

The improved genome of the nematode *Parapristionchus giblindavisi* provides insights into lineage-specific gene family evolution

Waltraud Röseler ¹, Maximilian Collenberg ², Kohta Yoshida ¹, Christa Lanz,² Ralf J. Sommer ¹, Christian Rödelsperger ^{1,*}

¹Department for Integrative Evolutionary Biology, Max Planck Institute for Biology, 72076 Tübingen, Germany,

²Department for Molecular Biology, Max Planck Institute for Biology, 72076 Tübingen, Germany

*Corresponding author: Department for Integrative Evolutionary Biology, Max Planck Institute for Biology, Max-Planck-Ring 9, 72076 Tübingen, Germany.
Email: christian.roedelsperger@tuebingen.mpg.de

Abstract

Nematodes such as *Caenorhabditis elegans* and *Pristionchus pacificus* are extremely successful model organisms for comparative biology. Several studies have shown that phenotypic novelty but also conserved processes are controlled by taxon-restricted genes. To trace back the evolution of such new or rapidly evolving genes, a robust phylogenomic framework is indispensable. Here, we present an improved version of the genome of *Parapristionchus giblindavisi* which is the only known member of the sister group of *Pristionchus*. Relative to the previous short-read assembly, the new genome is based on long reads and displays higher levels of contiguity, completeness, and correctness. Specifically, the number of contigs dropped from over 7,303 to 735 resulting in an N50 increase from 112 to 791 kb. We made use of the new genome to revisit the evolution of multiple gene families. This revealed *Pristionchus*-specific expansions of several environmentally responsive gene families and a *Pristionchus*-specific loss of the de novo purine biosynthesis pathway. Focusing on the evolution of sulfatases and sulfotransferases, which control the mouth form plasticity in *P. pacificus*, reveals differences in copy number and genomic configurations between the genera *Pristionchus* and *Parapristionchus*. Altogether, this demonstrates the utility of the *P. giblindavisi* genome to date and polarizes lineage-specific patterns.

Keywords: comparative genomics; purine; sulfatase; metabolism; duplication; *Caenorhabditis elegans*; *Pristionchus pacificus*

Introduction

The nematode *Pristionchus pacificus* had initially been introduced as a satellite model organism for comparative studies with the classical model organism *Caenorhabditis elegans* (Sommer 2006). This has extended our understanding about the evolution of single genes (Markov et al. 2016; Moreno et al. 2016), developmental networks (Casasa et al. 2021; Sun et al. 2021), and organismal phenotypes such as morphology (Rudel et al. 2005; Harry et al. 2022) and behavior (Wilecki et al. 2015; Moreno et al. 2016). More recently, *P. pacificus* developed into a stand-alone model organism to study the genetics of phenotypic plasticity (Sommer et al. 2017) and the origin of new genes (Rödelsperger et al. 2019). *Caenorhabditis elegans* and *P. pacificus* belong to 2 different nematode families of the order Rhabditida. Both species have been estimated to have diverged 130–310 million years ago (Howard et al. 2022). The diplogastrid family has more than 30 genera (Susoy et al. 2015) and includes the sister genera *Pristionchus* and *Parapristionchus*. While almost 50 species of the genus *Pristionchus* are known largely due to their common association with beetles and other insects (Kanzaki et al. 2021), the genus *Parapristionchus* currently has only a single member, *Parapristionchus giblindavisi* (Kanzaki et al. 2012). Also, only a single strain of *P. giblindavisi* was sampled from a beetle specimen and all subsequent attempts to

isolate additional strains and species remained unsuccessful. Nonetheless, *P. giblindavisi* displays a number of morphological features that distinguish it from *Pristionchus* nematodes. Furthermore, phylogenetic analysis revealed deep divergence from other genera but also supports monophyly with the genus *Pristionchus* (Kanzaki et al. 2012). The genome of *P. giblindavisi* was sequenced as part of a phylogenomic study that involved the genomes of 8 *Pristionchus* species and 2 outgroup species to investigate the origin and evolution of new genes (Prabh et al. 2018). Compared to the other diplogastrid genomes in this phylogenomic data set, the assembly showed the lowest degree of contiguity and completeness level as assessed by the benchmarking of universal single-copy orthologs (BUSCO) approach (Simão et al. 2015). The assembly also contained the highest fraction of ambiguous base calls that could result from remaining heterozygosity or from overcompression of repeats during the assembly of Illumina short reads (Barrière et al. 2009; Gnerre et al. 2011). As the genome of *P. giblindavisi* is of particular importance to polarize lineage-specific patterns in the *Pristionchus* genus, we decided to resequence its genome using single-molecule sequencing technology of the Pacific Bioscience platform. This reduced the number of contigs by a factor of 10 and leads to 7-fold increased N50 value which is a commonly used metric to assess the contiguity

Received: June 30, 2022. Accepted: August 08, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

of a genome assembly and which is defined as the minimal contig length in the set of largest contigs that cover at least 50% of the assembly. We further make use of the improved genome to investigate multiple lineage-specific trends of gene family evolution and to ask whether these patterns are *Pristionchus*-specific or shared with its sister genus *Parapristionchus*.

Methods

Worm culturing and DNA sequencing

Nematodes of the inbred *P. giblindavisi* strain RS5555 were grown on nematode growth medium. Worms were washed off of 100 fully grown plates using M9 buffer and were gently pelleted by centrifugation at 1,300× g for 1 min. The pellet was washed twice in M9 before worms were frozen in liquid nitrogen and ground to a fine powder using a mortar and pestle. The powder was directly transferred into the lysis buffer from the Qiagen genomic DNA extraction kit, which was used in combination with Qiagen genomic tip columns (500/G) (Qiagen, Hamburg, Germany). The protocol was performed following the manufacturer's instructions. All steps involving sample vortexing were replaced by sample inverting to limit unwanted DNA shearing. DNA quality and quantity were determined with a NanoDrop ND 1000 spectrometer (PeqLab, Erlangen, Germany), a Qubit 2.0 Fluorometer (Thermo Fisher Scientific, Waltham, USA), and by a Femto pulse system (Agilent, CA, USA). A total of 15 µg genomic DNA was sheared to a target fragment size of 13 kb using a Megaruptor 2 device (Diagenode, Denville, USA). A 13-kb template library was prepared using the BluePippin size-selection system according to the manufacturer's protocol (P/N 100-286-000-07, Pacific Biosciences, CA, USA). The final library was sequenced on half of an SMRT cell of a Pacific Biosciences Sequel II instrument following the Magbead loading protocol.

Genome assembly

PacBio raw reads were filtered for circular consensus sequencing (CCS) reads with >Q20 using PacBio ccs with `-min-rq 0.99` and `-chunk 10` (version 4.1.1.0). The chunked bam files were merged using `pbmerge` (v0.23.0) resulting in 15 Gb of CCS reads which would translate into around 60X coverage of the *P. giblindavisi* genome. BAM files were converted to FASTA format using `bam2-fasta` program (version 1.3.0) and the genome was assembled with the `Falcon-unzip2` program (version 1.8.1) while applying `genome_size = 170,000,000`, `input_type = preads`, and `overlap_filterin_setting = -min-idt 99.9` options (Chin et al. 2016). The de novo assembly was then polished by mapping the Q20 filtered CCS reads onto the primary contigs using `pbmm2` (version 1.2.0) while applying the option `-preset CCS`. The resulting SAM files were subsequently used to polish the genomes using `racon` (version 1.4.10) while applying default parameters (Vaser et al. 2017). We identified and softmasked repeats using `RepeatScout` (version 1.0.5; Price et al. 2005) with a seed length of 14 and the minimum number of repeats of 5 and `RepeatMasker` (ver. 4.1.1, <http://www.repeatmasker.org>; accessed 2022 Jun 1). The masked assembly was processed by `HaploMerger2` to build the haploid genome (Huang et al. 2017). For the evaluation of genome assembly quality, the BUSCO pipeline (version 3.1.0) was run in genome mode (`-m` genome option) against the nematode odb9 data set (N=982 orthologs; Simão et al. 2015). Furthermore, previously generated Illumina short reads were downloaded from the European nucleotide archive (accessions: ERR2114858, ERR2114859, ERR2114860, and ERR2114861) and were aligned against both assemblies *P. giblindavisi* assemblies using the `BWA mem` program (version

0.7.17-r1188; Li and Durbin 2009). The resulting alignment files were merged with the `samtools merge` command (version 0.1.18 r982:295; Li et al. 2009). Alignment statistics and ambiguous base calls were generated using the `samtools flagstat` and `mpileup` commands. From the final assembly, a single contig spanning 4.7 Mb was removed after scanning for *Escherichia coli* contamination by BLASTN analysis (*e*-value < 0.00001) against the *E. coli* genome (GenBank accession: NZ_SMT000000000.1; Urrutia et al. 2020).

Gene annotation

Evidence-based gene annotations for the *P. giblindavisi* genome were generated based on protein homology and transcriptome data using the PPCAC software (version 1.0; Athanasouli et al. 2020; Rödelisperger 2021). Specifically, the community-curated gene annotations of *P. pacificus* (El Paco gene annotations, version 3; Athanasouli et al. 2020) and a mixed-stage transcriptome assembly of *P. giblindavisi* (Rödelisperger et al. 2018) were transferred to the new *P. giblindavisi* genome assembly by alignment with the software `exonerate` (version 2.2.0; Slater and Birney 2005). For each 100-bp window, one representative gene model was chosen by selecting the gene model with the longest open reading frame (minimum protein length 60 amino acids, at least 3 exons). All other alternative gene models within a 100-bp window were discarded. To assess the quality of gene annotations, the BUSCO pipeline (version 3.1.0) was run in protein mode (`-m prot` option) against the nematode odb9 data set (N=982 orthologs; Simão et al. 2015). Protein-coding genes were scanned for the presence of protein domains by the `hmmsearch` program of the HMMER package (*e*-value cutoff option: `-E 0.001`, version 3.3) using the Pfam database (Pfam-A.hmm, version 3.1b2; Bateman et al. 1999). The `hmmsearch` program was run with the same settings on proteins of 9 other diplogastrid genomes (Athanasouli et al. 2020; Rödelisperger 2021) and on protein data from *C. elegans*, *Bursaphelenchus xylophilus*, *Brugia malayi*, and *Trichinella spiralis* as retrieved from WormBase ParaSite (version WBPS14; Howe et al. 2017). For the latter data sets, only the isoform giving rise to the longest protein sequence was analyzed. The program `OrthoFinder` (version 2.5.2, default mode) was used to complement the protein domain analysis (Emms and Kelly 2019).

Phylogenetic and synteny analysis

Candidate orthologs were identified based on either the protein domains or manual BLASTP searches on <http://www.pristionchus.org> (accessed 2022 Jun 1), Wormbase (<http://www.wormbase.org>; accessed 2022 Jun 1), and Wormbase ParaSite (version WBPS16; Howe et al. 2017). In addition to *P. giblindavisi*, *Pristionchus* (Rödelisperger 2021), *C. elegans* (*C. elegans* Sequencing Consortium 1998), and *Caenorhabditis briggsae* (Stein et al. 2003) protein sequences, we included data from *Haemonchus contortus* (Doyle et al. 2020), *Ancylostoma ceylanicum* (Schwarz et al. 2015), *Ascaris suum* (Jex et al. 2011), *B. malayi* (Foster et al. 2020), *B. xylophilus* (Kikuchi et al. 2011), *Panagrellus redivivus* (Srinivasan et al. 2013), and *Plectus sambesii* (Beltran et al. 2019) to increase taxon sampling for better phylogenetic resolution and as outgroup species. Protein sequences were aligned by the program `MUSCLE` (version 3.8.31; Edgar 2004) and maximum likelihood trees were reconstructed by the `phangorn` R package (version 3.4.4, LG substitution model with optimization of base frequencies and invariant sites, 100 bootstrap pseudoreplicates; Schliep 2011; R Core Team 2020). The phylogenetic trees were visualized with the `Dendroscope` software (version 3.6.3; Huson and Scornavacca 2012). Schematic species trees were defined based on

phylogenomic and phylotranscriptomic analysis (Rödelsperger et al. 2018; Smythe et al. 2019). For syntenic analysis of sulfatases and NAGLU genes, we first identified candidates based on phylogenetic analysis. Subsequently, we tested for exact collinearity by inspecting their genomic neighborhood in the genome browser.

Results and discussion

Single-molecule sequencing drastically increases assembly contiguity

We sequenced an inbred strain of *P. giblindavisi* RS5555 on the Pacific Bioscience platform (see Methods). This resulted in 15 Gb of circular consensus sequences (mean length: 13 kb) which would translate into around 60X coverage of the *P. giblindavisi* genome. These data were assembled into an initial draft assembly using the FALCON-Unzip software (Chin et al. 2016). This assembly spanned 340 Mb that were distributed across 1851 contigs with an N50 value of 600 kb. The BUSCO completeness was 86% (complete single copy and duplicated), but showed a high level of duplicates (11%), which pointed toward assembly problems with remaining heterozygosity (Barrière et al. 2009). We applied the HaploMerger2 software combined with repeat masking in order to build the haploid genome (Huang et al. 2017). This reduced the BUSCO percentage of duplicates to 3%. The final assembly spans 256 Mb, which is 55 Mb larger than the previous version of the *P. giblindavisi* genome (Prabh et al. 2018). At the same time, the number of contigs was reduced from 7,303 to 735 and the N50 increased from 112 to 791 kb (Fig. 1a). To further compare the quality of both versions, we realigned the previously generated short-read data (Prabh et al. 2018) against both assemblies and quantified the percentage of mapped reads, pairs that were mapped in the correct orientation, and the amount of ambiguous base calls (Fig. 1a). Although these data were used to generate

the initial version of the *P. giblindavisi* genome, the newly generated long-read assembly showed a much better representation of these short reads with higher fractions of mapped and correctly paired reads as well as lower fractions of ambiguous positions (Fig. 1a). These results suggest that the larger size of the new assembly is due to an overcompression of repeats in the previous assembly leading to incorrectly paired reads and higher fraction of ambiguous base calls. An assessment of genome completeness based on the BUSCO approach revealed an increase of completeness level from 79% to 86% (complete single copy and duplicated) for the raw genome assembly (Fig. 1b). Evidence-based gene annotations were generated by the PPCAC software that integrated transcriptome data from *P. giblindavisi* and protein homology with the community-curated gene annotations of *P. pacificus* (see Methods). This resulted in 22,488 gene models with a BUSCO completeness level of 86%, which also represents an improvement by 13% relative to the gene annotations of the previous assembly (Fig. 1c). Thus, the newly generated long-read assembly is superior to the old assembly in terms of contiguity, correctness, and completeness.

Pristionchus genomes show vast expansions of gene families associated with environmental responses

We made use of the improved *P. giblindavisi* genome to screen for major changes in large gene families. To this end, we compared the numbers of genes with a given protein domain in *P. giblindavisi* with the 8 available *Pristionchus* genomes (Bateman et al. 1999; Prabh et al. 2018). Specifically, we contrasted these protein domain counts in *P. giblindavisi* with the maximum count among the 8 *Pristionchus* genomes. This provides a simple method to screen for protein families that are larger in *P. giblindavisi* than in any *Pristionchus* species. However, this analysis did not reveal any

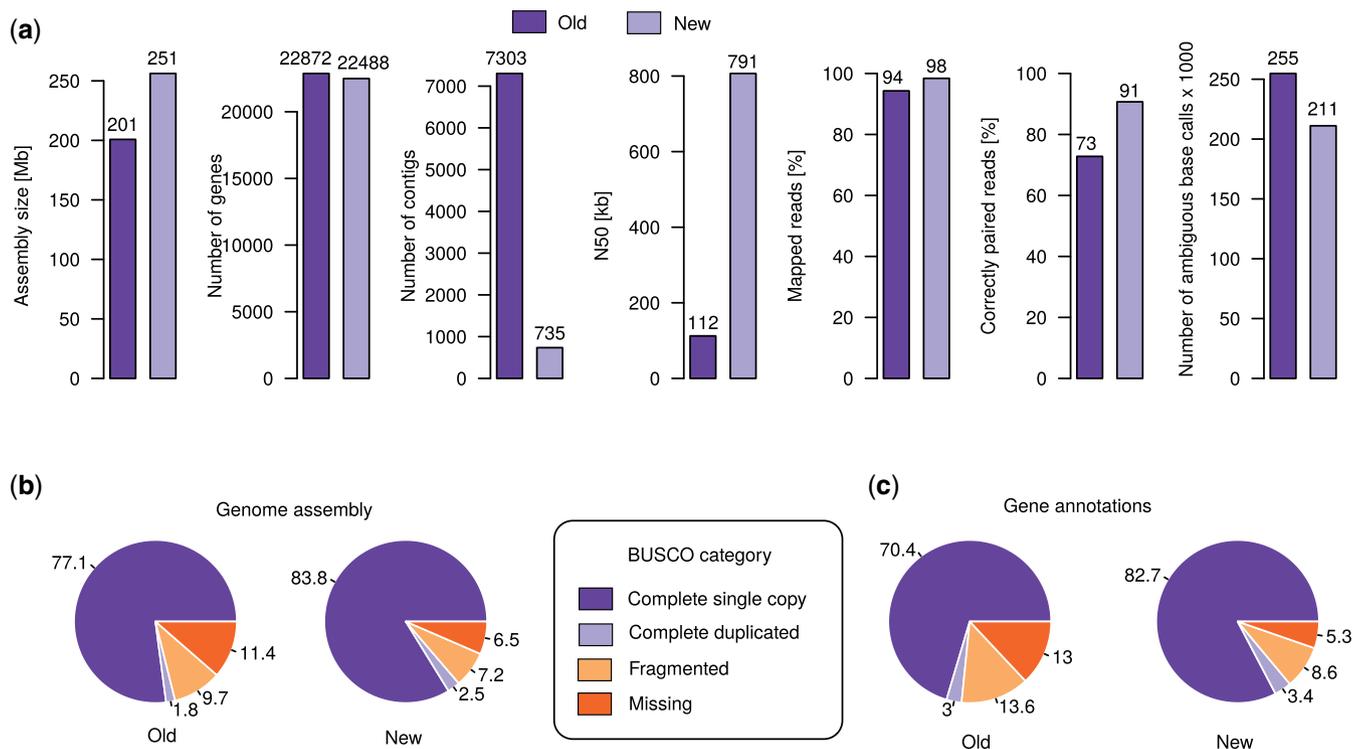


Fig. 1. Comparison of *P. giblindavisi* genome assemblies. a) The barplots show multiple features of the old and new *P. giblindavisi* genome. b) The piecharts show the evaluation of raw genome assemblies based on the BUSCO software. c) The pie charts show the BUSCO values for the protein-coding gene annotations.

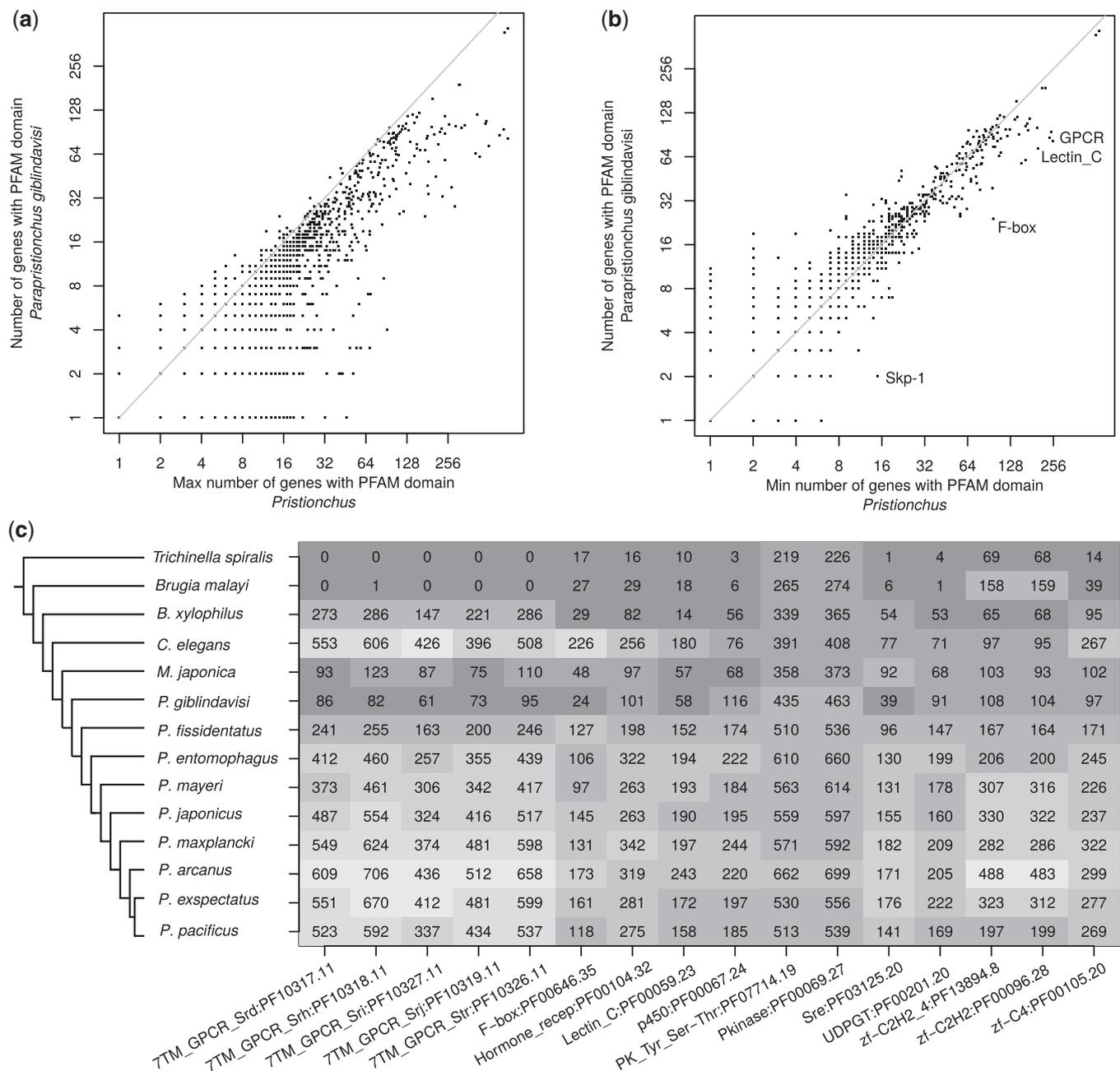


Fig. 2. Lineage-specific expansion of large gene families. a) The scatter plot shows the number of genes with a given protein domain in the *P. giblindavisi* genome against the maximum number in any *Pristionchus* genome. The general trend toward larger gene families in *Pristionchus* is expected since a single value for a gene family in *P. giblindavisi* is compared to the maximum value from 8 *Pristionchus* species. However, the absence of strong outliers in the upper left half suggests that there are no large *Parapristionchus*-specific expansions (b) The number of *P. giblindavisi* genes with a given protein domain is plotted against the minimum number in any *Pristionchus* genome. Outliers in the lower right half indicate gene families with higher numbers in all *Pristionchus* genomes when compared to *P. giblindavisi*. c) The heatmap shows the number of genes with a given Pfam domain for 14 nematode genomes. For all shown gene families, the domain count in any *Pristionchus* genome exceeds the domain count in the *P. giblindavisi* genome by at least 50.

large gene family expansion in *P. giblindavisi* (Fig. 2a). Hereby, we would arbitrarily define a large expansion as an increase of gene family size by at least 50 genes. A complementary analysis of orthogroups further supports the absence of large-scale expansions in *P. giblindavisi* (Supplementary Fig. 1). However, we could identify several candidates for smaller expansions such as the orthologs of the *C. elegans* glycosyltransferase *gly-7* which has one copy in most *Pristionchus* genomes but 16 copies in *P. giblindavisi* (Supplementary Fig. 1). In addition, we carried out a second screen comparing the number of protein domain counts in *P. giblindavisi* with the

minimum number of the *Pristionchus* genomes. This allows us to identify candidate gene families with consistently higher domain counts in all *Pristionchus* genomes when compared to *P. giblindavisi*. Specifically, we find several candidates for large *Pristionchus*-specific expansions which include F-box genes, C-type lectins, and G-protein coupled receptors (GPCRs; Fig. 2b). To further investigate these patterns, we selected *Pristionchus* gene families with a gene family size difference of 50 or more with regard to *P. giblindavisi* and visualized the domain counts in the 10 diplogastrid genomes using the genomes of more divergent *Micoletzkyia japonica* (Prabh et al. 2018),

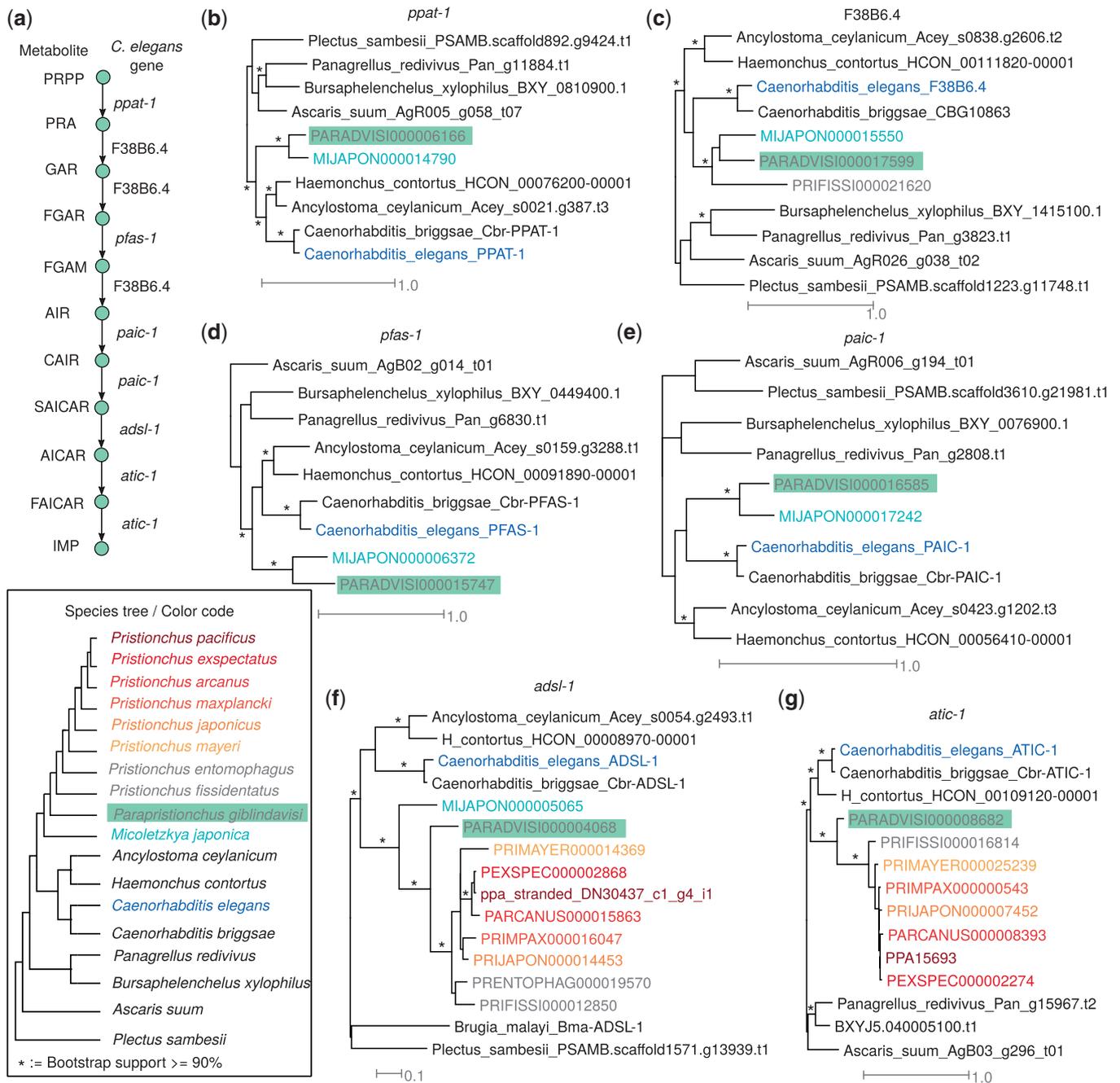


Fig. 3. *Pristionchus*-specific loss of de novo purine biosynthesis. a) The schematic pathway shows the biosynthesis of Inosine monophosphate (IMP) from Phosphoribosyl 1-pyrophosphate (PRPP) according to the KEGG database (Ogata et al. 1999). Four out of 6 genes in this pathway are missing in most *Pristionchus* species. However, all these genes have orthologs in *P. goblindavisi*. This is shown by phylogenetic analysis *ppat-1* (b), *F38B6.4* (c), *pfas-1* (d), *paic-1* (e), *adsl-1* (f), and *atic-1*.

C. elegans (*C. elegans* Sequencing Consortium 1998), *B. xylophilus* (Kikuchi et al. 2011), *B. malayi* (Foster et al. 2020), and *T. spiralis* (Mitreva et al. 2011) as outgroups (Fig. 2c). In addition to the above-mentioned gene families, this identified *Pristionchus*-specific expansions of nuclear hormone receptors, kinases, and zinc finger proteins. Many of these gene families integrate and respond to environmental signals. For example, C-type lectins, Cytochromes P450, and UDP-glucuronosyltransferase are thought to play key roles in detoxification of xenobiotics (Dieterich et al. 2008). Similarly, C-type lectins and F-box genes have been associated with immune response (Thomas 2006; Pees et al. 2016). Finally, GPCR and hormone receptors show enriched expression in neurons (Sural and Hobert

2021). These results point toward potential differences in the microenvironment of *Pristionchus* and *Parapristionchus* (e.g. exposure to different bacteria). However, comparative sampling efforts would be needed to experimentally test this hypothesis.

The de novo purine biosynthesis pathway has been lost specifically in *Pristionchus*

In order to screen for gene families that were specifically lost in the *Pristionchus* genus, we searched for protein domains that are missing in any *Pristionchus* species but are still present in *P. goblindavisi*. This identified 15 protein domains out of which 4 domains are involved in purine metabolism (Supplementary Fig. 2).

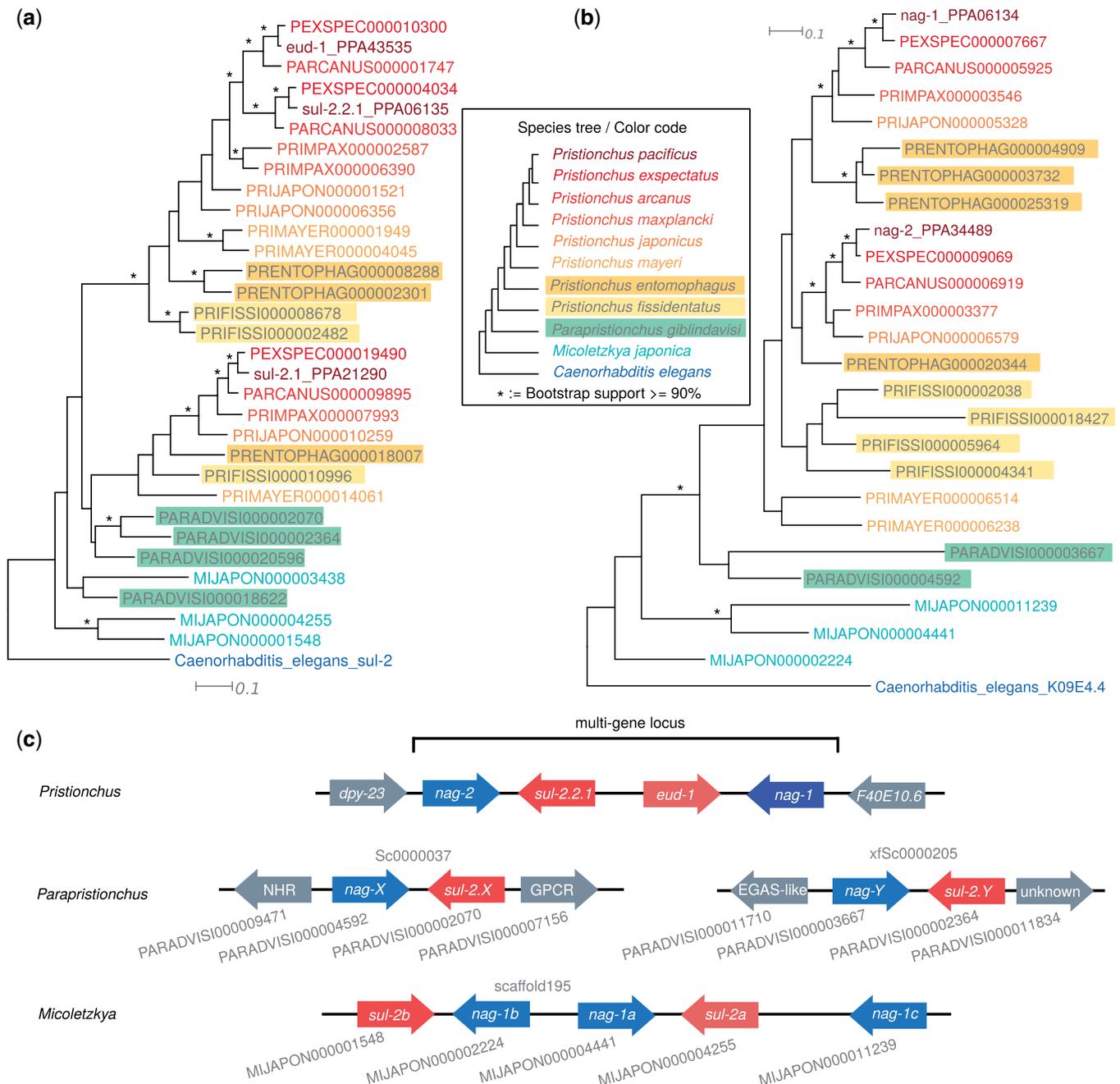


Fig. 4. Evolution of the multigene locus controlling the mouth form plasticity. a, b) Two sulfatases and 2 *nag* genes form a multigene locus in *P. pacificus* which controls its mouth form plasticity (Sieriebriennikov et al. 2018). The clustering of paralogous genes in the phylogenies of sulfatases (a) and *nag* genes (b) is consistent with homogenization of paralogs by gene conversion (Sieriebriennikov et al. 2018). c) Syntenic analysis of candidate orthologs implies that the specific arrangement of the multigene locus in *P. pacificus* is not conserved in genomes outside the genus *Pristionchus*.

Specifically, the Pfam domains Phosphoribosylglycinamide synthetase (PF01071), AIR carboxylases (PF00731), Mur ligase (PF02875), and SAICAR synthetase (PF01259) are all associated with purine metabolism (Bateman et al. 1999) and are missing from all *Pristionchus* species, but are found in *P. giblindavisi*, *M. japonica*, and *C. elegans*. Previous comparative genomic analyses proposed that *P. pacificus* and also other nematodes have lost the ability to synthesize purines de novo but that the loss might be compensated by the salvage pathway that synthesizes purines from other metabolites (Desjardins et al. 2013). Indeed, while the first 4 enzymes of this pathway are specific to the de novo

synthesis and are missing in *P. pacificus*, the last 2 enzymes encoded by the *C. elegans* genes *ads1-1* and *atic-1* also play a role in the salvage pathway and are present in *P. pacificus* (Desjardins et al. 2013). Our phylogenetic analysis confirms previous findings about the absence of these genes in *P. pacificus*, but adds novel data on the exact time point of this loss assigning it to the branch leading to the *Pristionchus* genus. As orthologs for all 6 enzymes could be detected in other diplogastrids such as *M. japonica* and *P. giblindavisi* the loss is specific to the genus *Pristionchus* (Fig. 3). This example demonstrates the utility of the *P. giblindavisi* genome to date lineage-specific gene losses.

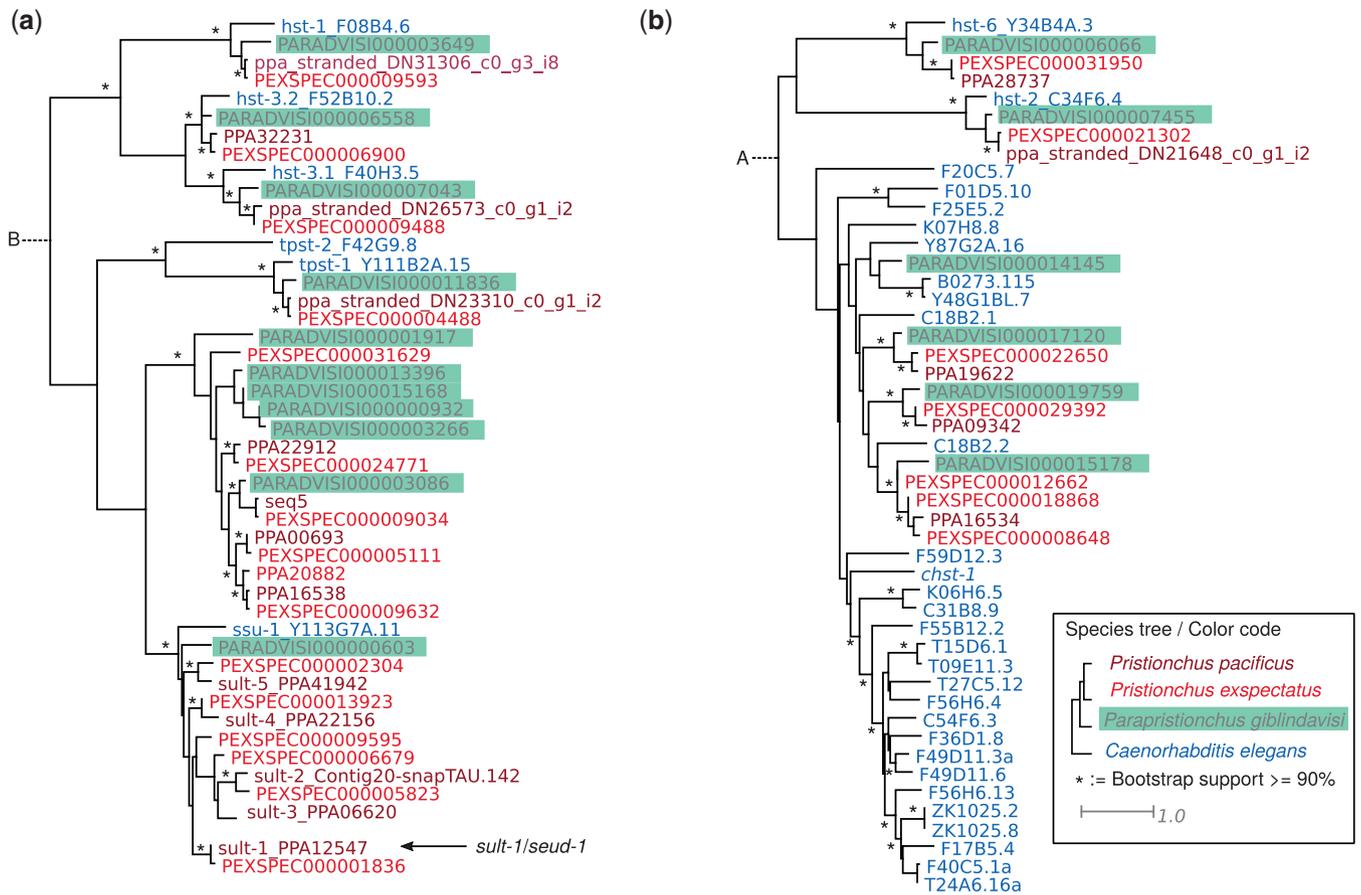


Fig. 5. Lineage-specific expansions of sulfotransferases. a) The phylogeny shows a maximum-likelihood tree of sulfotransferases type 1 (PF00685, e.g. *ssu-1*) and 3 (PF13469, e.g. *tpst-1*). The tree is rooted with the branch leading to type 2 sulfotransferases (PF03567, e.g. *chst-1*) which are shown in (b). While cytosolic sulfotransferases (orthologs of *ssu-1*) were expanded in *Pristionchus*, type 2 sulfotransferases were expanded in *Caenorhabditis*. The *P. pacificus* gene *sult-1* that has a well-characterized role in mouth form plasticity (Bui et al. 2018; Namdeo et al. 2018) likely arose from a *Pristionchus*-specific gene expansion and has no 1-1 ortholog in *P. giblindavisi*. This is in contrast to other highly conserved sulfotransferases such as *hst-1*, *hst-3.2*, *hst-3.1*, and *tpst-1*.

The multigene locus controlling mouth form plasticity in *P. pacificus* is not conserved outside the genus *Pristionchus*

Nematodes of the diplogastrid family have evolved teeth-like structures that allow them to predate and kill other nematodes including *C. elegans* (Susoy et al. 2015). Simultaneously, the development of this morphological innovation had been embedded into an ancestral regulatory network controlling the plastic responses to environmental cues (Casasa et al. 2021). This resulted in the manifestation of 2 distinct morphs, eury stomatous and stenostomatous, of which only the eury stomatous form is capable of killing other nematodes in *P. pacificus* (Wilecki et al. 2015). Genetic screens in *P. pacificus* have identified multiple components of this network that are centered around the sulfatase *eud-1* (Ragsdale et al. 2013; Kieninger et al. 2016; Sieriebriennikov et al. 2020). This includes also the sulfotransferase *sult-1/seud-1* (Namdeo et al. 2018; Bui et al. 2018), which further supports that this developmental switch is regulated by sulfation processes (Igreja and Sommer 2022). The switch gene *eud-1* is part of a multigene locus that was formed by a tandem duplication of a sulfatase and a neighboring N-acetyl-alpha-glucosaminidase (NAGLU) gene (Sieriebriennikov et al. 2018). All existing *Pristionchus* genomes support the exact conservation of gene order of this multigene locus and phylogenetic and experimental evidence support a homogenization between paralogous gene pairs by

means of gene conversion (Sieriebriennikov et al. 2018). While the locus had been rearranged in the genome of *M. japonica*, the genomic configuration of this locus in *P. giblindavisi* is not known. We performed phylogenetic analyses of sulfatases and NAGLUs to screen for potential orthologs in *P. giblindavisi* (Fig. 4, a and b) and tested whether these candidates are found in a syntenic region in the *P. giblindavisi* genome (Fig. 4c). This analysis revealed that the locus was split into 2 parts that reside on different contigs of *P. giblindavisi* assembly with several flanking genes at either side. These flanking genes are nonhomologous between the 2 *P. giblindavisi* loci (Fig. 4c), which supports that these 2 contigs are real and not a result of allelism (Barrière et al. 2009). In summary, these results suggest that the multigene locus was formed in the diplogastrid family, but was only preserved in the *Pristionchus* genus.

Pristionchus and *Caenorhabditis* exhibit lineage-specific expansions in different types of sulfotransferases

The sulfatase *eud-1* and sulfotransferase *sult-1* are taxonomically restricted duplicates, each of which has a single ortholog in *C. elegans* (Ragsdale et al. 2013; Namdeo et al. 2018). To date the timing of the lineage-specific expansion of *sult-1*-related sulfotransferases, we performed a phylogenetic analysis of sulfotransferases (Fig. 5, a and b). This included sulfotransferases of type

1 (PF00685, e.g. the cytosolic sulfotransferase *ssu-1*), type 2 (PF03567, e.g. the carbohydrate cytosolic sulfotransferase *chst-1*), and type 3 (PF13469, e.g. the tyrosylprotein sulfotransferase *tpst-1*) (Bateman et al. 1999; Igreja and Sommer 2022). In total, we found 18 sulfotransferases in *P. pacificus*. Some of these genes such as *tpst-1*, *hst-1*, *hst-2*, *hst-3.1*, *hst-3.2*, and *hst-6* have been preserved as 1-1 orthologs between all analyzed nematodes. However, different clades of sulfotransferases were expanded in the *Caenorhabditis* and *Pristionchus* lineage, respectively. Specifically, while cytosolic sulfotransferases (orthologs of *ssu-1*) are expanded in the *Pristionchus* genus (Fig. 5a), carbohydrate sulfotransferases are expanded in *C. elegans* (Fig. 5b). Notably, *sult-1* has only a single ortholog in *P. giblandavisi*, which implies that the duplication events giving rise to *sult-1* paralogs in *P. pacificus* occurred specifically within the *Pristionchus* lineage. Further phylogenetic analysis shows that all *Pristionchus* species have several *sult-1* orthologs (Supplementary Fig. 3). Thus, *sult-1* likely originated from a *Pristionchus*-specific sulfotransferase expansion.

Conclusions

Forward genetic screens in *P. pacificus* have led to the discovery of novel genes that control development and behavior (Mayer et al. 2015; Lightfoot et al. 2019). Investigating the evolutionary history of these novel genes allows us to infer hypotheses about the evolution of associated traits. In this context, the phylogenomic framework of *Pristionchus* and other diplogastrid nematodes has been particularly helpful to date the origin of novel genes or other genomic features (Prabh et al. 2018; Casasa et al. 2021). The genome of *P. giblandavisi* has a key role in these types of analyses as it allows to answer whether genomic features are restricted to the genus *Pristionchus* or are shared with other diplogastrid nematodes. In this study, we made use of single-molecule sequencing data to generate an updated version of the *P. giblandavisi* assembly. Relative to the previously published genome (Prabh et al. 2018), our current assembly exhibits higher contiguity, correctness, and completeness. Furthermore, we have demonstrated multiple analyses that show how these data could be used to investigate lineage-specific patterns in *Pristionchus* nematodes. Thus, the improved *P. giblandavisi* genome lays the basis for future phylogenetic studies.

Data availability

The genome assembly and sequencing reads have been uploaded to the European Nucleotide archive under the study accession PRJEB53331. The *P. giblandavisi* data (version 2) is also available at <http://www.pristionchus.org> and was submitted to WormBase ParaSite.

Supplemental material is available at G3 online.

Acknowledgments

The authors would like to thank the whole Sommer lab for helpful discussions.

Funding

This work was funded by the Max Planck Society.

Conflicts of Interest

None declared.

Literature cited

- Athanasouli M, Witte H, Weiler C, Loschko T, Eberhardt G, Sommer RJ, Rödelsperger C. Comparative genomics and community curation further improve gene annotations in the nematode *Pristionchus pacificus*. *BMC Genomics*. 2020;21(1):708. doi:10.1186/s12864-020-07100-0.
- Barrière A, Yang S-P, Pekarek E, Thomas CG, Haag ES, Ruvinsky I. Detecting heterozygosity in shotgun genome assemblies: lessons from obligately outcrossing nematodes. *Genome Res*. 2009;19(3):470–480. doi:10.1101/gr.081851.108.
- Bateman A, Birney E, Durbin R, Eddy SR, Finn RD, Sonnhammer EL. Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res*. 1999;27(1):260–262. doi:10.1093/nar/27.1.260.
- Beltran T, Barroso C, Birkle TY, Stevens L, Schwartz HT, Sternberg PW, Fradin H, Gunsalus K, Piano F, Sharma G, et al. Comparative epigenomics reveals that RNA polymerase II pausing and chromatin domain organization control nematode piRNA biogenesis. *Dev Cell*. 2019;48(6):793–810.e6. doi:10.1016/j.devcel.2018.12.026.
- Bui LT, Ivers NA, Ragsdale EJ. A sulfotransferase dosage-dependently regulates mouthpart polyphenism in the nematode *Pristionchus pacificus*. *Nat Commun*. 2018;9(1):4835. doi:10.1038/s41467-018-07354-z.
- C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*. 1998;282(5396):2012–2018. doi:10.1126/science.282.5396.2012.
- Casasa S, Biddle JF, Koutsovoulos GD, Ragsdale EJ. Polyphenism of a novel trait integrated rapidly evolving genes into ancestrally plastic networks. *Mol Biol Evol*. 2021;38(2):331–343. doi:10.1093/molbev/msaa235.
- Chin C-S, Peluso P, Sedlazeck FJ, Nattestad M, Concepcion GT, Clum A, Dunn C, O'Malley R, Figueroa-Balderas R, Morales-Cruz A, et al. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat Methods*. 2016;13(12):1050–1054. doi:10.1038/nmeth.4035.
- Desjardins CA, Cerqueira GC, Goldberg JM, Dunning Hotopp JC, Haas BJ, Zucker J, Ribeiro JMC, Saif S, Levin JZ, Fan L, et al. Genomics of *Loa loa*, a Wolbachia-free filarial parasite of humans. *Nat Genet*. 2013;45(5):495–500. doi:10.1038/ng.2585.
- Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, Fulton L, Fulton R, Godfrey J, Minx P, et al. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat Genet*. 2008;40(10):1193–1198. doi:10.1038/ng.227.
- Doyle SR, Tracey A, Laing R, Holroyd N, Bartley D, Bazant W, Beasley H, Beech R, Britton C, Brooks K, et al. Genomic and transcriptomic variation defines the chromosome-scale assembly of *Haemonchus contortus*, a model gastrointestinal worm. *Commun Biol*. 2020;3(1):656. doi:10.1038/s42003-020-01377-3.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32(5):1792–1797. doi:10.1093/nar/gkh340.
- Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20(1):238. doi:10.1186/s13059-019-1832-y.
- Foster JM, Grote A, Mattick J, Tracey A, Tsai Y-C, Chung M, Cotton JA, Clark TA, Geber A, Holroyd N, et al. Sex chromosome evolution in parasitic nematodes of humans. *Nat Commun*. 2020;11(1):1964. doi:10.1038/s41467-020-15654-6.
- Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A*. 2011;108(4):1513–1518. doi:10.1073/pnas.1017351108.

- Harry CJ, Messar SM, Ragsdale EJ. Comparative reconstruction of the predatory feeding structures of the polyphenic nematode *Pristionchus pacificus*. *Evol Dev.* 2022;24(1–2):16–36. doi:10.1111/ede.12397.
- Howard RJ, Giacomelli M, Lozano-Fernandez J, Edgcombe GD, Fleming JF, Kristensen RM, Ma X, Olesen J, Sørensen MV, Thomsen PF, et al. The Ediacaran origin of Ecdysozoa: integrating fossil and phylogenomic data. *J Geol Soc.* 2022;179(4). doi:10.1144/jgs2021-107.
- Howe KL, Bolt BJ, Shafie M, Kersey P, Berriman M. WormBase ParaSite—a comprehensive resource for helminth genomics. *Mol Biochem Parasitol.* 2017;215:2–10. doi:10.1016/j.molbiopara.2016.11.005.
- Huang S, Kang M, Xu A. HaploMerger2: rebuilding both haploid sub-assemblies from high-heterozygosity diploid genome assembly. *Bioinformatics.* 2017;33(16):2577–2579. doi:10.1093/bioinformatics/btx220.
- Huson DH, Scornavacca C. Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. *Syst Biol.* 2012;61(6):1061–1067. doi:10.1093/sysbio/sys062.
- Igreja C, Sommer RJ. The role of sulfation in nematode development and phenotypic plasticity. *Front Mol Biosci.* 2022;9:838148. doi:10.3389/fmolb.2022.838148.
- Jex AR, Liu S, Li B, Young ND, Hall RS, Li Y, Yang L, Zeng N, Xu X, Xiong Z, et al. *Ascaris suum* draft genome. *Nature.* 2011;479(7374):529–533. doi:10.1038/nature10553.
- Kanzaki N, Herrmann M, Weiler C, Röseler W, Theska T, Berger J, Rödelsperger C, Sommer RJ. Nine new *Pristionchus* (Nematoda: Diplogastridae) species from China. *Zootaxa.* 2021;4943(1):zootaxa.4943.1.1. doi:10.11646/zootaxa.4943.1.1.
- Kanzaki N, Ragsdale EJ, Herrmann M, Mayer WE, Tanaka R, Sommer RJ. *Parapristionchus giblandavisi* n. gen., n. sp. (Rhabditida: Diplogastridae) isolated from stag beetles (Coleoptera: Lucanidae) in Japan. *Nematology.* 2012;14(8):933–947. doi:10.1163/156854112X635878.
- Kieninger MR, Ivers NA, Rödelsperger C, Markov GV, Sommer RJ, Ragsdale EJ. The nuclear hormone receptor NHR-40 acts downstream of the sulfatase EUD-1 as part of a developmental plasticity switch in *Pristionchus*. *Curr Biol.* 2016;26(16):2174–2179. doi:10.1016/j.cub.2016.06.018.
- Kikuchi T, Cotton JA, Dalzell JJ, Hasegawa K, Kanzaki N, McVeigh P, Takanashi T, Tsai IJ, Assefa SA, Cock PJA, et al. Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathog.* 2011;7(9):e1002219. doi:10.1371/journal.ppat.1002219.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25(14):1754–1760. doi:10.1093/bioinformatics/btp324.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352.
- Lightfoot JW, Wilecki M, Rödelsperger C, Moreno E, Susoy V, Witte H, Sommer RJ. Small peptide-mediated self-recognition prevents cannibalism in predatory nematodes. *Science.* 2019;364(6435):86–89. doi:10.1126/science.aav9856.
- Markov GV, Meyer JM, Panda O, Artyukhin AB, Claaßen M, Witte H, Schroeder FC, Sommer RJ. Functional conservation and divergence of daf-22 paralogs in *Pristionchus pacificus* dauer development. *Mol Biol Evol.* 2016;33(10):2506–2514. doi:10.1093/molbev/msw090.
- Mayer MG, Rödelsperger C, Witte H, Riebesell M, Sommer RJ. The orphan gene dauerless regulates dauer development and intraspecific competition in nematodes by copy number variation. *PLoS Genet.* 2015;11(6):e1005146. doi:10.1371/journal.pgen.1005146.
- Mitrevva M, Jasmer DP, Zarlenga DS, Wang Z, Abubucker S, Martin J, Taylor CM, Yin Y, Fulton L, Minx P, et al. The draft genome of the parasitic nematode *Trichinella spiralis*. *Nat Genet.* 2011;43(3):228–235. doi:10.1038/ng.769.
- Moreno E, McGaughan A, Rödelsperger C, Zimmer M, Sommer RJ. Oxygen-induced social behaviours in *Pristionchus pacificus* have a distinct evolutionary history and genetic regulation from *Caenorhabditis elegans*. *Proc Biol Sci.* 2016;283(1825):20152263. doi:10.1098/rspb.2015.2263.
- Namdeo S, Moreno E, Rödelsperger C, Baskaran P, Witte H, Sommer RJ. Two independent sulfation processes regulate mouth-form plasticity in the nematode. 2018;145(13):dev166272. doi:10.1242/dev.166272.
- Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 1999;27(1):29–34. doi:10.1093/nar/27.1.29.
- Pees B, Yang W, Zárate-Potes A, Schulenburg H, Dierking K. High innate immune specificity through diversified C-type lectin-like domain proteins in invertebrates. *J Innate Immun.* 2016;8(2):129–142. doi:10.1159/000441475.
- Prabh N, Roeseler W, Witte H, Eberhardt G, Sommer RJ, Rödelsperger C. Deep taxon sampling reveals the evolutionary dynamics of novel gene families in nematodes. *Genome Res.* 2018;28(11):1664–1674. doi:10.1101/gr.234971.118.
- Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics.* 2005;21(Suppl 1):i351–i358. doi:10.1093/bioinformatics/bti1018.
- R Core Team. R: A Language and Environment for Statistical Computing. Vienna; 2020. [accessed 2020 Jun 1]. <https://www.R-project.org/>.
- Ragsdale EJ, Müller MR, Rödelsperger C, Sommer RJ. A developmental switch coupled to the evolution of plasticity acts through a sulfatase. *Cell.* 2013;155(4):922–933. doi:10.1016/j.cell.2013.09.054.
- Rödelsperger C. The community-curated *Pristionchus pacificus* genome facilitates automated gene annotation improvement in related nematodes. *BMC Genomics.* 2021;22(1):216. doi:10.1186/s12864-021-07529-x.
- Rödelsperger C, Prabh N, Sommer RJ. New gene origin and deep taxon phylogenomics: opportunities and challenges. *Trends Genet.* 2019;35(12):914–922. doi:10.1016/j.tig.2019.08.007.
- Rödelsperger C, Röseler W, Prabh N, Yoshida K, Weiler C, Herrmann M, Sommer RJ. Phylotranscriptomics of *Pristionchus* nematodes reveals parallel gene loss in six hermaphroditic lineages. *Curr Biol.* 2018;28(19):3123–3127.e5. doi:10.1016/j.cub.2018.07.041.
- Rudel D, Riebesell M, Sommer RJ. Gonadogenesis in *Pristionchus pacificus* and organ evolution: development, adult morphology and cell-cell interactions in the hermaphrodite gonad. *Dev Biol.* 2005;277(1):200–221. doi:10.1016/j.ydbio.2004.09.021.
- Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics.* 2011;27(4):592–593. doi:10.1093/bioinformatics/btq706.
- Schwarz EM, Hu Y, Antoshechkin I, Miller MM, Sternberg PW, Aroian RV. The genome and transcriptome of the zoonotic hookworm *Ancylostoma ceylanicum* identify infection-specific gene families. *Nat Genet.* 2015;47(4):416–422. doi:10.1038/ng.3237.
- Sieriebriennikov B, Prabh N, Dardiry M, Witte H, Röseler W, Kieninger MR, Rödelsperger C, Sommer RJ. A developmental switch generating phenotypic plasticity is part of a conserved multi-gene locus. *Cell Rep.* 2018;23(10):2835–2843.e4. doi:10.1016/j.celrep.2018.05.008.

- Sieriebriennikov B, Sun S, Lightfoot JW, Witte H, Moreno E, Rödelsperger C, Sommer RJ. Conserved nuclear hormone receptors controlling a novel plastic trait target fast-evolving genes expressed in a single cell. *PLoS Genet.* 2020;16(4):e1008687. doi:[10.1371/journal.pgen.1008687](https://doi.org/10.1371/journal.pgen.1008687).
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics.* 2015;31(19):3210–3212. doi:[10.1093/bioinformatics/btv351](https://doi.org/10.1093/bioinformatics/btv351).
- Slater GSC, Birney E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics.* 2005;6:31. doi:[10.1186/1471-2105-6-31](https://doi.org/10.1186/1471-2105-6-31).
- Smythe AB, Holovachov O, Kocot KM. Improved phylogenomic sampling of free-living nematodes enhances resolution of higher-level nematode phylogeny. *BMC Evol Biol.* 2019;19(1):121.
- Sommer RJ. *Pristionchus pacificus*. *WormBook.* 2006;1–8. doi:[10.1895/wormbook.1.102.1](https://doi.org/10.1895/wormbook.1.102.1).
- Sommer RJ, Dardiry M, Lenuzzi M, Namdeo S, Renahan T, Sieriebriennikov B, Werner MS. The genetics of phenotypic plasticity in nematode feeding structures. *Open Biol.* 2017;7(3):160332. doi:[10.1098/rsob.160332](https://doi.org/10.1098/rsob.160332).
- Srinivasan J, Dillman AR, Macchietto MG, Heikkinen L, Lakso M, Fracchia KM, Antoshechkin I, Mortazavi A, Wong G, Sternberg PW, et al. The draft genome and transcriptome of *Panagrellus redivivus* are shaped by the harsh demands of a free-living lifestyle. *Genetics.* 2013;193(4):1279–1295. doi:[10.1534/genetics.112.148809](https://doi.org/10.1534/genetics.112.148809).
- Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, et al. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. *PLoS Biol.* 2003;1(2):E45. doi:[10.1371/journal.pbio.0000045](https://doi.org/10.1371/journal.pbio.0000045).
- Sun S, Rödelsperger C, Sommer RJ. Single worm transcriptomics identifies a developmental core network of oscillating genes with deep conservation across nematodes. *Genome Res.* 2021;31(9):1590–1601. doi:[10.1101/gr.275303.121](https://doi.org/10.1101/gr.275303.121).
- Sural S, Hobert O. Nematode nuclear receptors as integrators of sensory information. *Curr Biol.* 2021;31(19):4361–4366.e2. doi:[10.1016/j.cub.2021.07.019](https://doi.org/10.1016/j.cub.2021.07.019).
- Susoy V, Ragsdale EJ, Kanzaki N, Sommer RJ. Rapid diversification associated with a macroevolutionary pulse of developmental plasticity. *eLife.* 2015;4:e05463. doi:[10.7554/eLife.05463](https://doi.org/10.7554/eLife.05463).
- Thomas JH. Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. *Genome Res.* 2006;16(8):1017–1030. doi:[10.1101/gr.5089806](https://doi.org/10.1101/gr.5089806).
- Urrutia A, García-Angulo VA, Fuentes A, Caneo M, Legüe M, Urquiza S, Delgado SE, Ugalde J, Burdisso P, Calixto A, et al. Bacterially produced metabolites protect *C. elegans* neurons from degeneration. *PLoS Biol.* 2020;18(3):e3000638. doi:[10.1371/journal.pbio.3000638](https://doi.org/10.1371/journal.pbio.3000638).
- Vaser R, Sović L, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* 2017;27(5):737–746. doi:[10.1101/gr.214270.116](https://doi.org/10.1101/gr.214270.116).
- Wilecki M, Lightfoot JW, Susoy V, Sommer RJ. Predatory feeding behaviour in *Pristionchus* nematodes is dependent on phenotypic plasticity and induced by serotonin. *J Exp Biol.* 2015;218(Pt 9):1306–1313. doi:[10.1242/jeb.118620](https://doi.org/10.1242/jeb.118620).

Communicating editor: M. Rockman