Phylotranscriptomics of *Pristionchus* nematodes reveals parallel gene loss in six hermaphroditic lineages

Christian Rödelsperger[1,2,*], Waltraud Röseler[1], Neel Prabh[1], Kohta Yoshida[1], Christian Weiler[1], Matthias Herrmann[1], Ralf J. Sommer[1,*]

[1] Department for Evolutionary Biology, Max-Planck-Institute for Developmental Biology, Max-Planck-Ring 9, 72076 Tübingen, Germany

[2] Lead Contact

[*]Author for Correspondence: christian.roedelsperger@tuebingen.mpg.de

ralf.sommer@tuebingen.mpg.de

Keywords: comparative genomics, genome evolution, convergent evolution, reproductive mode, mating system, RNA-seq, sex-specific, *C. elegans*

Running title: Phylotranscriptomics of *Pristionchus* nematodes

Please cite as: Rödelsperger C, Röseler W, Prabh N, Yoshida K, Weiler C, Herrmann M, Sommer RJ (2018). Phylotranscriptomics of Pristionchus Nematodes Reveals Parallel Gene Loss in Six Hermaphroditic Lineages. Curr Biol. 28:3123-3127. DOI: 10.1016/j.cub.2018.07.041

Summary

Mutation and recombination are main drivers of phenotypic diversity, but the ability to create new allelic combinations is strongly dependent on the mode of reproduction. While most animals are dioecious (i.e. separated male and female sexes), in a number of evolutionary lineages females have gained the ability to self-fertilize [1,2] with drastic consequences on effective recombination rate, genetic diversity, and the efficacy of selection [3]. In the genus *Caenorhabditis*, such hermaphroditic or androdioecious lineages including *C. briggsae* and *C. tropicalis* display a genome shrinkage relative to their dioecious sister species *C. nigoni* and *C. brenneri*, respectively [4,5]. However, common consequences of reproductive modes on nematode genomes remain unknown, because most taxa contain single or few androdioecious species. One exception is the genus *Pristionchus* with seven androdioecious species. *Pristionchus* worms are found in association with scarab beetles in worldwide samplings, resulting in deep taxon sampling and currently 39 culturable and available species. Here, we use phylotranscriptomics of all 39 *Pristionchus* species to provide a robust phylogeny based on an alignment of more than 2000 orthologous clusters, which indicates that the seven androdioecious species represent six independent lineages. We show that gene loss is more prevalent in all hermaphroditic lineages than in dioecious relatives and that the majority of lost genes evolved recently in the *Pristionchus* genus. Further, we provide evidence that genes with male-biased expression are preferentially lost in hermaphroditic lineages. This supports a contribution of adaptive gene loss to shaping nematode genomes following the evolution of hermaphroditism.

Results

To study the phylogenetic relationships within the genus *Pristionchus*, and to test if gene loss represents a general consequence of switching reproductive modes, we sequenced RNA-seq libraries of all 39 culturable *Pristionchus* species. These include nine undescribed species that were found on recent sampling trips in east Asia and that are currently under description. In addition, we sequenced the transcriptomes of two outgroups, *Parapristionchus giblindavisi* and *Micoletzkya japonica*. RNA-seq libraries of mixed-stage cultures yielded a median of 14 million paired reads (interquartile range (IQR):11-19 million reads), which were assembled *de novo* into a median of ~70,000 gene fragments (IQR:62,000-83,000, Table S1) with Trinity software [6]. The quality of the resulting transcriptomes was assessed by an approach for benchmarking universal single-copy orthologs (BUSCO) [7]. With a few exceptions, most transcriptomes show a comparably high level of BUSCO completeness (IQR:89-93%), comparable assembly sizes, and numbers of clustered proteins (Table S1).

We implemented a bioinformatic pipeline (see *Methods*) to produce a concatenated protein sequence alignment (2,092 orthologous gene clusters, ~ 350,000 amino acids) of orthologous genes and computed a maximum likelihood tree representing the phylogenetic relationships between species (Figure 1A). The resulting phylotranscriptomic species tree is largely consistent with previous phylogenetic analysis [8], identified all previously defined *Pristionchus* clades and allowed us to determine the phylogenetic position of all nine novel species, seven of which fall into the previously defined “pacificus”-group. Another novel species, RS5964 n. sp., extends the group around *P. triformis* and the novel species RS5918 n. sp. is a new sister species to the androdioecious *P. fissidentatus*. Most importantly, our new phylogeny confirms that hermaphroditism has indeed evolved at least six times in the history of the genus *Pristionchus*. *P. boliviae* and *P. mayeri*, belonging to the only hermaphroditic lineage with two species, are at a greater phylogenetic distance than

other species pairs and their reproductive isolation has been experimentally confirmed [9].

The emergence of hermaphroditic lineages in different parts of the phylogeny effectively represents six independent biological replicates, which facilitates a broad investigation of the genomic impact of changes in reproductive modes. Therefore, we analyzed if hermaphroditic *Pristionchus* lineages show a stronger trend towards gene loss with respect to their dioecious sister species. We focused on the analysis of gene families (defined as orthologous gene clusters [10]) rather than counting total number of genes, because gene numbers can more easily be influenced by a few large gene expansions. Comparison of all six hermaphroditic lineages with one closely related and one outgroup species, both of which are dioecious, showed 6-77% more losses of gene families in all hermaphroditic lineages than in the dioecious sister species (Figure 1B). With the exception of presumably the most recent event in the "pacificus"-group (P=0.01, binomial test) all other comparisons showed highly significant trends towards higher gene loss in hermaphroditic lineages ($P < 10^{-7}$). Given that gene absence in a transcriptome might be simply due to expression changes, we performed an analysis (see *Methods*) combining transcriptomic and genome data for *P. pacificus* [11] and *P. exspectatus [12]* as well as the transcriptome assembly of *P. arcanus* to test if predicted losses in *P. pacificus* and *P. exspectatus* lacked orthologous sequences in genomic data. Even though further analysis will be needed to confirm gene losses in individual cases [13,14], our genome-wide analysis shows that 80% of predicted gene losses were indeed supported by genomic data (Figure 1C). Thus, our phylotranscriptomic analysis strongly supports the conclusion of an overall higher rate of gene loss in hermaphroditic lineages as originally observed in *Caenorhabditis* [4,5,15]. However, our study shows this trend at much higher phylogenetic resolution including a close sister species for each of the six hermaphroditic *Pristionchus* lineages.

To test the current model that genes with male-specific functions are a primary source of loss [4], we generated sex-specific RNA-seq data of young adult worms for *P.*

pacificus, *P. exspectatus*, and *P. arcanus*. This resulted in the identification of 1,775-3,841 male-biased genes across the three species. The comparison of expression ratios between males and females/hermaphrodites for conserved and lost genes indicates that genes which were lost in the androdioecious *P. pacificus* are significantly enriched among genes with male-biased expression in *P. arcanus* ($P = 10^{-14}$, Wilcoxon rank sum test, Figure 1D) and *P. exspectatus* ($P = 10^{-4}$, Figure 1E). However, this trend does not hold true for genes which were lost in the dioecious *P. exspectatus* (Figure 1 D,F). This demonstrates that the tendency of genes with male-specific functions to be lost in hermaphroditic lineages is broadly observed across the nematode phylum.

Next, we studied whether the same gene families have been repeatedly lost in androdioecious species. For this, we analyzed the number of represented species in a given gene family as a proxy for gene age (Fig. 2A). We found a division of gene families into one older class of genes that are shared across almost all species and another class of young genes, each of which is restricted to a subset of *Pristionchus* species (Fig. 2A). However, gene families with recent losses in either dioecious or hermaphroditic lineages tend to be highly taxonomically restricted and almost completely lack the class representing older gene families ($P < 10^{-62}$, Wilcoxon rank sum test, Fig. 2B). This indicates that it is a general characteristic of gene loss to primarily affect young genes, rather than that members of old gene families are repeatedly lost.

Finally, we characterized those genes that were lost in hermaphroditic lineages by testing for overrepresentation of protein domains (Figure 2B). Consistent with a recent observation in two androdioecious *Caenorhabditis* species [16], we found BTB domain containing proteins to be significantly enriched among gene losses in hermaphroditic lineages in at least four out of six comparisons. Additionally, we found preferential loss of C-type lectins, and genes with a DUF316-type chymotrypsin domain in hermaphroditic lineages. However, other groups of proteins such as kinases and zinc finger motif containing

proteins tend to be lost in dioecious as well as hermaphroditic lineages. Together with the above-mentioned tendency of young genes to be lost in genomes of both reproductive modes, this further highlights the fact that many gene losses are not associated with hermaphroditism and simply reflect general trends of nematode genome evolution.

Discussion

While the overwhelming majority of free-living nematodes, in particular rhabditid nematodes, are dioecious, self-fertilizing hermaphrodites have evolved in at least seven genera, i.e. *Caenorhabditis*, *Oscheius*, *Heterorhabditis*, *Auanema*, *Allodiplogaster*, *Diplogasteroides*, and *Pristionchus* [17]. As most taxa contain few or only one androdioecious species, the genus *Pristionchus* with seven androdioecious species is ideally suited to study broad genomic changes that are associated with changes in reproductive modes. Here, we used phylotranscriptomics [18,19] of the *Pristionchus* genus not only to resolve the evolutionary relationships between species, but also to compare patterns of gene loss between dioecious and androdioecious sister lineages. Even though the assembled transcriptomes do not cover all existing genes, they still represent large portions of the total gene content and combined analysis of genomic and transcriptomic data for *P. pacificus* and *P. exspectatus* estimates 80% of losses to be true (Figure 1C). Our analysis demonstrates a higher rate of gene loss in all six hermaphroditic *Pristionchus* lineages with respect to their dioecious sister species. Various processes have been proposed to drive the gene loss in androdioecious species, such as the loss of genes with male-specific functions most likely due to relaxed evolutionary constraint [4,15], the loss of genes due to hitchhiking of deleterious alleles, which should be more pronounced in partially selfing species with lower effective recombination rate [20], and the transmission bias of smaller chromosomes in hermaphrodites [21]. The analysis of sex-specific transcriptomes of *Pristionchus* nematodes confirms previous findings of loss of genes with male-biased expression in *Caenorhabditis* [4] and is also consistent with the

finding of rapid turnover of male-biased genes [22]. However, gene loss may also be a consequence of mechanisms with little selectivity. For example, young genes are generally under weaker purifying selection [23] and have a higher propensity for being lost (Figure 2A). This holds true in hermaphroditic as well as dioecious lineages.

The current phylogeny of the *Pristionchus* genus suggests that different hermaphroditic lineages are of varying age, ranging from a very recent lineage such as the one consisting of *P. pacificus* to what seems like a considerably older lineage (*P. boliviae* and *P. mayeri)* where speciation may have occurred within an androdioecious branch. If further sampling efforts would allow more robust age estimates, future studies can address if the rate of gene loss remains constant over time, and if the age of a hermaphroditic lineage correlates with increasing degeneration of male-specific functions leading to an ultimate extinction. This could explain the apparently young age of all extant hermaphroditic lineages.

Acknowledgments

We would like to thank Dr. Jan M. Falcke for help in the culturing of worms and the extraction of RNA and Dr. Michael S. Werner for critically reading the manuscript. This project was
funded by the Max-Planck Society.

Authors contributions

Conceptualization, C.R.; Investigation W.R.; Formal Analysis, C.R.; Resources, M.H., C.W., N.P., and K.Y.; Writing – Original Draft, C.R., Writing – Review & Editing, C. R and R.J.S; Funding Acquisition, R.J.S.

Declaration of Interests

The authors declare no competing interests.

Figure legends

Figure 1. Phylotranscriptomics reveals loss of male-biased genes

A) Transcriptomes of 41 species (Table S1) were used to reconstruct a phylogenetic tree of the *Pristionchus* genus. Branch lengths denote the number of substitutions per site and numbers at internal nodes indicate bootstrap support (* := 100/100). The distribution of reproductive modes (indicated by pictograms) shows that hermaphroditism evolved at least six times independently. B) Each hermaphroditic lineage was compared with two dioecious species and gene families with specific species distribution patterns were counted. The squares indicate categories of gene families that are either present (green check) or absent (red cross) in a given species. The underlying numbers indicate the number of gene families that follow a given pattern. C) Based on genomic data for *P. pacificus* and *P. exspectatus*, gene losses were classified into 'true losses' or cases of 'no expression'. The bars show the median and range of the class proportions across ten data sets. D-F) The histograms show the distribution of male-biased expression in *P. arcanus* (D), *P. exspectatus* (E), and *P. pacificus* (F) for either lost or conserved genes. Conserved genes denote genes with orthologs in all three species.

Figure 2. Age distribution and protein domains of lost genes

A) The gray histogram shows the distribution of the number of species in gene families that are represented in at least the two dioecious species of interest. The left peak represents taxonomically restricted genes that arose recently in the *Pristionchus* genus. The right peak indicates old gene families that are broadly conserved across the genus. The two curves indicate the species distribution patterns for gene families with losses in dioecious (blue) and

androdioecious (red) species. B) The grid shows overrepresented protein domains among gene families with specific species distribution patterns. The color code indicates combination of enrichment factors and FDR corrected p-values.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Christian Rödelsperger (christian.roedelsperger@tuebingen.mpg.de).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Worm culturing

Pristionchus and diplogastrid strains were thawed from nitrogen long-term storage and put on nematode growing media (NGM) plates seeded with *E. coli* OP50 at 20°C. If necessary, worm plates were cleaned by bleaching [24]. Clean worms were transferred with a wormpick to fresh NGM plates and allowed to recover for several generations.

METHOD DETAILS

RNA extraction and library preparation

For each species, we rinsed 2-3 mixed-stage plates with M9 buffer and collected worm

pellets by slow centrifugation at 1300 rpm for 3 minutes at 4°C. Total RNA was isolated using standard Trizol extraction following the manufacturers' instructions (Zymo Research, CA, USA). RNA concentration was quantified using Qubit (Invitrogen Life technologies, CA, USA) and Nanodrop (PEQLAB Biotechnologie GmbH, Erlangen, Germany) measurements. RNAseq libraries were prepared using TruSeq RNA library preparation kit v2 (Illumina Inc, CA, USA) according to the manufacturer's instructions from 1 µg of total RNA in each sample. Libraries were quantified using Qubit and Bioanalyzer measurements (Agilent Technologies, CA, USA) and normalized to 2.5 nM. Samples were sequenced as 150bp paired end reads on multiplexed lanes of an Illumina HiSeq3000 (Illumina Inc, CA, USA).

Sex-specific RNA-seq

To generate separate transcriptomes for males and females in *P. exspectatus* and *P. arcanus*, as well as separate transcriptomes for males and hermaphrodites in *P. pacificus*, worms of a specific sex were picked individually to a fresh plate without *E. coli* OP50 in order to reduce the level of bacterial contamination. After a second confirmation of their sex, 200 J3 to young adult worms were transferred with a wormpick to a 1,5ml Eppendorf tube containing 1 ml of M9. Worms were washed twice with fresh M9 in order to remove remaining OP50. After another short centrifugation (tabletop centrifuge), the supernatant was removed and the worm pellet was immediately frozen at -80°C. Total RNA was isolated using standard Trizol extraction following the manufacturers' instructions (Zymo Research, CA, USA). RNA concentration was quantified using Qubit (Invitrogen Life technologies, CA, USA) and Nanodrop (PEQLAB Biotechnologie GmbH, Erlangen, Germany) measurements. RNAseq libraries were prepared using TruSeq RNA library preparation kit v2 (Illumina Inc, CA, USA) according to the manufacturer's instructions from 1 µg of total RNA in each sample. Libraries were quantified using Qubit and Bioanalyzer measurements (Agilent Technologies, CA, USA) and normalized to 2.5 nM. For each sex and species, three

biological replicates were made, which were multiplexed and sequenced as 150bp paired end reads on a single lane of an Illumina HiSeq3000 (Illumina Inc, CA, USA).

Transcriptome assembly

RNA-seq data sets were assembled into transcriptomes by the Trinity software (version v2.2.0 with default options) [6]. Trinity assigns contigs into isoforms, genes, and clusters. Initial assemblies were evaluated by the BUSCO approach [7].

Phylotranscriptomic pipeline

We condensed isoform information in the de novo transcriptomes by arbitrarily selecting only the first isoform that was reported by Trinity assembler (_i1) and we called potential ORFs by choosing the longest complete or partial ORF among all six reading frames if they were at least 60 amino acids long.

We ran a clustering approach on protein sequences as implemented in the Software cd-hit (version 4.3, default options) to remove recent duplicates and unresolved isoforms from the sets of predicted protein sequences [25]. We then ran all pairwise BLASTP searches (version 2.2.31 [26], e-value < 10^{-5}) and generated orthologous gene clusters by the help of the orthAgogue and mcl (version 1:14-137-1) programs [10,27], which represents a more efficient implementation of the widely employed orthoMCL approach [28]. We aligned orthologous clusters with at most one protein per species by MUSCLE (version 3.8.31, [29]). To reduce the effect of spurious alignments on the resulting phylogeny, we only considered alignments with data from at least 14 species and with at least 50 amino acid positions with coverage in all represented species. All alignments, passing these filters were then

concatenated and the final species tree was constructed by rapid bootstrap analysis and search for the best scoring maximum-likelihood tree with RAxML (version 8.2.9, options -m PROTGAMMAILG -f a -N 100) [30]. We chose the LG substitution model, because previous tests of various substitution models for dozens of *Pristionchus* gene families recurrently identified combinations including this substitution model as the best [31,32]. To test if intraspecies polymorphism could affect the resulting phylogeny, we sequenced RNA-seq libraries of six divergent strains of *P. pacificus* (RSA016, RSB001, PS1843, RS5266, RS5297, and RSB020). Rerunning our pipeline with these strains included, placed all of them at them same position as the reference strain *P. pacificus* RS2333, suggesting that our pipeline is robust to moderate levels of intraspecies diversity.

Analysis of gene loss

To study gene loss, we focused on orthologous gene clusters. These clusters were defined by orthAgogue and mcl programs taking pairwise BLASTP results as input (see above). We treated each orthologous cluster as individual gene family and counted the number of gene families with a particular species distribution pattern. For each hermaphroditic lineage, we chose a close sister species from the same group and an outgroup with a high quality transcriptome assembly to screen for patterns that would reflect losses at specific branches. This was based on the most parsimonious interpretation of the species distribution in a given gene family. We chose to count losses of gene families instead of individual genes, because this makes our analysis more robust with regard to large expansion in individual gene families. To define a baseline distribution of the number of represented species across gene families, we first calculated the species distribution patterns of all gene families that include genes from the dioecious sister and outgroup species (Figure 2A). This distribution was compared with gene families that are present in the outgroup but missing in one of the other species. To estimate the fraction of true losses vs. missing transcriptomic evidence, we

screened for further support of predicted gene loss in genomic data of *P. pacificus* [11] and *P. exspectatus* [12]. This was done by running orthAgogue on a combined genomic and transcriptomic data set of *P. pacificus* (gene annotations, transcriptome assembly), *P. exspectatus* (gene annotations, transcriptome assembly), and *P. arcanus* (transcriptome assembly only). To obtain measures of robustness, we downsampled this orthology data into ten random subsets, each of which contained 10,000 gene families, and analysed each subset separately. Gene families with a member from the outgroup species *P. arcanus* were evaluated, whether missing evidence in transcriptomic data of one species was also reflected by missing genomic evidence of the same species. Such gene families were classified as 'true loss'. Conversely, if a gene family did not contain an assembled transcript but had a gene annotation from a given species, this was classified as 'no expression'. The median percentage of the classes 'true loss' and 'no expression' together with the range among the ten downsampled data sets is shown in Figure 1C.

QUANTIFICATION AND STATISTICAL ANALYSIS

Quantification of expression levels

For the quantification of expression levels in sex-specific transcriptomes of *P. pacificus*, *P. exspectatus*, and *P. arcanus*, we aligned raw reads to the transcriptome assemblies of the corresponding species and calculated expression levels with the align_and_estimate_abundance.pl script of the Trinity pipeline (version v2.2.0, options: --est_method RSEM --aln_method bowtie2). Median expression values were calculated from all three biological replicates per sex and species and $\log_2$ expression fold changes between male and females/hermaphrodites were used to test for preferential loss of male-biased genes (Figure 1 D-F). To avoid division by zero, genes with expression evidence (FPKM >0)

in all three male samples and no evidence of expression (FPKM=0) in all three female/hermaphrodite samples were assigned a $\log_2$ expression fold change of 11. Using an arbitrary cutoff of four times higher expression in males vs. females/hermaphrodites, we identified 3,841 male-biased genes in *P. arcanus* and 2,779 and 1,775 male-biased genes in *P. exspectatus* and *P. pacificus*, respectively. To test for the significance of the differential male-bias between lost and conserved genes, a Wilcoxon rank sum test was applied on the distributions of expression fold changes within both sets of genes.

Overrepresentation analysis of protein domains

Protein domains in protein sequences were annotated by the hmmsearch program (v3.1b2, -E 0.001) [33]. Overrepresentation of protein domains in gene sets, defined by specific patterns in gene family data, was tested by a Fisher's exact test against a background of all proteins and corrected p-values with the FDR method as implemented in the p.adjust function of R. Enrichment values were calculated as the ratio between the observed fraction of genes with a specific species distribution pattern and a given protein domain, and the product of the marginal probabilities.

DATA AND SOFTWARE AVAILABILITY

Raw reads and transcriptome assemblies have been submitted to the European nucleotide archive under the study accession PRJEB20959.

KEY RESOURCES TABLE

References

1. Denver, D.R., Clark, K.A., and Raboin, M.J. (2011). Reproductive mode evolution in nematodes: insights from molecular phylogenies and recently discovered species. Mol. Phylogenet. Evol. *61*, 584–592.

2. Avise, J.C., and Mank, J.E. (2009). Evolutionary perspectives on hermaphroditism in fishes. Sex Dev. *3*, 152–163.

3. Glémin, S., and Galtier, N. (2012). Genome evolution in outcrossing versus selfing versus asexual species. Methods Mol. Biol. *855*, 311–335.

4. Yin, D., Schwarz, E.M., Thomas, C.G., Felde, R.L., Korf, I.F., Cutter, A.D., Schartner, C.M., Ralston, E.J., Meyer, B.J., and Haag, E.S. (2018). Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. Science *359*, 55–61.

5. Fierst, J.L., Willis, J.H., Thomas, C.G., Wang, W., Reynolds, R.M., Ahearne, T.E., Cutter, A.D., and Phillips, P.C. (2015). Reproductive Mode and the Evolution of Genome Size and Structure in Caenorhabditis Nematodes. PLoS Genet. *11*, e1005323.

6. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., *et al.* (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

7. Waterhouse, R.M., Seppey, M., Simão, F.A., Manni, M., Ioannidis, P., Klioutchnikov, G., Kriventseva, E.V., and Zdobnov, E.M. (2017). BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol. Biol. Evol. Available at: http://dx.doi.org/10.1093/molbev/msx319.

8. Susoy, V., Herrmann, M., Kanzaki, N., Kruger, M., Nguyen, C.N., Rödelsperger, C., Röseler, W., Weiler, C., Giblin-Davis, R.M., Ragsdale, E.J., *et al.* (2016). Large-scale diversification without genetic isolation in nematode symbionts of figs. Sci Adv *2*, e1501031.

9. Kanzaki, N., Ragsdale, E.J., Herrmann, M., Susoy, V., and Sommer, R.J. (2013). Two androdioecious and one dioecious new species of pristionchus (nematoda: diplogastridae): new reference points for the evolution of reproductive mode. J. Nematol. *45*, 172–194.

10. Ekseth, O.K., Kuiper, M., and Mironov, V. (2014). orthAgogue: an agile tool for the rapid prediction of orthology relations. Bioinformatics *30*, 734–736.

11. Rödelsperger, C., Meyer, J.M., Prabh, N., Lanz, C., Bemm, F., and Sommer, R.J. (2017). Single-Molecule Sequencing Reveals the Chromosome-Scale Genomic Architecture of the Nematode Model Organism Pristionchus pacificus. Cell Rep. *21*, 834–844.

12. Rödelsperger, C., Neher, R.A., Weller, A.M., Eberhardt, G., Witte, H., Mayer, W.E., Dieterich, C., and Sommer, R.J. (2014). Characterization of genetic diversity in the nematode Pristionchus pacificus from population-scale resequencing data. Genetics *196*, 1153–1165.

13. Gilabert, A., Curran, D.M., Harvey, S.C., and Wasmuth, J.D. (2016). Expanding the view on the evolution of the nematode dauer signalling pathways: refinement through gene gain and pathway co-option. BMC Genomics *17*, 476.

14. Rödelsperger, C. (2018). Comparative Genomics of Gene Loss and Gain in Caenorhabditis and Other Nematodes. Methods Mol. Biol. *1704*, 419–432.

15. Thomas, C.G., Li, R., Smith, H.E., Woodruff, G.C., Oliver, B., and Haag, E.S. (2012). Simplification and Desexualization of Gene Expression in Self-Fertile Nematodes. Curr. Biol. *22*, 2167–2172.

16. Yin, D., Schwarz, E.M., Thomas, C.G., Felde, R.L., Korf, I.F., Cutter, A.D., Schartner, C.M., Ralston, E.J., Meyer, B.J., and Haag, E.S. (2018). Rapid genome shrinkage in a self-fertile nematode reveals sperm competition proteins. Science *359*, 55–61.

17. Denver, D.R., Clark, K.A., and Raboin, M.J. (2011). Reproductive mode evolution in nematodes: insights from molecular phylogenies and recently discovered species. Mol. Phylogenet. Evol. *61*, 584–592.

18. Egger, B., Lapraz, F., Tomiczek, B., Müller, S., Dessimoz, C., Girstmair, J., Škunca, N., Rawlinson, K.A., Cameron, C.B., Beli, E., *et al.* (2015). A transcriptomic-phylogenomic analysis of the evolutionary relationships of flatworms. Curr. Biol. *25*, 1347–1353.

19. Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., *et al.* (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc. Natl. Acad. Sci. U. S. A. *111*, E4859–68.

20. Cutter, A.D., and Jovelin, R. (2015). When natural selection gives gene function the cold shoulder. Bioessays *37*, 1169–1173.

21. Wang, J., Chen, P.-J., Wang, G.J., and Keller, L. (2010). Chromosome size differences may affect meiosis and genome size. Science *329*, 293.

22. Harrison, P.W., Wright, A.E., Zimmer, F., Dean, R., Montgomery, S.H., Pointer, M.A., and Mank, J.E. (2015). Sexual selection drives evolution and rapid turnover of male gene expression. Proc. Natl. Acad. Sci. U. S. A. *112*, 4393–4398.

23. Prabh, N., and Rödelsperger, C. (2016). Are orphan genes protein-coding, prediction artifacts, or non-coding RNAs? BMC Bioinformatics *17*, 226.

24. Stiernagle, T. (2006). Maintenance of C. elegans. WormBook, 1–11.

25. Huang, Y., Niu, B., Gao, Y., Fu, L., and Li, W. (2010). CD-HIT Suite: a web server for clustering and comparing biological sequences. Bioinformatics *26*, 680–682.

26. Altschul, S. (1990). Basic Local Alignment Search Tool. J. Mol. Biol. *215*, 403–410.

27. Enright, A.J., Van Dongen, S., and Ouzounis, C.A. (2002). An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. *30*, 1575–1584.

28. Li, L., Stoeckert, C.J., Jr, and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.

29. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

30. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.

31. Baskaran, P., Rödelsperger, C., Prabh, N., Serobyan, V., Markov, G.V., Hirsekorn, A., and Dieterich, C. (2015). Ancient gene duplications have shaped developmental stage-specific expression in Pristionchus pacificus. BMC Evol. Biol. *15*, 185.

32. Darriba, D., Taboada, G.L., Doallo, R., and Posada, D. (2011). ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics *27*, 1164–1165.

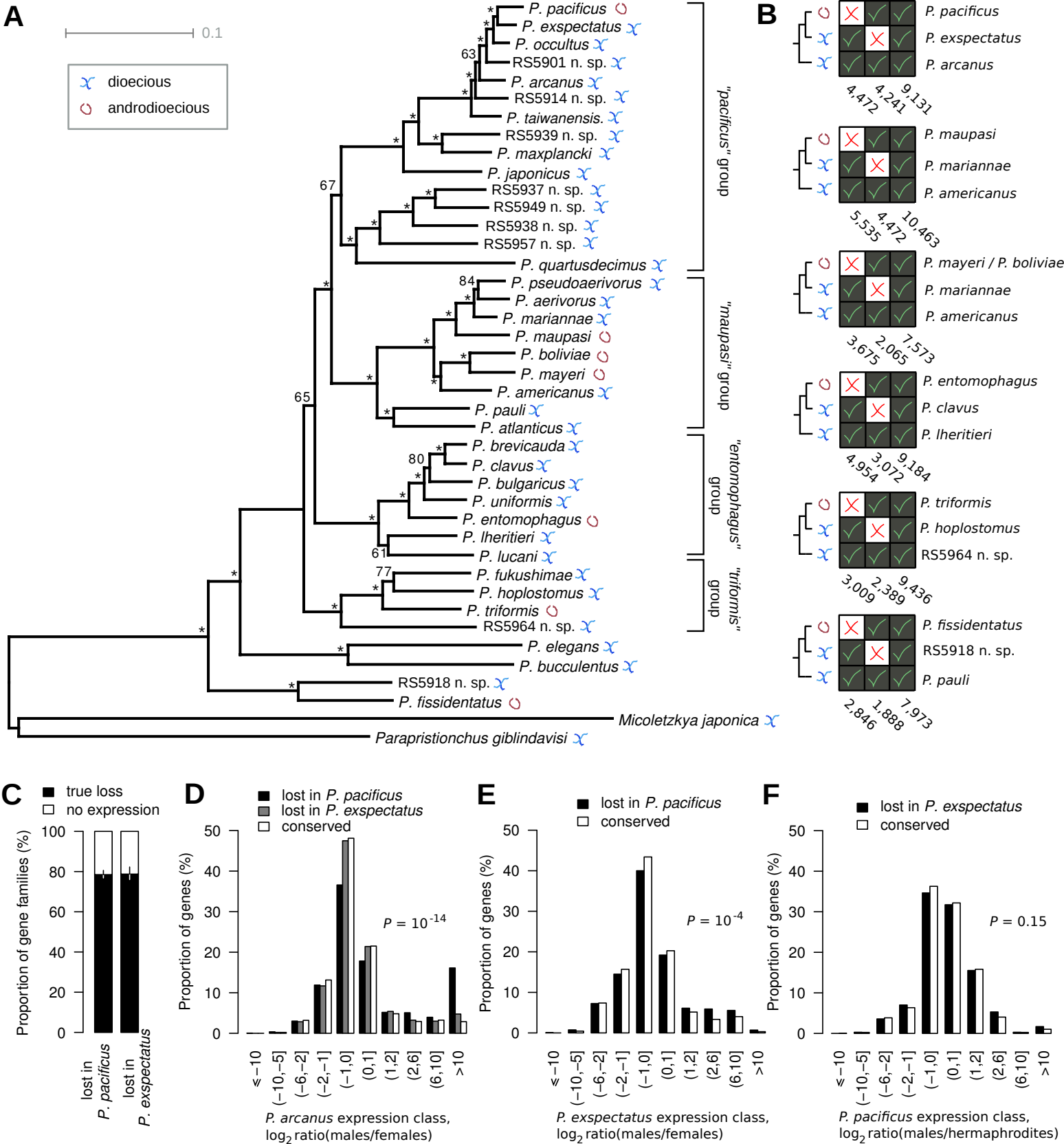
33. Mistry, J., Finn, R.D., Eddy, S.R., Bateman, A., and Punta, M. (2013). Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res. *41*, e121.

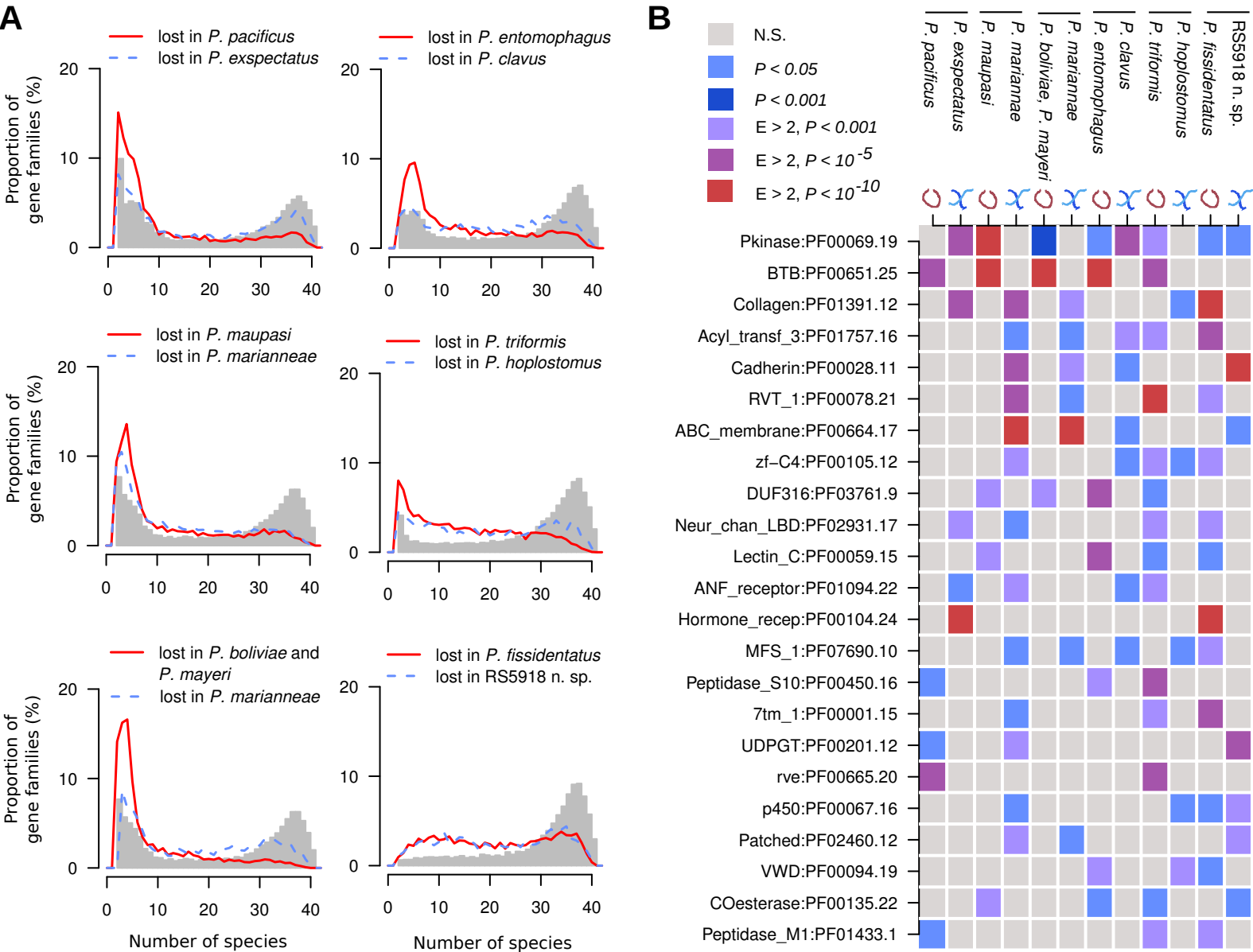
KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chemicals, Peptides, and Recombinant Proteins		
Direct-zol RNA MiniPrep	Zymo Research	R2050
TruSeq RNA library preparation kit v2	Illumina Inc	RS-122-2001
Deposited Data		
P. aerivorus	This study	ERS1802350
P. americanus	This study	ERS2071081
P. arcanus	This study	ERS2070939-ERS2070943 ERS2531917-ERS2531922
P. atlanticus	This study	ERS2076430
P. boliviae	This study	ERS2071083
P. brevicauda	This study	ERS2071082
P. bucculentus	This study	ERS2032584
P. bulgaricus	This study	ERS2032579
P. clavus	This study	ERS2071084
P. elegans	This study	ERS2071089
P. entomophagus	This study	ERS2071086
P. exspectatus	This study	ERS2070934-ERS2070938 ERS2531911-ERS2531916
P. fissidentatus	This study	ERS2071087
P. fukushimae	This study	ERS2071085
P. hoplostomus	This study	ERS2073557
P. japonicus	This study	ERS2071088
P. lheritieri	This study	ERS2032578
P. lucani	This study	ERS2076429
P. mariannae	This study	ERS1802349
P. maupasi	This study	ERS2304716
P. maxplancki	This study	ERS2071090
P. mayeri	This study	ERS1802353
P. occultus	This study	ERS2076428
P. pacificus	This study	ERS2070929-ERS2070933 ERS2531905-ERS2531910
P. pauli	This study	ERS2032580
P. pseudoaerivorus	This study	ERS2032581
P. quartusdecimus	This study	ERS2032582
P. taiwanensis	This study	ERS2076427
P. triformis	This study	ERS2032583
P. uniformis	This study	ERS2071079
Parapristinchus giblindavisi	This study	ERS2076425
Micoletzkya japonica	This study	ERS1755909
RS5901 n. sp.	This study	ERS2411366
RS5914 n. sp.	This study	ERS2411364

RS5918 n. sp.	This study	ERS2411359
RS5937 n. sp.	This study	ERS2411362
RS5938 n. sp.	This study	ERS2411363
RS5939 n. sp.	This study	ERS2411365
RS5949 n. sp.	This study	ERS2411361
RS5957 n. sp.	This study	ERS2411360
RS5964 n. sp.	This study	ERS2411902
RSA016	This study	ERS2535843
RSB001	This study	ERS2535844
PS1843	This study	ERS2535845
RS5266	This study	ERS2535846
RS5297	This study	ERS2535847
RSB020	This study	ERS2535848
Experimental Models: Organisms/Strains		
P. aerivorus	Sommer lab	RS5106
P. americanus	Sommer lab	RS5140
P. arcanus	Sommer lab	RS5527
P. atlanticus	Sommer lab	CZ3975ATL
P. boliviae	Sommer lab	RS5262
P. brevicauda	Sommer lab	RS5231
P. bucculentus	Sommer lab	RS5596
P. bulgaricus	Sommer lab	RS5283
P. clavus	Sommer lab	RS5284
P. elegans	Sommer lab	RS5229
P. entomophagus	Sommer lab	RS0144
P. exspectatus	Sommer lab	RS5522
P. fissidentatus	Sommer lab	RS5133
P. fukushimae	Sommer lab	RS5595
P. hoplostomus	Sommer lab	JU1090
P. japonicus	Sommer lab	RS5228
P. lheritieri	Sommer lab	SB245
P. lucani	Sommer lab	RS5050
P. mariannae	Sommer lab	RS5108
P. maupasi	Sommer lab	RS5861
P. maxplancki	Sommer lab	RS5594
P. mayeri	Sommer lab	RS5460
P. occultus	Sommer lab	RS5811
P. pacificus	Sommer lab	RS2333
P. pauli	Sommer lab	RS5130
P. pseudoaerivorus	Sommer lab	RS5139
P. quartusdecimus	Sommer lab	RS5230
P. taiwanensis	Sommer lab	RS5797
P. triformis	Sommer lab	RS5233
P. uniformis	Sommer lab	RS0141
Parapristinchus giblindavisi	Sommer lab	RS5555
Micoletzkya japonica	Sommer lab	RS5540
RS5901 n. sp.	Sommer lab	RS5901
RS5914 n. sp.	Sommer lab	RS5914

RS5918 n. sp.	Sommer lab	RS5918
RS5937 n. sp.	Sommer lab	RS5937
RS5938 n. sp.	Sommer lab	RS5938
RS5939 n. sp.	Sommer lab	RS5939
RS5949 n. sp.	Sommer lab	RS5949
RS5957 n. sp.	Sommer lab	RS5957
RS5964 n. sp.	Sommer lab	RS5964
RSA016	Sommer lab	RSA016
RSB001	Sommer lab	RSB001
PS1843	Sommer lab	PS1843
RS5266	Sommer lab	RS5266
RS5297	Sommer lab	RS5297
RSB020	Sommer lab	RSB020
Software and Algorithms		
Trinity (v2.2.0)	[6]	https://github.com/trinityrnaseq/trinityrnaseq/releases
BUSCO (v3.0.1)	[7]	http://busco.ezlab.org/
cd-hit (v 4.3)	[25]	http://weizhongli-lab.org/cd-hit/
BLASTP (v 2.2.31)	[26]	ftp://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/ LATEST/
orthAgogue	[10]	https://github.com/guyleonard/orthagogue
mcl (v 1:14-137-1)	[27]	https://micans.org/mcl/
MUSCLE (v3.8.31)	[29]	https://www.drive5.com/muscle/
RAxML (v 8.2.9)	[30]	https://sco.h-its.org/exelixis/software.html
hmmsearch (v3.1.b2)	[33]	http://hmmer.org/


A
0.1
dioecious
androdioecious
P. pacificus
P. exspectatus
P. occultus
RS5901 n. sp.
P. arcanus
RS5914 n. sp.
P. taiwanensis.
RS5939 n. sp.
P. maxplancki
P. japonicus
RS5937 n. sp.
RS5949 n. sp.
RS5938 n. sp.
RS5957 n. sp.
P. quartusdecimus
P. pseudoaerivorus
P. aerivorus
P. mariannae
P. maupasi
P. boliviae
P. mayeri
P. americanus
P. pauli
P. atlanticus
P. brevicauda
P. clavus
P. bulgaricus
P. uniformis
P. entomophagus
P. lheritieri
P. lucani
P. fukushimae
P. hoplostomus
P. triformis
RS5964 n. sp.
P. elegans
P. bucculentus
RS5918 n. sp.
P. fissidentatus
Micoletzkya japonica
Parapristionchus giblindavisi
63
67
84
65
80
61
77
"pacificus" group
"maupasi" group
"entomophagus" group
"triformis" group
B
P. pacificus
P. exspectatus
P. arcanus
4,472
4,241
9,131
P. maupasi
P. mariannae
P. americanus
5,535
4,472
10,463
P. mayeri / P. boliviae
P. mariannae
P. americanus
3,675
2,065
7,573
P. entomophagus
P. clavus
P. lheritieri
4,954
3,072
9,184
P. triformis
P. hoplostomus
RS5964 n. sp.
3,009
2,389
9,436
P. fissidentatus
RS5918 n. sp.
P. pauli
2,846
1,888
7,973
C
true loss
no expression
Proportion of gene families (%)
lost in P. pacificus
lost in P. exspectatus
D
lost in P. pacificus
lost in P. exspectatus
conserved
Proportion of genes (%)
P = 10-14
≤ −10
(−10,−5]
(−6,−2]
(−2,−1]
(−1,0]
(0,1]
(1,2]
(2,6]
(6,10]
>10
P. arcanus expression class, log2 ratio(males/females)
E
lost in P. pacificus
conserved
Proportion of genes (%)
P = 10-4
P. exspectatus expression class, log2 ratio(males/females)
F
lost in P. exspectatus
conserved
Proportion of genes (%)
P = 0.15
P. pacificus expression class, log2 ratio(males/hermaphrodites)


A
lost in P. pacificus
lost in P. exspectatus
lost in P. entomophagus
lost in P. clavus
lost in P. maupasi
lost in P. marianneae
lost in P. triformis
lost in P. hoplostomus
lost in P. boliviae and P. mayeri
lost in P. marianneae
lost in P. fissidentatus
lost in RS5918 n. sp.
Proportion of gene families (%)
Number of species
B
N.S.
P < 0.05
P < 0.001
E > 2, P < 0.001
E > 2, P < 10^-5
E > 2, P < 10^-10
P. pacificus
P. exspectatus
P. maupasi
P. marianneae
P. boliviae, P. mayeri
P. marianneae
P. entomophagus
P. clavus
P. triformis
P. hoplostomus
P. fissidentatus
RS5918 n. sp.
Pkinase:PF00069.19
BTB:PF00651.25
Collagen:PF01391.12
Acyl_transf_3:PF01757.16
Cadherin:PF00028.11
RVT_1:PF00078.21
ABC_membrane:PF00664.17
zf-C4:PF00105.12
DUF316:PF03761.9
Neur_chan_LBD:PF02931.17
Lectin_C:PF00059.15
ANF_receptor:PF01094.22
Hormone_recep:PF00104.24
MFS_1:PF07690.10
Peptidase_S10:PF00450.16
7tm_1:PF00001.15
UDPGT:PF00201.12
rve:PF00665.20
p450:PF00067.16
Patched:PF02460.12
VWD:PF00094.19
COesterase:PF00135.22
Peptidase_M1:PF01433.1