Libertas Academica
FREEDOM TO RESEARCH

Open Access: Full open access to this and thousands of other papers at http://www.la-press.com.

Evolutionary Bioinformatics

The Origin of Parasitism Gene in Nematodes: Evolutionary Analysis Through the Construction of Domain Trees

Yizi Yang[1] and Damin Luo[1,2]

[1]School of Life Sciences, Xiamen University, Xiamen, Fujian, China. [2]State Key Laboratory of Cellular Stress Biology, Xiamen University, Xiamen, Fujian, China.

ABSTRACT: Inferring evolutionary history of parasitism genes is important to understand how evolutionary mechanisms affect the occurrences of parasitism genes. In this study, we constructed multiple domain trees for parasitism genes and genes under free-living conditions. Further analyses of horizontal gene transfer (HGT)-like phylogenetic incongruences, duplications, and speciations were performed based on these trees. By comparing these analyses, the contributions of pre-adaptations were found to be more important to the evolution of parasitism genes than those of duplications, and pre-adaptations are as crucial as previously reported HGTs to parasitism. Furthermore, speciation may also affect the evolution of parasitism genes. In addition, Pristionchus pacificus was suggested to be a common model organism for studies of parasitic nematodes, including root-knot species. These analyses provided information regarding mechanisms that may have contributed to the evolution of parasitism genes.

KEYWORDS: domain, nematode, parasitism gene, evolution, Pristionchus pacificus

CITATION: Yang and Luo. The Origin of Parasitism Gene in Nematodes: Evolutionary Analysis Through the Construction of Domain Trees. *Evolutionary Bioinformatics* 2013:9 453–466 doi:10.4137/EBO.S13032.

TYPE: Original Research

FUNDING: This project was supported by the Natural Science Foundation of China (Grant No. 81171595) and the Natural Science Foundation of Fujian Province (Grant No. 2010 J01229).

COMPETING INTERESTS: Author(s) disclose no potential conflicts of interest.

CORRESPONDENCE: dmluo@xmu.edu.cn

Introduction

A stumbling block for nematode research in parasitism is a relative lack of understanding of the origin of parasitism. Several scientific approaches have been used to improve this situation, particularly hypotheses regarding for the evolution of parasitism and analysis methods using a series of statistical and experimental analyses. Some of these hypotheses are concerned with parasitism genes. Parasitism genes and their encoded proteins are crucial to the life cycle of parasitic nematodes. These genes are endued with special phylogenetic and biological roles during the process of nematode evolution, and therefore may provide clues to the evolution of parasitic nematodes and even their choice of host species. For this reason, considerable efforts have been directed towards identification and classification of parasitism genes in the last 15 years. After the identification of the first nematode parasitism gene, beta-1.4-endoglucanase gene, in 1998,[1] hundreds of parasitism genes and parasitism gene candidates have been consequently identified. At least 48 genes were suggested as parasitism gene candidates (PGCs) in the root-knot nematodes Meloidogyne incognita during 2001–2003 and more than 60 PGCs in the cyst nematode Heterodera glycines from 2003–2004.[2]

With the increase in the numbers and categories of parasitism genes, our understanding of the evolution of parasitic nematodes has improved in both depth and breadth. In 2008, Ying[3] classified known parasitism genes into eight categories based on their functions in her master's thesis. Unfortunately, the concept of the parasitism gene has not been clearly defined for description of these genes. In 2000, Davis et al[4] defined parasitism genes in the manner of genetic origins, whereas Baum[5] described parasitism genes as genetic determinants in 2007. These definitions offered little insight into the molecular characteristics of parasitism genes, and much more analysis is required to reveal their unique molecular traits together with the mechanisms causing these traits.

Although the identification of various parasitism genes provided clues for inferring likely mechanisms, such as horizontal gene transfer (HGT), adaptation, or duplication, and these hypothesized mechanisms received much support from previous studies,[4,6–9] how these mechanisms specifically contribute to the evolution of parasitism genes is not been well-understood because of a lack of enough comparison among multiple studies. Despite widespread occurrence and the critical role of HGTs, their contributions to parasitism genes have not been compared with those of other mechanisms. No inferred mechanism has been suggested to explain the evolution of "pioneer" parasitism genes, a group of parasitism genes that lack homology with other species. Moreover, little information regarding combinational effects of these mechanisms on the evolution of parasitism is known. Previous studies have suggested that HGTs events followed by gene duplication and early gain of introns promote plant parasitism ability in nematodes.[10] However, whether other combinations (eg, adaptations followed by duplications) are able to promote the evolution of parasitism genes remains unclear. These limitations require a more comprehensive analysis of the evolution of parasitism genes.

Using the HelmCoP database and the new version of PhyML software,[11–13] evolutionary analysis can be achieved through construction of domain trees in a fast and effective manner. Orthologous groups, ie, orthologs as well as their lineage-specific duplications,[14] collected in the HelmCoP database provide the opportunity to investigate evolutionary relationships between parasitism genes and other homologs. The purpose of constructing domain trees is not only to search observable clues in trees but also to facilitate the analyses based on these trees and consequently to obtain clear direct evidence from these analyses. These analyses may shed light on the evolution of parasitism genes and increase the understanding of how three inferred mechanisms specifically contribute to the evolution of parasitism genes, and in turn clarify questions regarding the origins of parasitism genes.

Methods

Protein sequences and orthologous groups. A total of 236 protein sequences encoded by known and candidate parasitism genes were downloaded from the NCBI database in accordance with the summaries by Mitchum et al and Ying.[2,3] These protein sequences were then used as search queries in the HelmCoP database accessed through BLASTP. Significant matches (lowest E-value and less than 1×10^{-4}) were recorded, and their corresponding gene names were used to search for the orthologous groups that they belonged to. The hit sequences with 100% identity to known or candidate parasitism genes were recognized as "known parasitism genes" in their orthologous groups. The hit sequences showing high ($\geq$ 75%) but not 100% identity to known or candidate parasitism genes were recognized as "potential parasitism genes". These identification criteria are more stringent than 40% for ordinary homolog detection or high threshold (70%) for gene ontology annotation.[15]

All protein sequences in the detected orthologous groups were downloaded from the HelmCoP database. Furthermore, some orthologous groups, which contained no "known parasitism genes" or "potential parasitism genes", were also downloaded from HelmCoP as control groups. Orthologous groups were numbered in the form of "Ortho17taxa" group_number, such as Ortho 17taxa1004. In addition, multiple bacterial sequences were downloaded from the NCBI database as outgroup sequences for analysis. Random sequences were generated using the Sequence Manipulation Suite in case that no appropriate bacterial sequences were available.[16]

Primary sequence analysis. Names of downloaded sequences were short-chopped manually. All protein sequences downloaded from HelmCoP were submitted to the CD-HIT webserver to remove data redundancy.[17,18] These sequences were clustered at 100% identity and 99% identity in two CD-HIT runs. Both Pfam and CD-Search were used to predict domain boundaries for all protein sequences.[19,20] Two prediction results were compared with each other to determine which method was more suitable. Predicted domains in protein sequences were selected depending on their relationships with main gene functions that facilitate parasite infection. Sequences of selected domains were extracted from full protein sequences using Jemboss tools.[21,22]

Evolutionary analysis of parasitism genes. Domain sequences were aligned using MAFFT (E-INS-i).[23] Resulting alignments were evaluated using Guidance.[24] Unreliable columns below a cut-off value of 0.93 were removed from the alignments. ProtTest was used to select the best-fitted evolutionary model for all alignments of orthologous groups.[25] Phylogenetic trees were constructed using PhyML.[13] The aBayes method of the approximate likelihood ratio test was used to estimate branch support.[12] A BioNJ tree and five random trees were set as starting trees, and the BEST method was set for tree topology search strategy. The output was the best of inferred trees, and then was visualized and modified using TreeView or FigTree.[26,27] Protein subfamilies in orthologous groups were inferred using the Secator program.[28]

Species trees of all orthologous groups were constructed using Interactive Tree of Life (ITOL), an online tool for the display and manipulation of phylogenetic trees,[29,30] and were then modified using Mesquite.[31] Phylogenetic incongruences (PIs) lie at the center of HGT detection, and HGTs occur if there is strong conflict between the phylogenies of the gene and of the organisms.[32] Not only HGTs, but also gene duplication, adaptive molecular evolution, and poor sampling can cause phylogenetic incongruences.[33,34] T-REX was used to detect HGT-like PIs by building a HGT network, since the program cannot predict differences between different species and different gene copies of the same species. Both modified species tree and domain tree were submitted to the T-REX webserver to infer HGT-like PIs, a transformative analysis of

HGTs.[35–37] Duplications are easily observed in trees, but they can be confounded by speciation in some circumstances. The species overlap method can detect both duplication and speciation events, and therefore, these events were detected using this method implemented in the ETE toolkit.[38]

Results

Characteristics of orthologous groups and protein sequences. Different orthologous groups showed different situations, regardless of whether a parasitism gene was identified. After removing data redundancy using CD-HIT, more than five hundred sequences were found in some orthologous groups, while a few groups contains only 2 or 3 sequences. Some groups, such as the ortho17taxa1012 group, included sequences of all 18 HelmCoP species (6 free-living nematode species (Caenorhabditis elegans, Caenorhabditis brenneri, Caenorhabditis briggsae, Caenorhabditis japonica, Caenorhabditis remanei, and Pristionchus pacificus), 4 parasitic nematode species (Brugia malayi, Meloidogyne hapla, Meloidogyne incognita, and Trichinella spiralis), 2 flatworms (Schistosoma japonicum and Schistosoma mansoni), 6 host or outgroup species (Arabidopsis thaliana, Drosophila melanogaster, Glycine max, Homo sapiens, Mus musculus, and Saccharomyces cerevisiae), but most groups contained fewer than 100 sequences from several species.

In most cases, domain boundaries predicted by both CD-Search and Pfam were comparable, and the CD-Search predictions were used to determine extraction of domain sequences. However, Pfam predictions can differ from those of the CD-Search in some cases. Moreover, in some extreme situations, domain types within particular sequences were only detected by either CD-Search or Pfam alone. Thus, comparisons of these domains and further selection of appropriate domains were required and was performed according to domain descriptions from either CD-Search or Pfam websites and GO IDs provided by HelmCoP. In principle, functions of chosen domains were directly related to gene functions that facilitate completion of the life cycle of parasitic nematodes.[3] A list of all orthologous groups, chosen domains, and constructed trees are presented in Table S1 (Supplementary Material).

Inferences of evolutionary histories of orthologous groups. Topologies of domain trees have some degree of connections, providing clues for the evolution of parasitism genes. Nearly all trees were unbalanced, and some clades in trees contained many more domains than their sister clades. Domains of known and potential parasitism genes exhibited various clustering patterns in inferred trees and therefore provided multiple observable clues.

To examine this data, we analyzed the 1648_tree (Fig. 1). Eight potential parasitism genes and their locations in clades in this tree provided 5 clues for the evolution of parasitism genes encoding peroxiredoxin. Four of 5 clues were divided into two types: the first type was the clustering pattern of genes from the same species, indicating gene duplication (DUP clue), and the second type was the clustering pattern of genes from parasitic species (may not be parasitism genes) with genes under free-living conditions, particularly with genes from P. pacificus or D. melanogaster species, a clue commonly found in trees but requiring further phylogenetic analysis (PI clue). Additionally, a vague clue perhaps for a closed phylogenetic relationship was also identified. In the 1648_tree, a clade composed of two potential parasitism genes, MI11766 and MI13944 of M. incognita and a clade composed of one potential parasitism genes TSP00400 together with another gene of T. spiralis species, were all examples of tree topologies providing the clues for duplications. In contrast, a clade formed by a potential parasitism gene MI02762a with other 6 genes of free-living nematodes as well as another clade composed of SMP04470_1 and DM0082927 offered the clues, which required phylogenetic analysis.

Both DUP and PI clues were also found in many other domain trees, and most clades that provided these clues were strongly supported by aBayes branch values. However, DU clues were different from PI clues with straightforward connections and hypothesized mechanisms behind them. The connection between DU clues and duplication was clear, whereas the connection between PI clues and hypothesized mechanisms, perhaps adaptation or HGT, requires further examination. In addition, vague clues such as that provided

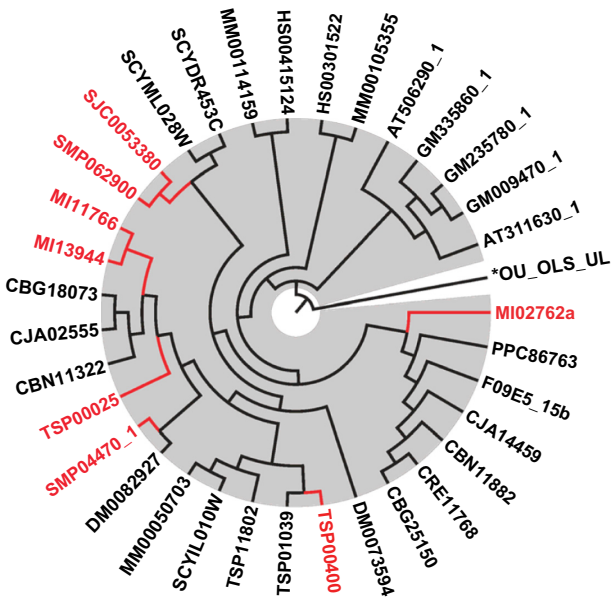

Figure 1. Cladogram of the 1648_tree. Red nodes and branches refer to known or potential parasitism genes and their closely related evolutions. A highlighted area with gray color refers to the unique subfamily inferred by Secator.

Abbreviations: CJA, C. japonica; CRE, C. remanei; CBN, C. brenneri; CBG, C. briggsae; PPC, P. pacificus; TSP, T. spiralis; MI, M. incognita; HS, H. sapiens; MM, M. musculus; DM, D. melanogaster; AT, A. thaliana; GM, G. max; SC, S. cerevisiae; SJC, S. japonicum; SMP, S. mansoni; Other, C. elegans; OU, outgroup sequence; BM, B. malayi (not shown in figure); MH, M. hapla (not shown in figure).

by a clade composed of SMP062900 and SJC0053380 in the 1648_tree could be found in other trees as well, which also requires further analysis.

Clustering patterns of known or potential parasitism genes in other trees, in which genes also encoded peroxiredoxins, were quite different from those in the 1648_tree (Fig. 2). However, one of these trees showed DUP clues. A clade composed of two B. malayi genes, i.e. BM37795_06 and BM37795_54, in the 20295_2 tree (Fig. 2: D) provided such a clue. In contrast, no DUP clue was found in the 2539_tree, the 6462_2_tree and the 7951_tree (Fig. 2: A, B, and C). In addition, one or more subfamilies were inferred in the 1648_tree and other trees, but no relationship was identified between inferred subfamilies and parasitism genes.

"Pioneer" parasitism genes and their domain trees provided many DUP clues. These parasitism genes encoded various proteins, and two or more DUP clues were found in these trees. The 1931_tree illustrates these duplications (Fig. 3: A). Known or potential parasitism genes were identified in all 4 major clades, which were also 4 subfamilies inferred by Secator. The occurrences of these parasitism genes varied across clades in this tree, regardless of where DUP clues were found. Similarly, no absolute relationship between the occurrence of parasitism genes and DUP clues was discovered in the other 5 domain trees of "pioneer" parasitism genes.

A series of DUP clues were found in similar clades in the 10515_2_tree (Fig. 3: B), the 7833_tree (Fig. 3: C), and the 11424_2_tree (Fig. 3: E). However, neither the occurrences nor the numbers of parasitism genes were the same in these clades. In the clade of the 10515_2_tree, two potential parasitism genes, i.e. MH26467094 and MH19052586, were related to two continuous duplication events. An analogous situation

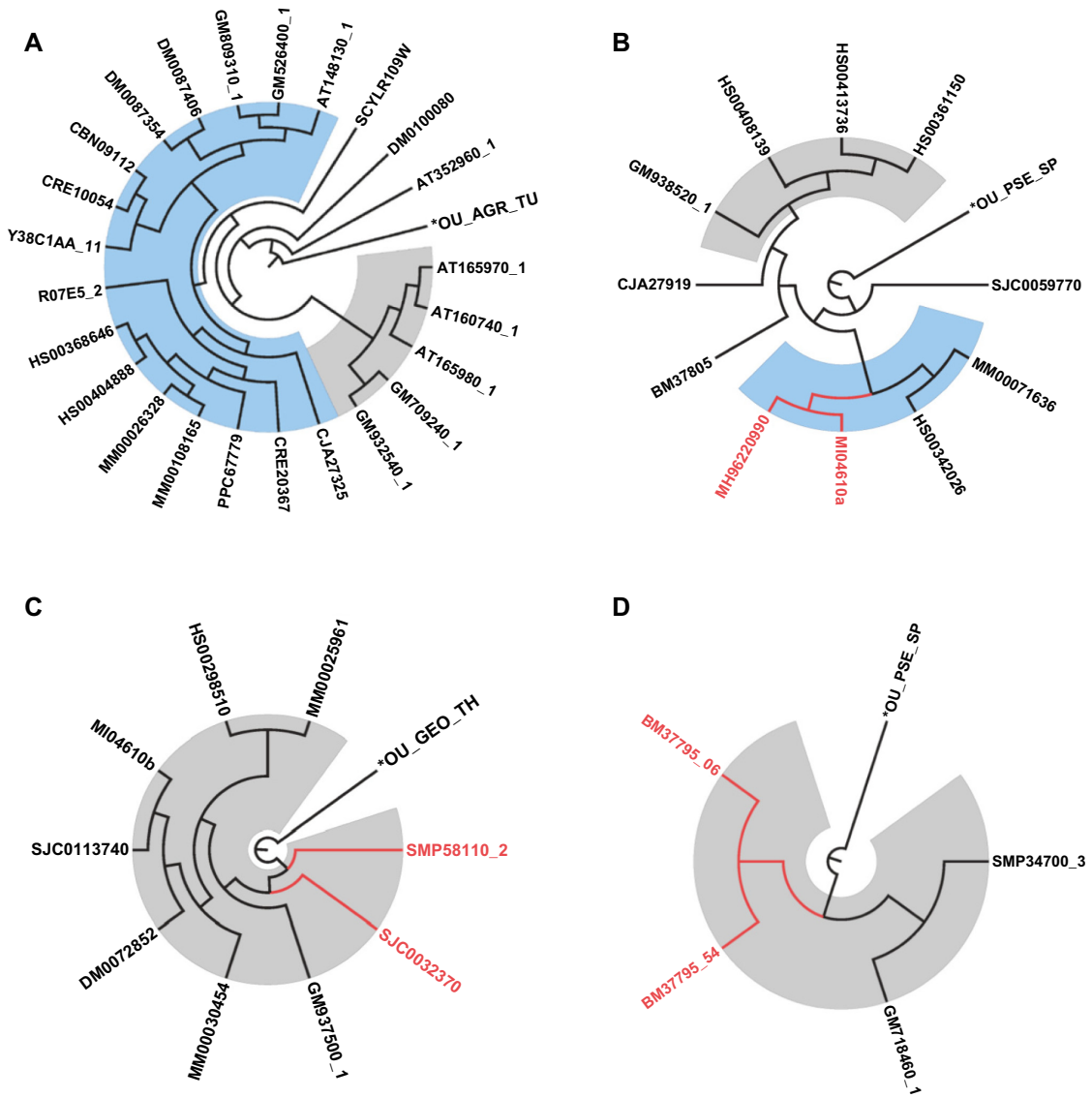

Figure 2. Cladograms of four domain trees. Red nodes and branches refer to known or potential parasitism genes and their closely related evolutions. Highlighted areas refer to the inferred subfamilies by Secator. The gray area is the first subfamily, whereas the powderblue refers to the second. Unhighlighted area refers to uncertainty in subfamily inference, except the outgroup. Gene names follow the description in Figure 1. (**A**) Cladogram of the 2539_tree; (**B**) Cladogram of the 6462_2_tree; (**C**) Cladogram of the 7951_tree; (**D**) Cladogram of the 20295_2_tree.

was observed in the clade of the 11424_2_tree, but it became 3 duplication events for 3 parasitism genes. In contrast, two parasitism genes, ie, MI09298c and MI13221a, were caused by duplication in the clade of the 7833_tree.

Furthermore, when we look at parasitism genes and their positions in other two trees, i.e. the 10119_2_tree (Fig. 3: D) and 11188_2_tree (Fig. 3: F), no direct relationship was found between them and duplication events. Similar to the 1648_tree, some vague clues, such as the clue offered by the clade consisting of 1 M. hapla gene and 3 M. incognita genes were also found in the 1931_tree, but no clear relationship can be inferred for the occurrences of parasitism genes and the mechanisms offering these clues.

The 1012_tree effectively illustrates the variety of clustering patterns offering PI clues (Fig. S1, Supplementary Material). Twenty-six known or potential parasitism genes and the clades where they were found offered a large number of DUP clues. Notably, three clades in this tree consisting of one P. pacificus gene and one gene from parasitic nematodes or bacterial species were observed. In addition to P. pacificus genes, all three genes, ie BM40590, MH77012204, and OU_LAC_VE, did not belong to identified parasitism gene families. However, the clues provided by the clades showed some degree of connections with the occurrences of parasitism genes of B. malayi, M. hapla, and M. incognita species in the 1012_tree. In contrast to the species with parasitism genes, there was no identified parasitism gene in the parasitic nematode T. spiralis, although there were multiple DUP clues to duplication events in this species. In addition, similarly to the 1648_tree and 1931_tree, vague clues from multiple clades formed by M. hapla genes and M. incognita genes were found in this tree.

In summary, two types of observable clues were discovered in several domain trees containing parasitism genes. DUP clues illustrated a strong connection with duplication events, whereas the mechanisms responsible for PI clues, perhaps adaptation and HGT, require further analyses. Furthermore, the connection between DUP clues and related parasitism genes was not absolute. In contrast, the occurrence of PI clues indicated the emergence of parasitism genes in domain trees under most situations. Additionally, vague clues perhaps relating to phylogenetic relationships were also found in multiple trees, but neither their associated mechanisms nor their relationships with the occurrence of parasitism genes are understood. Finally, other information, such as the specific genes involved in the events of three inferred mechanisms or the amount of gene subfamilies, showed no absolute relationship with the occurrence of parasitism genes in domain trees.

Detection of HGT-like phylogenetic incongruences. Analysis of HGT-like PIs was performed to reveal the mechanisms contributing to PI clues. A large number of PIs were detected in domain trees, which were displayed in abbreviated forms. For example, a PPCgene PI referred to incongruence between a P. pacificus gene and one gene of any other species, whereas a $DMgene_1$-$SMPgene_3$ PI was incongruence between a clade formed by 1 D. melanogaster gene and another clade composed of 3 S. mansoni genes. Some detected PIs were written in full forms, such as a PPC71126-MH44025137 PI that indicated incongruence between a P. pacificus gene and a M. hapla gene. Additionally, this PI can also be displayed as a $PPCgene_1$-$MHgene_1$ PI, which would be clear under most situations. Parasitic nematode species, parasitic flatworm species, parasitic species, remaining free-living eukaryotes except P. pacificus and D. melanogaster, free-living nematode species, bacterial outgroups, and gene duplication in one species were abbreviated as PNS, PFS, PS, FLS, FNS, BO, and DUP, respectively. (see also in Table 1). After removing many redundant PIs caused by the T-REX algorithm, the remaining PIs were classified into multiple large groups: I the BOgenes PIs; II the PPCgenes PIs, the DMgenes PIs, and the FLSgenes PIs; III the DUPgene PIs; IV the PSgenes PIs.

Let us take the 1377_tree as an example (Fig. 4). The detected PIs in this tree provided multiple examples of the PIs belonging to these four groups. The MI05094-BO_STR_LI PI (Group I), the MI03688-(MI02439, MI00991) PI (Group III), and the occurrence of parasitism gene MI04482 provided an evidence for HGT followed by duplications, which were in accordance with previous studies of parasitism genes.[10] Moreover, the MI06116-CBN16179 PI (Group II) and the (SMP006520, DM0083337, MI04482)-CBG15541 PI (Group II), and the SMP006520-MI04482 PI (Group IV) also provided weak evidence for other mechanisms such as adaptation. Directions were also estimated for these PIs since they were analyzed through a transformative test of HGTs. However, as shown in Figure 4, genes and species involved in these PIs rather than directions, suggested the mechanisms behind parasitism genes evolution. Detected PIs facilitated the search and comparison of various pieces of evidence from topological structures of all domain trees.

Strong evidence for associated mechanisms were obtained when detected PIs were collected from the results of all domain trees. The ($PPCgene_{n1}$)-($PNSgene_{n2}$) PIs (n1, $n2 \geq 1$), the PIs in a subgroup of group II, were detected in many domain trees containing one or more parasitism genes. The PIs in 30 domain trees belonging to this subgroup provided evidence for their connections with identified parasitism genes of nematode species. For example, the detected $PPCgene_1$-$MHgene_1$ PI and the $PPCgene_1$-$BMgene_1$ PI and multiple parasitism genes of B. malayi and M. hapla species were found in the 1012_tree. The simplest and most common PIs were the $PPCgene_1$-$PNSgene_1$ PIs, and more complex and derived PIs including $PPCgene_1$-($PNSgene_{n1}$) PIs, the ($PPCgene_{n1}$)-$PNSgene_1$ PIs, the ($PPCgene_{n1}$)-($PNSgene_{n2}$) PIs, and the ($PPCgene_1$, $FNSgene_1$)-($PNSgene_{n1}$) PIs, among others. All of these PIs showed strong correlations with the occurrence of parasitism genes in domain trees. A list of several PPC genes PIs and their associated parasitism genes can

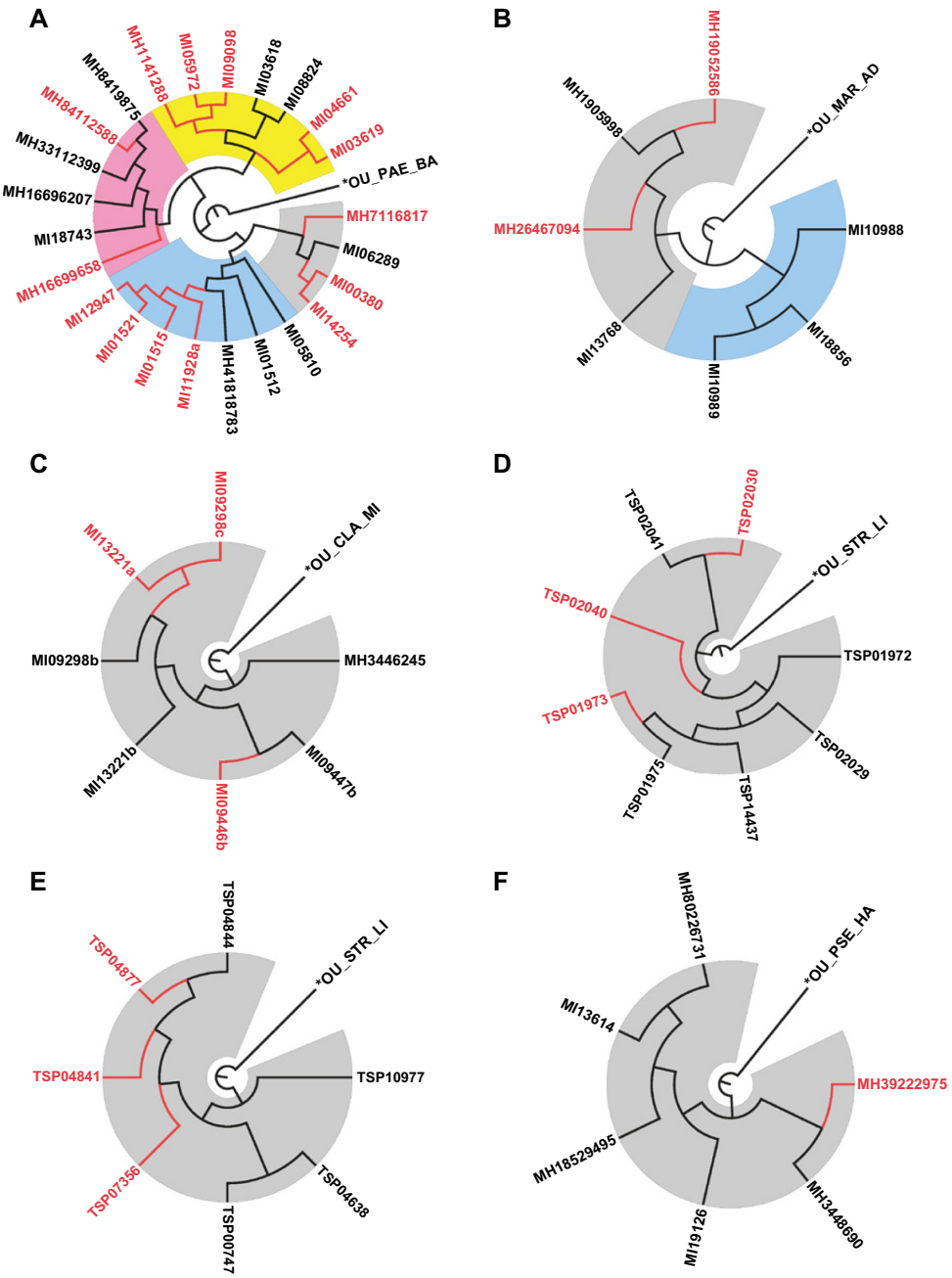

Figure 3. Cladograms of six domain trees containing "pioneer" parasitism genes. Red nodes and branches refer to known or potential parasitism genes and their closely related evolutions. Highlighted areas refer to the inferred subfamilies by Secator. The gray area is the first subfamily, whereas the powderblue refers to the second, the light pink to the third, and the yellow to the fourth. Gene names follow the description in Figure 1. (**A**) Cladogram of the 1931_ tree; (**B**) Cladogram of the 10515_2_tree; (**C**) Cladogram of the 7833_tree; (**D**) Cladogram of the 10119_2_tree; (**E**) Cladogram of the 11424_2_tree; (**F**) Cladogram of the 11882_2_tree.

be found in Table 2, and all representative PIs were collected in Table S2 (Supplementary Material).

Similarly to the above PIs, the detected ($DMgene_{n1}$)-($PFSgene_{n2}$) PIs (n1, n2 $\geq$ 1) exhibited strong connections with identified parasitism genes of flatworm species. These DMgenes PIs and corresponding parasitism genes in 6 trees provided good examples for this relationship, and the simplest PIs were the $DMgene_1$-$PFSgene_1$ PIs. Two examples of the PPCgenes PIs and the DMgenes PIs are shown in Figure S2 (Supplementary Material). Interestingly, in some situations the PPCgenes PIs indicated their connection with parasitism genes of flatworm species, and so were the DMgenes PIs for their connections with nematode parasitism genes, such as the $PPCgene_1$-($SMPgene_2$) PI in the 1035_tree and the $DMgene_1$-$MIgene_1$ PI in the 1078_2_tree. The PPCgenes in 2 trees and the DMgenes in 4 trees corresponded to this situation. Furthermore, the ($PPCgene_1$, $DMgene_1$)-$PNSgene_1$ PIs in 2 domain trees and the ($PPCgene_1$, $DMgene_1$)-($PFSgene_2$) in 1 domain tree also provided the evidence for their connections with parasitism genes in nematodes and flatworms.

In contrast, the PPCgenes PIs and the DMgenes PIs were also discovered in several trees of control groups, in which no parasitism gene was identified. However, these PIs were then found to either be analysis results of weak supported clades or from potentially false control groups. Some detected PPCgenes PIs, such as the $PPCgene_1$-$MIgene_1$ PI in the 7416_tree were not detected in the 7416_2_tree, which was based on the Pfam domain sequences. Other PPCgenes PIs, the $PPCgene_1$-$BMgene_1$ PI in the 1004_2_tree, and the $PPCgene_1$-$BMgene_1$ PI in the 1197_tree, were simultaneously found with duplicated genes of parasitic species in the same orthologous groups, suggesting potential un-identifications of parasitism genes in these groups. Therefore, these PIs actually did not provide persuasive opposing evidence for the correlation of the PPCgenes and DMgenes PIs with parasitism genes.

In addition, a few PPCgenes or DMgenes PIs were identified as redundant PIs, such as the $PPCgene_1$-($BMgene_1$, $DMgene_1$, $MMgene_1$) PI and the $DMgene_1$-($HSgene_1$, $SMPgene_1$) PI in the 3818_tree, which also provided some degree of evidence for their connections with parasitism genes. In general, the PPCgenes or DMgenes PIs and their association with parasitism genes supported the pre-adaptation mechanism described in the pre-adaptation hypothesis, and their related topological structures in trees were responsible for most PI clues discovered. Thus, P. pacificus may be a good model for many parasitic nematodes.

Not only the PPCgenes PIs and the DMgenes PIs but also other PIs were related to PI clues. In addition to the $MIgene_1$-$BOgene_1$ PI in the 1377_tree, no other similar PI was found. However, the $TSPgene_1$-($GMgene_2$, $BOgene_2$) PI in the 1010_2_tree offered evidence for some degree of connection with parasitism genes of T. spiralis species. Furthermore, the $PPCgene_1$-$BOgene_1$ PI and the ($DMgene_2$)-$BOgene_1$ PI were respectively found in the results of the 1012_tree and the 1030_tree. They indicated potential HGTs between these two species and bacteria.

The FLSgenes PIs and their connections with parasitism genes were also observed. The $ATgene_1$-$MIgene_1$ PI in the 3835_2_tree followed the description of adaptation of parasitism genes to their hosts. Eight similar PIs, including the ($HSgene_2$)-($BMgene_3$) PI in the 2468_2_tree, the $HSgene_1$-($SMPgene_1$, $SJCgene_1$) PI in the 10788_2_tree, and the ($HSgene_2$, $MMgene_3$)-$TSPgene_1$ PI in the 1875_tree, among others, also provided evidences for their connections with general adaptation. All aforementioned PIs and their associated PI clues highlight the potential roles of pre-adaptations, general adaptations, and HGTs in the evolution of many parasitism genes.

Additionally, the analysis of HGT-like PIs also detected multiple PSgenes PIs, such as the $MIgene_1$-$MHgene_1$ PI in the 2536_2_tree and the $TSPgene_1$-$BMgene_1$ PI in the 8474_tree, and these PIs and vague clues that they provided showed some degree of connection with the occurrence of parasitism genes in some trees. However, the true mechanism responsible for these PSgene PIs cannot be determined through this analysis. Many DUPgenes PIs associated with DUP clues for duplications were detected as well, such as the $MIgene_1$-$MIgene_1$ PI and the $SMPgene_1$-$SMPgene_1$ PI in the 1016_tree, the the $MIgene_1$-($MIgene_2$) PI in the 1377_tree above etc. However, these test results showed loss of detection, since the analysis of HGT-like PIs were not designed to the detection of gene duplication. Duplication events of parasitism genes were observed in 54 domain trees, and the DUPgenes PIs relating to these genes were found in 31 of these trees only. Therefore, these PIs and their associated mechanisms were not sufficiently clear and stable to draw convincing conclusions.

Detection of duplication and speciation. Although DUP clues were clear in trees, detailed properties of duplications needed to be analyzed. The number of gene copies that arose by duplications and the specific contribution of each duplication event to the evolution of parasitism gene were important for understanding how duplication is responsible for the parasitism gene, and therefore both were focused on the results of the species overlap analysis. A large number of duplications and speculations were detected in domain trees. Similarly to the PIs, duplication and speciation were also represented using abbreviations. A $MIgene_1$-($MIgene_3$) DUP referred to a duplication event occurring at divergence time of two clades, while a $SMPgene_1$-($SJCgene_2$) SP meant the speciation event responsible for the divergence between one S. mansoni gene and two S. japonicum genes. It was expected that comparison of these DUPs would provide evidence for how DUPs affected the evolution of parasitism genes.

The evolutionary histories of "pioneer" parasitism genes shown in Figure 3 were characterized based on a series of DUP clues, and multiple DUPs were detected in trees containing these genes. Fifteen DUPs in total were detected in the 1931_tree, and a ($MIgene_9MHgene_2$)-($MIgene_7MHgene_6$) DUP

Table 1. Quick access table for abbreviations (Definition of abbreviation).

DEFINITION	ABBREVIATION
horizontal gene transfer	HGT
phylogenetic incongruence	PI
duplication	DUP
speciation	SP
Pristionchus pacificus	PPC
Drosophila melanogaster	DM
parasitic nematode species	PNS
parasitic flatworm species	PFS
parasitic species	PS
free-living species except *P. pacificus* and *D. melanogaster*	FLS
free-living nematode species	FNS
bacterial outgroup	BO

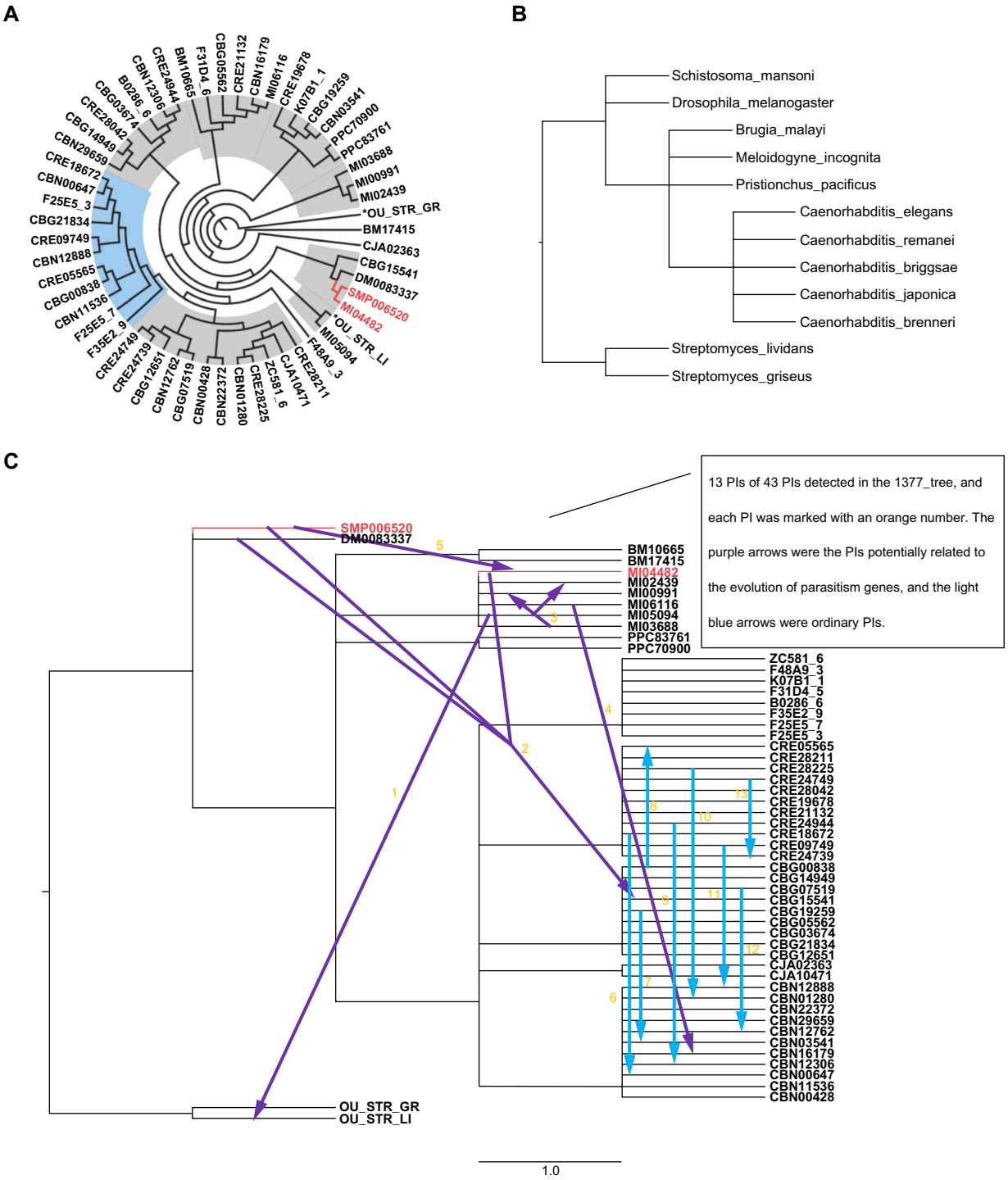

Figure 4. Two examples of trees submitted to T-REX. Red nodes and branches refer to known or potential parasitism genes and their closely related evolutions. Highlighted areas refer to the inferred subfamilies by Secator. The gray area is the first subfamily, whereas the powderblue refers to the second. Gene names follow the description in Figure 1. (**A**) Cladogram of the 1377_tree; (**B**) Corresponding species phylogeny; (**C**) Corresponding species tree modified using Mesquite.

showed the largest number of gene copies among all detected DUP events responsible for the evolution of parasitism genes. In addition to this DUP, other 13 DUPs of these 15 were also related to parasitism genes. This was similar to results of analysing other domain trees. Multiple DUPs detected in the 10515_2_tree, 10119_2_tree, and 7833_tree were related to the evolution of parasitism genes. In contrast, a few DUPs in the 11424_tree, 10515_2_tree, 11882_tree and the 1931_tree had no relationship with the parasitism gene.

According to the detected DUPs in these trees of "pioneer" parasitism genes, no absolute relationship was identified between DUPs and parasitism genes. However, some weak aBayes branch supports may undercut the credibility of evidences from DUP detections in these trees. Additional

Table 2. List of several PPCgenes PIs and their related parasitism genes.

DOMAIN TREE	PPCGENES PIS	PARASITISM GENES
1010_tree	From subtree (PPC71126) to subtree (MH44025137, MI00172a, MI03289)	MI10006a
1012_tree	From subtree (PPC63534) to subtree (MH77012204)	MH48161740; MH76222745; MH3753530; MH15247343; MH48155517
1035_tree	From subtree (CJA08401, PPC66583, PPC66584) to subtree (MH93362173, MI07779, MI08928); From subtree (TSP02932) to subtree (PPC77428)	MI10154 TSP02738
2215_2_tree	From subtree (C33C12_3a, C33C12_8, CBG17050, CBG17056, CBG19964, CBN00995, CBN17884, CJA06529, CJA14571, CRE06912, CRE06915, CRE13462, PPC59073, PPC67026, PPC73468, PPC73833, Y4C6B_6) to subtree (MI13184)	MI09178; MI03215
2468_2_tree	From subtree (PPC58498) to subtree (BM16375)	BM00065; BM26865; BM00095
5833_2_tree	From subtree (PPC82160) to subtree (TSP09554)	TSP09554
9659_tree	From subtree (MI04057) to subtree (CBG14965, CBN15633, CJA00897, CRE28051, PPC71128, Y44E3 A_2)	MI04057

Table 3. Number of detected events from the root prior to the emergence of the first parasitism gene in some domain trees.

DOMAIN TREE	NUMBER OF DETECTED EVENTS	RELATED PARASITISM GENES
1931_tree	3 duplications	MI04661; MI03619; MH16699658
1492_tree	3 duplications and 1 speciations	BM19970; BM02100; BM20385; BM53615; BM13265
1651_tree	5 duplications and 2 speciations	TSP07750; TSP01441; TSP01070; TSP08945; TSP00436; TSP01064
1796_tree	4 duplications and 3 speciations	MI18347
6953_2_tree	3 duplications and 2 speciations	MI03214
7833_tree	2 duplications	MI09446b; MI09298c; MI13221a
11424_2_tree	3 duplications	TSP07356

evidence was obtained from other trees. Multiple DUPs were also detected in large trees, such as the 1010_tree and the 1012_tree. The evolution of some parasitism genes, TSP07316 in the 1010_tree and MI09928a in the 1012_tree for instance, were clearly caused by DUP mechanism. In contrast, it was also found that DUPs were related to many genes of the same parasitic species such as TSP07309 and TSP07290 in the 1010_tree and MI12364 and MI17963 in the 1012_tree, but these genes were not identified parasitism genes. In general, although DUP events caused the evolution of many parasitism genes in various cases, the evidence from domain trees supported that there was no absolute relationship between them.

Additionally, the amounts of DUP events prior to the occurrence of the first parasitism gene in the trees of "pioneer" parasitism genes were compared with those in the trees containing ordinary parasitism genes. For "pioneer" parasitism genes, 2 or 3 DUPs were enough for the evolution of parasitism genes, such as the situation in the 1931_tree. In contrast, more ancient DUPs were required for the evolution of ordinary parasitism genes, such as the situations in the 1651_tree and the 1796_tree (Fig. S3 and Fig. S4, Supplementary Material). The amounts of detected events prior to the emergence of the first parasitism gene in several trees are listed in Table 3. Representative DUPs and SPs in all 141 domain trees can be found in Table S2 (Supplementary Material).

Considering the widespread occurrence of DUPs and their associated DUP clues in the trees that no parasitism gene was found in, a comparison between the trees containing parasitism genes and the trees without parasitism gene was necessary to understand how duplication affected the evolution of parasitism genes. The 5203_tree was compared with the 7833_tree as as shown in Figure 5. Both trees had 6 genes of M. incognita species and 1 gene of M. hapla species, but had different topological structures. There were 2 main clades in the 7833_tree, whereas the genes in the 5203_tree formed 1 major clade. It was suggested that early duplications and their contributions to large clades in trees could be related to the evolution of parasitism genes.

The contributions of DUPs to the evolution of parasitism genes were also determnined by comparing DUP events with detected PPCgenes and DMgenes PIs. Both PNSgenes DUPs and PPCgenes PIs can be found in many domain trees containing nematode parasitism genes. However, PPCgenes PIs but no PNSgenes DUPs were found in 9 domain trees with nematode parasitism genes. Analogously, the DMgenes PIs but not PFSgenes DUPs were found in 2 domains trees containing parasitism genes of flatworm species. In contrast, the PNSgenes DUPs with the absences of the PPCgenes or DMgenes PI were only related to the evolution of ordinary parasitism genes in 7 domain trees. Furthermore, the PPCgenes PIs and DMgenes PIs showed stronger correlations with the occurrences of parasitism genes than those of DUP detected in the same trees, according to the analysis results of the 1045_tree and 1137_tree, among others. In addition, parasitism genes of all 4 parasitic nematodes and 2 flatworms were shown to be related to the PPCgenes or DMgenes PIs, but no parasitism

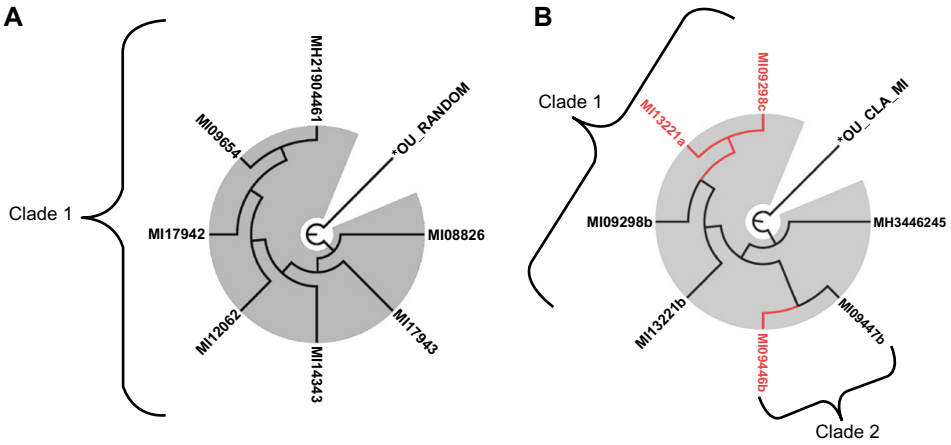

Figure 5. Cladograms of two domain trees. Red nodes and branches refer to known or potential parasitism genes and their closely related evolutions. The gray area refers to the unique subfamilies inferred by Secator. Gene names follow the description in Figure 1. (**A**) Cladogram of the 5203_2_tree; (**B**) Cladogram of the 7833_tree.

gene of B. malayi species was found to be caused by DUP. Two B. malayi genes in the 20295_2_tree were later found to be alternative splice forms but not true duplicates, because they both referred to the same gene. Overall, although DUPs were related to the evolution of parasitism genes in most cases, their roles were not as important as those of pre-adaptations to the final occurrences of parasitism genes.

The evolution of "pioneer" parasitism gene remains an academic problem, and thus cannot be simply explained using the detected DUPs in domain trees. For this reason, all 10 trees of "pioneer" domain trees and their associated results were not taken into consideration when the contributions of DUPs were compared with those of pre-adaptations and HGTs. Moreover, the SP events were also detected using the species overlap method when the DUPs were detected. Some of these events provided vague clues in trees, and consequently had some degree of connection with the evolution of parasitism genes. A $MIgene_1$-$MHgene_1$ SP was shown to be partly related to the evolution of a parasitism gene MH79012567 in the 23137_2_tree, and a $TSPgene_1$-($GMgene_5$, $ATgene_2$) SP was connected with a parasitism gene TSP06028 in the 9928_tree. Similarly, a ($SMPgene_1$, $SJCgene_1$)-($SCgene_2$) SP was partly responsible for two aforementioned parasitism genes, ie, SMP062900 and SJC0053380, in the 1648_tree. Therefore, the analysis using the species overlap method also provided some evidences for the contributions of SP mechanism to the evolution of parasitism genes, and thus made vague clues more clear than ever.

Discussion

Improvements in revealing the origin of parasitism gene. Understanding specific contributions of 3 inferred mechanisms could help to reveal the origin of parasitism gene. Two types of observable clues found widespread in domain trees and corresponding analyses to track these clues supported the contributions of all three inferred mechanisms to the evolution of parasitism genes. However, after comparing several lines of evidence from detections of PIs and DUP and SP events, an idea regarding important roles of both pre-adaptation and HGT was generated. This idea is discussed from following aspects: I pre-adaptation hypothesis; II widespread pre-adaptations and their associated PPCgenes PIs or DMgenes PIs; III common combinational pattern with DUP mechanism.

Unlike ordinary adaptations of parasite genes causing high similarity between parasitism genes and host genes, the pre-adaptations of ancestor genes for parasitic life, which was described by the pre-adaptation hypothesis, caused various traits of current parasitism genes.[39] These traits due to pre-adaptations could also be found in genes of P. pacificus species, since the same pre-adaptations for parasitism in its genome and genes were identified. Shared traits in genes and genomes may explain why many parasitism genes showed high similarities with the genes of P. pacificus species under free-living conditions. General adaptations to host genes were also observed in a few trees, but this was not as widespread as pre-adaptations in trees. Additionally, some evidence for general adaptation may suffer from the problem of uncertainty due to weak aBayes support in trees. These issues make it important to understand pre-adaptation on a deeper level.

The widespread PPCgenes PIs may show an important role of pre-adaptation in the evolution of parasitism genes. This supported the previous claim that pre-adaptations are crucial for the evolution of parasitism on the theoretical background.[40] Moreover, P. pacificus may be a good model for many parasitic nematodes, including root-knot nematodes as suggested by the results of the analysis of HGT-like PIs. This idea disagreed with the suggestions by a previous analysis,[41] but may receive supports from some recent studies, such as the host-finding behavior in P. pacificus.[42] In addition, D. melanogaster was also highlighted due to its potential association with the pre-adaptation and the evolution of parasitism genes in flatworms, since both the DMgenes PIs and PPCgenes PIs are two types of important PIs detected in

the analysis. To understand the role of D. melanogaster in the evolution of parasitism, a comparison of traits in P. pacificus and *D. melanogaster* was listed in Table 4. Despite some shared features between two species, there were significant differences in larvae development and phylogenetic positions for P. pacificus and D. melanogaster. This indicates that additional studies are required to reveal whether D. melanogaster is a model organism for parasitism.

Similarly to pre-adaptation, HGTs were previously known for their importance in the evolution of parasitism genes and their contribution to the evolution of many parasitism genes with bacterial origins, such as chorismate mutase, L-threonine aidolase, and Nod factors, among others.[1,43] HGTs in combination with DUP events clearly play critical roles in parasitism evolution. Compared with HGTs, various PPCgenes or DMgenes PIs and their relationship with DUP events suggest that pre-adaptations perhaps are as important as HGTs for the evolution of parasitism genes. Bacterial genes must gain eukaryotic features, such as intron and codon usage, as part of an adaptation process to host genome after acquisition by HGTs, according to the description by Haegeman and colleagues in 2011.[6] Therefore, the mechanism of HGT cannot be completely independent of the mechanism of adaptation. Overall, it is very likely that both pre-adaptation and HGT are crucial for the evolution of parasitism genes and play similar or related roles in the evolution of parasitism genes.

Compared with the contributions of pre-adaptation and HGT, evidence from the analyses in this study suggests a supporting but not necessary role of the DUP mechanism for the evolution of parasitism genes. The main difference between DUP mechanism and other two mechanisms lies in their relationship with the absolute occurrences of parasitism genes in trees. Although some researchers stated that DUPs can promote the emergence of new or more-specialized function through neo- or subfunctionalization,[44] it is more likely that DUPs contribute to maintaining or amplifying the evolutionary trend towards parasitism. Otherwise, it may be hard to explain unexpected occurrence of parasitism genes caused by DUPs in trees. In addition to the three inferred mechanisms, some evidence also suggests that speciation is related to the evolution of parasitism genes. Considering the relationship between speciation and cumulative effect of adaptation, some mechanisms, such as co-adaptation of some gene systems, may help to understand the role of speciation. A study in 2005 provided such evidence for the co-adaptation of the amylase gene system in D. melanogaster.[45]

In summary, the contributions of pre-adaptations are more important to the evolution of parasitism genes than those of DUPs, and they are as crucial as HGTs to parasitism. Identifying related events of pre-adaptations and HGTs would help to recognize many parasitism genes, and in turn to promote the understanding of the origin of parasitism gene. A common procedure for the evolution of all parasitism genes including "pioneer" parasitism genes may be deduced when more information are available. Unfortunately, current lack of detailed information about the evolution of parasitism gene caused immature understanding of this evolutionary process as displayed in Figure 6. However, this figure provides a global view of the evolution of parasitism genes as accurately as possible. It was noticed that "PPC" feathers obtained through adaptation steps refer to a series of features acquired due to pre-adaptation, making parasitism genes exhibit high similarities between them and their homologs in P. pacificus or D. melanogaster species.

Future developments of evolutionary analysis. This analysis was performed based on orthologous groups downloaded from HelmCoP, and all were the results of OrthoMCL clustering. Generally, this method has advantages over other methods on within-group consistency of protein function and domain architecture.[46] However, some simple algorithms, like cRBH, could be a better predictor for the inference of single-copy gene.[47] Dubious assignments of several protein sequences to certain orthologous groups were still found when prediction results of protein domain were available. However, genes within most groups were still reliable for evolutionary analysis because there are few single-copy genes but widespread gene duplications in genomes of nematodes and other eukaryotes.[48] Grouping of "pioneer" parasitism genes by OrthoMCL supported this idea. Lack of bacterial genes in orthologous groups from HelmCoP really caused some problems for evolutionary analysis. Homologs of parasitism genes in bacterial species could be critical for detecting HGTs in bacteria and nematodes. Furthermore, bacterial sequences are appropriate outgroups for tree inferences. Therefore, it is reasonable to construct orthologous groups containing bacterial genes in future analysis.

CD-HIT clustering is required to remove data redundancy, and generally homologous proteins are clustered at the level of 90% identity. In this analysis, clustering at this identity level would undoubtedly eliminate protein sequences of parasitism genes caused by duplication. Therefore, these sequences were clustered at the highest identity. However,

Table 4. Comparison of *P. pacificus* and *D. melanogaster*.

	P. PACIFICUS	D. MELANOGASTER
Life style	necromeny	necromeny
Food	*Escherichia coli*	microorganisms and fruits
Associated organism	scarab beetle	rotting fruits
Reproductive mode	self-fertilizing hermaphrodite	cross-fertilizing
Parasitic closed species	*Brugia malayi*; *Meloidogyne hapla*; *Meloidogyne incognita*; *Trichinella spiralis*; and so on	*Drosophila endobranchia*; *Drosophila carcinophila*; and so on
Phylogenetic position	Bilateria: Nematoda: Chromadoria	Bilateria: Coelomata

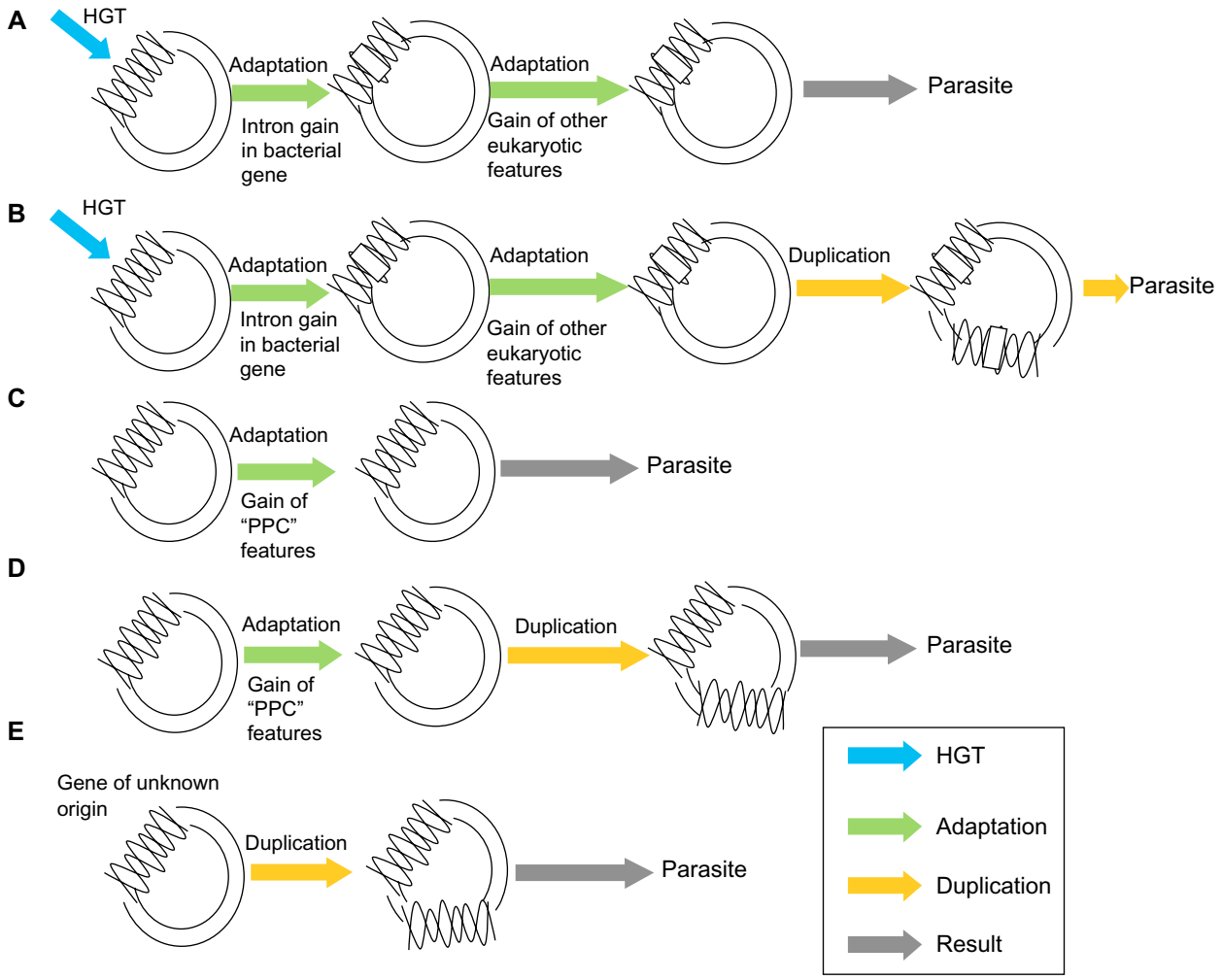

Figure 6. Five putative procedures for the evolution of parasitism gene. (**A**) and (**B**) are the HGT procedures; (**C**) and (**D**) are the adaptation procedures; (**E**) is the evolution of "pioneer" parasitism genes through duplication. The circular ring refers to the whole genome of host, not indicating a circular chromosome. The double helix refers to a gene, and the square refers to an intron.

sequences generated by such clustering may also be related to poor branch support and small branch length in trees. To solve this problem, domain trees for the same orthologous groups may be constructed using other fast and efficient methods, and comparing these trees could help to access the accuracy of tree topologies in future.

HGT-like PI, ie, the PI that can be detected as HGT by multiple HGT-detection programs, is an important concept proposed in this analysis, and can be detected in form of HGT by T-REX and other programs. Classifying HGT-like PIs is important for identifying the PPCgenes and the BOgenes PIs, and thus to detect parasitism gene from orthologous groups. In this analysis, these PIs were classified and identified by observation. How to automatically classify and identify HGT-like PIs remains a challenge.

Conclusion

Our analyses provided an evolutionary understanding of parasitism genes and their homologs in many species through the construction of domain trees. Tree topologies illustrated that parasitism genes can evolve through three inferred mechanisms under complicated backgrounds. Since information regarding the contributions of these mechanisms to the evolution of parasitism genes is limited, detection of HGT-like PIs together with DUPs and SPs provided insight about the specific contributions of these mechanisms. The PIs associated with pre-adaptations were found to be tightly related to the occurrences of parasitism genes. This suggests important roles of pre-adaptations for the evolution of parasitism genes, similarly to previously reported HGTs. In contrast, DUP may play a supporting role in the evolution towards parasitism. Furthermore, SP may also affect the evolution of some parasitism genes under certain situations. In addition, Pristionchus pacificus was suggested as a common model organism for the evolution of parasitism in most nematodes species. Pre-adaptations and HGTs can be detected as the PPCgenes PIs and the BOgenes PIs, and a method designed for taking into account these PIs may be able to detect many parasitism genes from multiple orthologous groups.

Acknowledgements

We would like to thank John Martin at Nematode.net and Anders Gorm Pedersen at CBS DTU for their help in our studies.

Author Contributions

Conceived and designed the experiments: YY. Analysed the data: YY. Wrote the first draft of the manuscript: YY. Contributed to the writing of the manuscript: YY, DL. Agree with manuscript results and conclusions: YY, DL. Jointly developed the structure and arguments for the paper: YY, DL. Made critical revisions and approved final version: YY, DL. All authors reviewed and approved of the final manuscript.

DISCLOSURES AND ETHICS

As a requirement of publication the authors have provided signed confirmation of their compliance with ethical and legal obligations including but not limited to compliance with ICMJE authorship and competing interests guidelines, that the article is neither under consideration for publication nor published elsewhere, of their compliance with legal and ethical guidelines concerning human and animal research participants (if applicable), and that permission has been obtained for reproduction of any copyrighted material. This article was subject to blind, independent, expert peer review. The reviewers reported no competing interests.

Supplementary Data

Supplementary Table 1. Orthologous groups, domains and their associated trees.

Supplementary Table 2. Identified parasitism genes, representative HGT-like PIs, duplication, and speciation events.

Supplementary Figure 1. Cladogram of the 1012_tree with aBayes support.

Supplementary Figure 2. Two species trees modified using Mesquite.

Supplementary Figure 3. Cladogram of the 1651_tree.

Supplementary Figure 4. Cladogram of the 1796_tree.

REFERENCES

1. Smant G, Stokkermans JPWG, Yan Y, et al. Endogenous cellulases in animals: isolation of beta-1,4-endoglucanase genes from two species of plant-parasitic cyst nematodes. *Proc Natl Acad Sci U S A*. 1998;95(9):4906–11.
2. Mitchum MG, Hussey RS, Davis EL, Baum TJ. Application of Biotechnology to Understand Pathogenesis in Nematode Plant Pathogens. In: Punja Z, De Boer SH, Sanfaçon H, eds. *Biotechnology and Plant Disease Management*. London, UK: CABI; 2007:580.
3. Ying X. *The Analysis of Characteristics of Parasitic Proteins and the Building of Clustering Model* [MS thesis dissertation]: School of Life Sciences, Xiamen Univeristy; 2008.
4. Baum TJ, Hussey RS, Davis EL. Root-knot and cyst nematode parasitism genes: the molecular basis of plant parasitism. *Genet Eng (N Y)*. 2007;28:17–43.
5. Davis EL, Hussey RS, Baum TJ, et al. NEMATODE PARASITISM GENES. *Annu Rev Phytopathol*. 2000;38:365–96.
6. Haegeman A, Jones JT, Danchin EG. Horizontal gene transfer in nematodes: a catalyst for plant parasitism? *Mol Plant Microbe Interact*. 2011;24(8):879–87.
7. Plantard O, Bouju-Albert A, Malard MA, Hermouet A, Capron G, Verheyden H. Detection of Wolbachia in the tick Ixodes ricinus is due to the presence of the hymenoptera endoparasitoid Ixodiphagus hookeri. *PLoS One*. 2012;7(1):e30692.
8. Guo Y, Ni J, Denver R, Wang X, Clark SE. Mechanisms of molecular mimicry of plant CLE peptide ligands by the parasitic nematode Globodera rostochiensis. *Plant Physiol*. 2011;157(1):476–84.
9. Coghlan A. Nematode genome evolution. 2005 Sep 7. In: WormBook: The Online Review of C. elegans Biology [Internet]. Pasadena (CA): WormBook; 2005. Available from: http://www.ncbi.nlm.nih.gov/books/NBK19768/.
10. Danchin EG, Rosso MN, Vieira P, et al. Multiple lateral gene transfers and duplications have promoted plant parasitism ability in nematodes. *Proc Natl Acad Sci U S A*. 2010;107(41):17651–6.
11. Abubucker S, Martin J, Taylor CM, Mitreva M. HelmCoP: an online resource for helminth functional genomics and drug and vaccine targets prioritization. *PloS one*. 2011;6(7):e21832.
12. Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol*. 2011;60(5):685–99.
13. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. 2010;59(3):307–21.
14. Koonin EV. Orthologs, paralogs, and evolutionary genomics. *Annu Rev Genet*. 2005;39(39):309–38.
15. Fontana P, Cestaro A, Velasco R, Formentin E, Toppo S. Rapid annotation of anonymous sequences from genome projects using semantic similarities and a weighting scheme in gene ontology. *PloS One*. 2009;4(2):e4619.
16. Stothard P. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *Biotechniques*. 2000;28(6):1102–4.
17. Huang Y, Niu B, Gao Y, Fu L, Li W. CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics*. 2010;26(5):680–2.
18. Li W, Jaroszewski L, Godzik A. Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics*. 2001;17(3):282–3.
19. Bateman A, Birney E, Cerruti L, et al. The Pfam protein families database. *Nucleic Acids Res*. 2002;30(1):276–80.
20. Marchler-Bauer A, Zheng C, Chitsaz F, et al. CDD: conserved domains and protein three-dimensional structure.*Nucleic Acids Res*. 2013;41(Database issue): D348–52.
21. Carver T, Bleasby A. The design of Jemboss: a graphical user interface to EMBOSS. *Bioinformatics*. 2003;19(14):1837–43.
22. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet*. 2000;16(6):276–7.
23. Katoh K, Kuma K, Miyata T, Toh H. Improvement in the accuracy of multiple sequence alignment program MAFFT. *Genome Inform*. 2005;16(1):22–33.
24. Penn O, Privman E, Ashkenazy H, Landan G, Graur D, Pupko T. GUIDANCE: a web server for assessing alignment confidence scores. *Nucleic Acids Res*. 2010;38(Web Server issue):W23–8.
25. Abascal F, Zardoya R, Posada D. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics*. 2005;21(9):2104–5.
26. Page RD. TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci*. 1996;12(4):357–8.
27. Rambaut A. FigTree. http://tree.bio.ed.ac.uk/software/figtree/. Updated Dec 5, 2012. Accessed May 4, 2013.
28. Wicker N, Perrin GR, Thierry JC, Poch O. Secator: a program for inferring protein subfamilies from phylogenetic trees. *Mol Biol Evol*. 2001;18(8):1435–41.
29. Letunic I, Bork P. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation.*Bioinformatics*. 2007;23(1):127–8.
30. Letunic I, Bork P. Interactive Tree Of Life v2: online annotation and display of phylogenetic trees made easy.*Nucleic Acids Res*. 2011;39(Web Server issue):W475–8.
31. Maddison W, Maddison D. Mesquite: a modular system for evolutionary analysis. http://mesquiteproject.org. Updated Sep 30, 2011. Accessed Jul 18, 2013.
32. Keeling PJ, Palmer JD. Horizontal gene transfer in eukaryotic evolution. *Nat Rev Genet*. 2008;9(8):605–18.
33. Ruths DA. *Application of phylogenetic incongruence to detecting and reconstructing interspecific recombination and horizontal gene transfer* [MS thesis dissertation]: Rice University; 2006.
34. Dávalos LM, Cirranello AL, Geisler JH, Simmons NB. Understanding phylogenetic incongruence: lessons from phyllostomid bats. *Biol Rev Camb Philos Soc*. 2012;87(4):991–1024.
35. Boc A, Makarenkov V. New efficient algorithm for detection of horizontal gene transfer events. In: Benson G, Page RDM, eds. *Algorithms in Bioinformatics*. Vol 2812: Springer Berlin Heidelberg; 2003:190–201.
36. Boc A, Philippe H, Makarenkov V. Inferring and validating horizontal gene transfer events using bipartition dissimilarity. *Syst Biol*. 2010;59(2):195–211.
37. Boc A, Diallo AB, Makarenkov V. T-REX: a web server for inferring, validating and visualizing phylogenetic trees and networks. *Nucleic Acids Res*. 2012;40(Web Server issue):W573–9.
38. Huerta-Cepas J, Dopazo J, Gabaldón T. ETE: a python Environment for Tree Exploration. *BMC Bioinformatics*. 2010;11:24.
39. Dieterich C, Sommer RJ. How to become a parasite—lessons from the genomes of nematodes. *Trends Genet*. 2009;25(5):203–9.
40. Poulin R. *Evolutionary Ecology of Parasites*. 2nd ed. Princeton, New Jersey, U S A: Princeton University Press; 2007.
41. Vieira P. *The Pinewood Nematode Bursaphelenchus xylophilus*. Vol 1. The Netherlands: Brill; 2004.
42. Brown FD, D'Anna I, Sommer RJ. Host-finding behaviour in the nematode Pristionchus pacificus. *Proc Biol Sci*. 2011;278(1722):3260–9.

43. Baldwin JG, Nadler SA, Adams BJ. Evolution of plant parasitism among nematodes. *Annu Rev Phytopathol*. 2004;42:83–105.
44. Lynch M, Conery JS. The evolutionary fate and consequences of duplicate genes. *Science*. 2000;290(5494):1151–5.
45. Araki H, Yoshizumi S, Inomata N, Yamazaki T. Genetic coadaptation of the amylase gene system in Drosophila melanogaster: evidence for the selective advantage of the lowest AMY activity and of its epistatic genetic background. *J Hered*. 2005;96(4):388–95.
46. Chen F, Mackey AJ, Vermunt JK, Roos DS. Assessing performance of orthology detection strategies applied to eukaryotic genomes. *PLoS One*. 2007;2(4):e383.
47. Salichos L, Rokas A. Evaluating ortholog prediction algorithms in a yeast model clade. *PLoS One*. 2011;6(4):e18755.
48. Woollard A. Gene duplications and genetic redundancy in C. elegans. *WormBook*. 2005:1–6.