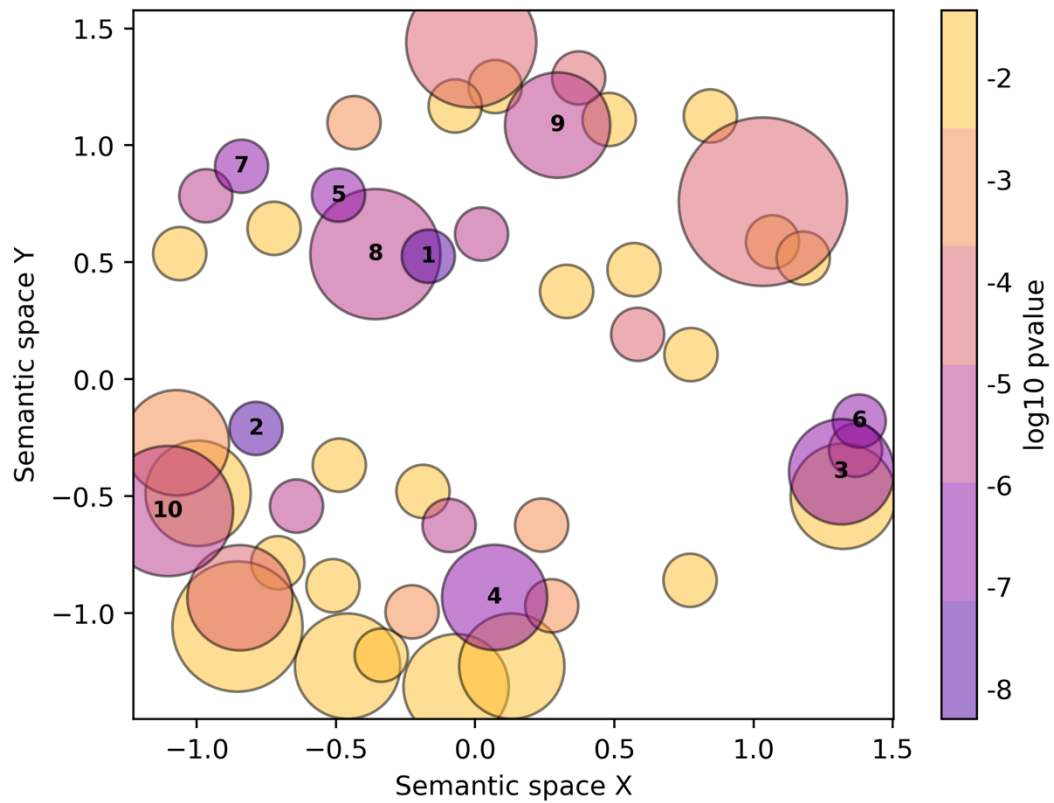
- | | |
|---|---|
| 1. heme transport | 6. porphyrin-containing compound biosy... |
| 2. cysteine biosynthetic process from ... | 7. clathrin-dependent endocytosis invo... |
| 3. detoxification of copper ion | 8. succinyl-CoA metabolic process |
| 4. cellular iron ion homeostasis | 9. wing disc dorsal/ventral pattern fo... |
| 5. dsRNA transport | 10. phagosome-lysosome fusion |

Figure 23 The representative terms of HGT genes originated from plants in *B. mucronatus*.



- | | |
|---|---|
| 1. manganese ion transmembrane transpo... | 6. regulation of pharyngeal pumping |
| 2. negative regulation of translationa... | 7. regulation of production of small R... |
| 3. calcium ion transmembrane transport | 8. negative regulation of sprouting an... |
| 4. subsynaptic reticulum organization | 9. positive regulation of potassium io... |
| 5. negative regulation of extrinsic ap... | 10. bicellular tight junction assembly |

Figure 24 The representative terms of 597 families were gained in Clade IV.



- | | |
|---|--|
| 1. bradykinin catabolic process | 6. response to dietary excess |
| 2. ubiquitin recycling | 7. insulin metabolic process |
| 3. amyloid-beta clearance | 8. cellular response to histamine |
| 4. positive regulation of protein olig... | 9. protein heterooligomerization |
| 5. trehalose metabolic process | 10. negative regulation of phosphatidyl... |

Figure 25 The representative terms of 60 families were lost in Clade IV.

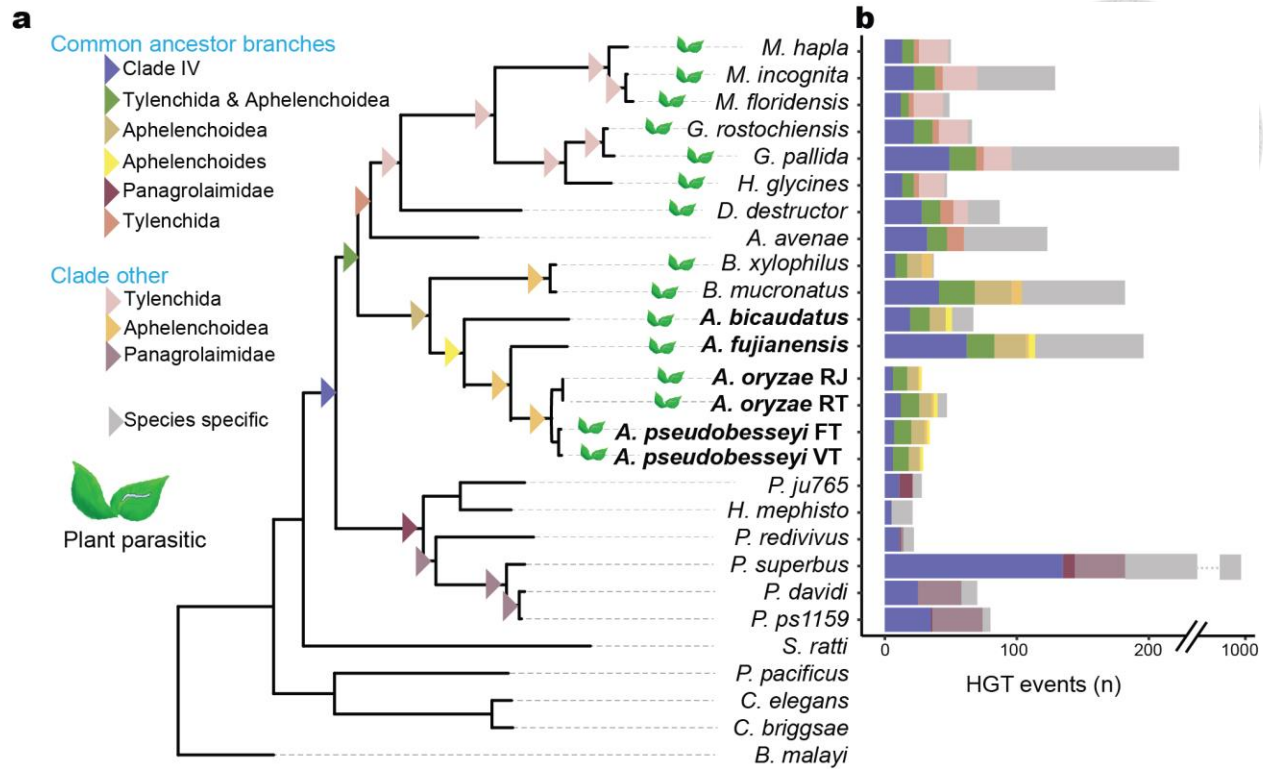


Figure 26 HGT families occurred time points among Clade IV nematodes.

a. HGT events took place among branches of species phylogeny were inferred according to the dollop results, the branches containing HGT families are labelled with the triangle marked. **b.** Number of HGT events corresponding to the label of branches in phylogeny.

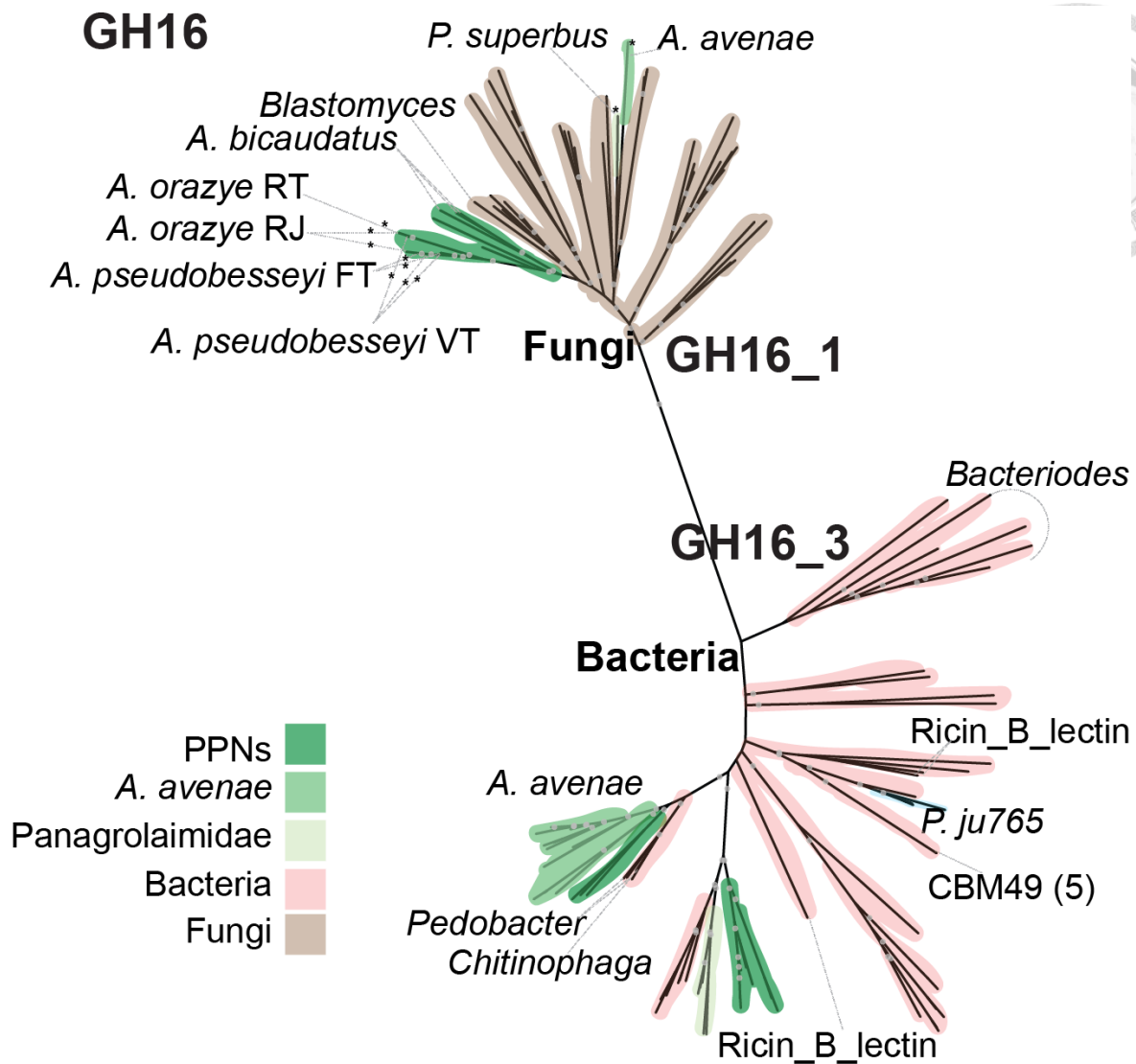


Figure 27 Phylogeny of GH16 in Clade IV nematodes.

Two subfamilies of GH16 copies were divided into two clades identifying to be origin from Bacteria and Fungi, respectively. **Different colours denote different kingdoms and species as shown in legend.**

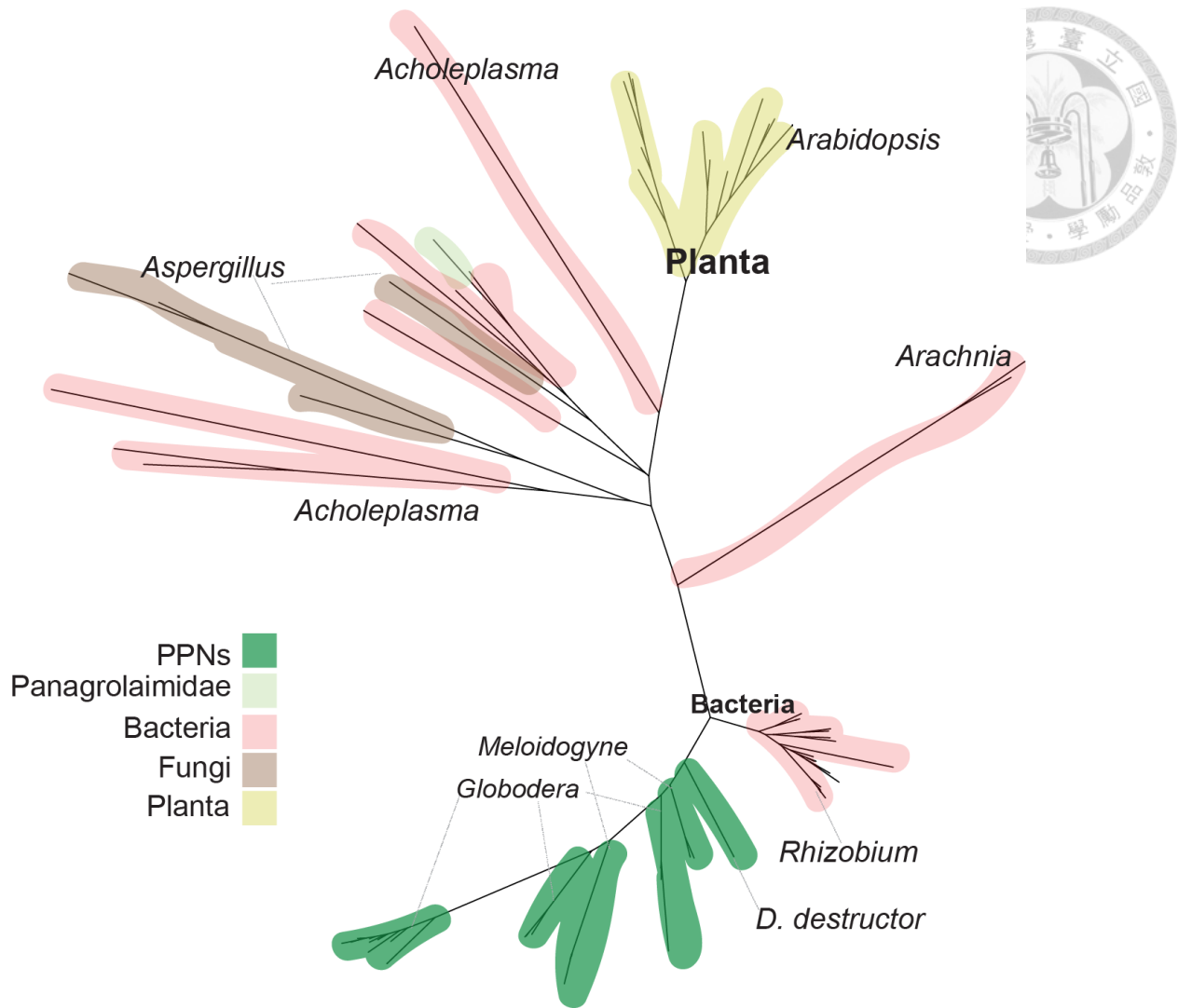


Figure 28 Phylogenetic relationship between GH32 copies in PPNs and *Panagrolaimomorpha* nematodes. Different colours denote different kingdoms and species as shown in legend.

GH43

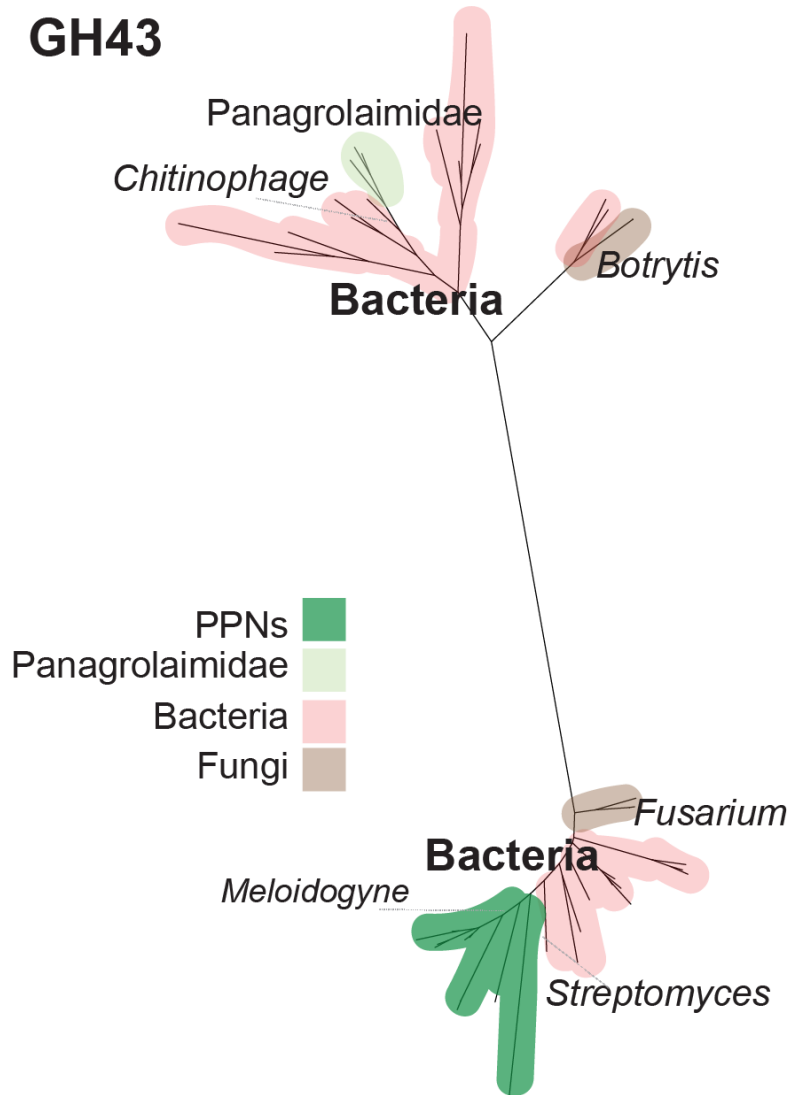


Figure 29 Phylogenetic relationship between GH43 copies in PPNs and *Panagrolaimomorpha* nematodes.

Different colours denote different kingdoms and species as shown in legend.

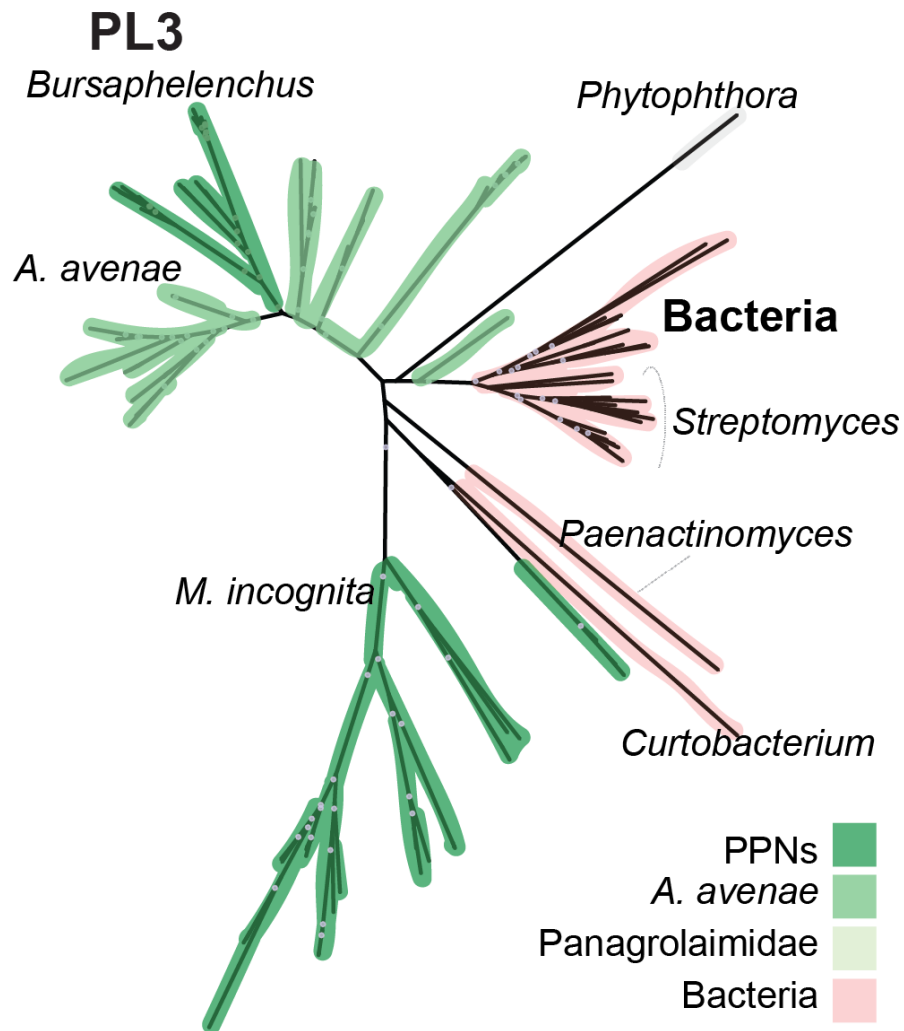


Figure 30 Phylogenetic relationship between PL3 copies in PPNs and their donor origin copies.

Different colours denote different kingdoms and species as shown in legend.

GH45

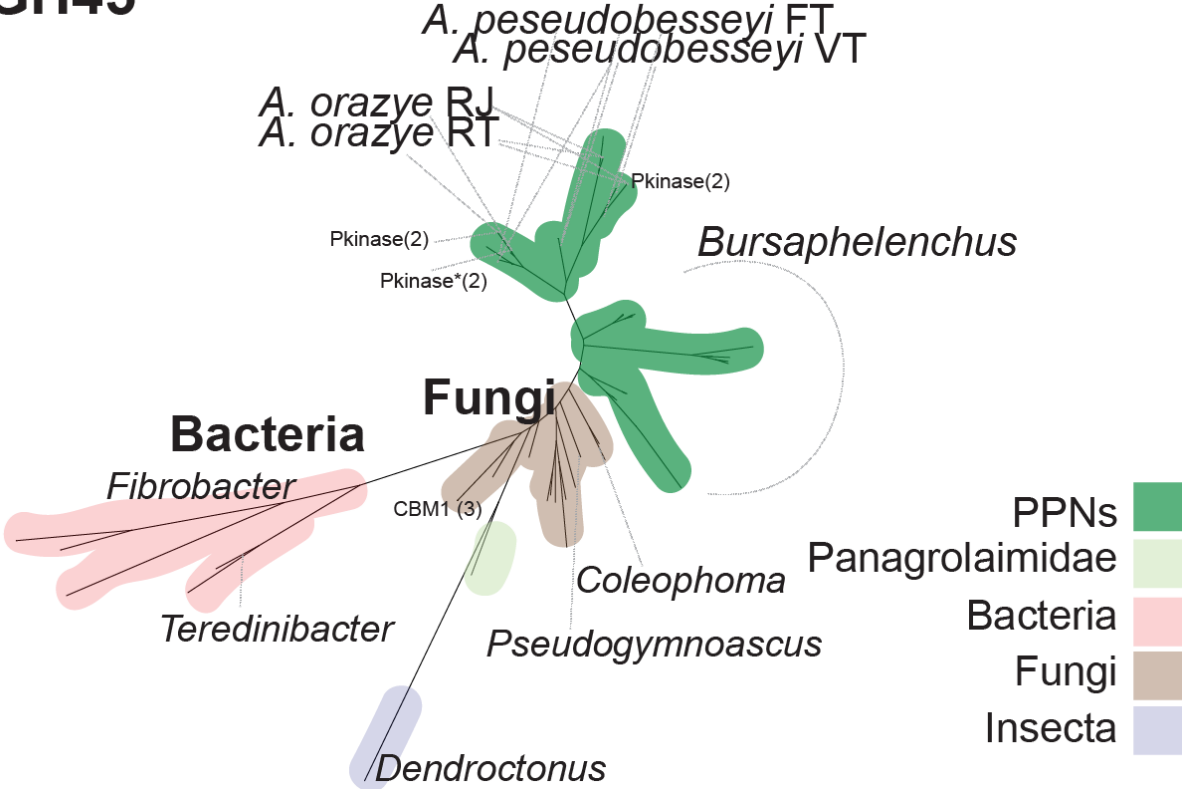


Figure 31 GH45 relationship between *Aphelenchoidea* nematodes and their candidate origin denoted with different colours.

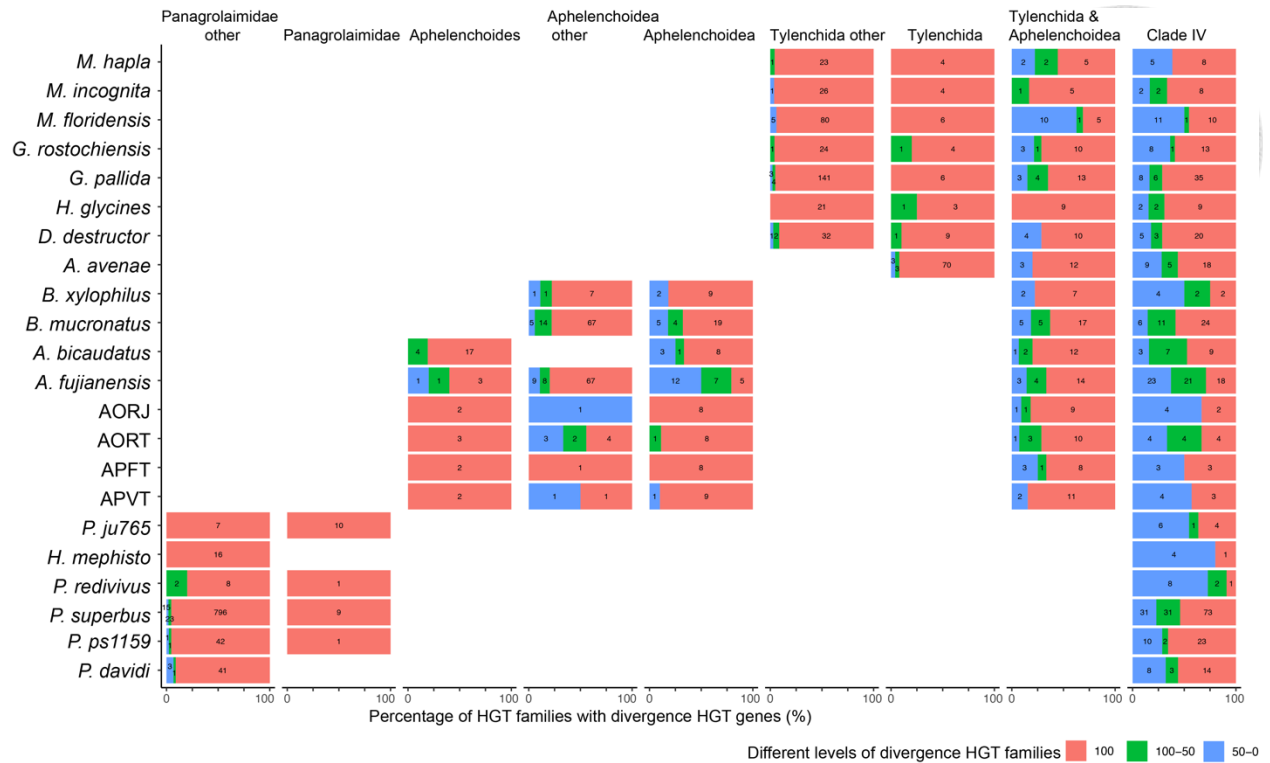


Figure 32 Different levels of HGT families having HGT genes.

The proportion of genes identified as HGT in HGT families were furtherly combined with the HGT took place information, allowing us to observe the fate of HGT genes after being transferred from its donor species

Tables

Table 1 Reads information of *Aphelenchoides* used in this study.



Organism	Sequencing platform	Data type	Num. seqs	total length	avg. length	depth
<i>Aphelenchoides pseudobesseyi</i> FT	Pacbio	gDNA	1,233,626	6,604,279,819	5,354	148
<i>Aphelenchoides oryzae</i> RT	Pacbio	gDNA	468,883	3,302,554,445	7,044	70
<i>Aphelenchoides fujianensis</i> Dali	Oxford Nanopore	gDNA	3,262,577	28,410,132,614	8,708	198
<i>Aphelenchoides bicaudatus</i> Fsh	Oxford Nanopore	gDNA	2,270,737	19,605,212,642	8,634	422
<i>Aphelenchoides pseudobesseyi</i> VT	Oxford Nanopore	gDNA	1,563,051	5,045,207,193	3,228	113
<i>Aphelenchoides oryzae</i> RJ	Oxford Nanopore	gDNA	1,529,061	8,790,197,869	5,749	188
<i>Aphelenchoides pseudobesseyi</i> FT	Illumina	gDNA	16,852,727	2,500,146,809	150	56
<i>Aphelenchoides oryzae</i> RT	Illumina	gDNA	8,585,567	1,275,850,852	150	27
<i>Aphelenchoides fujianensis</i> Dali	Illumina	gDNA	11,414,155	1,709,470,422	150	12
<i>Aphelenchoides bicaudatus</i> Fsh	Illumina	gDNA	10,648,060	1,592,638,316	150	34
<i>Aphelenchoides pseudobesseyi</i> VT	Illumina	gDNA	11,304,143	1,691,579,704	150	38
<i>Aphelenchoides oryzae</i> RJ	Illumina	gDNA	8,450,881	1,264,219,242	150	27
<i>Aphelenchoides pseudobesseyi</i> FT	Illumina	mRNA	17,745,043	2,417,489,232	150	54
<i>Aphelenchoides oryzae</i> RT	Illumina	mRNA	20,219,016	3,021,879,538	150	64
<i>Aphelenchoides fujianensis</i> Dali	Illumina	mRNA	9,115,682	1,297,890,588	150	9
<i>Aphelenchoides bicaudatus</i> Fsh	Illumina	mRNA	9,551,535	1,348,142,645	150	29
<i>Aphelenchoides pseudobesseyi</i> VT	Illumina	mRNA	10,578,794	1,479,646,550	150	33
<i>Aphelenchoides oryzae</i> RJ	Illumina	mRNA	11,779,704	1,656,862,488	150	35

Table 2 *Bursaphelenchus xylophilus* telomere identified in four chromosomes.

Chromosome	Region	START	END	Num. copy	Type
BXYJ5_Chr1	START	1	1,674	279	TTAGGC
BXYJ5_Chr3	END	12,729,358	12,730,644	214	TTAGGC
BXYJ5_Chr4	START	15	1,311	216	TTAGGC
BXYJ5_Chr4	END	12,792,168	12,794,859	448	TTAGGC
BXYJ5_Chr6	END	12,821,428	12,822,217	129	TTAGGC

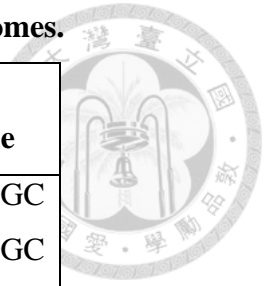


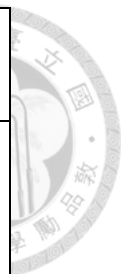
Table 3 Genome statistic of 27 representative nematodes.

scientific name	<i>A. pseudobesseyi</i>		<i>A. oryzae</i>		<i>A. bicaudatus</i>	<i>A. fujianensis</i>
	APFT	APVT	AORT	AORJ		
assembly size	44,700,143	44,715,555	47,421,222	46,756,815	46,428,495	143,834,666
Num.scaffolds	80	28 (3 Chr)	109	61	106	782
longest scaffold length	11,439,185	17,135,672	4,526,991	6,283,131	12,153,480	4,270,062
scaffold average	558,752	1,596,984	435,057	766,505	438,005	183,932
scaffold median	108,987	5,000	137,696	14,464	13,353	73,426
N50	2,203,724	16,939,329	1,080,231	4,086,162	11,905,581	553,027
N90	295,630	10,346,542	196,117	874,258	486,805	94,252
L50	5	2	10	5	2	68
L90	26	3	43	14	10	320
Num.genes	12,916	12,948	12,427	11,652	11,701	20,051
gene length	26,885,937	25,543,038	25,360,769	24,336,843	24,629,164	42,825,561
exon length	18,533,664	17,156,448	17,740,380	16,407,717	15,669,537	25,966,002
exon coverage	41.5	38.4	37.4	35.1	33.7	18.1
intron region	8,352,273	8,386,590	7,620,389	7,929,126	8,959,627	16,859,559
intron coverage	18.7	18.8	16.1	17.0	19.3	11.7
intergenic region	17,814,206	19,172,517	22,060,453	22,419,972	21,799,331	101,009,105
intergenic coverage	39.9	42.9	46.5	48.0	47.0	70.2
BUSCO completeness	81.3	78.2	81.2	78.9	76.4	66.6





scientific name	<i>B. xylophilus</i>	<i>B. mucronatus</i>	<i>M. floridensis</i>	<i>M. hapla</i>	<i>M. incognita</i>	<i>G. pallida</i>
assembly size	78,274,012	73,053,838	74,846,012	53,017,507	183,531,997	123,625,196
Num.scaffolds	11 (6Chr)	72 (6 Chr)	8,887	3,452	12,091	6,873
longest scaffold length	15,101,617	11,930,853	88,393	360,446	391,312	599,721
scaffold average	7,115,819	1,014,637	8,422	15,359	15,179	17,987
scaffold median	12,140,115	40,957	5,460	5,799	5,576	1,699
N50	12,794,883	11,524,788	13,261	37,608	38,588	120,481
N90	12,140,115	11,262,134	3,816	6,538	5,237	10,665
L50	3	4	1,685	372	1,209	296
L90	6	6	5,769	1,607	6,296	1,437
Num.genes	15,884	13,696	12,762	14,419	43,718	16,403
gene length	42,968,487	48,820,938	29,784,141	26,381,312	87,804,348	39,553,007
exon length	20,740,098	38,526,538	13,523,574	15,103,235	43,654,679	17,623,861
exon coverage	26.5	52.7	18.1	28.5	23.8	14.3
intron region	22,228,389	10,294,400	16,260,567	11,278,077	44,149,669	21,929,146
intron coverage	28.4	14.1	21.7	21.3	24.1	17.7
intergenic region	35,305,525	24,232,900	45,061,871	26,636,195	95,727,649	84,072,189
intergenic coverage	45.1	33.2	60.2	50.2	52.2	68.0
BUSCO completeness	83.5	89.4	46.5	59.8	75.4	50.9



scientific name	<i>G. rostochiensis</i>	<i>D. destructor</i>	<i>H. glycines</i>	<i>A. avenae</i>	<i>P. ps1159</i>	<i>P. superbus</i>
assembly size	95,876,286	111,138,200	141,354,287	264,011,204	84,957,092	76,664,811
Num.scaffolds	4,281	1,761	267	18,660	17,628	53,192
longest scaffold length	688,384	3,556,246	21,025,901	5,486,223	142,873	29,093
scaffold average	22,396	63,111	529,417	14,149	4,819	1,441
scaffold median	2,432	3,200	8,484	3,073	2,278	889
N50	88,688	561,030	16,265,615	141,150	9,924	1,894
N90	16,837	64,313	7,602,578	3,282	2,005	642
L50	278	47	4	385	2,232	10,051
L90	1,175	260	9	8,251	9,419	39,638
Num.genes	14,308	13,931	11,882	43,185	24,802	19,887
gene length	39,571,094	57,122,052	41,632,132	74,228,574	34,241,233	22,121,968
exon length	18,198,124	17,903,483	14,661,405	37,876,380	27,215,315	17,207,361
exon coverage	19.0	16.1	10.4	14.3	32.0	22.4
intron region	21,372,970	39,218,569	26,970,727	36,352,194	7,025,918	4,914,607
intron coverage	22.3	35.3	19.1	13.8	8.3	6.4
intergenic region	56,305,192	54,016,148	99,722,155	189,782,630	50,715,859	54,542,843
intergenic coverage	58.7	48.6	70.5	71.9	59.7	71.1
BUSCO completeness	71.0	73.8	57.3	79.1	26.8	57.8



scientific name	<i>P. ju765</i>	<i>H. mephisto</i>	<i>P. redivivus</i>	<i>P. davidi</i>	<i>P. redivivus</i>	<i>S. rattii</i>
assembly size	64,262,255	61,428,783	65,093,147	118,052,509	43,166,851	64,262,255
Num.scaffolds	13,268	880	940	39,598	136	13,268
longest scaffold length	907,536	2,546,343	2,280,433	42,214	16,759,152	907,536
scaffold average	4,843	69,805	69,248	2,981	317,403	4,843
scaffold median	1,902	4,110	12,561	2,013	3,408	1,902
N50	10,861	313,311	262,414	4,454	11,693,564	10,861
N90	1,881	73,741	48,805	1,376	748,239	1,881
L50	1,259	53	64	7,745	2	1,259
L90	6,668	193	289	26,333	7	6,668
Num.genes	20,756	12,612	24,249	25,806	12,464	20,756
gene length	29,788,750	47,428,429	38,057,022	27,701,286	21,972,961	29,788,750
exon length	27,088,072	46,305,972	26,892,733	20,888,203	17,528,583	27,088,072
exon coverage	42.2	75.4	41.3	17.7	40.6	42.2
intron region	2,700,678	1,122,457	11,164,289	6,813,083	4,444,378	2,700,678
intron coverage	4.2	1.8	17.2	5.8	10.3	4.2
intergenic region	34,473,505	14,000,354	27,036,125	90,351,223	21,193,890	34,473,505
intergenic coverage	53.6	22.8	41.5	76.5	49.1	53.6
BUSCO completeness	72.4	80.4	83.9	72.9	79.9	72.4



scientific name	<i>P. pacificus</i>	<i>C. elegans</i>	<i>C. briggsae</i>	<i>B. malayi</i>
assembly size	158,500,396	100,286,401	108,384,165	88,235,797
Num.scaffolds	47	7 (6 chr)	367	197
longest scaffold length	39,556,110	20,924,180	21,540,570	24,943,668
scaffold average	3,372,349	14,326,629	295,325	447,897
scaffold median	47,073	15,279,421	4,598	18,006
N50	23,915,096	17,493,829	17,485,439	14,214,749
N90	17,019,893	13,783,801	14,578,851	13,467,244
L50	3	3	3	3
L90	6	6	6	5
Num.genes	26,342	20,184	20,821	10,878
gene length	117,703,652	63,598,046	64,877,833	48,919,897
exon length	34,886,008	21,069,588	24,904,897	14,776,100
exon coverage	22.0	21.0	23.0	16.7
intron region	82,817,644	42,528,458	39,972,936	34,143,797
intron coverage	52.3	42.4	36.9	38.7
intergenic region	40,796,744	36,688,355	43,506,332	39,315,900
intergenic coverage	25.7	36.6	40.1	44.6
BUSCO completeness	91.1	99.7	98.9	98.9

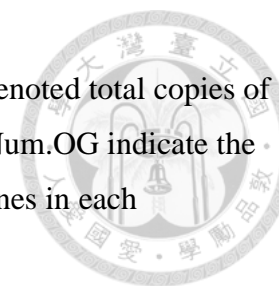
Table 4 GO enrichment of 27 families acquired from plants in *B. mucronatus*.

GO.ID	Term	Annotated	Significant	Expected	P values
GO:0015886	heme transport	9	4	0.09	9.80E-07
GO:0006535	cysteine biosynthetic process from serin...	5	3	0.05	8.60E-06
GO:0071585	detoxification of cadmium ion	8	3	0.08	4.70E-05
GO:0010273	detoxification of copper ion	9	3	0.09	7.10E-05
GO:0070574	cadmium ion transmembrane transport	10	3	0.1	0.0001
GO:0006879	cellular iron ion homeostasis	34	4	0.33	0.0003
GO:0033227	dsRNA transport	17	3	0.17	0.00054
GO:0006779	porphyrin-containing compound biosynthet...	20	3	0.2	0.00089
GO:0006104	succinyl-CoA metabolic process	5	2	0.05	0.00092
GO:0061883	clathrin-dependent endocytosis involved ...	5	2	0.05	0.00092
GO:0090383	phagosome acidification	5	2	0.05	0.00092
GO:0048190	wing disc dorsal/ventral pattern formati...	23	3	0.22	0.00135
GO:0051382	kinetochore assembly	7	2	0.07	0.0019
GO:1903543	positive regulation of exosomal secretio...	8	2	0.08	0.00252
GO:0006105	succinate metabolic process	8	2	0.08	0.00252
GO:0090385	phagosome-lysosome fusion	8	2	0.08	0.00252
GO:0006622	protein targeting to lysosome	9	2	0.09	0.00322
GO:0045732	positive regulation of protein catabolic...	130	4	1.27	0.00324
GO:0006334	nucleosome assembly	31	3	0.3	0.00325
GO:0032510	endosome to lysosome transport via multi...	10	2	0.1	0.004

GO:0061621	canonical glycolysis	11	2	0.11	0.00486
GO:0060628	regulation of ER to Golgi vesicle-mediat...	11	2	0.11	0.00486
GO:0019076	viral release from host cell	12	2	0.12	0.0058
GO:0070199	establishment of protein localization to...	12	2	0.12	0.0058
GO:0032042	mitochondrial DNA metabolic process	13	2	0.13	0.00681

Table 5 HGT enriched Pfam in two major nematodes HGT episodes.

The largest AI genes present Pfam were selected to representative the origin host in HGT families. Num.pfam denoted total copies of pfam identified in clade IV or PPN families. Num.sp indicated number of unique species contained this motif. Num.OG indicate the number of orthogroups contained this domain. Num.sp in each OGs denoted number of species present HGT genes in each orthogroups, and the bracket number indicated number of those species with identified domain.



Node	Pfam	Num.pfam	Num.sp	Num.OG	Num.sp in each OGs	The closest origin of NR genus
Clade IV	ABC_tran	27	8	5	2(1),5(4),3(3),2(2),2(1)	Agrobacterium,Caulobacter
Clade IV	Aldedh	9	3	4	2(1),3(3),2(1),2(2)	Caulobacter,Escherichia
Clade IV	ADH_N	51	15	3	14(14),2(2),5(4)	Escherichia,Caulobacter
Clade IV	DEAD	14	6	3	4(2),4(2),4(4)	Bremia,Brassica
Clade IV	GTP_EFTU	4	3	3	2(1),3(1),2(2)	Alcaligenes,Stenotrophomonas
Clade IV	Response_reg	8	2	3	2(2),2(2),2(2)	Escherichia,Sphingomonas
Clade IV	ADH_zinc_N	51	15	2	14(2),5(2)	Caulobacter,Nannizzia
Clade IV	Glyco_hydro_16	43	9	2	11(2),7(7)	Aspergillus,Stylonychia
Clade IV	Chromo	24	8	2	9(8),21(1)	Rhizoctonia,Parasponia
Clade IV	Helicase_C	13	7	2	4(2),4(1)	Brassica,Stenotrophomonas
Clade IV	Acetyltransf_1	10	6	2	4(4),3(3)	Escherichia,Labrys
Clade IV	RVT_1	14	5	2	4(1),6(4)	Rhizoctonia,Citrobacter
Clade IV	adh_short_C2	9	5	2	5(4),4(1)	Caulobacter,Candidatus
Clade IV	adh_short	8	5	2	3(3),5(4)	Caulobacter,Escherichia
Clade IV	Pkinase	8	3	2	4(1),4(3)	Rozella,Naegleria
Clade IV	Carboxyl_trans	7	3	2	2(2),2(2)	Rhodoblastus,NA
Clade IV	Glutaredoxin	6	3	2	2(2),3(3)	Alcaligenes,Caulobacter
Clade IV	AMP-binding	5	3	2	2(2),2(1)	Caenispirillum,Alcaligenes
Clade IV	GTP_EFTU_D2	4	3	2	2(1),3(1)	Caulobacter,Rhizobium
Clade IV	rve	4	3	2	4(2),9(1)	Ceratobasidium,Rhizoctonia

Clade IV	Acyl-CoA_dh_N	3	3	2	2(1),2(2)	Caulobacter,Alicyclobacillus
Clade IV	Ank_5	3	3	2	6(1),3(1)	Citrobacter,NA
Clade IV	CTP_synth_N	3	3	2	2(1),2(2)	Agrobacterium,Sphingopyxis
Clade IV	TPR_16	3	3	2	6(2),2(1)	Botryobasidium,Caulobacter
Clade IV	ABC_membrane	7	2	2	2(1),5(2)	Caulobacter,Caulobacter
Clade IV	Biotin_lipoyl	6	2	2	2(2),2(1)	Caulobacter,Caulobacter
Clade IV	Ras	6	2	2	2(2),2(2)	Absidia,Sesamum
Clade IV	Peptidase_M24	5	2	2	2(2),2(1)	Caulobacter,Alcaligenes
Clade IV	ketoacyl-synt	4	2	2	3(2),5(1)	Caulobacter,Alcaligenes
Clade IV	Pyr_redox_2	4	2	2	2(1),2(2)	,Dyadobacter
Clade IV	GATase	3	2	2	2(1),2(1)	Escherichia,Escherichia
Clade IV	Plug	3	2	2	2(2),2(1)	Stenotrophomonas,Dyadobacter
Clade IV	TPP_enzyme_N	3	2	2	2(2),2(1)	Sphingomonas,Sphingomonas
Clade IV	AAA_19	2	2	2	6(1),2(1)	Thielaviopsis,Pseudomonas
Clade IV	APH	2	2	2	2(1),2(1)	Ponticoccus,NA
Clade IV	PP-binding	2	1	2	3(1),5(1)	Caulobacter,Streptomyces
PPNs	AAA	12	3	3	2(2),3(1),2(1)	Sphingopyxis,Caulobacter
PPNs	JAB	9	8	2	2(2),7(7)	Quercus,Pararhizobium
PPNs	HK	8	5	2	4(1),4(4)	Caballeronia,Caballeronia
PPNs	UQ_con	9	4	2	3(3),3(3)	Cannabis,Entamoeba

References

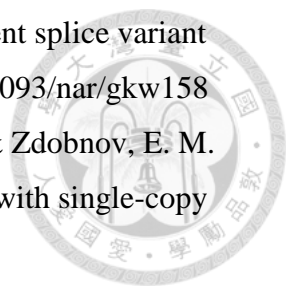
1. Lai, C.-K. *et al.* The Aphelenchoides genomes reveal major events of horizontal gene transfers in clade IV nematodes. *bioRxiv* 2022.09.13.507733 (2022). doi:10.1101/2022.09.13.507733
2. Blaxter, M. L. *et al.* A molecular evolutionary framework for the phylum Nematoda. *Nat.* 1998 3926671 **392**, 71–75 (1998).
3. Van Megen, H. *et al.* A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* **11**, 927–950 (2009).
4. Quist, C. W., Smant, G. & Helder, J. Evolution of Plant Parasitism in the Phylum Nematoda. *Annu. Rev. Phytopathol.* **53**, 289–310 (2015).
5. Grynberg, P. *et al.* Comparative genomics reveals novel target genes towards specific control of plant-parasitic nematodes. (2020). doi:10.20944/preprints202010.0449.v1
6. 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).
7. 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
8. 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).
9. Kent, W. J. BLAT—The BLAST-Like Alignment Tool. *Genome Res.* **12**, 656 (2002).
10. Wick, R. R., Judd, L. M. & Holt, K. E. Performance of neural network basecalling tools for Oxford Nanopore sequencing. *Genome Biol.* (2019). doi:10.1186/s13059-019-1727-y
11. 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
12. Vaser, R., Sović, I., Nagarajan, N. & Šikić, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* (2017). doi:10.1101/gr.214270.116
13. Medaka: Sequence correction by using ONT research.



<https://github.com/nanoporetech/medaka>

14. Walker, B. J. *et al.* Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* (2014). doi:10.1371/journal.pone.0112963
15. 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
16. Alonge, M. *et al.* RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. *Genome Biol.* **20**, 224 (2019).
17. Flynn, J. M. *et al.* RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.* (2020). doi:10.1073/pnas.1921046117
18. Edgar, R. C. & Bateman, A. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461 (2010).
19. Coghlan, A., Coghlan, A., Tsai, I. J. & Berriman, M. Creation of a comprehensive repeat library for a newly sequenced parasitic worm genome. *Protoc. Exch.* (2018). doi:10.1038/protex.2018.054
20. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics* (2009). doi:10.1002/0471250953.bi0410s25
21. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* (2014). doi:10.1093/bioinformatics/btu170
22. Dobin, A. & Gingeras, T. R. Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinforma.* (2015). doi:10.1002/0471250953.bi1114s51
23. 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).
24. Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* (2011). doi:10.1038/nbt.1883
25. 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
26. Pertea, M. *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* (2015). doi:10.1038/nbt.3122

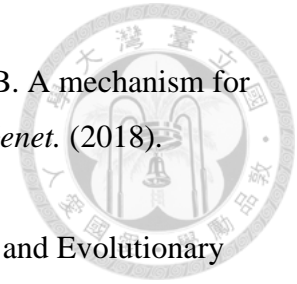


- 
27. Song, L., Sabunciyan, S. & Florea, L. CLASS2: Accurate and efficient splice variant annotation from RNA-seq reads. *Nucleic Acids Res.* (2016). doi:10.1093/nar/gkw158
28. Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* (2015). doi:10.1093/bioinformatics/btv351
29. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 1–9 (2004).
30. Stanke, M. *et al.* AUGUSTUS: A b initio prediction of alternative transcripts. *Nucleic Acids Res.* (2006). doi:10.1093/nar/gkl200
31. 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).
32. Davis, P. *et al.* WormBase in 2022—data, processes, and tools for analyzing *Caenorhabditis elegans*. *Genetics* **220**, (2022).
33. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
34. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* (2014). doi:10.1093/bioinformatics/btu033
35. Xu, J. -H. Carbohydrate Active Enzyme database. in *Catalysis from A to Z* (2020). doi:10.1002/9783527809080.catatz02801
36. Fast and accurate short read alignment with Burrows-Wheeler transform. - PubMed - NCBI. Available at: <https://www.ncbi.nlm.nih.gov/pubmed/19451168>. (Accessed: 1st November 2019)
37. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* (2011). doi:10.1093/bioinformatics/btr509
38. Nadachowska-Brzyska, K., Burri, R., Smeds, L. & Ellegren, H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Mol. Ecol.* **25**, 1058–1072 (2016).
39. Yoshida, K. K., Hasegawa, K., Mochiji, N. & Miwa, J. Early embryogenesis and anterior-posterior axis formation in the white-tip Nematode *Aphelenchoides besseyi* (Nematoda: Aphelenchoididae). *J. Nematol.* **41**, 17–22 (2009).
40. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and

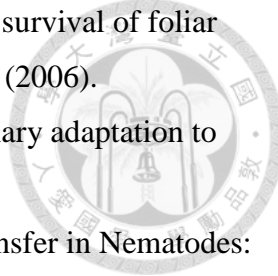
- Smudgeplots: Reference-free profiling of polyploid genomes. *bioRxiv* 747568 (2019). doi:10.1101/747568
41. The Evolution and Diversity of DNA Transposons in the Genome of the Lizard *Anolis carolinensis*. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3014272/>. (Accessed: 24th April 2020)
42. Thomas, J. & Pritham, E. J. Helitrons, the Eukaryotic Rolling-circle Transposable Elements. *Microbiol. Spectr.* **3**, (2015).
43. Cantarel, B. L. *et al.* MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* (2008). doi:10.1101/gr.6743907
44. Finn, R. D. *et al.* Pfam: The protein families database. *Nucleic Acids Research* (2014). doi:10.1093/nar/gkt1223
45. Kikuchi, T., Eves-Van Den Akker, S. & Jones, J. T. Genome Evolution of Plant-Parasitic Nematodes. *Annu. Rev. Phytopathol.* (2017). doi:10.1146/annurev-phyto-080516-035434
46. Mitreva, M. *et al.* The draft genome of the parasitic nematode *Trichinella spiralis*. *Nat. Genet.* (2011). doi:10.1038/ng.769
47. Tcherepanova, I., Bhattacharyya, L., Rubin, C. S. & Freedman, J. H. Aspartic proteases from the nematode *Caenorhabditis elegans*: Structural organization and developmental and cell-specific expression of *asp-1*. *J. Biol. Chem.* **275**, 26359–26369 (2000).
48. McKerrow, J. H. *et al.* *Strongyloides stercoralis*: Identification of a protease that facilitates penetration of skin by the infective larvae. *Exp. Parasitol.* **70**, 134–143 (1990).
49. Koch, B. J., Ryan, J. F. & Baxevanis, A. D. The Diversification of the LIM Superclass at the Base of the Metazoa Increased Subcellular Complexity and Promoted Multicellular Specialization. doi:10.1371/journal.pone.0033261
50. Dall, E. & Brandstetter, H. Structure and function of legumain in health and disease. *Biochimie* **122**, 126–150 (2016).
51. Holm Viborg, A. *et al.* A subfamily roadmap for functional glycomics of the evolutionarily diverse Glycoside Hydrolase Family 16 (GH16). doi:10.1074/jbc.RA119.010619
52. Coconi Linares, N., Dilokpimol, A., Stålbrand, H., Mäkelä, M. R. & de Vries, R. P. 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
53. 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).
54. Wang, F. *et al.* Transcriptomic analysis of the rice white tip nematode, *Aphelenchoides besseyi* (Nematoda: Aphelenchoididae). *PLoS One* **9**, (2014).
55. Stevens, L. *et al.* The Genome of *Caenorhabditis bovis*. *Curr. Biol.* **30**, 1023-1031.e4 (2020).
56. 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).
57. Fradin, H. *et al.* Genome Architecture and Evolution of a Unichromosomal Asexual Nematode. *Curr. Biol.* **27**, 2928-2939.e6 (2017).
58. de la Rosa, P. M. G. *et al.* A telomere-to-telomere assembly of *Oscheius tipulae* and the evolution of rhabditid nematode chromosomes. *G3 Genes, Genomes, Genet.* (2021). doi:10.1093/G3JOURNAL/JKAA020
59. Haas, B. J., Delcher, A. L., Wortman, J. R. & Salzberg, S. L. DAGChainer: A tool for mining segmental genome duplications and synteny. *Bioinformatics* (2004). doi:10.1093/bioinformatics/bth397
60. Pedersen, B. S. & Quinlan, A. R. Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics* (2018). doi:10.1093/bioinformatics/btx699
61. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* (2010). doi:10.1093/bioinformatics/btq033
62. Shinya, R. *et al.* Possible stochastic sex determination in *Bursaphelenchus* nematodes. *Nat. Commun.* **2022 131 13**, 1–14 (2022).
63. Woodruff, G. C. Patterns of putative gene loss suggest rampant developmental system drift in nematodes. *bioRxiv* 627620 (2019). doi:10.1101/627620
64. Meier, B. *et al.* The MRT-1 nuclease is required for DNA crosslink repair and telomerase activity in vivo in *Caenorhabditis elegans*. *EMBO J.* **28**, 3549–3563 (2009).
65. Shtessel, L. *et al.* *Caenorhabditis elegans* POT-1 and POT-2 repress telomere maintenance

- pathways. *G3 Genes, Genomes, Genet.* **3**, 305–313 (2013).
66. Ren, L., Huang, W., Cannon, E. K. S., Bertoli, D. J. & Cannon, S. B. A mechanism for genome size reduction following genomic rearrangements. *Front. Genet.* (2018). doi:10.3389/fgene.2018.00454
67. Cicconardi, F. *et al.* Chromosome Fusion Affects Genetic Diversity and Evolutionary Turnover of Functional Loci but Consistently Depends on Chromosome Size. *Mol. Biol. Evol.* **38**, 4449–4462 (2021).
68. Wang, J. *et al.* Comparative genome analysis of programmed DNA elimination in nematodes. *Genome Res.* **27**, 2001–2014 (2017).
69. Yin, D. *et al.* Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. *Science (80-.)*. **359**, 55–61 (2018).
70. Ali, M. A., Azeem, F., Li, H. & Bohlmann, H. Smart parasitic nematodes use multifaceted strategies to parasitize plants. *Frontiers in Plant Science* **8**, 1699 (2017).
71. Danchin, E. G. J., Guzeeva, E. A., Mantelin, S., Berepiki, A. & Jones, J. T. Horizontal Gene Transfer from Bacteria Has Enabled the Plant-Parasitic Nematode *Globodera pallida* to Feed on Host-Derived Sucrose. *Mol. Biol. Evol.* **33**, 1571–1579 (2016).
72. 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).
73. Jones, J. T., 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). doi:10.1163/156854105775142919
74. Grynberg, P. *et al.* Comparative genomics reveals novel target genes towards specific control of plant-parasitic nematodes. (2020). doi:10.20944/preprints202010.0449.v1
75. Schiffer, P. H. *et al.* Signatures of the Evolution of Parthenogenesis and Cryptobiosis in the Genomes of Panagrolaimid Nematodes. *iScience* **21**, 587–602 (2019).
76. 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).
77. Buchfink, B., Xie, C. & Huson, D. H. Fast and sensitive protein alignment using DIAMOND. *Nature Methods* (2014). doi:10.1038/nmeth.3176
78. 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
79. 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
80. 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
81. Alexa, A. & Rahnenfuhrer, J. topGO: enrichment analysis for Gene Ontology. R Packag. version 2.26.0. *R Packag. version 2.26.0* (2016).
82. Huerta-Cepas, J. *et al.* Fast genome-wide functional annotation through orthology assignment by eggNOG-mapper. *Mol. Biol. Evol.* (2017). doi:10.1093/molbev/msx148
83. Reijnders, M. J. M. F. & Waterhouse, R. M. Summary Visualizations of Gene Ontology Terms With GO-Figure! *Front. Bioinforma.* (2021). doi:10.3389/fbinf.2021.638255
84. Szitenberg, A. *et al.* Comparative genomics of apomictic root-knot nematodes: Hybridization, ploidy, and dynamic genome change. *Genome Biol. Evol.* **9**, 2844–2861 (2017).
85. Elling, A. A. *et al.* Sequence mining and transcript profiling to explore cyst nematode parasitism. *BMC Genomics* **10**, (2009).
86. 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).
87. Kooliyottil, R., Rao Gadhachanda, K., Solo, N. & Dandurand, L. M. ATP-Binding Cassette (ABC) Transporter Genes in Plant-Parasitic Nematodes: An Opinion for Development of Novel Control Strategy. *Front. Plant Sci.* (2020). doi:10.3389/fpls.2020.582424
88. 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).
89. 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)

- 
90. Jagdale, G. B. & Grewal, P. S. Infection behavior and overwintering survival of foliar nematodes, *Aphelenchoides fragariae*, on *Hosta*. *J. Nematol.* **38**, 130 (2006).
91. Wan, X. *et al.* The *Aphelenchus avenae* genome highlights evolutionary adaptation to desiccation. *Commun. Biol.* **4**, 1–8 (2021).
92. Haegeman, A., Jones, J. T. & Danchin, E. G. J. Horizontal Gene Transfer in Nematodes: A Catalyst for Plant Parasitism? <http://dx.doi.org/10.1094/MPMI-03-11-0055> **24**, 879–887 (2011).
93. Holterman, M., Schratzberger, M. & Helder, J. Nematodes as evolutionary commuters between marine, freshwater and terrestrial habitats. *Biol. J. Linn. Soc.* **128**, 756–767 (2019).
94. Brown, A. M. V. *et al.* Comparative genomics of wolbachia-cardinium dual endosymbiosis in a plant-parasitic nematode. *Front. Microbiol.* **9**, 2482 (2018).
95. Ma, J. *et al.* Major episodes of horizontal gene transfer drove the evolution of land plants. *Mol. Plant* **15**, 857–871 (2022).
96. Li, Y. *et al.* HGT is widespread in insects and contributes to male courtship in lepidopterans. *Cell* **185**, 2975–2987.e10 (2022).