### Title

The Draft Genome and Transcriptome of Panagrellus redivivus Are Shaped by the Harsh Demands of a Free-Living Lifestyle




### Journal

Genetics, 193(4)

### ISSN

0016-6731



### Authors

Srinivasan, Jagan
Dillman, Adler R
Macchietto, Marissa G
et al.



### Publication Date

2013-04-01

### DOI

10.1534/genetics.112.148809

Peer reviewed

# The Draft Genome and Transcriptome of *Panagrellus redivivus* Are Shaped by the Harsh Demands of a Free-Living Lifestyle

**Jagan Srinivasan,*,†,1,2 Adler R. Dillman,*,†,2 Marissa G. Macchietto,‡,§ Liisa Heikkinen,** Merja Lakso,** Kelley M. Fracchia,‡ Igor Antoshechkin,* Ali Mortazavi,‡,§ Garry Wong,** and Paul W. Sternberg*,†,3**

*Division of Biology, California Institute of Technology, Pasadena, California 91125, †Howard Hughes Medical Institute, Pasadena, California 91125, ‡Developmental and Cell Biology and §Center for Complex Biological Systems, University of California, Irvine, California 92697, and **Department of Neurobiology, A. I. Virtanen Institute, University of Eastern Finland, Kuopio 70211, Finland

**ABSTRACT** Nematodes compose an abundant and diverse invertebrate phylum with members inhabiting nearly every ecological niche. *Panagrellus redivivus* (the "microworm") is a free-living nematode frequently used to understand the evolution of developmental and behavioral processes given its phylogenetic distance to *Caenorhabditis elegans*. Here we report the *de novo* sequencing of the genome, transcriptome, and small RNAs of *P. redivivus*. Using a combination of automated gene finders and RNA-seq data, we predict 24,249 genes and 32,676 transcripts. Small RNA analysis revealed 248 microRNA (miRNA) hairpins, of which 63 had orthologs in other species. Fourteen miRNA clusters containing 42 miRNA precursors were found. The RNA interference, dauer development, and programmed cell death pathways are largely conserved. Analysis of protein family domain abundance revealed that *P. redivivus* has experienced a striking expansion of BTB domain-containing proteins and an unprecedented expansion of the cullin scaffold family of proteins involved in multi-subunit ubiquitin ligases, suggesting proteolytic plasticity and/or tighter regulation of protein turnover. The eukaryotic release factor protein family has also been dramatically expanded and suggests an ongoing evolutionary arms race with viruses and transposons. The *P. redivivus* genome provides a resource to advance our understanding of nematode evolution and biology and to further elucidate the genomic architecture leading to free-living lineages, taking advantage of the many fascinating features of this worm revealed by comparative studies.

NEMATODES are highly prolific organisms that originated during the Precambrian or Cambrian explosion over 500 million years ago and have subsequently evolved exquisite adaptations, allowing them to inhabit nearly all ecological niches (Malakhov and Hope 1994; Ayala and Rzhetsky 1998; Blaxter *et al.* 1998; Rodriguez-Trelles *et al.* 2002). Most nematodes are adapted to "free-living" lifestyles (*i.e.*, nonparasitic and not associated with plants or animals, or only transiently associated as in phoresy) in terrestrial, freshwater, and marine environments, while others have parasitic lifestyles (Malakhov and Hope 1994). Free-living nematodes such as *Caenorhabditis elegans* have proven to be invaluable models for elucidating developmental and behavioral processes, leading to major discoveries including the genetic pathways underlying programmed cell death and the discovery of microRNA (miRNAs) and RNA interference, among others (Ambros and Horvitz 1984; Yuan *et al.* 1993; Fire *et al.* 1998). There is a huge repertoire of culturable free-living species for comparative studies, potentially making it difficult to decide which to prioritize for sequencing (Blaxter *et al.* 1998).

The free-living nematode *Panagrellus redivivus* has been used as a model system since the days of Linnaeus and is an established free-living comparative taxon to *C. elegans* (Sternberg and Horvitz 1981, 1982; Srinivasan *et al.* 2008). Fascinating differences in cell lineages and in behavior have been observed between the two (Sternberg and Horvitz


doi: 10.1534/genetics.112.148809
Manuscript received December 15, 2012; accepted for publication January 8, 2013
Available freely online through the author-supported open access option.
Supporting information is available online at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.148809/-/DC1.
Arrays have been submitted to the GEO database at NCBI as series GSE44020. Small RNA sequence read archive accession no. SRX217718.
[1]Present address: Department of Biology and Biotechnology, Worcester Polytechnic Institute, Worcester, MA 01605.
[2]These authors contributed equally to this work.
[3]Corresponding author: California Institute of Technology, 1200 E. California Blvd., MC156-29, Pasadena, CA 91125. E-mail: pws@caltech.edu

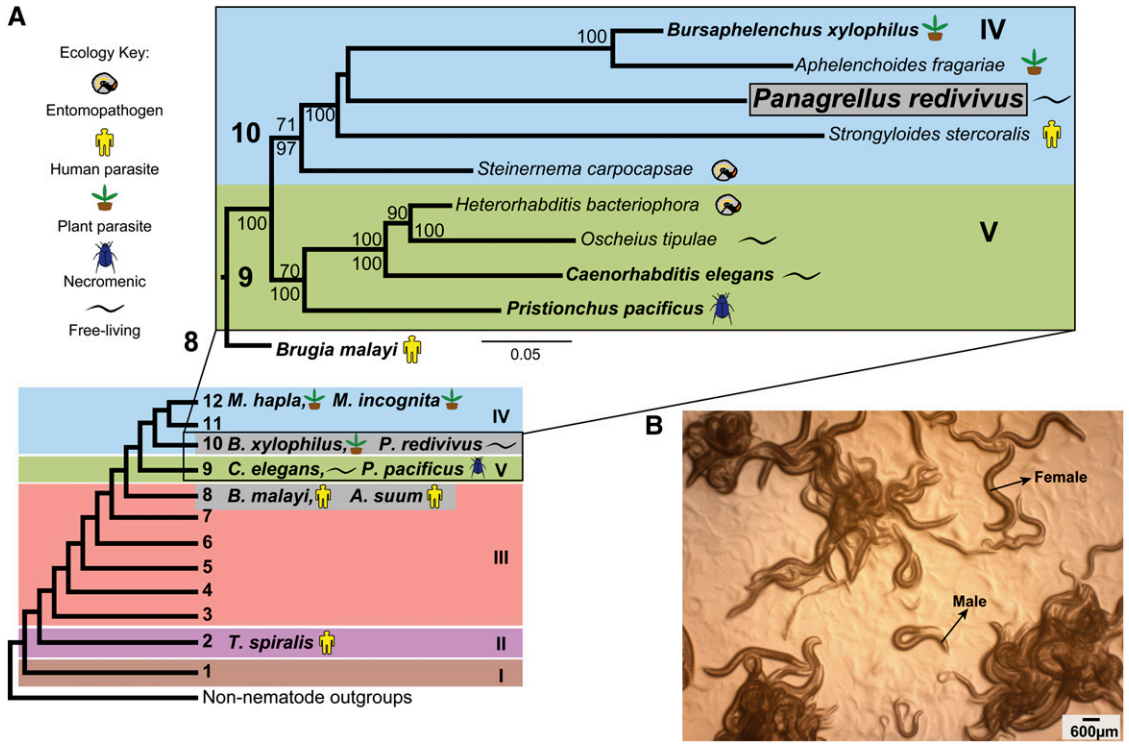


**Figure 1** Phylogenetic classification of the nematode phylum and the position of the nematode *P. redivivus*. (A) A schematic representation of the division of the phylum Nematoda into clades, with the 12-clade designation after Holterman *et al.* (2006) and the 5-clade designation after Blaxter *et al.* (1998) in Roman numerals. Blaxter clades are encompassed in colored boxes, and nematode ecologies are represented by colored icons. The diagram zooms into a maximum-likelihood (ML) tree of representative taxa in clades 9 and 10, based on small subunit ribosomal DNA sequences. ML bootstrap support values ≥70 are shown above nodes while Bayesian posterior probabilities ≥70, concordant with the ML analysis, are shown below supported nodes. The closest sequenced nematode neighbor to *P. redivivus* in clade 10 is the migratory endoparasitic nematode *B. xylophilus*. The scale bar shows the amount of nucleotide changes per site that have occurred across taxa. (B) *P. redivivus* is a gonochoristic species comprising males and females. It is larger in size than *C. elegans* and lays young ones instead of eggs.

1981, 1982; Sulston *et al.* 1983; Srinivasan *et al.* 2008). For example, *P. redivivus* exhibits several distinguishing morphological characteristics: it is a gonochoristic (male–female) species requiring both sexes for reproduction and is ovoviviparous eggs hatch *in utero* and the young larvae are subsequently released through the vulva (Figure 1B). The larvae undergo four molts post hatch. Males have 9 chromosomes while females have 10 (Hechler 1970; Sternberg and Horvitz 1981). *P. redivivus* adults average 2 mm in size, twice as long as *C. elegans* adults (Hechler 1970; Sternberg and Horvitz 1982).

Historically, free-living nematodes have served as useful models for understanding basic biology such as organ development and signal transduction (Sternberg and Horvitz 1981, 1982; Srinivasan *et al.* 2008). In addition to comparative developmental studies, *P. redivivus* has been used in aquatic and soil toxicity studies, revealing interesting insights into the effects of pollutants and toxins on reproduction, movement, and feeding (Ager *et al.* 1984; Debus and Niemann 1994; Hempel *et al.* 1995; Boyd and Williams 2003; Niu *et al.* 2010). Small metabolites isolated from several fungal species have been successfully tested for their nematocidal activity using *P. redivivus* (Li *et al.* 2005; Huang *et al.* 2009; da Cruz *et al.* 2011). *P. redivivus* has been used to isolate male and female sex pheromones (Choe *et al.* 2012). It has also been used as a model for studying infection using human bacterial pathogens (Laws *et al.* 2005). Hence, *P. redivivus* has been used as a model system extensively in many diverse fields of biology in addition to being a free-living comparative taxon with *C. elegans*, making it a standout among free-living nematode sequencing candidates.

A molecular phylogenetic approach based on small subunit ribosomal DNA suggests the presence of 12 monophyletic clades in Nematoda (Figure 1A) (Holterman *et al.* 2006; van Megan *et al*. 2009). According to this phylogeny, *P. redivivus* belongs to clade 10, whereas *C. elegans* belongs to clade 9 (Figure 1). Sequencing efforts have focused primarily on the crown clades of Chromadoria with >13 sequenced genomes. All of the sequenced free-living nematode genomes currently available are restricted to clade 9 and are within the *Caenorhabditis* genus (Dillman *et al.* 2012). Other than the caenorhabditids, nematode sequencing efforts have prioritized either plant or animal parasites—including some of the most devastating agricultural and human pathogens such as plant parasites within *Meloidogyne* and the human parasites *Brugia malayi* and *Trichinella spiralis* (Ghedin *et al.* 2007; Opperman *et al.* 2008; Mitreva *et al.* 2011), which cause elephantiasis and trichinosis,

respectively. *P. redivivus* represents the first noncaenorhabditid free-living nematode to be sequenced. Although little is known about its natural ecology, published literature suggests that *P. redivivus* has been isolated from a variety of environments, including felt beer hall mats, insect frass, slime from tree wounds, rotting fruit, insects, and wheat paste (Ferris 2009; Félix and Duveau 2012). These are acidic and nutrient-rich environments and have considerable overlap with the nutrient-rich natural habitats of *C. elegans*, which has also been isolated from rotting/decaying matter, especially rotting fruit (Kiontke and Sudhaus 2006; Félix and Duveau 2012). Given this ecological overlap, it is interesting to consider the architecture of free-living nematode genomes and how they might adapt to their respective niches. The phylogenetic position of *P. redivivus* and its ecological overlap with *C. elegans* make it an excellent species for studying the evolution of development, behavior, and adaptation (Figure 1A) (Blaxter *et al.* 1998; Holterman *et al.* 2006).

Here we describe the *de novo* assembly and characterization of a draft genome, transcriptome, and the complement of small RNAs of *P. redivivus*.

## Materials and Methods

### *Strain culturing and maintenance of P. redivivus*

For genomic and transcriptomic analysis, we used the *P. redivivus* strain PS2298/MT8872 (Sternberg and Horvitz 1981) originally obtained from D. J. Hooper (Rothamsted Experimental Station, Harpenden, Hertsfordshire, England). This strain was raised at 20° using standard methods.

### *Isolation of DNA and RNA*

*P. redivivus* worms were grown on 5–10, 10-cm nutrient agar dishes containing *Escherichia coli* OP50 plates until near starvation. The worms were rinsed and collected with M9 buffer and washed multiple times to remove any *E. coli*. After the last wash in M9, the worms were suspended in M9 for 15–30 min. The worms were then snap-frozen in liquid nitrogen in ~100-μl aliquots and stored at −80°. Worms were thawed and refrozen two to three times to break the cuticle before extracting either genomic DNA or bulk RNA. Genomic DNA was extracted using two rounds of proteinase K digestion followed by phenol-chloroform extraction. The genomic DNA was then treated with RNase A for digestion of any RNAs present in the sample. Bulk RNA was extracted using the Qiagen RNeasy mini kit.

### *Genomic and RNA-Seq library construction*

A genomic library (library ID 12193) was constructed using Illumina Paired End DNA Sample Preparation Kit according to the manufacturer's instructions. Briefly, 3 μg of genomic DNA were fragmented using nebulization. The fragments were end-repaired, 3′ adenylated, and ligated to Illumina's paired-end adaptors. The ligation products were size-selected on an agarose gel to yield fragments of ~350 bp. These fragments were then PCR-amplified to produce the finished library. Mate pair, a.k.a. "jumping" library (library ID 13185), was prepared using Illumina Mate Pair Library Preparation kit v2. Briefly, 7.5 μg of genomic DNA was fragmented using HydroShear device (Genomic Instrumentation Services) to generate fragments of ~2.2 kb. Following end repair and biotinylation, the 2.2-kb fragment was gel-purified and circularized. Circular DNA was fragmented using Bioruptor NGS (Diagenode), and biotin-containing fragments were isolated using Dynabeads (Invitrogen). The fragments were end-repaired, 3′ adenylated, and ligated to NEBNext Multiplex Adaptors (NEB). The ligation products were PCR-amplified and size-selected using AMPure XP beads (Beckman Coulter) to generate the finished library of ~450 bp in length. The RNA-Seq mixed-stage, poly(A)-selected library was created from 10 μg of total RNA using a standard unstranded protocol (Mortazavi *et al.* 2008, 2010). Libraries were quantified using a Qubit fluorometer (Invitrogen), and size distributions were verified using an Agilent Bioanalyzer and the High Sensitivity DNA Kit. Genomic and RNA-seq libraries were sequenced on Illumina Genome Analyzer IIx sequencer in paired-end mode with the read length of 76 nt. The jumping library was sequenced on Illumina HiSeq2000 in paired-end mode with the read length of 100 nt.

### *Genome assembly and annotation*

The genomic libraries were built, sequenced, assembled, filtered, and repeat-masked as previously described (Mortazavi *et al.* 2010) using Velvet 1.2.07 and RepeatModeler 1.0.5, RepeatMasker 3.0.3, recon 1.70, and RepeatScout 1.0.5. The mixed-stage transcriptome was sequenced as previously described (Mortazavi *et al.* 2010) and assembled into complementary DNAs (cDNAs) using Oases 0.2.6 (Schulz *et al.* 2012). RNA-seq reads were submitted to the Sequence Read Archive under accession no. GSM1076725. This Whole Genome Shotgun project has been deposited at DNA Data Bank of Japan/ EMBL/GenBank under accession no. AOMH00000000. The version described in this article is the first version, AOMH01000000.

Assembled cDNAs were mapped onto the genome with blat and used as hints for gene finding using Augustus 2.6 with *C. elegans* settings (Stanke *et al.* 2008). Separately, RNA-seq reads were mapped onto the genome using TopHat 1.4 (Trapnell *et al.* 2009), assembled into transcripts using Cufflinks 2.02 (Trapnell *et al.* 2010). Candidate single nucleotide variations (SNVs) in the genome and transcriptome mapped reads were called using the samtools 0.1.13 (Li *et al*. 2009) pileup and varFilter options (Supporting Information, Figure S1). Candidate SNVs in the transcriptome that fell within 5 bp of exon junctions were filtered out as likely splicing artifacts.

### *Generation of the small RNA library*

Small RNAs were isolated from mixed cultures of *P. redivivus* using the miRVana kit (Ambion) according to the manufacturer's instructions. A small RNA library was then produced from

the isolated RNAs using NEBNext small RNA sample prep Set 1 (New England Biolabs). The library was then size-selected on a 6% PAGE gel with the cut band corresponding to ~90–120 bp. Library quality and size were confirmed prior to sequencing on a Bioanalyzer (Agilent).

For additional methods involving analyses of sequence information of the genome, see Supporting Information.

## Results and Discussion

### *P. redivivus genomic assembly and transcript annotation*

We sequenced 34 million, 350-bp fragments and 52 million, 2200-bp fragments of genomic *P. redivivus* DNA using paired-end 75-bp reads and 100-bp reads, respectively, and assembled them as described in *Materials and Methods*. After filtering out *E. coli* genomic DNA, the *P. redivivus* 64.4-Mb assembly had an N50 of 262.4 kb, with a maximum scaffold size of 2318 kb. The assembly has a GC content of 44.25%, and 7.01% of the genome was repeat-masked (Table 1). The *Caenorhabditis* genomes have a lower average GC content of 37%, which makes the genome of *P. redivivus* more similar to the necromenic nematode *Pristionchus pacificus* (43% GC) with respect to GC content.

We collected RNA from ~100,000 mixed-stage worms and sequenced 35 million, 200-bp cDNA fragments using paired-end 75-bp reads that were assembled into 18,298 distinct cDNAs with an N50 of 2.0 kb (Table S1) that were mapped onto the genome to assist the Augustus gene finder. Augustus identified 26,372 transcripts in 24,249 genes with 78,945 splice junctions. To extend the Augustus predictions of protein-coding genes, we used TopHat and Cufflinks (Trapnell *et al.* 2010) to map the RNA-seq data set onto the assembled genome. Cufflinks assembled 32,676 transcripts in 24,178 genes. Augustus predicted 32.9% of the consolidated gene models, whereas 19.3% of the models came from Cufflinks only (Figure 2A). Novel splice isoforms represented the bulk (11.5%) of the new transcripts identified by Cufflinks, while novel intergenic transcripts accounted for only 1.5% (Figure 2B). A survey of expression levels revealed that the novel splice isoform and non-Augustus gene models were highly expressed (Figure 2C). Thus, *de novo* protein-coding gene prediction (Augustus) and *de novo* transcript assembly on the genome (Cufflinks) are complementary methods that can be combined to obtain a more complete annotation of genes. We estimate that this draft of the *P. redivivus* genome is ~98.2% complete, based on protein-clustering analysis of the *P. redivivus* proteome with the *C. elegans* Core Eukaryotic Genes Mapping Approach (CEGMA) protein set (Parra *et al.* 2007) (Figure 2D).

### *Analysis of small RNAs*

A small RNA library was prepared by first isolating <200 nucleotide RNAs from whole mixed-stage animals by column chromatography. 5′ RNA and 3′ DNA adaptors were added using T4 RNA ligase I and T4 RNA ligase II (New England Biolabs), respectively. Both enzymes require 5′ phosphate present in the donor molecule and 3′ hydroxyl (OH) group in the acceptor for activity. The library was amplified and size selected for ~90–120 bp fragments corresponding to inserts ~20–50 bp in size. We sequenced 24 million reads from a small RNA library generated from mixed-stage animals (Table S2). We identified 248 miRNA precursors with at least 10 reads that support the presence of a mature miRNA derived from the hairpin precursor (Table 2 and File S2). For 218 hairpins, both 5′ and 3′ mature miRNAs were present with at least one read, and for 116 hairpins both were supported with at least 10 reads. In 157 miRNA genes, the dominantly expressed mature miRNA is located in the 3′ arm of the hairpin, a phenomenon that has also been observed in other nematodes (de Wit *et al.* 2009). In a few cases, there were two miRNAs expressed from the same miRNA loci, one from the plus strand and the other from the minus strand, suggesting the existence of antisense miRNA transcription (Ruby *et al.* 2007). We considered miRNA hairpins located within 500 bp from each other to be clustered; thus we found 14 miRNA clusters each containing two to seven miRNAs (Figure S2 and Figure S3C). In total, 42 miRNAs were located in these clusters and were likely derived from multicistronic precursors. Seventeen miRNAs came from multiple loci (Table 2).

Using conservation of both mature miRNA and its hairpin sequence as criteria, we found orthologs for *P. redivivus* miRNAs in humans (46 of the 1527 miRBase miRNAs), *Drosophila* (31/240), *C. elegans* (46/223), *Caenorhabditis briggsae* (28/140), *Caenorhabditis remanei* (29/109), *P. pacificus* (20/124), *B. malayi* (20/32), and *Ascaris suum* (50/97) (Table S3). Among these were the well-studied and broadly conserved miRNAs *let-7*, *miR-1*, and *miR-124* and the first miRNA identified, *lin-4* (Lee *et al.* 1993). Altogether, 63 *P. redivivus* miRNAs have at least one ortholog among the species studied. Hierarchical clustering was used to visualize the distribution and conservation of these miRNAs, separating those highly conserved from miRNAs with only one or two orthologs (Figure 3). The most highly expressed miRNA was *prd-21808-8719-5p* (34%), for which we found no orthologs, whereas the second, *prd-miR-51-5p*, was conserved in six species (*C. elegans*, *C. remanei*, *B. malayi*, *A. suum*, *D. melanogaster*, and *Homo sapiens*). In addition, *prd5043_2650-3p* and *prd17878_7454-5p* were conserved only in *A. suum*. In all, 10 of the 20 most abundant miRNAs from *P. redivivus* had an ortholog in *C. elegans* (Figure 4). *lin-4* (4.7% of all miRNA reads) and *miR-1* (2.1%) were also among the 20 most abundantly expressed miRNAs in the data set (Figure 4).

In addition to miRNAs, we also found evidence for the presence of endogenous small interfering RNAs (siRNAs) through identification of a cluster of nonhairpin-derived small RNAs. These consisted of reads that we tentatively identify as belonging to 21U, 22G, and 26G classes. The cluster in contig Pred1187 spanning nucleotides 15–486, consisted of 132 21U RNA reads (U first nucleotide, 21

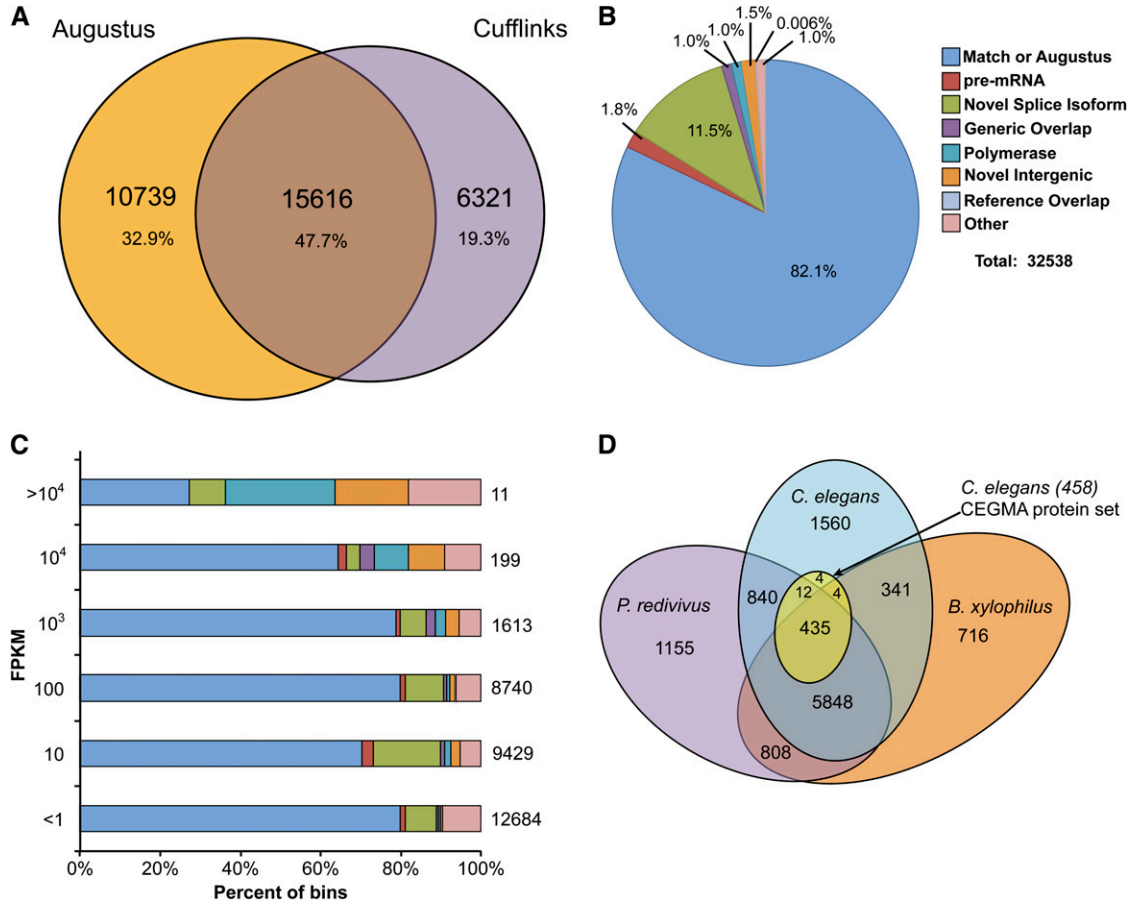


**Figure 2** Improved gene annotations using RNA sequencing. (A) Venn diagram capturing the differences between gene-finder-based annotations (Augustus) and RNA-seq-based transcript models (Cufflinks). All percentages are based on 32,676 consolidated transcript models that do not include the small categories in B. (B) Different match classes for Cufflinks + Augustus consolidated annotations to the original Augustus transcripts. Representative population size corresponds to all 30,601 models reported by cufflinks. (C) The distribution of transcripts detected in, or specific to, each fragments per kilobase of exon per million fragments mapped (FPKM) range and cumulative totals for all corresponding class annotations. (D) Venn Diagram capturing protein clusters between *P. redivivus* and the CEGMA protein set (Parra *et al.* 2007).

nucleotides in length), 94 22G RNA reads (G first nucleotide, 22 nucleotides in length), and 78 26G RNA reads (G first nucleotide, 26 nucleotides in length) complementary to a 1-kb region of the predicted gene pred1_g624, a 221-amino-acid protein with two transmembrane domains and no obvious orthologs in other species. These RNAs were spaced at varying distances in both 5′ and 3′ directions (Figure S3A). We were not able to find large clusters of 21U-RNAs, similar to those described in *C. elegans* (15,722 unique 21U-RNA species expressed over a 200-kb region) (Ruby *et al.* 2007; Batista *et al.* 2008). *C. elegans* 21U-RNA species are expressed in the germline, are bound by Argonaut subfamily *piwi*-related protein PRG-1, and are thought to be the nematode equivalent of Piwi-interacting RNAs (piRNAs) found in *Drosophila* and humans. We were also unable to identify PRG-1 or PRG-2 orthologs in the transcriptome

**Table 1 Features of the *P. redivivus* genome and transcriptome**

| Genome characteristics | *P. redivivus* | *C. elegans* | *B. xylophilus* | *A. suum* |
|---|---|---|---|---|
| Estimated genome size (Mb) | 64.4 | 100 | 74.6 | 272 |
| N50 (bp) | 26,2414 | [a] | 1,158,000 | 407,899 |
| GC content (%) | 44.25 | 35.4 | 40.4 | 37.9 |
| Repetitive sequences (%) | 7.1 | | 22 | 4.4 |
| Average intron length (bp) | 163 | 320 | 153 | 1,081 |
| Average exon length (bp) | 288 | 201.6 | 288.9 | 153 |
| Average no. of exons per gene | 4 | 6.5 | 4.5 | 6 |

The total estimated genome size of *P. redivivus* is 64.4 Mb based on our sequence data from the genome and transcriptome.
[a] Fully sequenced genome, end-to-end.

**Table 2 Summary of miRNAs discovered in *P. redivivus***

| | |
|---|---|
| 248 | Confirmed miRNAs: criteria 1, computationally predicted hairpin; criteria 2, >10 reads |
| 116 | miRNA[a] sequences with >10 reads |
| 63 | Orthologs in related species |
| 16 | Located in exons |
| 2 | Located in gene UTR |
| 3 | Pairs of hairpins expressed from both strands |
| 14 | miRNA clusters |
| 17 | miRNAs have multiple origins in the genome |

[a] "star" sequence, or less abundant mature miRNA molecule processed from the hairpin in miRNA nomenclature.

study (Table S4), suggesting that this class of small RNA may not be utilized in *P. redivivus*. By contrast, we did find evidence for the 22G-RNA class of small RNAs. We observed 18 clusters of 22G-RNA, which were defined by at least 20 22G-RNA reads (criteria for a cluster were multiple reads with <10 copies each, length 21–23 nt, and spaced by at most 200 nucleotides). Sixteen of these clusters were located on the opposite strand of the target gene. Within these clusters were only 22G-RNAs (G first nucleotide, 21–23 nucleotides in length). An example of a 22G cluster from *P. redivivus* is shown in Figure S2. The 22G-RNA class is further divided into two subclasses, one bound by CSR-1 and required for holocentric chromosome segregation (Claycomb *et al.* 2009), and the other subclass bound by worm-specific AGOs and playing an important role in transposon, pseudogene, cryptic locus, and protein-coding gene silencing (Gu *et al.* 2009). The *P. redivivus* genome has five CSR-1 orthologs (Table S4), suggesting that the CSR-1 pathway may be more elaborated in *P. redivivus* compared to *C. elegans*.

### *Orthology analysis*

We analyzed protein orthology to explore the architecture of the *P. redivivus* proteome. We compared 24,249 *P. redivivus* proteins to the proteomes of seven other nematode species and an insect outgroup: *P. redivivus*, *C. elegans*, *P. pacificus*,

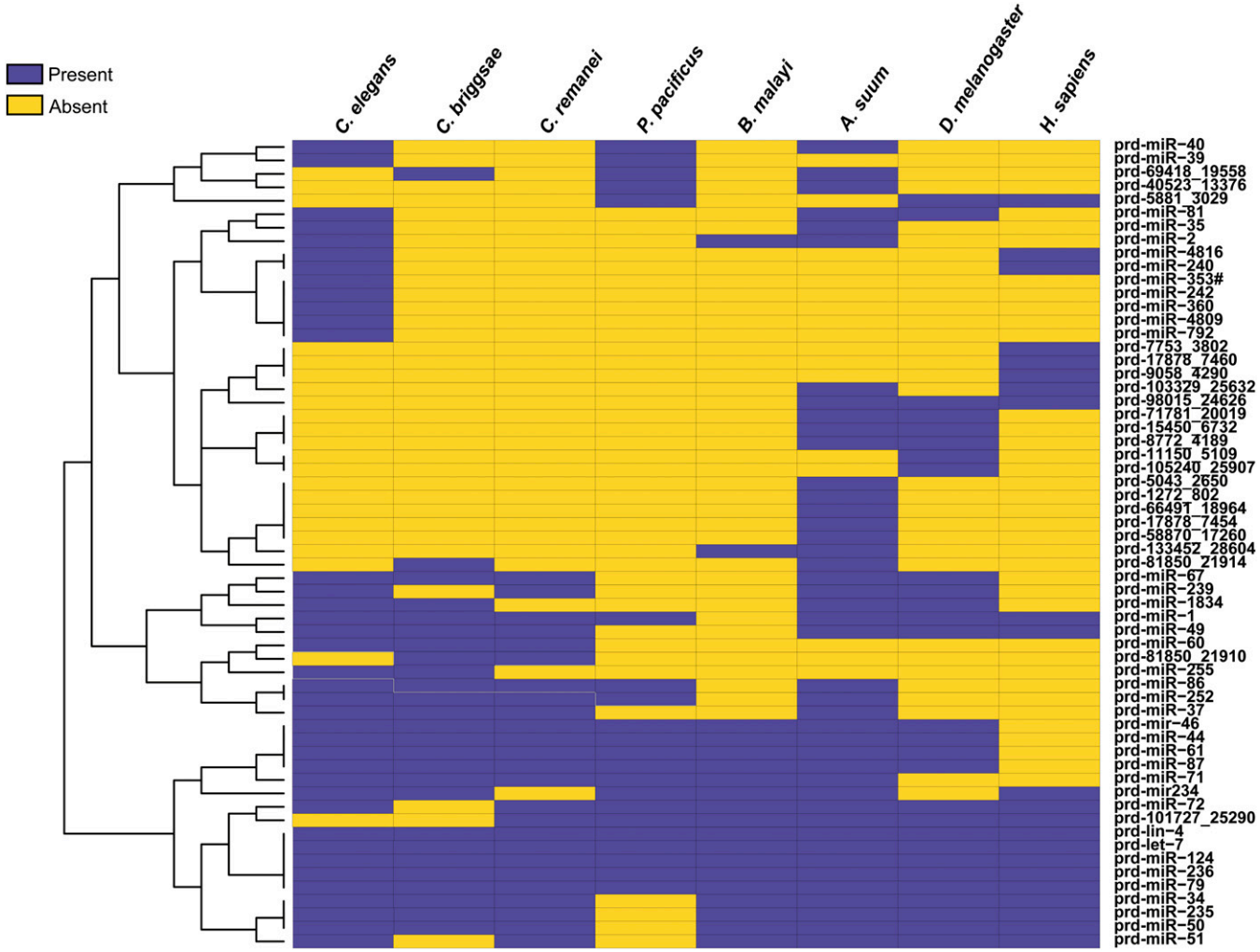


**Figure 3** Clustered heat map of orthologous miRNAs from different species of nematodes, *Drosophila* and humans. The figure shows the spread of conservation of the *P. redivivus* miRNAs in the studied species. miRNAs with more orthologs are located in clusters at the bottom of the map, while those miRNAs with only a few orthologs are found at the top.

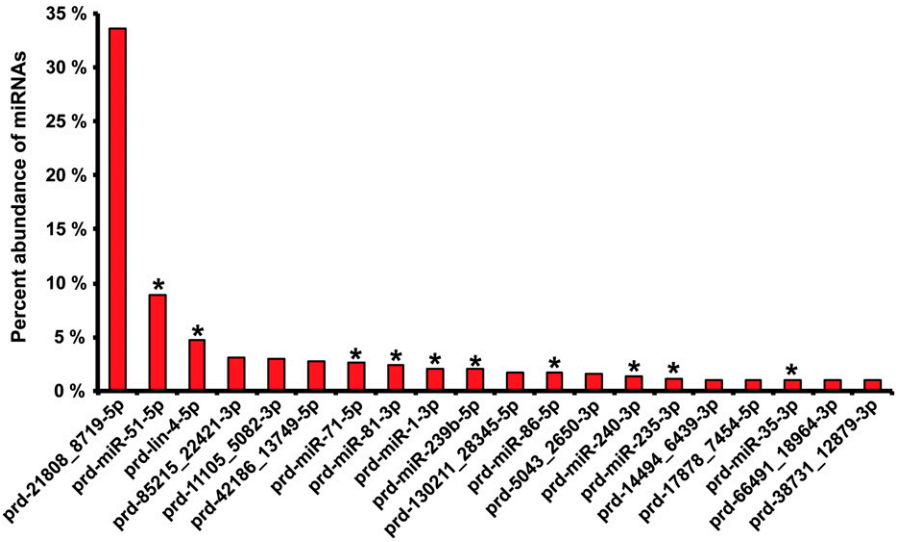


**Figure 4** Relative abundance of different miRNAs in the *P. redivivus* and their conservation in *C. elegans*. The proportions of the 20 most highly expressed *P. redivivus* miRNAs in our analysis. Ten of these miRNAs are conserved in *C. elegans* and are marked with asterisk at the top of the bar.

*Meloidogyne hapla*, *Bursaphelenchus xylophilus*, *Brugia malayi*, *A. suum*, *T. spiralis*, and the parasitoid wasp *Nasonia vitripennis* (*C. elegans* Genome Sequencing Consortium 1998; Ghedin *et al.* 2007; Dieterich *et al.* 2008; Opperman *et al.* 2008; Werren *et al.* 2010; Jex *et al.* 2011; Kikuchi *et al.* 2011; Mitreva *et al.* 2011) (Table 3 and File S1). An important caveat of such orthology analyses is that the accuracy of the results relies on the quality and completeness of the proteomes used. We found a total of 9156 orthology clusters that included 17,415 *P. redivivus* proteins; 281 of these were found exclusively in nematodes. A total of 1,664 orthology clusters included at least one protein from each of the nine taxa that we analyzed (N:N), 521 of which were strictly conserved at a 1:1 ratio across all taxa (Table 3). These highly conserved proteins provide a candidate list of additional potential phylogenetic markers that could increase the signal-to-noise ratio in future phylum-wide phylogenetic analyses (Holterman *et al.* 2006; van Megan *et al.* 2009). *P. redivivus* had 6834 orphan proteins that did not cluster with any examined proteins, suggesting that they are uniquely derived in *P. redivivus* or that they are sufficiently divergent from their orthologs so as to not be recognizably related by sequence similarity alone. We find it remarkable that, despite using eight nematode proteomes, representing only 4 of the 12 clades (Figure 1), we still find that >20% of the protein-coding genes in each species are orphans, with little-to-no sequence homology with other proteins in the analysis (Table S5). This suggests that a tremendous diversity of proteins underlies the superficial similarity of nematode morphology and that many novel proteins may yet remain to be discovered with additional genome sequencing of these wonderfully adaptable worms.

### *Signaling and regulatory pathways in P. redivivus*

Organisms often display remarkable plasticity during their life cycle and are capable of adapting to different conditions by sensing their environment and physiological status. Behavioral and metabolic changes are the most common forms of plasticity in response to environmental changes (Fielenbach and Antebi 2008). Both these processes are rapid responses to the environment and help the organism maintain homeostasis. Since the nematode *P. redivivus* is free-living and appears to inhabit nutrient-rich environments, we examined whether changes in components of signaling or developmental pathways reflect its adaptation to such a lifestyle. We also screened the assembled genome for the conservation of important biological pathways including the dauer, cell-death, and RNA interference (RNAi) pathways.

### *Dauer formation pathway*

One of the most extensively studied molecular pathways in *C. elegans* is the dauer formation pathway (Fielenbach and Antebi 2008). The dauer diapause represents a long-lived life stage, which is a developmental response to stressful environmental conditions such as low availability of food and high population density. Detailed molecular and genetic analyses in *C. elegans* have revealed how the worm senses its environment and reacts to changing environmental conditions by activating conserved signaling pathways to initiate entry into the dauer stage (Fielenbach and Antebi 2008; Schaedel *et al.* 2012).

**Table 3 *P. redivivus* orthology statistics**

| | |
|---|---|
| Predicted proteins in *P. redivivus* | 24,249 |
| *P. redivivus* proteins in clusters | 17,415 |
| Clusters with *P. redivivus* proteins | 9,156 |
| Clusters without *P. redivivus* proteins | 8,794 |
| *P. redivivus* orphan proteins (unclustered) | 6,834 |
| N:N orthologous protein clusters | 1,664 |
| 1:1 orthologous protein clusters | 521 |
| Orthologous protein clusters including all nematode taxa but no insect ortholog | 281 |

The *P. redivivus* genome and transcriptome reveal 27,266 proteins. Of these, 443 proteins are conserved at a 1:1 ratio across seven other nematodes, listed in Table 6, and the insect outgroup *N. vitripennis*. Only 266 orthologous protein clusters in this data set were exclusively nematode proteins.

**Table 4 Conservation of the dauer pathway**

| *C. elegans* protein | *Cele* | *Ppac* | *Pred* | *Bxyl* | *Mhap* | *Bmal* | *Asuu* | *Tspi* | *Nvit* |
|---|---|---|---|---|---|---|---|---|---|
| Pheromone | | | | | | | | | |
| DAF-22 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | x | 1 |
| DAF-6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | x | x |
| Guanylyl cyclases | | | | | | | | | |
| DAF-11 | 1 | 1 | 1 | 1 | x | 1 | 1 | x | x |
| DAF-21 | 1 | 1 | 1 | x | 1 | 1 | 1 | 1 | 2 |
| DAF-10 | 1 | x | 1 | 1 | x | 1 | 1 | 1 | 1 |
| TGFβ-like pathway | | | | | | | | | |
| DAF-1 | 1 | 1 | 1 | 1 | x | 1 | 1 | 2 | 2 |
| DAF-4 | 1 | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| DAF-7 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 |
| DAF-8 | 3 | 1 | 3 | 3 | 1 | 3 | 3 | 2 | 3 |
| DAF-14 | 1 | x | x | x | x | x | x | x | x |
| DAF-3 | 2 | 5 | 1 | 2 | 1 | 1 | 2 | 2 | 1 |
| DAF-5 | 1 | x | x | x | x | x | x | x | x |
| Insulin-like pathway | | | | | | | | | |
| DAF-2 | 1 | 1 | 2 | 2 | x | 4 | 2 | 2 | 1 |
| DAF-16 | 1 | 1[a] | 1 | 2 | 1 | 2 | 1 | 1 | 2 |
| DAF-23/AGE-1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | x | 1 |
| Steroid hormone | | | | | | | | | |
| DAF-9 | 1 | 1 | 2 | 1 | x | 1 | 1 | x | x |
| DAF-12 | 1 | 1 | 1 | 1 | x | 1 | 2 | x | x |
| Other effectors | | | | | | | | | |
| DAF-15 | 1 | 1 | 1 | 1 | x | 1 | 1 | 1 | 1 |
| DAF-19 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 |
| DAF-18 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| DAF-28 | 8 | x | x | x | x | x | x | x | x |
| DAF-36 | 1 | 1 | x | x | x | 4 | 1 | x | 3 |
| TAX-2 | 1 | 1 | 1 | 1 | 1 | x | 1 | x | x |
| TAX-4 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 |
| EGL-4 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 |

Shown are the number of proteins from each species analyzed that cluster as orthologs with the known *C. elegans* protein. "x" indicates that no proteins from the proteome used clustered with the known *C. elegans* protein. Descriptive labels for certain pathway components known in *C. elegans* are given in the first column. Results are based on an orthology analysis using the available proteomes and OrthoMCL (see Supporting Information).

[a] Note that, while *daf-16* was not present in the version of the proteome that we used, it is known to be present in the *P. pacificus* genome (Ogawa *et al.* 2011).

Several researchers have suggested that *P. redivivus* does not form dauers (Hechler 1970; Stock and Nadler 2006). This observation is surprising as the genome of *P. redivivus* encodes nearly all the major components of the dauer pathway, with 21 of the 25 proteins that we examined being conserved (Table 4). Howerver, *Panagrellus* isolates do form dauers in nature (Félix and Duveau 2012), and it is possible that this species or this laboratory strain lost the ability to form dauers under standard laboraotory conditions. Several of the pathway components apparently absent in *P. redivivus* appear to be specific to the caenorhabditids.

### *Cell death pathway*

Cell death is a critical process during development because animals need to eliminate unwanted cells in a regulated and timely manner (Horvitz 2003). In *C. elegans*, programmed cell death is believed to be molecularly initiated by the activation of a core cell-death pathway, consisting of *egl-1*, *ced-9*, *ced-4*, and *ced-3* (Metzstein *et al.* 1998). Cell death is highly abundant during *P. redivivus* germline and somatic development (Sternberg and Horvitz 1981, 1982). In addition, *P. redivivus* development undergoes specific cell deaths that are evolutionarily derived (*e.g.*, the female gonadal posterior distal tip cell). Given these observations, we examined whether core components of the cell-death pathway are present in *P. redivivus*. Most known effectors are conserved; however, CED-9 appears to be absent (Table S6). We assume that this is a result of the draft nature of the genome or possibly the result of divergent sequence rather than the actual absence of this gene. CED-9 is central in the regulation and prevention of cell death in many species and is highly conserved from *C. elegans* to humans (Metzstein *et al.* 1998).

### *RNAi pathway*

The RNAi pathway has become a valuable experimental tool to perturb individual or groups of genes to uncover their specific function(s), although its application and reliability across different nematodes has been inconsistent (Urwin *et al.* 2002; Viney and Thompson 2008; Dalzell *et al.* 2010, 2011). An obvious potential explanation for this is a conspicuous lack of certain RNAi effectors in some nematode species. However, we note that other factors, such as culturing conditions, rather than the disparity of RNAi effectors across species may better explain RNAi competencies among nematodes (Dalzell *et al.* 2011). We found that many RNAi effector genes are conserved in *P. redivivus* (Table S4 and Table S7). We found 56 RNAi effector proteins that cluster with known effectors in *C. elegans*, including at least 16 Argonaute-like proteins, in the *P. redivivus* genome (Table S4 and Table S7), suggesting that *P. redivivus* has more of the known RNAi pathway conserved with *C. elegans* than any other noncaenorhabditid nematode that has been sequenced; however, this is due in part to *P. redivivus* expansions in certain orthologous clusters such as CSR-1- and EKL-1-like proteins (Dalzell *et al.* 2011). Despite the high number of orthologous effectors in *C. elegans* and *P. redivivus*, 21 of the 73 RNAi effectors that we examined appear to be specific to the *C. elegans* lineage, having no apparent orthologs in any of the taxa that we analyzed (Table S4 and Table S7). We are hopeful that *P. redivivus* will be susceptible to experimental RNAi, given the apparent conservation of so many effectors that need to be tested.

A novel small RNA pathway required for the production and/or function of germline small RNA(s) in *C. elegans* includes four regulator genes (*csr-1*, *drh-3*, *ego-1*, and *ekl-1*) (Rocheleau *et al.* 2008). We found that three of these genes have expanded to small families in the *P. redivivus* lineage, with five paralogs of CSR-1, three paralogs of DRH-3, and six paralogs of EKL-1. All three genes are required for RNAi in the *C. elegans* germline and share a chromosome segregation-defective and embryonic-lethal phenotype (Grishok *et al.* 2001, 2005; Kim *et al.* 2005; Robert *et al.* 2005; Duchaine *et al.* 2006). An unusual group of

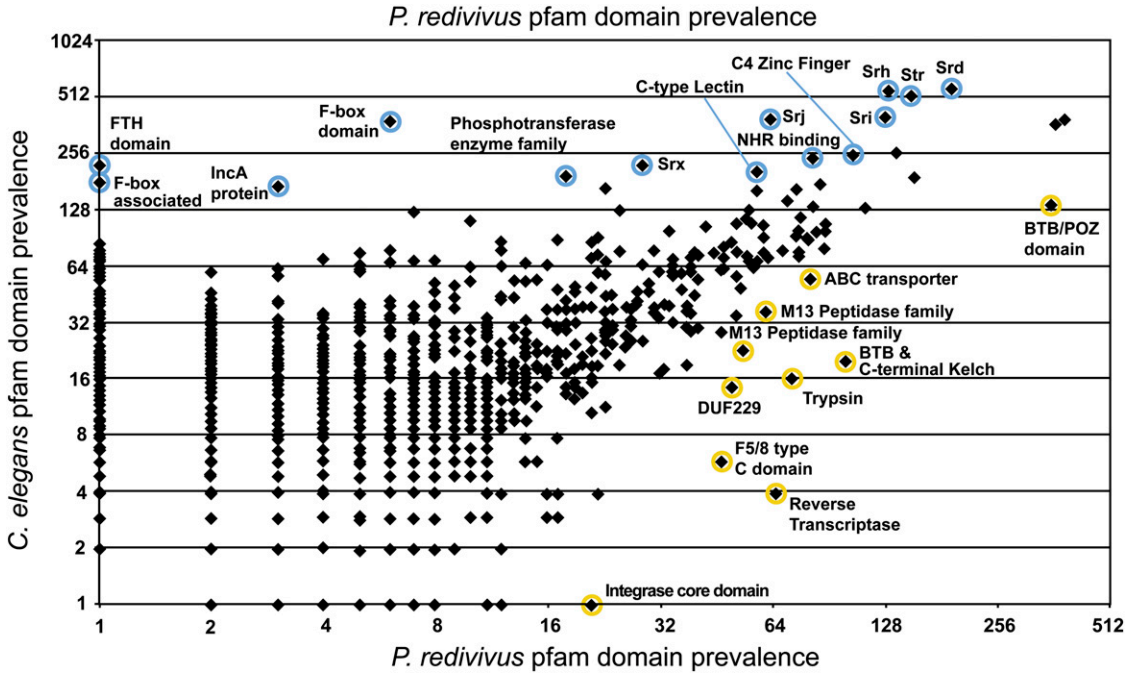


**Figure 5** A scatterplot showing the abundance of Pfam protein family domains in the *P. redivivus* and *C. elegans* genomes. The 14 most enriched Pfam domains in *C. elegans* relative to *P. redivivus* are highlighted in blue while those seemingly enriched in *P. redivivus* relative to *C. elegans* are highlighted in yellow. The genome of *C. elegans* is enriched in serpentine family domain GPCRs and F-box domains. In contrast, *P. redivivus* is highly enriched in BTB domains.

retrotransposons, named PAT elements, was previously identified in *P. redivivus*. These retrotransposons have contributed to a higher spontaneous mutation rate in *P. redivivus* compared to *C. elegans* (Link *et al.* 1987; de Chastonay *et al.* 1992). Although we could not precisely determine the number of PAT element copies in the *P. redivivus* genome, an HMMER Pfam analysis indicates the presence of at least 65 copies of reverse transcriptase and 23 copies of integrase, suggesting at least 23–65 retroelements (Table S8) (Finn *et al.* 2011). The apparent expansion of *csr-1*, *drh-3*, and *ekl-1* and the abundance of retrotransposons in the genome suggest pronounced regulation of transposons in the germline of *P. redivivus*.

### *Protein family domain abundance*

An analysis of domain abundance of various protein families provides an unbiased approach to exploring the vast sea of genomic data and reveals striking differences between the free-living nematodes *C. elegans* and *P. redivivus* (Figure 5; Figure S5). The *C. elegans* genome is greatly enriched in F-box, F-box-associated, FTH, and C-type lectin domains, among others, when compared to the *P. redivivus* genome. By contrast, the *P. redivivus* genome is enriched in BTB/POZ, BTB and C-terminal Kelch, trypsin, reverse transcriptase, and integrase core domains, among others (Figure 5). Both F-box and BTB domains are structural motifs that mediate protein–protein interactions, and proteins containing these domains are associated with signal transduction, cell-cycle regulation, and other cellular functions (Craig and Tyers 1999; Kipreos and Pagano 2000; Pintard *et al.* 2003; Stogios *et al.* 2005). The few members of these protein families that have been well-studied function as adaptors that determine the substrate specificity of E3 ubiquitin ligases, targeting substrates for proteolysis (Bai *et al.* 1996; Craig and Tyers 1999; Gagne *et al.* 2002; Furukawa *et al.* 2003; Pintard *et al.* 2003). Both F-box and BTB proteins are thought to play an important role in nematode immunity, with certain substrate-binding motifs having undergone heavy positive selection to target bacterial and viral peptides in the ever-escalating host–pathogen arms race (Dawkins and Krebs 1979; Thomas 2006). A detailed examination of the family of BTB domain-containing proteins in the *P. redivivus* genome revealed that there are large lineage-specific clades that appear to be rapidly evolving, suggestive of their involvement in immune responses (Tables 5 and 6; Figure 6). In addition, the presence of smaller, conserved orthology clusters of BTB/POZ and BTB/C-terminal Kelch proteins suggests that these likely target endogenous proteins, possibly for degradation in an E3 ligase proteolysis pathway (Figure 6; Figure S4) (Petroski and Deshaies 2005). This pattern of expansion and conservation of *P. redivivus* BTB domain-containing proteins, with many seemingly fast-evolving lineage-specific clusters, is consistent with observations from the *C. elegans* genome that F-box and BTB domain-containing proteins likely function in immunity and proteolysis (Thomas 2006).

The extent of variation in the number of F-box and BTB domains between *P redivivus* and *C. elegans* is striking. We pursued this observation further by evaluating the prevalence of F-box and BTB proteins across many nematodes

**Table 5 Selected domain prevalence among nematodes**

| Species | F-box/region/associated | SOCS/BC Box | BTB/POZ/C-terminal Kelch | Cullin |
|---|---|---|---|---|
| *M. hapla* | 10 | 2 | 54 | 3 |
| *B. xylophilus* | 15 | 1 | 51 | 9 |
| *P. redivivus* | 7 | 4 | 368 | 16 |
| *C. elegans* | 299 | 5 | 107 | 7 |
| *P. pacificus* | 17 | 0 | 78 | 7 |
| *A. suum* | 10 | 5 | 39 | 8 |
| *B. malayi* | 15 | 5 | 47 | 8 |
| *T. spiralis* | 11 | 1 | 15 | 8 |

Free-living nematodes have a dramatic expansion of these domains throughout their genomes.

and found that the free-living nematodes are outliers, having far more of either of these protein domains than any other nematodes, including the necromenic nematode *P. pacificus* as well as *B. xylophilus*, a plant-parasitic member of clade 10 along with *P. redivivus* (Figure 1 and Table 5). We also note that the trend across nematodes seems to favor BTBs over F-box proteins, with the exception of *C. elegans*, which has far more F-box proteins. Expanding this analysis across eukaryotes reveals that metazoans generally have more BTB proteins than F-box proteins, with plants and *C. elegans* being the exceptions (Table 5 and Table 6).

Due to the dramatic disparity of F-box domains between the free-living nematodes *P. redivivus* and *C. elegans*, we investigated the conservation of F-box domain-containing proteins across nematodes (Table 5 and Table S9). We found few F-box domain-containing proteins broadly conserved in nematodes and insects (Table S9) (Jin *et al.* 2004). We have yet to find the highly conserved SEL-10 (CDC4), known for its role in Skp, Cullin, F-box containing complex (or SCF complex)-mediated proteolysis in our genome or transciptome.

We suggest that the apparent evolution of F-box and BTB proteins in *P. redivivus* could be a response to viral susceptibility. Both BTB and F-box domain-containing proteins are traditionally known for their roles as the substrate-specifying subunits of multisubunit **c**ullin-**R**ING ubiquitin **l**igases (CRLs) (Feldman *et al.* 1997; Michel and Xiong 1998; Pintard *et al.* 2003; Petroski and Deshaies 2005; Sarikas *et al.* 2011). These ligases are modular and are responsible for targeting a wide variety of substrates for proteolysis by ubiquitylating them. These complexes are assembled on a cullin scaffold, which tethers a RING protein to a substrate-specifying subunit, usually through an adaptor protein as in the case of the canonical $SCF^{Cdc4}$ CRL. There are a variety of different CRLs, each associating with a specific cullin protein and possessing different specificity, depending on the adaptor and/or substrate-specifying subunit used (Petroski and Deshaies 2005). For example, *C. elegans* is known to have seven CRLs with ubiquitin-ligase activity, each built on a distinguishing cullin scaffold (*cul-1* through *cul-6* and *apc-2*) (Sarikas *et al.* 2011). CUL-1 CRLs use F-box proteins for substrate specificity, while those with a CUL-3 cullin scaffold use BTB proteins for substrate specificity. CRL machinery is widely exploited by viruses as a method of immune evasion, with most known examples targeting aspects of the CUL-1 CRL (Barry and Fruh 2006).

In support of this hypothesis, we observed an unprecedented expansion of cullin proteins in the *P. redivivus* genome (Table 5 and Figure 7). We explored this expansion by constructing a gene tree of all cullin homology domain-containing proteins across sequenced nematode genomes (Figure 7). Because of the dramatic expansion of BTB proteins, we expected an accompanying expansion of the CUL-3 family in *P. redivivus*, but paradoxically found an expansion of CUL-1-like proteins, which are known to use F-box proteins for substrate specificity in other metazoans (Petroski and Deshaies 2005; Sarikas *et al.* 2011). We also identified a number of novel cullin proteins, with little similarity to any described families, including several that appear to have arisen due to recent tandem duplications (Figure 7). The apparent absence of any CUL-2 ortholog and the abundance of novel cullin proteins suggest a surprising amount of regulatory and proteolytic plasticity in *P. redivivus*, which may be shaped by the stressful demands of the free-living lifestyle,

**Table 6 Selected domain prevalence among animals**

| Species | F-box/region/associated | SOCS/BC BOX | BTB/POZ/C-terminal Kelch | Cullin |
|---|---|---|---|---|
| *H. sapiens* | 68 | 59 | 194 | 9 |
| *Mus musculus* | 82 | 45 | 202 | 9 |
| *Tribolium castaneum* | 15 | 11 | 66 | 7 |
| *N. vitripennis* | 27 | 6 | 148 | 7 |
| *D. melanogaster* | 26 | 21 | 219 | 7 |
| *Arabidopsis thaliana* | 307 | 0 | 51 | 10 |
| *Saccharomyces cerevisiae* | 4 | 0 | 1 | 4 |

Pfam domain abundance of F-box, SOCS, BTB, and cullin domain-containing proteins across eukaryotes. Most eukaryotes have more BTB-domain-containing proteins than F-box-domain-containing proteins.

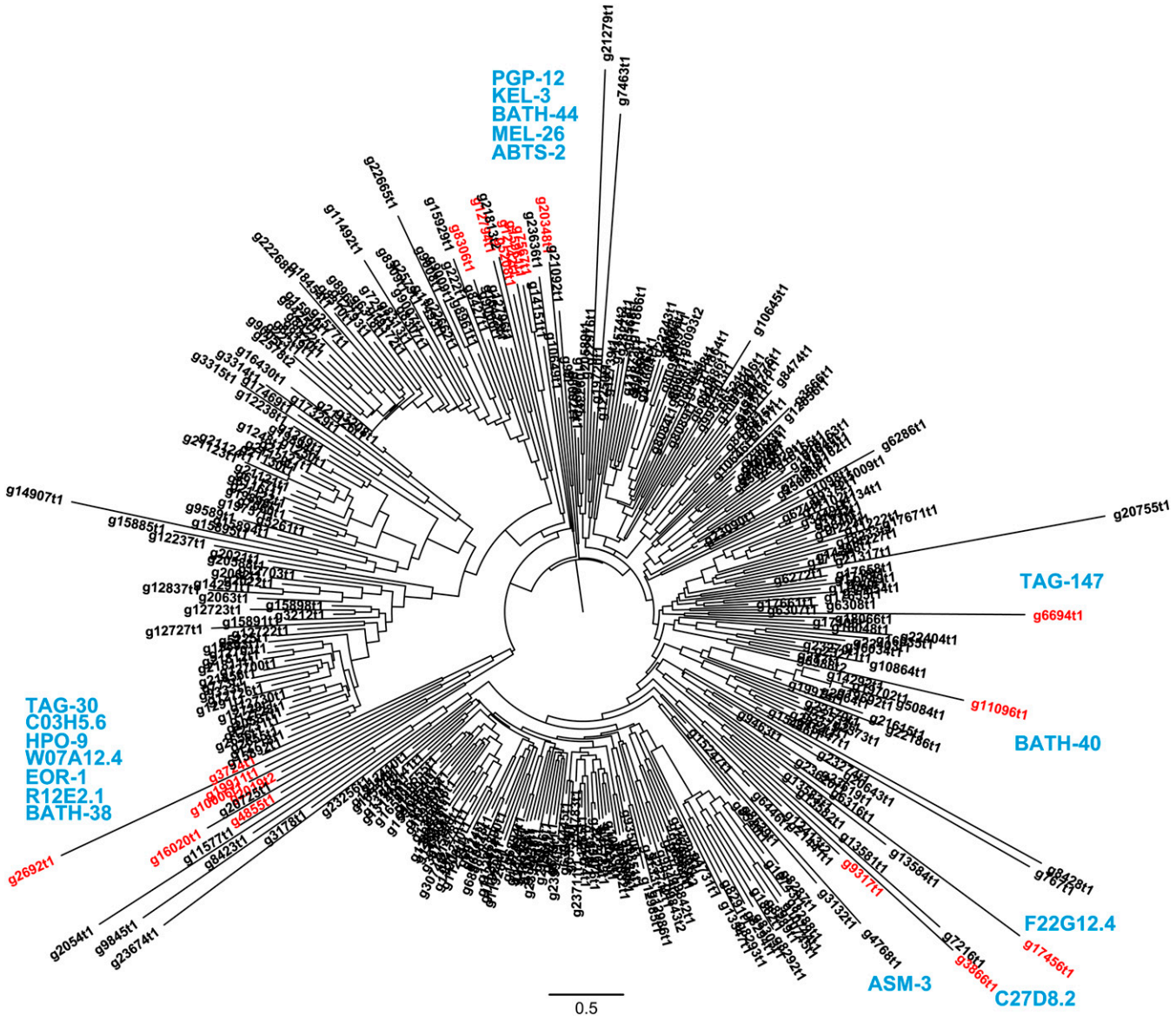


**Figure 6** Protein neighbor-joining tree of the BTB-domain-containing proteins in *P. redivivus*. This neighbor-joining gene tree shows all of the 367 BTB-domain-containing proteins in the *P. redivivus* proteome. Only 21 of these are conserved in at least two other nematode species, forming a total of 17 orthology clusters. These conserved proteins are highlighted in red, with the *C. elegans* ortholog names written in blue, where known. There are 22 genes conserved in at least one other nematode species, forming a total of 18 orthology clusters (see File S1). The 354 lineage-specific BTB proteins are suggested to be rapdily evolving and could function in binding non-endogenous proteins such as bacterial pattern recognition proteins.

immunological or otherwise. It is not uncommon for components of the ubiquitin system to be adapted to expand the immune system (*e.g.*, Han *et al.* 2011; Yewdell 2005). Although beyond the scope of this current work, our data suggest that exploring the role of these cullins and the function of the CRL complexes that they form would increase our understanding of the adaptative changes that *P. redivivus* has undergone to cope with the stresses of its ecological niche.

### *G-protein-coupled receptors*

Genomes of different organisms encode families of chemoreceptors, the size and diversity of which can reflect the niche inhabited by the organism (Thomas and Robertson 2008). These proteins mediate the first step in the transduction of chemical and other types of stimuli such as taste and pheromone signals. We screened the G-protein-coupled receptor (GPCR) repertoire in the genome of *P. redivivus* to better understand how the evolution of this large family of receptors might reflect its life history and how it compares to *C. elegans*. Not unexpectedly, both *C. elegans* and *P. redivivus* have an abundance of serpentine family domains (Srh, Str, Srd, etc.) belonging to the GPCR superfamily (Table 7). We found that, although the *P. redivivus* genome possesses a variety of GPCR proteins, it is far less abundant than what we found in *C. elegans* (Table 7). We observed that *P. redivivus*

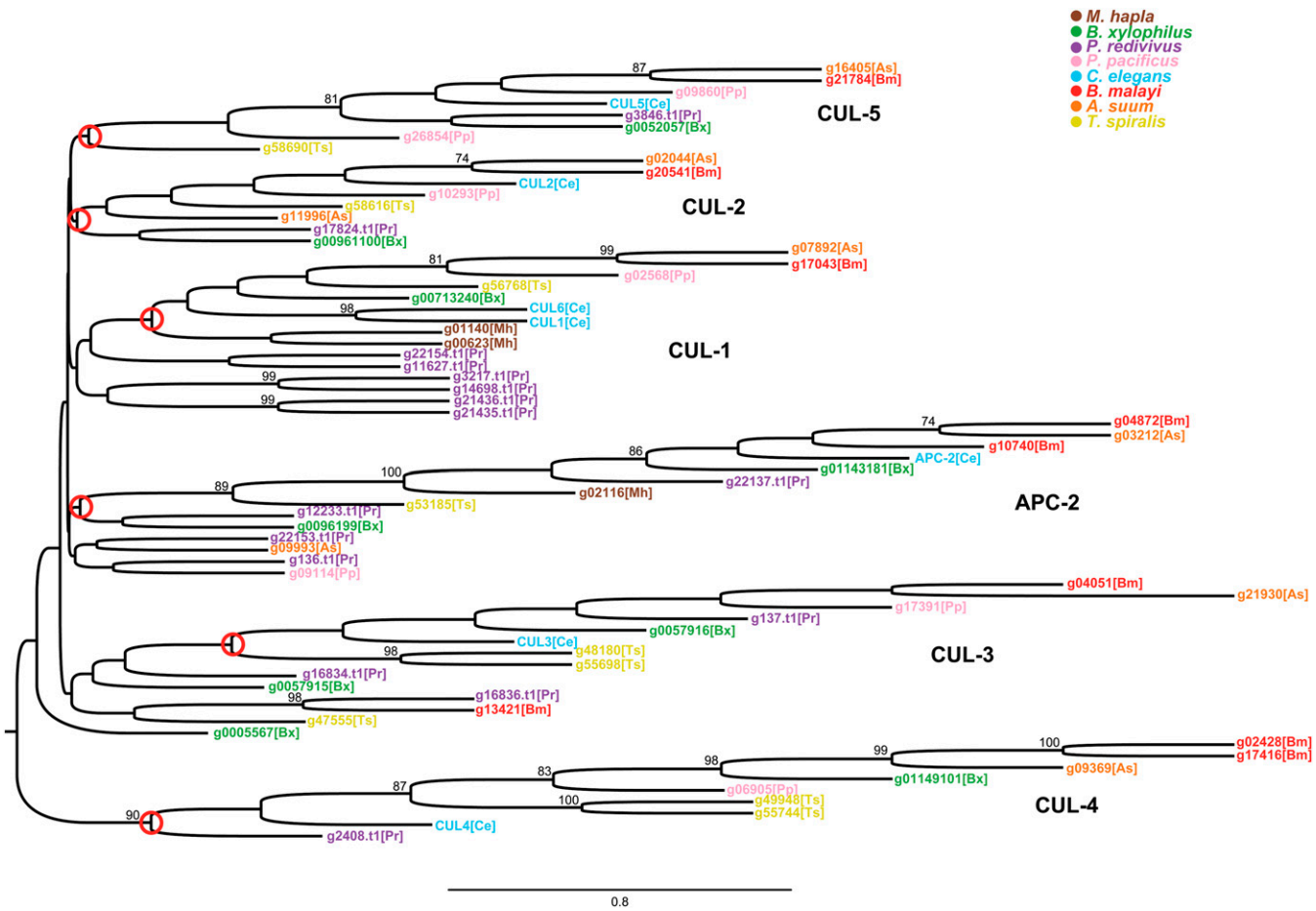


**Figure 7** Distance-based protein tree of cullin homology domain-containing proteins among nematodes. All five major nematode cullin families are monophyletic with their putative origins circled in red. *P. redivivus* has six CUL-1-like proteins, two APC-2-like proteins, and four other cullin homology domain-containing proteins with little sequence similarity to known families.

had 1075 serpentine domains compared to the 3259 domains of *C. elegans*. It has been suggested that *C. elegans* requires an abundance of chemoreceptors to navigate and interpret the nutrient-rich environments in which it lives (Robertson and Thomas 2006).

The number of serpentine GPCR domains in *P. redivivus* is similar to that of its clade mate, the migratory endoparasitic *B. xylophilus*. We did find that, of the nematodes that we analyzed, the animal parasites (*A. suum*, *B. malayi*, *and T. spiralis*) have far fewer GPCR domains compared to their free-living counterparts. Our data suggest that, in the specialized environments that these worms inhabit inside their hosts, they do not need a large repertoire of receptors, whereas the free-living nematodes, and nematodes that spend more foraging time in complex soil environments (*e.g.*, *P. pacificus* and *B. xylophilus*), require a larger set so that they can better navigate and interpret their environment (Table 7).

### *ABC transporters*

ATP-binding cassette (ABC) transporters provide a means for a wide variety of substrates to be actively transported across membranes, hydrolyzing ATP in the process (Davidson *et al.* 2008; Sundaram *et al.* 2008). *C. elegans* has 61 ABC transporters, representing ~0.3% of its protein-coding gene repertoire (Sheps *et al.* 2004). We found 94 putative ABC transporters in the *P. redivivus* genome, representing ~0.4% of the total number of genes that we report in this draft genome (Table S10 and File S1). We found *P. redivivus* orthologs for 52 of the 61 *C. elegans* ABC transporters, indicating a high level of conservation (Table S10). In addition to having lineage-specific ABC transporters, we see expansions of *hmt-1*-like and *pgp*-like ABC transporters (Table S10). Unsurprisingly, both of these families of ABC transporters are involved in tolerance of heavy metals and other toxins. *hmt-1* functions in heavy metal tolerance, mitigating the toxic effects of arsenic, cadmium, and copper on *C. elegans*, while *pgp-5* is involved in resistance to heavy metals and bacterial toxins (Kurz *et al.* 2007; Schwartz *et al.* 2010). Expansions in these particular families of ABC transporters may explain the high level of copper tolerance reported in *P. redivivus*, which has been shown to have higher tolerance to copper than *C. elegans* or *P. pacificus* (Boyd and Williams 2003).

**Table 7 pfam GPCR abundance and diversity**

| Serpentine receptor class | Pfam domain | *Cele* | *Pred* | *Bxyl* | *Ppac* | *Mhap* | *Asuu* | *Bmal* | *Tspi* |
|---|---|---|---|---|---|---|---|---|---|
| 7TM GPCR chemoreceptor Srh | PF10318.2 | 593 | 196 | 221 | 154 | *2* | 9 | 0 | 0 |
| 7TM GPCR chemoreceptor Srd | PF10317.2 | 446 | 131 | 206 | 146 | 28 | 16 | 0 | 0 |
| 7TM GPCR chemoreceptor Str | PF10326.2 | 441 | 152 | 238 | 176 | 20 | 13 | 0 | 0 |
| 7TM GPCR chemoreceptor Sri | PF10327.2 | 367 | 129 | 90 | 61 | 0 | 1 | 0 | 0 |
| 7TM GPCR chemoreceptor Srj | PF10319.2 | 338 | 63 | 143 | 131 | 2 | 6 | 0 | 0 |
| 7TM GPCR chemoreceptor Srx | PF10328.2 | 138 | 28 | 23 | 50 | 9 | 20 | 9 | 2 |
| 7TM GPCR chemoreceptor Srw | PF10324.2 | 170 | 22 | 27 | 15 | 5 | 23 | 7 | 4 |
| 7TM GPCR chemoreceptor Srsx | PF10320.2 | 82 | 70 | 62 | 41 | 60 | 35 | 2 | 18 |
| Srg family chemoreceptor | PF02118.14 | 101 | 82 | 13 | 39 | 8 | 1 | 0 | 0 |
| 7TM GPCR chemoreceptor Srbc | PF10316.2 | 88 | 9 | 4 | 4 | 0 | 4 | 0 | 0 |
| 7TM GPCR receptor class ab chemoreceptor | PF10292.2 | 80 | 42 | 51 | 25 | 13 | 5 | 5 | 1 |
| Sre G-protein-coupled chemoreceptor | PF03125.11 | 68 | 51 | 30 | 38 | 9 | 5 | 2 | 0 |
| 7TM GPCR chemoreceptor Srv | PF10323.2 | 60 | 44 | 10 | 36 | 1 | 2 | 0 | 4 |
| 7TM GPCR chemoreceptor Srz | PF10325.2 | 72 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| 7TM GPCR chemoreceptor Srt | PF10321.2 | 63 | 40 | 33 | 35 | 21 | 3 | 0 | 1 |
| 7TM GPCR chemoreceptor Sra | PF02117.9 | 59 | 4 | 6 | 5 | 2 | 1 | 3 | 0 |
| 7TM GPCR chemoreceptor Sru | PF10322.2 | 47 | 11 | 0 | 1 | 1 | 1 | 0 | 0 |
| 7TM GPCR chemoreceptor Srb | PF02175.9 | 30 | 0 | 0 | 4 | 0 | 1 | 0 | 0 |
| Serpentine receptor-like protein, class xa | PF03383.8 | 16 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Total serpentine GPCR Domains | | 3259 | 1075 | 1157 | 961 | 191 | 146 | 28 | 31 |
| No. of proteins with multiple GPCR domains | | 768 | 276 | 296 | 229 | 34 | 24 | 3 | 3 |
| No. of proteins with only one GPCR domain | | 746 | 290 | 198 | 307 | 115 | 78 | 20 | 25 |
| Total no. of GPCR domain-containing proteins | | 1514 | 566 | 494 | 536 | 149 | 102 | 23 | 28 |

Repertoire of seven transmembrane receptor domain families across various nematode species identified by hmmscan and pfam is shown (Finn *et al.* 2011). *C. elegans* exhibits a dramatic expansion of various subfamilies of these proteins. In contrast, the parasitic species do not display a large number of domains for this class of genes.

### *Eukaryotic release factor*

The expanded protein domain families in *P. redivivus* compared to *C. elegans* include the eukaryotic release factor domains 1, 2, and 3. These three domains are found in the eukaryotic release factor 1 protein (eRF1), which is highly conserved from yeast to humans and plays a key role in translational termination (Frolova *et al.* 1994). eRF1 recognizes termination codons contained in messenger RNA (mRNA) and competes with suppressor transfer RNA(s) for the ribosomal A site (Drugeon *et al.* 1997). Most animals have 2–3 eRF1 orthologs while *P. redivivus* seems to have a striking expansion of 15 (Tables 5 and 6). While one of these is a putative ortholog of ETF-1 in *C. elegans*, the rest are quite diverse and appear to be scattered throughout the genome (Figure S6). How would a nematode or any other animal make use of an expansion of eRF1-like proteins and what might that reveal about the life history or natural ecology of *P. redivivus*?

Suppression of translational termination is a common strategy of animal and plant viruses and is necessary for the replication of some viruses (ten Dam *et al.* 1990). The expansion of eRF1 proteins in *P. redivivus* could represent an enhanced arsenal against viral assault, providing evidence of an historical or ongoing arms race between *P. redivivus* and viral antagonists. eRF1 levels are important for translational termination such that overexpressing eRF1 reduces readthrough (Drugeon *et al.* 1997; Le Goff *et al.* 1997) and depleting eRF1 increases readthrough (Stansfield *et al.* 1996). It is known that targeted depletion of eRF1 is a strategy employed by some viruses, such as the murine leukemia virus, whose reverse transcriptase interacts with eRF1 to increase translational readthrough, leading to efficient replication of the virus (Orlova *et al.* 2003). Perhaps additional copies of eRF1 ensure peptide chain termination, preventing or decreasing translational readthrough and/or ribosomal frameshifting and thus conferring resistance or immunity to some viruses. Little is known about the breadth and diversity of viruses that infect nematodes, especially noncaenorhabditid nematodes. There are no reports regarding viral infection of *P. redivivus*; however, *P. redivivus* is known to have a relatively high load of unusual retrotransposons, designated as PAT retroid elements and thought to be distantly related to the Gypsy family of retrotransposons (Link *et al.* 1987; de Chastonay *et al.* 1992). While most retroid elements produce GAG and Pol genes by translational readthrough or ribosomal framshifting from the same mRNA transcript, PAT elements are thought to generate separate transcripts for GAG and Pol genes, respectively (Link *et al.* 1987; de Chastonay *et al.* 1992). This implies that they have evolved to regulate GAG and Pol ratios at the transcriptional level, bypassing the need for translational readthrough or ribosomal frameshifting. We speculate that this could represent a vivid example of an evolutionary arms race, with *P. redivivus* evolving an expanded repertoire of eRF1 genes to ensure translational termination while PAT elements, the only known active retroelements in *P. redivivus*, have shifted to generate their genes in discrete transcripts, thus overcoming the host genome's defenses, although this remains to be explored.

### *Concluding remarks*

The annotated draft genome and transcriptome of *P. redivivus* provides a powerful resource in evolutionary and ecological

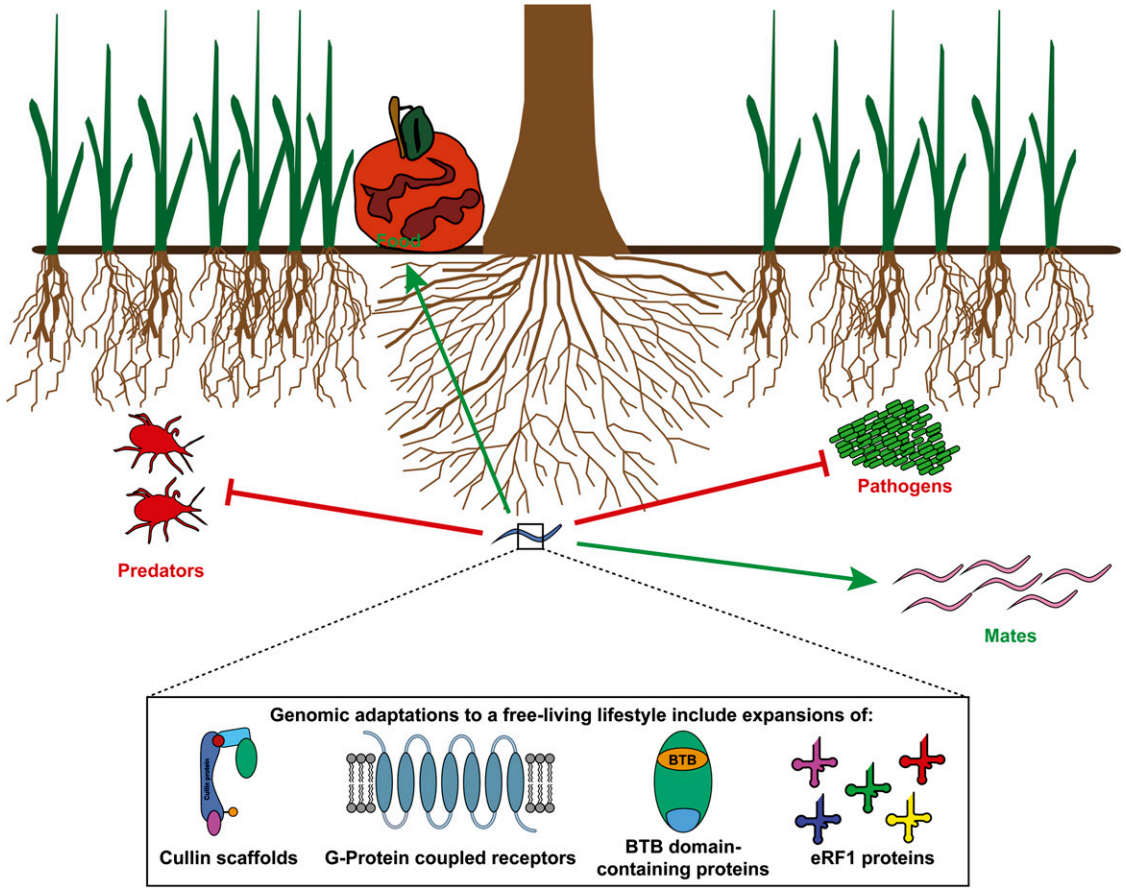


**Figure 8** Summary graphic showing the free-living lifestyle of *P. redivivus* and several genomic adaptations that facilitate it. As a free-living nematode, *P. redivivus* must seek out mates and find food resources while avoiding pathogens and predators in the complex environments in which it lives. The sequenced genome shows expansions of cullin proteins, GPCRs, BTB-domain-containing proteins, and immune effectors. These features of the *P. redivivus* genome appear to be adaptations to the free-living niche it occupies.

comparative genomics. As it is the first free-living genome outside of the *Caenorhabditis* family to be sequenced, the genome highlights features that are common with the *C. elegans* genome. This may reflect common constraints and adaptations resulting from the free-living lifestyle. Free-living worms live in a complex and dynamic environment and must be able to generate appropriate responses to different stimuli and protect themselves against exogenous threats such as predators and pathogens, while still managing to find food and mates (Figure 8). Our analyses suggest some common genomic and transcriptomic features between the *P. redivivus* and *C. elegans* genomes. These include a large complement of GPCRs for interpreting and navigating nutrient-rich environments and an expansion of immune-related proteins to combat the abundant pathogens found in such environments. We also observe unexpected novelties, such as an unprecedented expansion of cullin scaffold proteins in *P. redivivus* and an unprecedented expansion of eRF1 orthologs. Some of the genomic features that we have described such as expansions in certain ABC transporters and eRF1 proteins may explain previous observations regarding toxin tolerance and the unusual PAT retroelements present in the *P. redivivus* genome (Link *et al.* 1987; de Chastonay *et al.* 1992; Boyd and Williams 2003) . These findings potentially encourage the development of new avenues in nematode research (Figure 8).

Comparative nematode genomics has come a long way since the release of the first whole nematode genome sequence (*C. elegans* Genome Sequencing Consortium 1998). Many additional nematode genomes have been sequenced, and the continuing drop in cost will ensure that even more will be sequenced (Stein *et al.* 2003; Ghedin *et al.* 2007; Abad *et al.* 2008; Dieterich *et al.* 2008; Mortazavi *et al.* 2010; Jex *et al.* 2011; Kikuchi *et al.* 2011; Mitreva *et al.* 2011; Sommer and Streit 2011). In addition, sequencing the genomes of nematode pests is providing researchers an avenue for identifying pharmacological targets that could be useful in the development of novel drugs against these parasitic nematodes (*e.g.*, Jex *et al.* 2011). Comparison of genes involved in parasitism across various nematode clades expands our knowledge of the role played by processes such as horizontal gene transfer in the evolution of parasitism by nematodes (Mayer *et al.* 2011). Although there are 13 nematode genome sequences available, with many more in preparation,

sequencing efforts have focused primarily on the crown clades of Chromadoria, heavily covering clade 9, and many of these projects have focused (appropriately) on parasites; however, we believe that our understanding of development, gene regulation, and niche partitioning among nematodes, as well as parasitism, will be greatly enhanced by studying the free-living ancestors from which parasites evolved (Dillman *et al.* 2012). This comparative analysis highlights some of the potential selective pressures on free-living nematodes and the adaptations that allow them to thrive in the natural world.

## Acknowledgments

We thank Hillel Schwartz, Margaret C. Ho, Alexei Aravin, Claire C. de la Cova, Iva Greenwald, Kay Hofmann, and Raymond Deshaies for discussions. We thank Alicia Rogers and Zane Goodwin for their patience and help with some of the computations. This work was supported by a National Institutes of Health U.S. Public Health Service Training Grant (T32GM07616) to A.R.D. and by the Howard Hughes Medical Institute (with which P.W.S. is an Investigator). Sequencing was carried out at the Millard and Muriel Jacobs Genome Facility at the California Institute of Technology.

## Literature Cited

Abad, P., J. Gouzy, J. M. Aury, P. Castagnone-Sereno, E. G. Danchin *et al.*, 2008 Genome sequence of the metazoan plant-parasitic nematode *Meloidogyne incognita*. Nat. Biotechnol. 26: 909–915.

Ager, D., M. R. Samoiloff, and A. Petkau, 1984 Effect of dose-rate on the frequency of X-linked lethal mutation in the nematode *Panagrellus redivivus*. Mutat. Res. 140: 107–110.

Ambros, V., and H. R. Horvitz, 1984 Heterochronic mutants of the nematode *Caenorhabditis elegans*. Science 226: 409–416.

Ayala, F. J., and A. Rzhetsky, 1998 Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. Proc. Natl. Acad. Sci. USA 95: 606–611.

Bai, C., P. Sen, K. Hofmann, L. Ma, M. Goebl *et al.*, 1996 SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell 86: 263–274.

Barry, M., and K. Fruh, 2006 Viral modulators of cullin RING ubiquitin ligases: culling the host defense. Sci. STKE 2006: pe21.

Batista, P. J., J. G. Ruby, J. M. Claycomb, R. Chiang, N. Fahlgren *et al.*, 2008 PRG-1 and 21U-RNAs interact to form the piRNA complex required for fertility in *C. elegans*. Mol. Cell 31: 67–78.

Blaxter, M. L., P. De Ley, J. R. Garey, L. X. Liu, P. Scheldeman *et al.*, 1998 A molecular evolutionary framework for the phylum Nematoda. Nature 392: 71–75.

Boyd, W. A., and P. L. Williams, 2003 Comparison of the sensitivity of three nematode species to copper and their utility in aquatic and soil toxicity tests. Environ. Toxicol. Chem. 22: 2768–2774.

*C. elegans* Genome Sequencing Consortium, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282: 2012–2018.

Choe, A., T. Chuman, S. H. von Reuss, A. T. Dossey, J. J. Yim *et al.*, 2012 Sex-specific mating pheromones in the nematode *Panagrellus redivivus*. Proc. Natl. Acad. Sci. USA 109: 20949–20954.

Claycomb, J. M., P. J. Batista, K. M. Pang, W. Gu, J. J. Vasale *et al.*, 2009 The Argonaute CSR-1 and its 22G-RNA cofactors are required for holocentric chromosome segregation. Cell 139: 123–134.

Craig, K. L., and M. Tyers, 1999 The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. Prog. Biophys. Mol. Biol. 72: 299–328.

da Cruz, D. G., F. B. Araujo, M. B. Molento, R. A. Damatta, and C. de Paula Santos, 2011 Kinetics of capture and infection of infective larvae of trichostrongylides and free-living nematodes *Panagrellus* sp. by *Duddingtonia flagrans*. Parasitol. Res. 109: 1085–1091.

Dalzell, J. J., S. McMaster, C. C. Fleming, and A. G. Maule, 2010 Short interfering RNA-mediated gene silencing in Globodera pallida and Meloidogyne incognita infective stage juveniles. Int. J. Parasitol. 40: 91–100.

Dalzell, J. J., P. McVeigh, N. D. Warnock, M. Mitreva, D. M. Bird *et al.*, 2011 RNAi effector diversity in nematodes. PLoS Negl. Trop. Dis. 5: e1176.

Davidson, A. L., E. Dassa, C. Orelle, and J. Chen, 2008 Structure, function, and evolution of bacterial ATP-binding cassette systems. Microbiol. Mol. Biol. Rev. 72: 317–364.

Dawkins, R., and J. R. Krebs, 1979 Arms races between and within species. Proc. R. Soc. Lond. B Biol. Sci. 205: 489–511.

Debus, R., and R. Niemann, 1994 Nematode test to estimate the hazard potential of solved contaminations. Chemosphere 29: 611–621.

de Chastonay, Y., H. Felder, C. Link, P. Aeby, H. Tobler *et al.*, 1992 Unusual features of the retroid element PAT from the nematode *Panagrellus redivivus*. Nucleic Acids Res. 20: 1623–1628.

de Wit, E., S. E. Linsen, E. Cuppen, and E. Berezikov, 2009 Repertoire and evolution of miRNA genes in four divergent nematode species. Genome Res. 19: 2064–2074.

Dieterich, C., S. W. Clifton, L. N. Schuster, A. Chinwalla, K. Delehaunty *et al.*, 2008 The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. Nat. Genet. 40: 1193–1198.

Dillman, A. R., A. Mortazavi, and P. W. Sternberg, 2012 Incorporating genomics into the toolkit of nematology. J. Nematol. 44(2): 191–205.

Drugeon, G., O. Jean-Jean, L. Frolova, X. Le Goff, M. Philippe *et al.*, 1997 Eukaryotic release factor 1 (eRF1) abolishes readthrough and competes with suppressor tRNAs at all three termination codons in messenger RNA. Nucleic Acids Res. 25: 2254–2258.

Duchaine, T. F., J. A. Wohlschlegel, S. Kennedy, Y. Bei, D. Conte Jr., *et al.*, 2006 Functional proteomics reveals the biochemical niche of *C. elegans* DCR-1 in multiple small-RNA-mediated pathways. Cell 124: 343–354.

Feldman, R. M., C. C. Correll, K. B. Kaplan, and R. J. Deshaies, 1997 A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. Cell 91: 221–230.

Félix, M. A., and F. Duveau, 2012 Population dynamics and habitat sharing of natural populations of Caenorhabditis elegans and C. briggsae. BMC Biol. 2012 10: 59. doi: 10.1186/1741-7007-10-59.

Ferris, H., 2009 The beer mat nematode, *Panagrellus redivivus*: a study of the connectedness of scientific discovery. J. Nematode Morphol. Syst. 12: 19–25.

Fielenbach, N., and A. Antebi, 2008 *C. elegans* dauer formation and the molecular basis of plasticity. Genes Dev. 22: 2149–2165.

Finn, R. D., J. Clements, and S. R. Eddy, 2011 HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39: W29–W37.

Fire, A., S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver *et al.*, 1998 Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391: 806–811.

Frolova, L., X. Le Goff, H. H. Rasmussen, S. Cheperegin, G. Drugeon *et al.*, 1994 A highly conserved eukaryotic protein family possessing properties of polypeptide chain release factor. Nature 372: 701–703.

Furukawa, M., Y. J. He, C. Borchers, and Y. Xiong, 2003 Targeting of protein ubiquitination by BTB-Cullin 3-Roc1 ubiquitin ligases. Nat. Cell Biol. 5: 1001–1007.

Gagne, J. M., B. P. Downes, S. H. Shiu, A. M. Durski, and R. D. Vierstra, 2002 The F-box subunit of the SCF E3 complex is encoded by a diverse superfamily of genes in *Arabidopsis*. Proc. Natl. Acad. Sci. USA 99: 11519–11524.

Ghedin, E., S. Wang, D. Spiro, E. Caler, Q. Zhao *et al.*, 2007 Draft genome of the filarial nematode parasite *Brugia malayi*. Science 317: 1756–1760.

Grishok, A., A. E. Pasquinelli, D. Conte, N. Li, S. Parrish *et al.*, 2001 Genes and mechanisms related to RNA interference regulate expression of the small temporal RNAs that control *C. elegans* developmental timing. Cell 106: 23–34.

Grishok, A., J. L. Sinskey, and P. A. Sharp, 2005 Transcriptional silencing of a transgene by RNAi in the soma of *C. elegans*. Genes Dev. 19: 683–696.

Gu, W., M. Shirayama, D. Conte Jr., J. Vasale, P. J. Batista *et al.*, 2009 Distinct argonaute-mediated 22G-RNA pathways direct genome surveillance in the *C. elegans* germline. Mol. Cell 36: 231–244.

Han, K., D. I. Lou, and S. L. Sawyer, 2011 Identification of a genomic reservoir for new TRIM genes in primate genomes. PLoS Genet. 7: e1002388.

Hechler, H. C., 1970 Reproduction, chromosome number, and postembryonic development of *Panagrellus redivivus* (Nematoda: Cephalobidae). J. Nematol. 2: 355–361.

Hempel, M., Y. K. Chau, B. J. Dutka, R. McInnis, K. K. Kwan *et al.*, 1995 Toxicity of organomercury compounds: bioassay results as a basis for risk assessment. Analyst (Lond.) 120: 721–724.

Holterman, M., A. van der Wurff, S. van den Elsen, H. van Megen, T. Bongers *et al.*, 2006 Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades. Mol. Biol. Evol. 23: 1792–1800.

Horvitz, H. R., 2003 Nobel lecture. Worms, life and death. Biosci. Rep. 23: 239–303.

Huang, X., J. Liu, J. Ding, Q. He, R. Xiong *et al.*, 2009 The investigation of nematocidal activity in *Stenotrophomonas maltophilia* G2 and characterization of a novel virulence serine protease. Can. J. Microbiol. 55: 934–942.

Jex, A. R., S. Liu, B. Li, N. D. Young, R. S. Hall *et al.*, 2011 *Ascaris suum* draft genome. Nature 479: 529–533.

Jin, J., T. Cardozo, R. C. Lovering, S. J. Elledge, M. Pagano *et al.*, 2004 Systematic analysis and nomenclature of mammalian F-box proteins. Genes Dev. 18: 2573–2580.

Kikuchi, T., J. A. Cotton, J. J. Dalzell, K. Hasegawa, N. Kanzaki *et al.*, 2011 Genomic insights into the origin of parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. PLoS Pathog. 7: e1002219.

Kim, J. K., H. W. Gabel, R. S. Kamath, M. Tewari, A. Pasquinelli *et al.*, 2005 Functional genomic analysis of RNA interference in *C. elegans*. Science 308: 1164–1167.

Kiontke, K., and W. Sudhaus, 2006 Ecology of *Caenorhabditis* species (January 9, 2006), *WormBook*, ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.37.1, http://www.wormbook.org.

Kipreos, E. T., and M. Pagano, 2000 The F-box protein family. Genome Biol. 1: REVIEWS3002.

Kurz, C. L., M. Shapira, K. Chen, D. L. Baillie, and M. W. Tan, 2007 *Caenorhabditis elegans pgp-5* is involved in resistance to bacterial infection and heavy metal and its regulation requires TIR-1 and a p38 map kinase cascade. Biochem. Biophys. Res. Commun. 363: 438–443.

Laws, T. R., S. A. Smith, M. P. Smith, S. V. Harding, T. P. Atkins *et al.*, 2005 The nematode *Panagrellus redivivus* is susceptible to killing by human pathogens at 37 degrees C. FEMS Microbiol. Lett. 250: 77–83.

Lee, R. C., R. L. Feinbaum, and V. Ambros, 1993 The *C. elegans* heterochronic gene lin-4 encodes small RNAs with antisense complementarity to *lin-14*. Cell 75: 843–854.

Li, G. H., Y. M. Shen, and K. Q. Zhang, 2005 Nematicidal activity and chemical component of Poria cocos. J. Microbiol. 43: 17–20.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 1000 Genome Project Data Processing Subgroup. Bioinformatics 25: 2078–2079.

Link, C. D., J. Graf-Whitsel, and W. B. Wood, 1987 Isolation and characterization of a nematode transposable element from *Panagrellus redivivus*. Proc. Natl. Acad. Sci. USA 84: 5325–5329.

Malakhov, V. V., and W. D. Hope, 1994 *Nematodes: Structure, Development, Classification, and Phylogeny.* Smithsonian Institution Press, Washington, DC.

Mayer, W. E., L. N. Schuster, G. Bartelmes, C. Dieterich, and R. J. Sommer, 2011 Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. BMC Evol. Biol. 11: 13.

Metzstein, M. M., G. M. Stanfield, and H. R. Horvitz, 1998 Genetics of programmed cell death in *C. elegans*: past, present and future. Trends Genet. 14: 410–416.

Michel, J. J., and Y. Xiong, 1998 Human CUL-1, but not other cullin family members, selectively interacts with SKP1 to form a complex with SKP2 and cyclin A. Cell Growth Differ. 9: 435–449.

Mitreva, M., D. P. Jasmer, D. S. Zarlenga, Z. Wang, S. Abubucker *et al.*, 2011 The draft genome of the parasitic nematode *Trichinella spiralis*. Nat. Genet. 43: 228–235.

Mortazavi, A., B. A. Williams, K. McCue, L. Schaeffer, and B. Wold, 2008 Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5: 621–628.

Mortazavi, A., E. M. Schwarz, B. Williams, L. Schaeffer, I. Antoshechkin *et al.*, 2010 Scaffolding a *Caenorhabditis* nematode genome with RNA-seq. Genome Res. 20: 1740–1747.

Niu, X. M., Y. L. Wang, Y. S. Chu, H. X. Xue, N. Li *et al.*, 2010 Nematodetoxic aurovertin-type metabolites from a root-knot nematode parasitic fungus Pochonia *chlamydosporia*. J. Agric. Food Chem. 58: 828–834.

Ogawa, A., G. Bento, G. Bartelmes, C. Dieterich, and R. J. Sommer, 2011 *Pristionchus pacificus daf-16* is essential for dauer formation but dispensable for mouth form dimorphism. Development 138: 1281–1284.

Opperman, C. H., D. M. Bird, V. M. Williamson, D. S. Rokhsar, M. Burke *et al.*, 2008 Sequence and genetic map of *Meloidogyne hapla*: a compact nematode genome for plant parasitism. Proc. Natl. Acad. Sci. USA 105: 14802–14807.

Orlova, M., A. Yueh, J. Leung, and S. P. Goff, 2003 Reverse transcriptase of Moloney murine leukemia virus binds to eukaryotic release factor 1 to modulate suppression of translational termination. Cell 115: 319–331.

Parra, G., K. Bradnam, and I. Korf, 2007 CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23: 1061–1067.

Petroski, M. D., and R. J. Deshaies, 2005 Function and regulation of cullin-RING ubiquitin ligases. Nat. Rev. Mol. Cell Biol. 6: 9–20.

Pintard, L., J. H. Willis, A. Willems, J. L. Johnson, M. Srayko *et al.*, 2003 The BTB protein MEL-26 is a substrate-specific adaptor of the CUL-3 ubiquitin-ligase. Nature 425: 311–316.

Robert, V. J., T. Sijen, J. van Wolfswinkel, and R. H. Plasterk, 2005 Chromatin and RNAi factors protect the *C. elegans* germline against repetitive sequences. Genes Dev. 19: 782–787.

Robertson, H. M., and J. H. Thomas, 2006 The putative chemoreceptor families of *C. elegans* (January 6, 2006), *WormBook*,

ed. The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.66.1, http://www.wormbook.org.

Rocheleau, C. E., K. Cullison, K. Huang, Y. Bernstein, A. C. Spilker *et al.*, 2008 The *Caenorhabditis elegans ekl* (enhancer of ksr-1 lethality) genes include putative components of a germline small RNA pathway. Genetics 178: 1431–1443.

Rodriguez-Trelles, F., R. Tarrio, and F. J. Ayala, 2002 A methodological bias toward overestimation of molecular evolutionary time scales. Proc. Natl. Acad. Sci. USA 99: 8112–8115.

Ruby, J. G., C. H. Jan, and D. P. Bartel, 2007 Intronic microRNA precursors that bypass Drosha processing. Nature 448: 83–86.

Sarikas, A., T. Hartmann, and Z. Q. Pan, 2011 The cullin protein family. Genome Biol. 12: 220.

Schaedel, O. N., B. Gerisch, A. Antebi, and P. W. Sternberg, 2012 Hormonal signal amplification mediates environmental conditions during development and controls an irreversible commitment to adulthood. PLoS Biol. 10: e1001306.

Schulz, M. H., D. R. Zerbino, M. Vingron, and E. Birney, 2012 Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28: 1086–1092.

Schwartz, M. S., J. L. Benci, D. S. Selote, A. K. Sharma, A. G. Chen *et al.*, 2010 Detoxification of multiple heavy metals by a half-molecule ABC transporter, HMT-1, and coelomocytes of *Caenorhabditis elegans*. PLoS ONE 5: e9564.

Sheps, J. A., S. Ralph, Z. Zhao, D. L. Baillie, and V. Ling, 2004 The ABC transporter gene family of *Caenorhabditis elegans* has implications for the evolutionary dynamics of multidrug resistance in eukaryotes. Genome Biol. 5: R15.

Sommer, R. J., and A. Streit, 2011 Comparative genetics and genomics of nematodes: genome structure, development, and lifestyle. Annu. Rev. Genet. 45: 1–20.

Srinivasan, J., O. Durak, and P. W. Sternberg, 2008 Evolution of a polymodal sensory response network. BMC Biol. 6: 52.

Stanke, M., M. Diekhans, R. Baertsch, and D. Haussler, 2008 Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24: 637–644.

Stansfield, I., L. Eurwilaichitr, Akhmaloka, and M. F. Tuite, 1996 Depletion in the levels of the release factor eRF1 causes a reduction in the efficiency of translation termination in yeast. Mol. Microbiol. 20: 1135–1143.

Stein, L. D., Z. Bao, D. Blasiar, T. Blumenthal, M. R. Brent *et al.*, 2003 The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. 1: E45.

Sternberg, P. W., and H. R. Horvitz, 1981 Gonadal cell lineages of the nematode *Panagrellus redivivus* and implications for evolution by the modification of cell lineage. Dev. Biol. 88: 147–166.

Sternberg, P. W., and H. R. Horvitz, 1982 Postembryonic nongonadal cell lineages of the nematode Panagrellus redivivus: description and comparison with those of *Caenorhabditis elegans*. Dev. Biol. 93: 181–205.

Stock, P. S., and S. A. Nadler, 2006 Morphological and molecular characterisation of *Panagrellus* spp. (Cephalobina: Panagrolaimidae): taxonomic status and phylogenetic relationships. Nematology 8: 921–938.

Stogios, P. J., G. S. Downs, J. J. Jauhal, S. K. Nandra, and G. G. Prive, 2005 Sequence and structural analysis of BTB domain proteins. Genome Biol. 6: R82.

Sulston, J. E., E. Schierenberg, J. G. White, and J. N. Thomson, 1983 The embryonic cell lineage of the nematode *Caenorhabditis elegans*. Dev. Biol. 100: 64–119.

Sundaram, P., W. Han, N. Cohen, B. Echalier, J. Albin *et al.*, 2008 *Caenorhabditis elegans* ABCRNAi transporters interact genetically with *rde-2* and *mut-7*. Genetics 178: 801–814.

ten Dam, E. B., C. W. A. Pleij, and L. Bosch, 1990 RNA pseudoknots: translational frameshifting and readthrough of viral RNAs. Virus Genes 4: 121–136.

Thomas, J. H., 2006 Adaptive evolution in two large families of ubiquitin-ligase adapters in nematodes and plants. Genome Res. 16: 1017–1030.

Thomas, J. H., and H. M. Robertson, 2008 The *Caenorhabditis* chemoreceptor gene families. BMC Biol. 6: 42.

Trapnell, C., L. Pachter, and S. L. Salzberg, 2009 TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105–1111.

Trapnell, C., B. A. Williams, G. Pertea, A. Mortazavi, G. Kwan *et al.*, 2010 Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat. Biotechnol. 28: 511–515.

Urwin, P. E., C. J. Lilley, and H. J. Atkinson, 2002 Ingestion of double-stranded RNA by preparasitic juvenile cyst nematodes leads to RNA interference. Molecular plant-microbe interactions. Mol. Plant Microbe Interact. 15: 747–752.

van Megan, H., S. van den Elsen, M. Holterman, G. Karssen, P. Mooyman *et al.*, 2009 A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. Nematology 11: 927–950.

Viney, M. E., and F. J. Thompson, 2008 Two hypotheses to explain why RNA interference does not work in animal parasitic nematodes. Int. J. Parasitol. 38: 43–47.

Werren, J. H., S. Richards, C. A. Desjardins, O. Niehuis, J. Gadau *et al.*, 2010 Functional and evolutionary insights from the genomes of three parasitoid *Nasonia* species. Science 327: 343–348.

Yewdell, J. W., 2005 Immunoproteasomes: regulating the regulator. Proc. Natl. Acad. Sci. USA 102: 9089–9090.

Yuan, J., S. Shaham, S. Ledoux, H. M. Ellis, and H. R. Horvitz, 1993 The *C. elegans* cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 beta-converting enzyme. Cell 75: 641–652.

*Communicating editor: O. Hobert*

GENETICS

**Supporting Information**
http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.148809/-/DC1

# The Draft Genome and Transcriptome of *Panagrellus redivivus* Are Shaped by the Harsh Demands of a Free-Living Lifestyle

**Jagan Srinivasan, Adler R. Dillman, Marissa G. Macchietto, Liisa Heikkinen, Merja Lakso, Kelley M. Fracchia, Igor Antoshechkin, Ali Mortazavi, Garry Wong, and Paul W. Sternberg**


DOI: 10.1534/genetics.112.148809

**File S1**

**Supporting Methods**

**Strain culturing and maintenance of *P. redivivus.*** We used the *P. redivivus* strain (PS2298/MT8872) (STERNBERG and HORVITZ 1981) in our genomic and transcriptomic analysis. This strain was raised at 20$^{o}$C using standard methods.

**Isolation of DNA and RNA.** *P redivivus* worms were grown on five to ten 10-cm nutrient agar containing *E. coli* OP50 plates till near starvation. They were washed and collected with M9 buffer and washed multiple times to remove any *E. coli.* After the last wash in M9, the worms were suspended in M9 for 15-30 minutes. The worms were then snap-frozen in liquid nitrogen in ~100-µL aliquots and stored at −80°C. Worms were thawed and refrozen two to three times to break the cuticle before extracting either genomic DNA or bulk RNA. Genomic DNA was extracted using two rounds of proteinase K digestion followed by phenol-chloroform extraction. The genomic DNA was then treated RNase A for digestion of any RNAs present in the sample. Bulk RNA was extracted using the Qiagen RNeasy mini kit.

**Genomic and RNA-Seq library construction.** Genomic library (Library ID 11628) was constructed using Illumina Paired End DNA Sample Preparation Kit according to the manufacturer's instructions. Briefly, 3 µg of genomic DNA were fragmented using nebulization. The fragments were end repaired, 3' adenylated and ligated to Illumina's paired end adaptors. The ligation products were size selected on an agarose gel to yield fragments of approximate length of 350 bp and PCR amplified to produce the finished library. RNA-Seq library was created from 10 µg of total RNA. mRNA was purified using Dynal magnetic oligo(dT) beads (Invitrogen) and fragmented with 40mM Tris-acetate, pH 8.1, 100 mM KOAc, 30 mM MgOAc buffer for 4 min at 94C. First and second cDNA strands were synthesized using random primers and SuperScript II RT (Invitrogen), and RNaseH and DNA Pol I, respectively. The rest of the procedure was identical to that used for the genomic library preparation, except that the gel cut for the RNA-seq library was ~ 300 bp. Libraries were quantified using Qubit fluorometer (Invitrogen) and size distributions were verified using Agilent Bioanalyzer and the High Sensitivity DNA Kit. Libraries were sequenced on Illumina Genome Analyzer IIx sequencer in paired-end mode with the read length of 76 nt.

**Genome assembly and annotation.** Both the genomic and the mixed-stage transcriptome libraries were built, sequenced, assembled, filtered, and repeat-masked as previously described (MORTAZAVI *et al.* 2010) using Velvet 1.0.9. Genome and RNA-seq reads were submitted to the Sequence Read Archive under the accession number GSE44020.

Assembled cDNA was used to train Augustus 2.5 (STANKE *et al.* 2008) for protein-coding gene finding. Separately, RNA-seq reads were mapped onto the genome using TopHat 1.3.1 (TRAPNELL *et al.* 2009), assembled intro transcripts using Cufflinks 1.2.0 (TRAPNELL *et al.* 2010) and merged with the Augustus annotations using the RABT method (ROBERTS *et al.* 2011). Candidate SNVs in the genome and transcriptome mapped reads were called using the samtools (LI *et al.* 2009) pileup and varFilter options. Candidate SNVs in the transcriptome that fell within 5 bp of exon junctions were filtered out as likely splicing artifacts.

**Generation of the small RNA library.** Small RNAs were isolated from mixed cultures of *P. redivivus* using miRVana kit (Ambion) according to the manufacturer's instructions. A small RNA library was then produced from the isolated RNAs using NEBNext small RNA sample prep Set 1 (New England Biolabs). The library was then size selected on a 6% PAGE gel with the cut band corresponding to ~90-120 bp. Library quality and size was confirmed prior to sequencing on a Bioanalyzer (Agilent).

**Small RNA sequence data analysis.** 3' adapters and polyA tails were trimmed from the reads using an in-house script. Reads that were primer dimers, as well as reads matching to *E. coli OP50* genome, were discarded. Further, *P. redivivus* tRNAs were predicted using Aragorn (LASLETT and CANBACK 2004), and reads exactly mapping to these sequences were removed from the data set. Following trimming, all reads from 10 to 28 nt were used for miRNA prediction as described below. Reads were first mapped against the *P. redivivus* genome with Bowtie (LANGMEAD *et al.* 2009) allowing no mismatches and reporting only alignments for reads that had less than ten perfect matches to the genome. Using these alignments, reads in overlapping genomic locations were clustered together. Potential miRNA precursors were then excised from the genome using these clusters. First, all 60-100 nt long sequences, comprised of one or two adjacent clusters were extracted from the genome. Then, for all read clusters shorter than 60 nt, two putative miRNA precursor sequences were extracted, once including the flanking sequence 40 nt downstream and once including the flanking sequence 40 nt upstream of the cluster. Altogether, this procedure yielded 759 potential miRNA hairpins encompassing more than ten reads each. To classify a sequence as a miRNA hairpin, we used the following criteria: the lowest energy secondary structure of the sequence calculated with RNAfold (HOFACKER 2003) is a hairpin, the most abundant read mapped to the sequence area (i.e. putative mature miRNA) has at least ten occurrences and is located in the other arm of the hairpin, and there is strong base pairing between the mature miRNA and the opposite arm of the hairpin. We also supplemented this list with miRNAs found using miRDeep2 (FRIEDLANDER *et al.* 2008). From both search methods, all such hairpins where the mature miRNA sequence was present with at least ten reads were included and provided a final list of 248 miRNAs. These miRNAs were searched for orthologs using a similar procedure as described by Wang *et al* (WANG *et al.* 2011). First, all mature miR sequences were downloaded from miRBase release 18 (KOZOMARA and GRIFFITHS-JONES 2011) for *C. elegans*, *Caenorhabditis. briggsae*, *Caenorhabditis remanei*, *Pristionchus pacificus*, *Brugia malayi*, *Ascaris suum*,

*Drosophila melanogaster* and *Homo sapiens*, and the pairs of miRNAs that shared the 7 nt seed region (nucleotides 2-8) were searched. All these seed match miR pairs and the corresponding hairpin sequence pairs were then aligned using EMBOSS Needle with its default scoring matrix (match 5, mismatch -4, gap -10, gap extension -0.5, (http://emboss.sourceforge.net/apps/cvs/emboss/apps/needle.html). The similarity of two sequences was measured by the ratio of the alignment score over the alignment length. To get the cutoff ratio for high similarity, all miRNA pairs that did not share the 7nt seed sequences were used as background, and the median value for their alignment score ratios were calculated (0.565 for mature miRNA alignment and 0.550 for hairpin alignment). Then, alignments of the matching seed miRNAs 2-fold or more above background (1.13 for mature miRNA and 1.10 for hairpin) were considered to share high sequence similarity and thus to be orthologs.

**Orthology analyses**

To study the evolution of gene families across nematodes, we used the available predicted protein datasets from WormBase release WS225 (www.wormbase.org)—*Brugia malayi*, *Caenorhabditis elegans*, *Meloidogyne hapla*, *Pristionchus pacificus*, and *Trichinella spiralis*. We also included the *Ascaris suum* and *Bursaphelenchus xylophilus* predicted proteome data sets from WormBase release WS229. For outgroup and comparative analysis we used the predicted protein datasets of the *Arabidopsis thaliana* (vGNOMON 7/9/07), *Drosophila melanogaster* (v10/30/11), *Homosapiens sapiens* (v9/7/11), *Mus musculus* (v3/4/11), *Nasonia vitripennis* (v1.2), *Saccharomyces cerevisiae* (v2/3/11), and *Tribolium castaneum* (vTcas 3.0) genome projects, obtained from the NCBI/NIH repository (ftp://ftp.ncbi.nih.gov/genomes). Version 1.4 of the OrthoMCL pipeline was used to cluster proteins into families of orthologous genes, with default settings and the BLAST parameters recommended in the OrthoMCL documentation (Li *et al.* 2003).

**Analysis of genome completeness**

Genome completeness was determined by clustering the Augustus-predicted *P.redivivus* protein set with a core set of eukaryotic proteins (CEGMA) using OrthoMCL 1.4. *P. redivivus* showed orthology with 447 out of 455 proteins within the CEGMA protein set, which translates to an estimated *P. redivivus* assembly completeness of 98.2%.

**Protein domain analyses**

To evaluate the prevalence of protein domains in the proteome of *Panagrellus redivivus* and other species, we used the hmmscan program from the latest version of HMMER (3.0) software package, which implements probabilistic profile hidden Markov models (Finn *et al.* 2011). We set our threshold *E*-value criterion at $10^{-6}$, so that no known false-positive matches would

be detected in assigning Pfam domain identities. We ran this analysis on the proteomes mentioned above and filtered out splice isoforms from the *C. elegans* proteome.

**Gene tree analyses**

Some protein families were further explored by evaluating gene trees either with whole protein sequences or by protein domain sequences. To do these analyses we aligned protein sequences using MUSCLE (EDGAR 2004). Aligned protein sequences were then evaluated by distance analysis using the JTT matrix and a subsequent Neighbor-joining tree was created using the PHYLIP software package version 3.68, using the protdist and neighbor programs, and seqboot where bootstrap values are reported (FELSENSTEIN 2005).

**Supporting Figures**

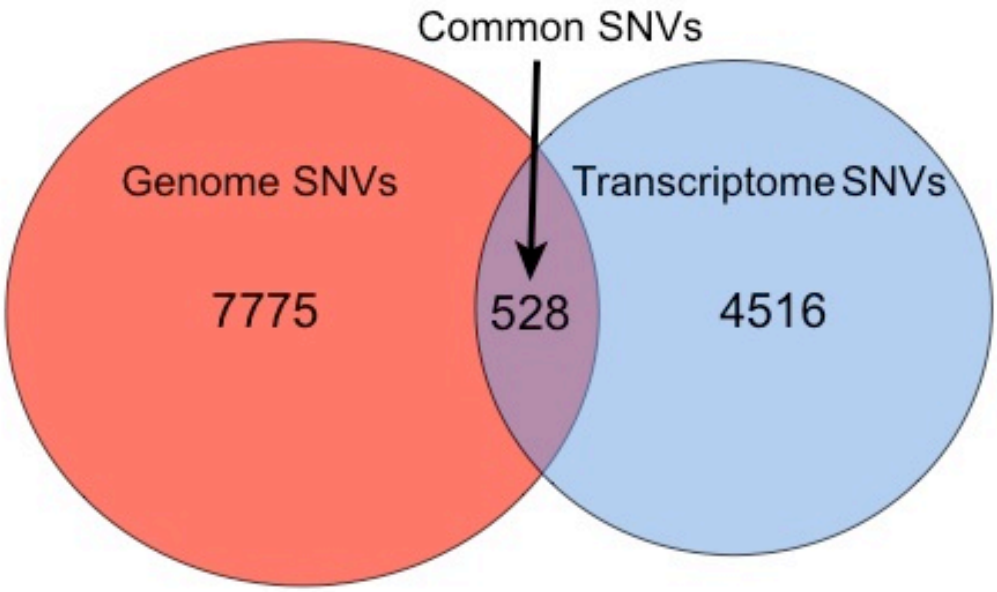


**Figure S1** Analysis of SNVs called separately from the genome and from the transcriptome shows minimal overlap.

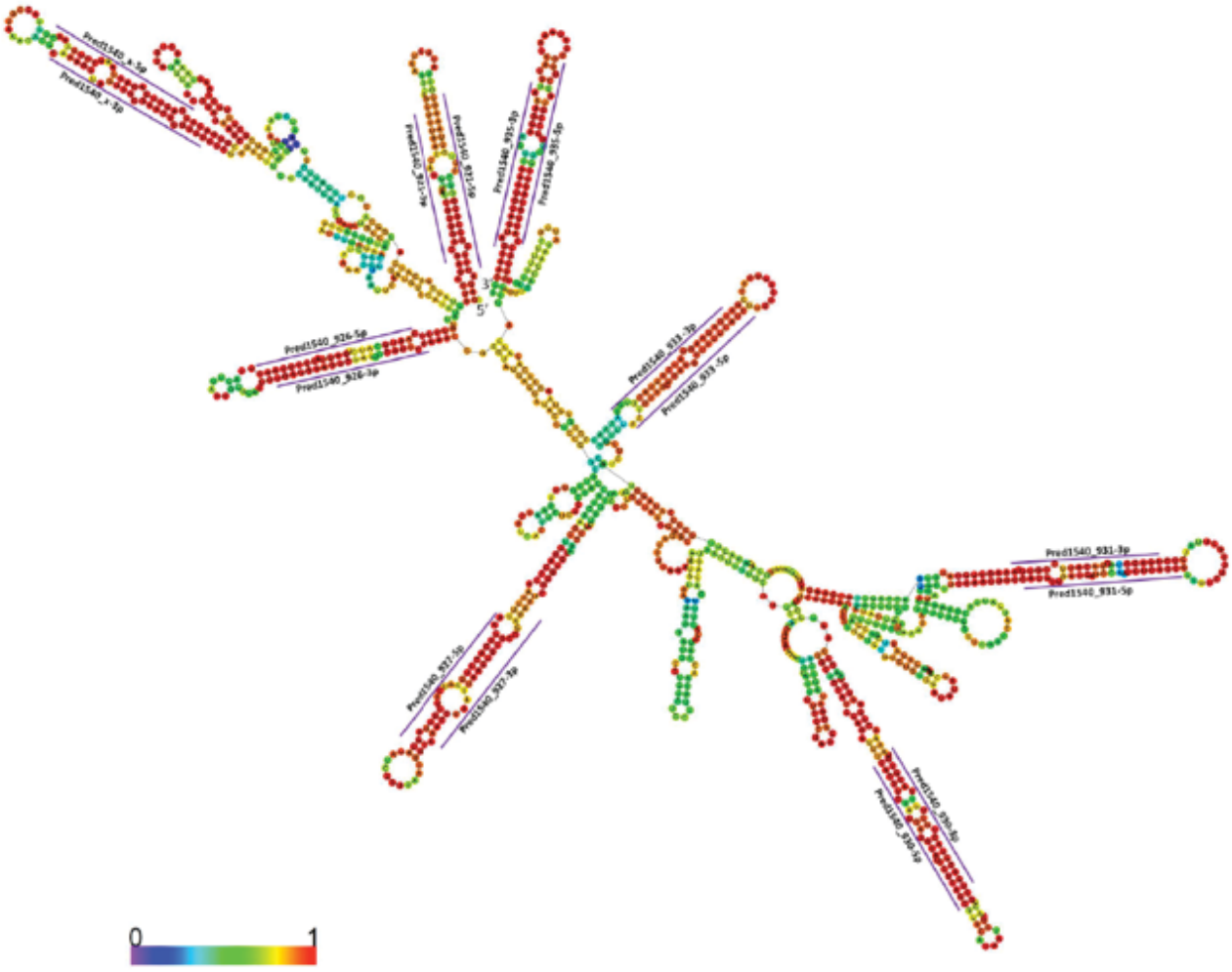


**Figure S2** The predicted minimum free energy (MFE) secondary structure of an 8 miRNA cluster, located in contig Pred1540:35496-36679. The structure was drawn using RNAFold program and is colored by base-pairing probabilities scaled from 0-1. For unpaired regions the color denotes the probability of being unpaired. Black lines denote mature miRNA sequences within each hairpin. One of the eight miRNAs in the cluster, Pred1540_x, is not included to the set of predicted *P. redivivus* miRNAs because of too few supporting reads. 5' and 3' nucleotide termini are denoted.

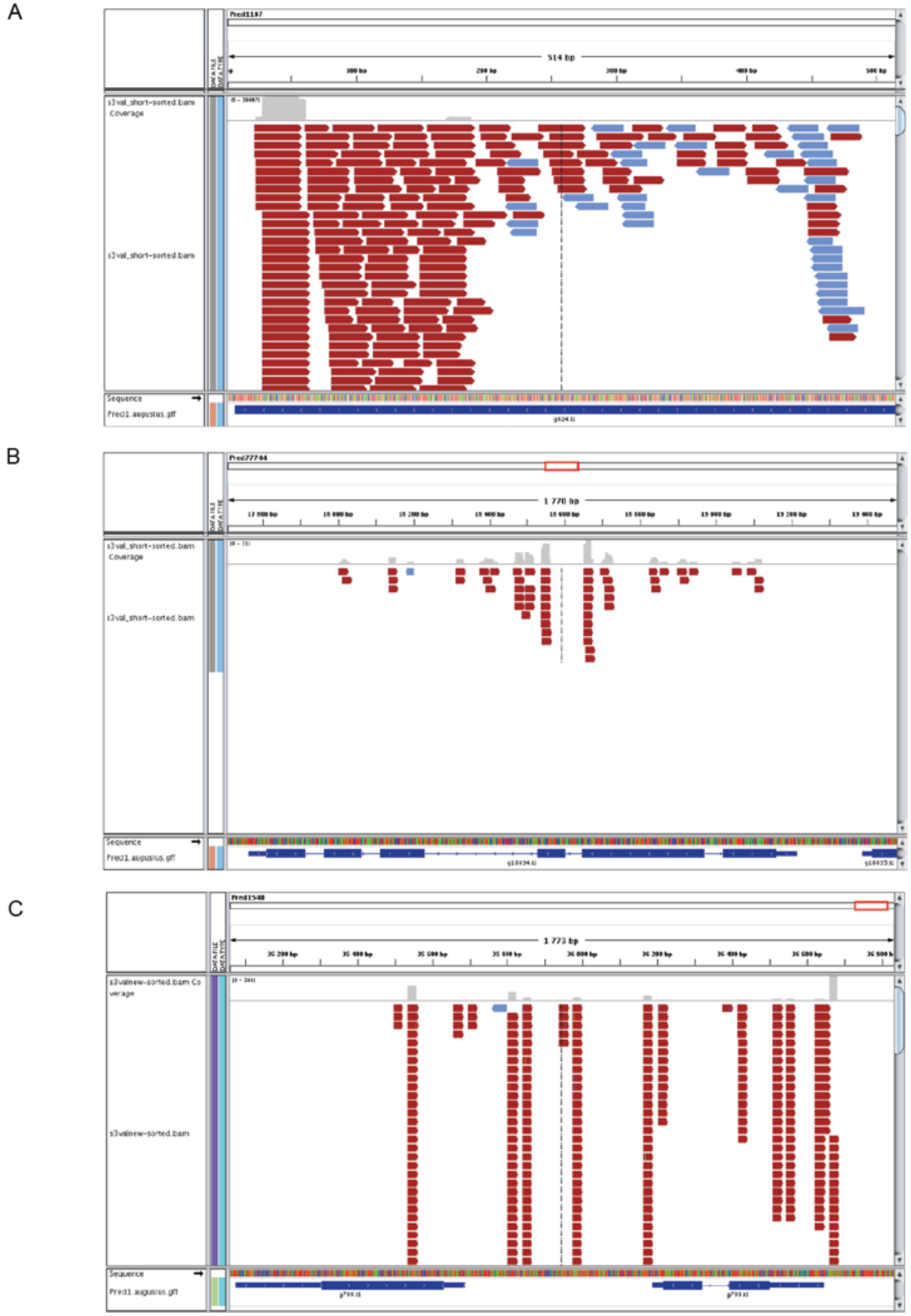


**Figure S3** Integrated Genome Viewer 2.0 displays of three different small RNA classes in *P. redivivus*. Contig names and size in bp are shown at the top. The predicted transcript from Augustus is shown at the bottom. A) Cluster showing mixture of 21U, 22G, and 26G RNA reads aligned in 5' and 3' directions and offset at small variable distances of the coding region of transcript g624.t1. B) 22G RNA reads clustered along the coding and noncoding regions of transcript 18894.t1. All reads begin with G and were 21-23 nucleotides in length. C) miRNA cluster showing eight different miRNA hairpins, each miRNA 65-70 nucleotides in

size. The cluster spans Augustus genes g788.t1 and g789.t1.The predicted secondary structure of this cluster is depicted in Figure S1).

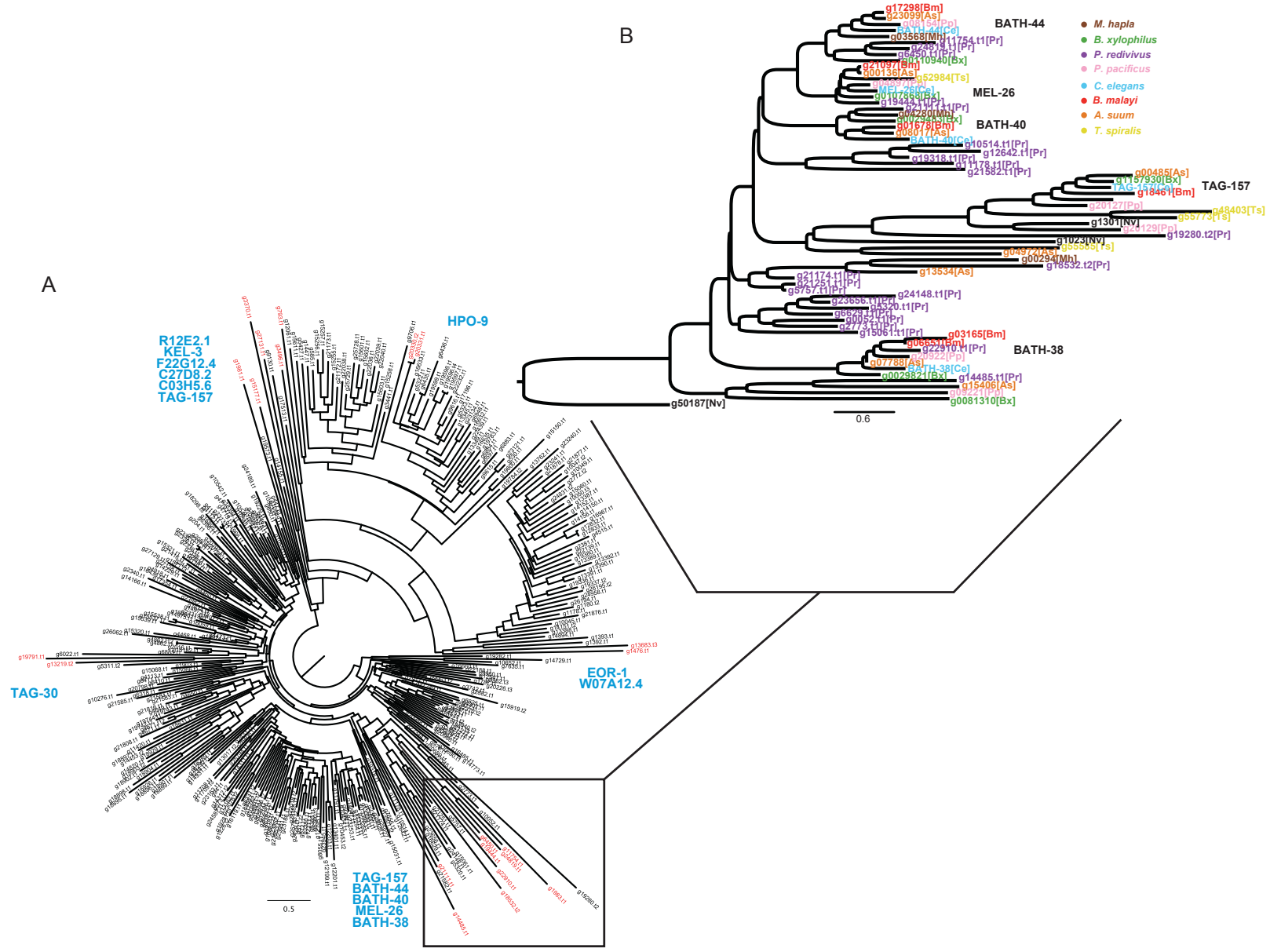


**Figure S4** Venn Diagram capturing protein clusters between *P. redivivus* and the Core Eukaryotic Gene Mapping Approach (CEGMA) protein set (PARRA *et al.* 2007).

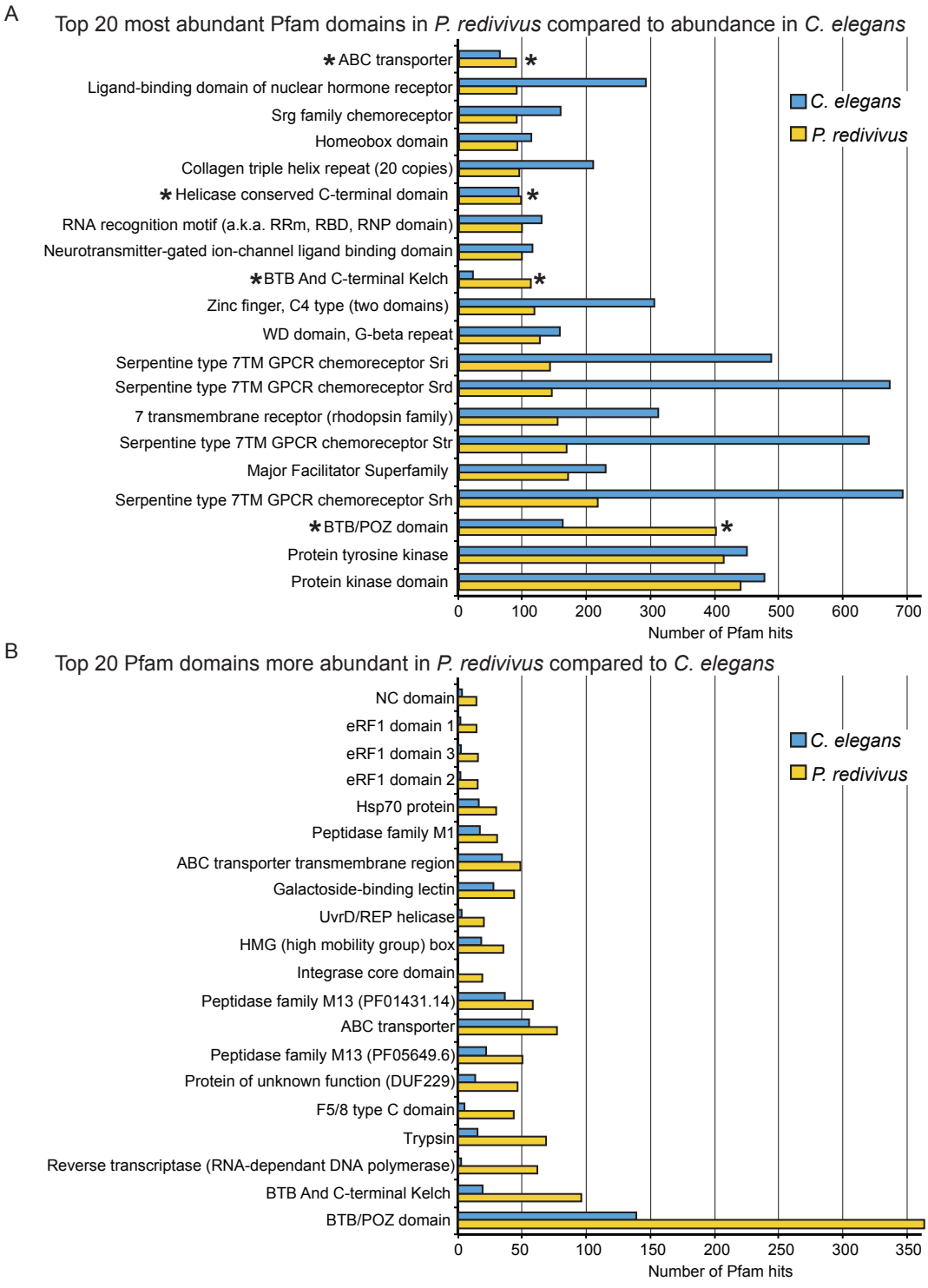


**Figure S5** Top 20 most abundant Pfam domains present in *P. redivivus* and their abundance in *C. elegans*. These genomes seem highly enriched in serpentine family domain G-protein-coupled receptors (GPCRs), though the *C. elegans* genome has a much larger complement of these protein domains. The *P. redivivus* genome is highly enriched in BTB-associated domains, ABC transporters, and several other protein families. * represents domains that are more abundant in *P. redivivus* compared to *C. elegans*.

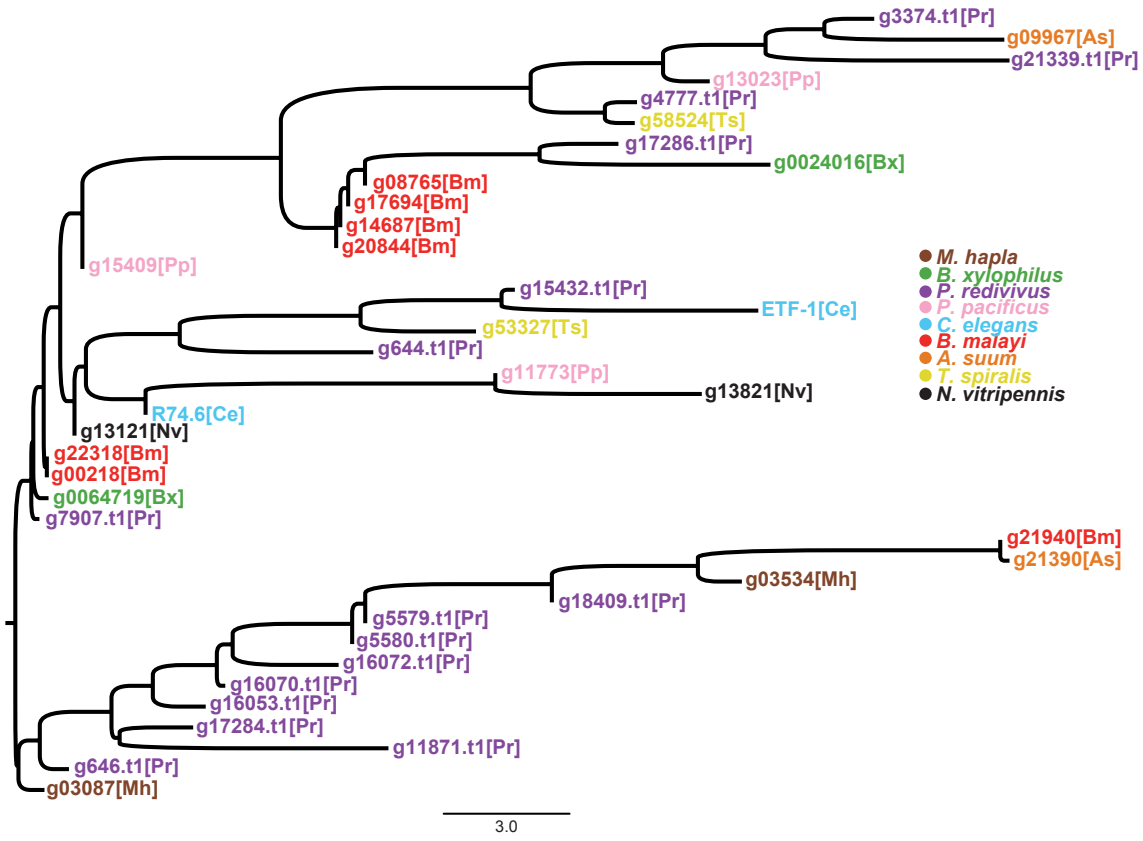


**Figure S6** Protein neighbor-joining tree of the eRF1 domain-containing proteins in *P. redivivus* and other nematodes.

**Supporting Tables**

**Table S1** RNAseq analysis of the transcriptome of *P. redivivus*.

**Table S2** Bioinformatics workflow of the miRNA-seq data and the number of obtained reads.

**Table S3** Summary of conservation of miRNAs across different species in the animal kingdom. miRNA orthologs were identified according to criteria described in methods. The species is shown in the first row and the number of *P. redivivus* orthologs identified and the total number of miRNAs for the species is shown in the second row. The number of rows and the number of orthologs may not match due to miRNA families. Only miRNA families are shown in each row. Symbols: #, 3 miRNA family; + ortholog; - ortholog not found.

**Table S4** A table of the number of proteins from each species analyzed that cluster as orthologs with the known *C. elegans* protein. 'X' indicates that no proteins from the proteome used clustered with the known *C. elegans* protein. Proteins that were lineage specific have a horizontal line drawn through all taxa except *C. elegans*. Descriptive labels for certain pathway components known in *C. elegans* are given on the far left. Results are based on an orthology analysis using the available proteomes and OrthoMCL, see supporting methods above. Protein names in brackets appear in the same orthology clusters. * Proteins that were $\leq$115 amino acids long and are not likely to be found using sequence similarity analysis. † Pseudogene in *C. elegans*, so there was no protein sequence available to use in a sequence similarity search.

**Table S5** A summary of the two orthology analyses described, showing the total number of genes analyzed in each proteome, how many of them clustered with other proteins in the analysis and how many were unclustered orphans, showing little to no homology with other proteins included in the analyses.

**Table S6** A table of the number of conserved cell death proteins from each species analyzed that cluster as orthologs with the known *C. elegans* protein. 'X' indicates that no proteins from the proteome used clustered with the known *C. elegans* protein. Proteins that were lineage specific have a horizontal line drawn through all taxa except *C. elegans*. Descriptive labels for certain pathway components known in *C. elegans* are given on the far left. Results are based on an orthology analysis using the available proteomes and OrthoMCL. (See Supporting Methods). * Proteins that were $\leq$115 amino acids long and are not likely to be found using sequence similarity analysis.

**Table S7** A table showing the conservation of the RNAi pathway in *C. elegans* and other nematodes. 'X' indicates that no proteins from the proteome used clustered with the known *C. elegans* protein. Proteins that were lineage specific have a horizontal line drawn through all taxa except *C. elegans*. Descriptive labels for certain pathway components known in *C. elegans* are given on the far left. Results are based on an orthology analysis using the available proteomes and OrthoMCL (see Methods).

**Table S8** The putative retroelement Pol genes identified by pfam using hmmscan (FINN *et al.* 2011).

**Table S9** F-box domain-containing proteins across eight nematode taxa with an insect outgroup. The table shows the presence and number of proteins in orthologous clusters across these taxa. Proteins in brackets are in the same orthology cluster (i.e. g23006.t1 and g14039.t1).

**Table S10** A table of all the ABC transporters in *C. elegans* and their orthologs in *P. redivivus*. Numbers in parentheses indicate the total number of proteins in that particular orthology cluster for that species. For instance, there are 25 *P. redivivus* orthologs and 13 *C. elegans* orthologs that show up in the large PGP cluster. There are 9 *C. elegans* ABC transporters with no apparent orthologs in *P. redivivus*, shown at the bottom of the table.

**References**

EDGAR, R. C., 2004 MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC bioinformatics 5: 113.

FELSENSTEIN, J., 2005 PHYLIP (Phylogeny Inference Package) version 3.6. , pp.

FINN, R. D., J. CLEMENTS and S. R. EDDY, 2011 HMMER web server: interactive sequence similarity searching. Nucleic acids research 39: W29-37.

FRIEDLANDER, M. R., W. CHEN, C. ADAMIDI, J. MAASKOLA, R. EINSPANIER *et al.*, 2008 Discovering microRNAs from deep sequencing data using miRDeep. Nature biotechnology 26: 407-415.

HOFACKER, I. L., 2003 Vienna RNA secondary structure server. Nucleic acids research 31: 3429-3431.

KOZOMARA, A., and S. GRIFFITHS-JONES, 2011 miRBase: integrating microRNA annotation and deep-sequencing data. Nucleic acids research 39: D152-157.

LANGMEAD, B., C. TRAPNELL, M. POP and S. L. SALZBERG, 2009 Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology 10: R25.

LASLETT, D., and B. CANBACK, 2004 ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic acids research 32: 11-16.

LI, H., B. HANDSAKER, W. A., F. T., J. RUAN *et al.*, 2009 1000 Genome Project Data Processing Subgroup. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079.

LI, L., C. J. STOECKERT, JR. and D. S. ROOS, 2003 OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome research 13: 2178-2189.

MORTAZAVI, A., E. M. SCHWARZ, B. WILLIAMS, L. SCHAEFFER, I. ANTOSHECHKIN *et al.*, 2010 Scaffolding a Caenorhabditis nematode genome with RNA-seq. Genome research 20: 1740-1747.

PARRA, G., K. BRADNAM and I. KORF, 2007 CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23: 1061-1067.

ROBERTS, A., H. PIMENTEL, C. TRAPNELL and L. PACHTER, 2011 Identification of novel transcripts in annotated genomes using RNA-Seq. Bioinformatics 27: 2325-2329.

STANKE, M., M. DIEKHANS, R. BAERTSCH and D. HAUSSLER, 2008 Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24: 637-644.

STERNBERG, P. W., and H. R. HORVITZ, 1981 Gonadal cell lineages of the nematode Panagrellus redivivus and implications for evolution by the modification of cell lineage. Developmental biology 88: 147-166.

TRAPNELL, C., L. PACHTER and S. L. SALZBERG, 2009 TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111.

TRAPNELL, C., B. A. WILLIAMS, G. PERTEA, A. MORTAZAVI, G. KWAN *et al.*, 2010 Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology 28: 511-515.

WANG, J., B. CZECH, A. CRUNK, A. WALLACE, M. MITREVA *et al.*, 2011 Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles. Genome research 21: 1462-1477.

**Supporting Data**

**BTB-domain-containing protein clusters from *P. redivivus*. Clusters are named by the *C. elegans* orthologous proteins, where present.**
**All BTB-domain-containing proteins are highlighted in yellow and clusters with many *P. redivivius* orthologs where only some are BTB proteins have the number of BTB proteins listed in brackets.**

**EOR-1**
ORTHOMCL5317(8 genes,7 taxa): BM05527(bmal) GS_14267(asum) PPA20445(ppac) WBGene00001324(cele) bxyl_g00116384(bxyl) redi_g16020.t1(zred) tspi_g48373(tspi) tspi_g60681(tspi)

**W07A12.4**
ORTHOMCL3142(9 genes,8 taxa): GS_16469(asum) PPA31007(ppac) PPA32996(ppac) WBGene00012322(cele) bxyl_g01109618(bxyl) mhap_03236(mhap) nvit_g18565(nvit) redi_g19911.t1(zred) tspi_g50234(tspi)

**TAG-147**
[1]ORTHOMCL1803(11 genes,8 taxa): BM18461(bmal) GS_00485(asum) PPA20127(ppac) PPA20129(ppac) WBGene00006493(cele) bxyl_g00579306(bxyl) nvit_g13012(nvit) redi_g6692.t1(zred) redi_g6694.t1(zred) tspi_g48403(tspi) tspi_g55773(tspi)

**ASM-3**
ORTHOMCL8495(5 genes,4 taxa): PPA24121(ppac) WBGene00000213(cele) bxyl_g0012819(bxyl) redi_g16191.t1(zred) redi_g9317.t1(zred)

**KEL-3**
ORTHOMCL1975(11 genes,8 taxa): BM03876(bmal) BM17531(bmal) GS_18704(asum) PPA00662(ppac) PPA00663(ppac) PPA13609(ppac) WBGene00002185(cele) bxyl_g01254252(bxyl) nvit_g18288(nvit) redi_g20348.t1(zred) tspi_g57230(tspi)

**R12E2.1**
ORTHOMCL5247(8 genes,7 taxa): BM06592(bmal) BM18835(bmal) GS_09496(asum) PPA26160(ppac) WBGene00020030(cele) nvit_g16033(nvit) redi_g2019.t2(zred) tspi_g59362(tspi)

**HPO-9**
ORTHOMCL1979(11 genes,8 taxa): BM03687(bmal) BM07675(bmal) BM17061(bmal) GS_13356(asum) GS_22443(asum) PPA14021(ppac) WBGene00015463(cele) bxyl_g01143104(bxyl) mhap_00974(mhap) nvit_g16480(nvit) redi_g3724.t1(zred)

**BATH-40**
ORTHOMCL8101(6 genes,6 taxa): BM01678(bmal) GS_08017(asum) WBGene00013689(cele) bxyl_g0029483(bxyl) mhap_04280(mhap) redi_g11096.t1(zred)

**BATH-38**
ORTHOMCL6829(7 genes,6 taxa): BM03165(bmal) BM06651(bmal) GS_07788(asum) PPA20922(ppac) WBGene00009945(cele) bxyl_g00298214(bxyl) redi_g10006.t1(zred)

**BATH-44**
ORTHOMCL3657(9 genes,7 taxa): BM17298(bmal) GS_23099(asum) PPA08154(ppac) WBGene00015567(cele) bxyl_g01109404(bxyl) mhap_03568(mhap) redi_g12142.t1(zred) redi_g15961.t1(zred) redi_g7567.t1(zred)

**F22G12.4**

ORTHOMCL6288(7 genes,7 taxa): BM21024(bmal) GS_19257(asum) WBGene00009064(cele) bxyl_g01078102(bxyl) nvit_g15044(nvit) redi_g17456.t1(zred) tspi_g54200(tspi)

**C03H5.6**
ORTHOMCL9113(5 genes,5 taxa): BM17469(bmal) GS_05918(asum) WBGene00015408(cele) bxyl_g00713271(bxyl) redi_g2692.t1(zred)

**MEL-26**
ORTHOMCL6278(7 genes,7 taxa): BM21097(bmal) GS_00136(asum) PPA04897(ppac) WBGene00003209(cele) bxyl_g0107868(bxyl) redi_g5208.t1(zred) tspi_g52984(tspi)

**TAG-30**
ORTHOMCL2378(10 genes,9 taxa): BM19162(bmal) GS_04235(asum) PPA02105(ppac) WBGene00006415(cele) bxyl_g01653240(bxyl) mhap_01817(mhap) nvit_g14598(nvit) nvit_g25739(nvit) redi_g4855.t1(zred) tspi_g59021(tspi)

**C27D8.2**
ORTHOMCL7650(6 genes,5 taxa): BM21778(bmal) GS_00955(asum) GS_02692(asum) PPA00736(ppac) WBGene00007778(cele) redi_g3866.t1(zred)

**ABTS-2**
ORTHOMCL8568(5 genes,4 taxa): PPA04976(ppac) WBGene00009929(cele) bxyl_g00713372(bxyl) redi_g10749.t1(zred) redi_g8306.t1(zred)

**PGP-12**
[1]ORTHOMCL79(72 genes,7 taxa): BM06293(bmal) BM17379(bmal) BM20113(bmal) GS_00985(asum) GS_01681(asum) GS_07518(asum) GS_08285(asum) GS_12341(asum) GS_19586(asum) GS_20427(asum) GS_21361(asum) GS_22685(asum) PPA03557(ppac) PPA04690(ppac) PPA07555(ppac) PPA15485(ppac) PPA16243(ppac) PPA17189(ppac) PPA17954(ppac) PPA19458(ppac) PPA24272(ppac) PPA24275(ppac) PPA25898(ppac) WBGene00003995(cele)[pgp-1] WBGene00003996(cele)[pgp-2] WBGene00003997(cele)[pgp-3] WBGene00003998(cele)[pgp-4] WBGene00003999(cele)[pgp-5] WBGene00004000(cele)[pgp-6] WBGene00004001(cele)[pgp-7] WBGene00004002(cele)[pgo-8] WBGene00004003(cele) WBGene00004005(cele) WBGene00004006(cele) WBGene00004007(cele) WBGene00004008(cele) bxyl_g00116315(bxyl) bxyl_g00116473(bxyl) bxyl_g00116844(bxyl) bxyl_g0036416(bxyl) bxyl_g0036420(bxyl) bxyl_g0050856(bxyl) bxyl_g0050857(bxyl) bxyl_g00579212(bxyl) bxyl_g01109228(bxyl) bxyl_g01109473(bxyl) nvit_g50599(nvit) redi_g10895.t1(zred) redi_g11074.t1(zred) redi_g12160.t1(zred) redi_g12794.t1(zred) redi_g14521.t1(zred) redi_g17132.t1(zred) redi_g17150.t2(zred) redi_g18108.t1(zred) redi_g18526.t1(zred) redi_g18578.t1(zred) redi_g18582.t1(zred) redi_g19718.t1(zred) redi_g19719.t1(zred) redi_g19721.t1(zred) redi_g19722.t1(zred) redi_g2208.t1(zred) redi_g22296.t1(zred) redi_g22409.t1(zred) redi_g3036.t1(zred) redi_g4627.t1(zred) redi_g5577.t1(zred) redi_g603.t1(zred) redi_g8824.t1(zred) redi_g9667.t1(zred) redi_g9732.t1(zred)

**Conserved but not in elegans**
ORTHOMCL6158(7 genes,6 taxa): GS_04972(asum) GS_13534(asum) bxyl_g007134(bxyl) mhap_00294(mhap) nvit_g10235(nvit) redi_g2054.t1(zred) tspi_g55585(tspi)

ORTHOMCL9299(5 genes,5 taxa): BM01546(bmal) GS_23132(asum) bxyl_g00579416(bxyl) mhap_00838(mhap) redi_g11577.t1(zred)

[1]ORTHOMCL4443(8 genes,3 taxa): bxyl_g00333183(bxyl) bxyl_g00422693(bxyl) bxyl_g0125434(bxyl) mhap_00665(mhap) redi_g1224.t1(redv) redi_g15878.t1(redv) redi_g15879.t1(redv) redi_g25540.t1(redv)

ORTHOMCL8748(5 genes,5 taxa): GS_15406(asum) PPA09221(ppac) bxyl_g00813105(bxyl) nvit_g50187(nvit) redi_g7463.t1(zred)

**Conserved in 1 other nematode**

ORTHOMCL14989(2 genes,2 taxa): bxyl_g01147108(bxyl) redi_g13581.t1(zred)

ORTHOMCL16699(2 genes,2 taxa): GS_24285(asum) redi_g7216.t1(zred)

**Panagrellus redivivus specific BTB proteins**

>redi_g12237.t1
NOT FOUND
>redi_g14151.t1
NOT FOUND
>redi_g20725.t1
NOT FOUND
>redi_g6316.t1
NOT FOUND
>redi_g8423.t1
NOT FOUND
>redi_g9589.t1
NOT FOUND

[4]ORTHOMCL5789(7 genes,1 taxa): redi_g16801.t1(redv) redi_g16811.t1(redv) redi_g16812.t1(redv) redi_g17701.t1(redv) redi_g17890.t1(redv) redi_g22216.t1(redv) redi_g3415.t1(redv)

ORTHOMCL13667(2 genes,1 taxa): redi_g25501.t1(redv) redi_g25502.t1(redv)

ORTHOMCL11020(3 genes,1 taxa): redi_g17841.t1(redv) redi_g3588.t1(redv) redi_g3589.t1(redv)

ORTHOMCL13981(2 genes,1 taxa): redi_g15945.t1(redv) redi_g21130.t1(redv)

[5]ORTHOMCL2107(10 genes,1 taxa): redi_g10099.t1(zred) redi_g14292.t1(zred) redi_g1879.t1(zred) redi_g19692.t1(zred) redi_g19702.t1(zred) redi_g19974.t1(zred) redi_g19976.t1(zred) redi_g2064.t1(zred) redi_g5084.t1(zred) redi_g9976.t1(zred)

[56]ORTHOMCL82(66 genes,1 taxa): redi_g1245.t1(zred) redi_g1248.t1(zred) redi_g1249.t1(zred) redi_g1250.t1(zred) redi_g1252.t1(zred) redi_g1253.t1(zred) redi_g12627.t1(zred) redi_g12700.t1(zred) redi_g12701.t1(zred) redi_g12703.t1(zred) redi_g12718.t1(zred) redi_g12722.t1(zred) redi_g12723.t1(zred) redi_g12726.t1(zred) redi_g12727.t1(zred) redi_g12728.t1(zred) redi_g12729.t1(zred) redi_g12730.t1(zred) redi_g12731.t1(zred) redi_g12837.t1(zred) redi_g1291.t1(zred) redi_g14291.t1(zred) redi_g14914.t1(zred) redi_g15891.t1(zred) redi_g15892.t1(zred) redi_g15893.t1(zred) redi_g15894.t1(zred) redi_g15895.t1(zred) redi_g15898.t1(zred) redi_g15899.t1(zred) redi_g19757.t1(zred) redi_g19894.t1(zred) redi_g19896.t1(zred) redi_g2021.t1(zred) redi_g2022.t1(zred) redi_g20587.t1(zred) redi_g20588.t1(zred) redi_g20589.t1(zred) redi_g2062.t1(zred) redi_g2063.t1(zred) redi_g21121.t1(zred) redi_g21122.t1(zred) redi_g21123.t1(zred) redi_g21124.t1(zred) redi_g21127.t1(zred) redi_g21129.t1(zred) redi_g21130.t1(zred) redi_g21811.t1(zred) redi_g21812.t1(zred) redi_g21813.t2(zred) redi_g21856.t1(zred) redi_g21906.t1(zred) redi_g2655.t1(zred) redi_g2656.t1(zred) redi_g3212.t1(zred) redi_g3213.t1(zred) redi_g5225.t1(zred) redi_g5261.t1(zred) redi_g6116.t1(zred) redi_g6117.t1(zred) redi_g7060.t1(zred) redi_g767.t1(zred) redi_g768.t1(zred) redi_g771.t1(zred) redi_g773.t1(zred) redi_g9333.t1(zred)

[1]ORTHOMCL568(18 genes,1 taxa): redi_g1.t1(zred) redi_g11666.t1(zred) redi_g13515.t1(zred) redi_g1387.t1(zred) redi_g14866.t1(zred) redi_g15509.t1(zred) redi_g15885.t1(zred) redi_g16928.t1(zred) redi_g22609.t1(zred) redi_g358.t1(zred)

redi_g5001.t1(zred) redi_g5857.t1(zred) redi_g6437.t1(zred) redi_g8179.t1(zred) redi_g8251.t1(zred) redi_g865.t1(zred) redi_g885.t1(zred) redi_g91.t1(zred)

[225]ORTHOMCL7(272 genes,1 taxa): redi_g10032.t1(zred) redi_g10033.t1(zred) redi_g10034.t1(zred) redi_g10036.t1(zred) redi_g10645.t1(zred) redi_g10646.t1(zred) redi_g10649.t1(zred) redi_g10650.t1(zred) redi_g10743.t1(zred) redi_g10744.t1(zred) redi_g10745.t1(zred) redi_g10841.t1(zred) redi_g10842.t1(zred) redi_g10863.t1(zred) redi_g10864.t1(zred) redi_g10886.t1(zred) redi_g10971.t1(zred) redi_g10972.t1(zred) redi_g10973.t1(zred) redi_g1098.t1(zred) redi_g11209.t1(zred) redi_g11210.t1(zred) redi_g11211.t1(zred) redi_g11217.t1(zred) redi_g11222.t1(zred) redi_g11223.t1(zred) redi_g11290.t1(zred) redi_g11574.t2(zred) redi_g11575.t1(zred) redi_g11674.t1(zred) redi_g11862.t1(zred) redi_g11863.t1(zred) redi_g11864.t1(zred) redi_g11865.t1(zred) redi_g11866.t1(zred) redi_g11872.t1(zred) redi_g11928.t1(zred) redi_g11929.t1(zred) redi_g12250.t1(zred) redi_g12255.t1(zred) redi_g12286.t1(zred) redi_g12397.t1(zred) redi_g12410.t1(zred) redi_g12413.t2(zred) redi_g12473.t1(zred) redi_g12474.t1(zred) redi_g12476.t1(zred) redi_g12556.t1(zred) redi_g12557.t1(zred) redi_g12739.t1(zred) redi_g12745.t1(zred) redi_g12749.t1(zred) redi_g12985.t1(zred) redi_g12986.t1(zred) redi_g12997.t1(zred) redi_g12998.t1(zred) redi_g12999.t1(zred) redi_g13136.t1(zred) redi_g13140.t1(zred) redi_g13141.t1(zred) redi_g13142.t1(zred) redi_g13143.t1(zred) redi_g13144.t1(zred) redi_g13206.t1(zred) redi_g13547.t1(zred) redi_g13582.t1(zred) redi_g13583.t1(zred) redi_g13584.t1(zred) redi_g14194.t1(zred) redi_g14227.t1(zred) redi_g14228.t1(zred) redi_g14781.t1(zred) redi_g14783.t1(zred) redi_g14895.t1(zred) redi_g15009.t1(zred) redi_g1503.t1(zred) redi_g15247.t1(zred) redi_g15248.t1(zred) redi_g15446.t1(zred) redi_g15447.t1(zred) redi_g15783.t1(zred) redi_g15903.t1(zred) redi_g15904.t1(zred) redi_g16034.t1(zred) redi_g16035.t1(zred) redi_g16036.t1(zred) redi_g16048.t1(zred) redi_g16161.t1(zred) redi_g16162.t1(zred) redi_g16163.t1(zred) redi_g16164.t1(zred) redi_g16165.t1(zred) redi_g16390.t1(zred) redi_g16549.t1(zred) redi_g16550.t1(zred) redi_g16551.t1(zred) redi_g16554.t1(zred) redi_g16555.t1(zred) redi_g16557.t1(zred) redi_g16916.t1(zred) redi_g17089.t1(zred) redi_g17102.t1(zred) redi_g17105.t1(zred) redi_g1715.t1(zred) redi_g1716.t1(zred) redi_g1718.t1(zred) redi_g17486.t1(zred) redi_g17506.t1(zred) redi_g17509.t1(zred) redi_g17654.t1(zred) redi_g17655.t1(zred) redi_g17658.t1(zred) redi_g17660.t1(zred) redi_g17661.t1(zred) redi_g17671.t1(zred) redi_g17775.t1(zred) redi_g17776.t1(zred) redi_g17985.t1(zred) redi_g18066.t1(zred) redi_g18067.t1(zred) redi_g18233.t1(zred) redi_g1882.t1(zred) redi_g1884.t1(zred) redi_g1885.t1(zred) redi_g19110.t1(zred) redi_g19317.t1(zred) redi_g19388.t1(zred) redi_g19720.t1(zred) redi_g19826.t1(zred) redi_g20643.t1(zred) redi_g20869.t2(zred) redi_g20929.t1(zred) redi_g21092.t1(zred) redi_g21317.t1(zred) redi_g214.t1(zred) redi_g21441.t1(zred) redi_g216.t1(zred) redi_g21615.t1(zred) redi_g217.t1(zred) redi_g21797.t1(zred) redi_g21807.t1(zred) redi_g2196.t1(zred) redi_g22179.t1(zred) redi_g22186.t1(zred) redi_g22226.t1(zred) redi_g22243.t1(zred) redi_g22327.t1(zred) redi_g22403.t1(zred) redi_g22404.t1(zred) redi_g22531.t1(zred) redi_g22839.t1(zred) redi_g22976.t1(zred) redi_g23023.t1(zred) redi_g23024.t1(zred) redi_g23025.t1(zred) redi_g23026.t1(zred) redi_g23088.t1(zred) redi_g23089.t1(zred) redi_g23090.t1(zred) redi_g23256.t1(zred) redi_g23270.t1(zred) redi_g23271.t1(zred) redi_g23272.t1(zred) redi_g23274.t1(zred) redi_g23604.t1(zred) redi_g23605.t1(zred) redi_g23619.t1(zred) redi_g23620.t1(zred) redi_g23675.t1(zred) redi_g2371.t1(zred) redi_g23711.t1(zred) redi_g23712.t1(zred) redi_g2372.t1(zred) redi_g2373.t1(zred) redi_g23994.t1(zred) redi_g23995.t1(zred) redi_g2575.t1(zred) redi_g3074.t2(zred) redi_g3075.t1(zred) redi_g3132.t1(zred) redi_g3133.t1(zred) redi_g3178.t1(zred) redi_g3211.t1(zred) redi_g337.t1(zred) redi_g340.t1(zred) redi_g3449.t1(zred) redi_g3474.t1(zred) redi_g3475.t1(zred) redi_g3476.t1(zred) redi_g3478.t1(zred) redi_g3666.t1(zred) redi_g4134.t1(zred) redi_g4445.t1(zred) redi_g4731.t1(zred) redi_g4768.t1(zred) redi_g5002.t1(zred) redi_g5358.t1(zred) redi_g5360.t1(zred) redi_g5718.t1(zred) redi_g5922.t1(zred) redi_g6092.t1(zred) redi_g6248.t1(zred) redi_g6272.t1(zred) redi_g6286.t1(zred) redi_g6307.t1(zred) redi_g6308.t1(zred) redi_g6333.t1(zred) redi_g6334.t1(zred) redi_g6446.t1(zred) redi_g6521.t1(zred) redi_g6568.t1(zred) redi_g679.t1(zred) redi_g6876.t1(zred) redi_g6884.t1(zred) redi_g6888.t1(zred) redi_g6890.t1(zred) redi_g6891.t1(zred) redi_g6894.t1(zred) redi_g6895.t1(zred) redi_g6924.t1(zred) redi_g6926.t1(zred) redi_g6927.t1(zred) redi_g6955.t1(zred) redi_g6960.t1(zred) redi_g7315.t1(zred) redi_g7632.t1(zred) redi_g8084.t1(zred) redi_g8089.t1(zred) redi_g8090.t1(zred) redi_g8091.t1(zred) redi_g8092.t1(zred) redi_g8093.t2(zred) redi_g8094.t1(zred) redi_g8095.t1(zred) redi_g8096.t1(zred) redi_g8097.t1(zred) redi_g8287.t1(zred) redi_g8288.t1(zred) redi_g8289.t1(zred) redi_g8291.t1(zred) redi_g8292.t1(zred) redi_g8293.t1(zred) redi_g8294.t1(zred) redi_g8295.t1(zred) redi_g8301.t1(zred) redi_g8302.t1(zred) redi_g8303.t1(zred) redi_g8304.t1(zred) redi_g8474.t1(zred) redi_g8475.t1(zred) redi_g8476.t1(zred) redi_g8477.t1(zred) redi_g8539.t1(zred) redi_g8540.t1(zred) redi_g8566.t1(zred) redi_g86.t1(zred) redi_g8988.t2(zred) redi_g9.t1(zred)

redi_g9158.t1(zred) redi_g9227.t1(zred) redi_g9350.t1(zred) redi_g9351.t1(zred) redi_g9372.t1(zred) redi_g9463.t1(zred) redi_g9749.t1(zred) redi_g9840.t1(zred) redi_g9841.t1(zred) redi_g9842.t1(zred) redi_g9843.t2(zred) redi_g9844.t1(zred) redi_g9982.t1(zred) redi_g9983.t1(zred)

[37]ORTHOMCL92(59 genes,1 taxa): redi_g10933.t1(zred) redi_g11492.t1(zred) redi_g11493.t1(zred) redi_g12238.t1(zred) redi_g12793.t1(zred) redi_g12795.t1(zred) redi_g12796.t1(zred) redi_g15197.t1(zred) redi_g15929.t1(zred) redi_g15930.t1(zred) redi_g16423.t1(zred) redi_g16919.t1(zred) redi_g18454.t1(zred) redi_g19372.t1(zred) redi_g222.t1(zred) redi_g22268.t1(zred) redi_g223.t1(zred) redi_g22661.t1(zred) redi_g22662.t1(zred) redi_g22663.t1(zred) redi_g22665.t1(zred) redi_g2577.t1(zred) redi_g2578.t2(zred) redi_g2579.t1(zred) redi_g4157.t1(zred) redi_g4160.t2(zred) redi_g4161.t1(zred) redi_g4169.t1(zred) redi_g5732.t1(zred) redi_g6317.t1(zred) redi_g6318.t1(zred) redi_g6319.t1(zred) redi_g7211.t1(zred) redi_g8307.t1(zred) redi_g8308.t1(zred) redi_g8309.t1(zred) redi_g8310.t1(zred) redi_g8311.t1(zred) redi_g8312.t1(zred) redi_g8313.t1(zred) redi_g8316.t1(zred) redi_g8317.t1(zred) redi_g8419.t1(zred) redi_g8420.t1(zred) redi_g8422.t1(zred) redi_g8424.t1(zred) redi_g8425.t1(zred) redi_g8427.t1(zred) redi_g8428.t1(zred) redi_g8961.t1(zred) redi_g8963.t1(zred) redi_g9005.t1(zred) redi_g9006.t1(zred) redi_g9007.t1(zred) redi_g9008.t1(zred) redi_g9009.t1(zred) redi_g9010.t1(zred) redi_g9349.t1(zred) redi_g9584.t1(zred)

[1]ORTHOMCL24(150 genes,1 taxa): redi_g10020.t1(zred) redi_g10021.t1(zred) redi_g10022.t1(zred) redi_g10314.t1(zred) redi_g10341.t1(zred) redi_g10404.t1(zred) redi_g10405.t1(zred) redi_g10406.t1(zred) redi_g10407.t1(zred) redi_g10410.t1(zred) redi_g10411.t1(zred) redi_g10439.t1(zred) redi_g10671.t1(zred) redi_g10689.t1(zred) redi_g11036.t1(zred) redi_g11102.t1(zred) redi_g11387.t1(zred) redi_g11497.t1(zred) redi_g11498.t1(zred) redi_g1216.t1(zred) redi_g12265.t1(zred) redi_g12411.t1(zred) redi_g12540.t1(zred) redi_g12541.t1(zred) redi_g12542.t1(zred) redi_g13179.t1(zred) redi_g13696.t1(zred) redi_g1411.t1(zred) redi_g14119.t1(zred) redi_g14263.t1(zred) redi_g14907.t1(zred) redi_g15163.t1(zred) redi_g15299.t1(zred) redi_g15317.t1(zred) redi_g15318.t1(zred) redi_g15830.t1(zred) redi_g15869.t1(zred) redi_g16298.t1(zred) redi_g16349.t1(zred) redi_g16350.t1(zred) redi_g16351.t1(zred) redi_g16358.t1(zred) redi_g16359.t1(zred) redi_g16361.t1(zred) redi_g16362.t1(zred) redi_g16363.t1(zred) redi_g16364.t1(zred) redi_g16926.t1(zred) redi_g17050.t1(zred) redi_g17297.t1(zred) redi_g17314.t1(zred) redi_g17316.t1(zred) redi_g17317.t1(zred) redi_g17406.t1(zred) redi_g17407.t1(zred) redi_g17885.t1(zred) redi_g17886.t1(zred) redi_g17889.t1(zred) redi_g18023.t1(zred) redi_g1804.t1(zred) redi_g1805.t1(zred) redi_g1806.t1(zred) redi_g1807.t1(zred) redi_g1813.t1(zred) redi_g18873.t1(zred) redi_g19868.t1(zred) redi_g20406.t1(zred) redi_g2098.t1(zred) redi_g21074.t1(zred) redi_g21145.t1(zred) redi_g213.t1(zred) redi_g21343.t1(zred) redi_g21344.t1(zred) redi_g21352.t1(zred) redi_g21420.t1(zred) redi_g21454.t1(zred) redi_g2154.t1(zred) redi_g21609.t1(zred) redi_g22225.t1(zred) redi_g2224.t1(zred) redi_g2225.t1(zred) redi_g2226.t1(zred) redi_g2227.t1(zred) redi_g2229.t1(zred) redi_g22462.t1(zred) redi_g2250.t1(zred) redi_g22765.t1(zred) redi_g22768.t1(zred) redi_g22769.t2(zred) redi_g22770.t1(zred) redi_g22946.t1(zred) redi_g22984.t1(zred) redi_g23019.t1(zred) redi_g23484.t1(zred) redi_g23606.t1(zred) redi_g23607.t1(zred) redi_g23637.t1(zred) redi_g23659.t1(zred) redi_g23661.t1(zred) redi_g23682.t1(zred) redi_g23683.t1(zred) redi_g3013.t1(zred) redi_g3029.t1(zred) redi_g3033.t1(zred) redi_g3041.t1(zred) redi_g3072.t1(zred) redi_g3165.t1(zred) redi_g3167.t1(zred) redi_g3258.t1(zred) redi_g3473.t1(zred) redi_g3477.t1(zred) redi_g3479.t1(zred) redi_g353.t1(zred) redi_g3561.t1(zred) redi_g363.t1(zred) redi_g364.t1(zred) redi_g365.t1(zred) redi_g3788.t2(zred) redi_g4155.t1(zred) redi_g4164.t1(zred) redi_g4171.t1(zred) redi_g4237.t1(zred) redi_g4301.t1(zred) redi_g4302.t1(zred) redi_g4364.t1(zred) redi_g4409.t1(zred) redi_g4411.t1(zred) redi_g4434.t1(zred) redi_g4448.t1(zred) redi_g4603.t1(zred) redi_g5074.t1(zred) redi_g5808.t1(zred) redi_g5996.t1(zred) redi_g6054.t1(zred) redi_g6664.t1(zred) redi_g6850.t1(zred) redi_g7224.t1(zred) redi_g7720.t1(zred) redi_g7724.t1(zred) redi_g7730.t1(zred) redi_g7786.t1(zred) redi_g7809.t1(zred) redi_g7812.t1(zred) redi_g8330.t1(zred) redi_g8349.t1(zred) redi_g8487.t1(zred) redi_g8948.t1(zred) redi_g9488.t1(zred) redi_g9576.t1(zred) redi_g97.t1(zred)

[1]ORTHOMCL115(51 genes,1 taxa): redi_g10254.t1(redv) redi_g10271.t1(redv) redi_g10494.t1(redv) redi_g11359.t1(redv) redi_g11364.t1(redv) redi_g11784.t2(redv) redi_g11897.t1(redv) redi_g12357.t1(redv) redi_g12805.t1(redv) redi_g13572.t1(redv) redi_g13573.t1(redv) redi_g15358.t1(redv) redi_g1591.t1(redv) redi_g16048.t1(redv) redi_g16442.t1(redv) redi_g16445.t1(redv) redi_g18691.t1(redv) redi_g18693.t1(redv) redi_g18694.t1(redv) redi_g1896.t1(redv) redi_g1897.t1(redv) redi_g22495.t1(redv) redi_g22582.t1(redv) redi_g22802.t1(redv) redi_g24033.t1(redv)

redi_g24083.t1(redv) redi_g24242.t1(redv) redi_g2479.t1(redv) redi_g25450.t1(redv) redi_g25466.t1(redv)
redi_g25467.t1(redv) redi_g25468.t1(redv) redi_g25790.t1(redv) redi_g25865.t1(redv) redi_g3549.t1(redv) redi_g3964.t1(redv)
redi_g3970.t1(redv) redi_g3971.t1(redv) redi_g4139.t1(redv) redi_g5623.t1(redv) redi_g587.t1(redv) redi_g588.t1(redv)
redi_g589.t1(redv) redi_g8467.t2(redv) redi_g9077.t1(redv) redi_g910.t1(redv) redi_g9187.t1(redv) redi_g9835.t1(redv)
redi_g9838.t1(redv) redi_g9839.t1(redv) redi_g9840.t1(redv)

[1]ORTHOMCL1(595 genes,1 taxa): redi_g10100.t1(redv) redi_g10120.t1(redv) redi_g10123.t1(redv) redi_g10164.t1(redv)
redi_g10165.t1(redv) redi_g10166.t1(redv) redi_g10167.t1(redv) redi_g10172.t1(redv) redi_g10177.t1(redv)
redi_g10193.t1(redv) redi_g10210.t1(redv) redi_g10211.t1(redv) redi_g10212.t1(redv) redi_g10216.t1(redv)
redi_g10244.t1(redv) redi_g10304.t1(redv) redi_g10305.t1(redv) redi_g10342.t1(redv) redi_g10343.t1(redv)
redi_g10397.t1(redv) redi_g10400.t1(redv) redi_g10437.t1(redv) redi_g10524.t1(redv) redi_g10532.t1(redv)
redi_g10533.t1(redv) redi_g10553.t1(redv) redi_g10554.t1(redv) redi_g10596.t1(redv) redi_g10801.t1(redv)
redi_g10832.t1(redv) redi_g10833.t1(redv) redi_g10834.t1(redv) redi_g10839.t1(redv) redi_g1084.t1(redv)
redi_g10898.t1(redv) redi_g10964.t1(redv) redi_g11148.t1(redv) redi_g11162.t1(redv) redi_g11204.t1(redv)
redi_g11205.t1(redv) redi_g11233.t1(redv) redi_g11252.t1(redv) redi_g11442.t2(redv) redi_g11492.t1(redv)
redi_g11509.t1(redv) redi_g11513.t1(redv) redi_g11519.t1(redv) redi_g1154.t1(redv) redi_g11548.t1(redv)
redi_g11549.t1(redv) redi_g11552.t1(redv) redi_g11553.t1(redv) redi_g11571.t1(redv) redi_g11601.t1(redv)
redi_g11660.t1(redv) redi_g11753.t1(redv) redi_g11793.t1(redv) redi_g11827.t1(redv) redi_g11891.t1(redv)
redi_g11892.t1(redv) redi_g11895.t1(redv) redi_g11900.t1(redv) redi_g11913.t1(redv) redi_g11914.t1(redv)
redi_g11919.t1(redv) redi_g11960.t1(redv) redi_g11975.t1(redv) redi_g12018.t1(redv) redi_g12019.t1(redv)
redi_g12101.t1(redv) redi_g12138.t1(redv) redi_g1216.t1(redv) redi_g12166.t1(redv) redi_g12167.t1(redv)
redi_g12168.t1(redv) redi_g12518.t1(redv) redi_g12519.t1(redv) redi_g12520.t1(redv) redi_g12521.t1(redv)
redi_g12522.t1(redv) redi_g12524.t1(redv) redi_g12526.t1(redv) redi_g12649.t1(redv) redi_g12651.t1(redv)
redi_g12694.t1(redv) redi_g12713.t1(redv) redi_g12771.t1(redv) redi_g12772.t1(redv) redi_g12773.t1(redv)
redi_g12779.t1(redv) redi_g12793.t1(redv) redi_g12876.t1(redv) redi_g13032.t1(redv) redi_g13033.t1(redv)
redi_g13119.t1(redv) redi_g13388.t1(redv) redi_g13389.t1(redv) redi_g13535.t1(redv) redi_g13584.t1(redv)
redi_g13585.t1(redv) redi_g13586.t1(redv) redi_g13595.t1(redv) redi_g13613.t1(redv) redi_g13677.t1(redv)
redi_g13678.t1(redv) redi_g1373.t1(redv) redi_g1375.t1(redv) redi_g1380.t1(redv) redi_g13924.t1(redv) redi_g13925.t1(redv)
redi_g14002.t1(redv) redi_g1408.t1(redv) redi_g1409.t1(redv) redi_g14153.t1(redv) redi_g1416.t1(redv) redi_g1417.t1(redv)
redi_g14181.t1(redv) redi_g14215.t1(redv) redi_g1434.t1(redv) redi_g14343.t1(redv) redi_g14408.t1(redv)
redi_g14415.t1(redv) redi_g14416.t1(redv) redi_g1445.t1(redv) redi_g14480.t1(redv) redi_g14481.t1(redv)
redi_g14483.t1(redv) redi_g14494.t1(redv) redi_g14535.t1(redv) redi_g14536.t1(redv) redi_g14537.t1(redv)
redi_g14538.t1(redv) redi_g14542.t1(redv) redi_g14591.t1(redv) redi_g14629.t1(redv) redi_g14630.t1(redv)
redi_g14634.t1(redv) redi_g14635.t1(redv) redi_g14636.t1(redv) redi_g14637.t1(redv) redi_g14645.t1(redv)
redi_g14648.t1(redv) redi_g14649.t1(redv) redi_g14680.t1(redv) redi_g15084.t1(redv) redi_g15091.t1(redv)
redi_g15092.t1(redv) redi_g15093.t1(redv) redi_g15119.t1(redv) redi_g15265.t1(redv) redi_g15330.t1(redv)
redi_g15331.t1(redv) redi_g15350.t1(redv) redi_g15377.t1(redv) redi_g15380.t1(redv) redi_g15484.t1(redv)
redi_g15485.t1(redv) redi_g15496.t1(redv) redi_g15744.t1(redv) redi_g15944.t1(redv) redi_g15946.t1(redv)
redi_g16113.t1(redv) redi_g16122.t1(redv) redi_g16124.t1(redv) redi_g16126.t1(redv) redi_g16127.t1(redv)
redi_g16129.t1(redv) redi_g16130.t1(redv) redi_g16169.t1(redv) redi_g16240.t1(redv) redi_g16241.t1(redv)
redi_g16293.t1(redv) redi_g16294.t1(redv) redi_g16311.t1(redv) redi_g1636.t1(redv) redi_g16529.t1(redv)
redi_g16683.t1(redv) redi_g16720.t1(redv) redi_g16793.t1(redv) redi_g17007.t1(redv) redi_g17012.t1(redv)
redi_g17014.t1(redv) redi_g17021.t1(redv) redi_g17119.t1(redv) redi_g17120.t1(redv) redi_g17202.t1(redv)
redi_g17400.t1(redv) redi_g17403.t1(redv) redi_g1744.t1(redv) redi_g1754.t1(redv) redi_g17560.t1(redv) redi_g17620.t1(redv)
redi_g17621.t1(redv) redi_g17623.t1(redv) redi_g1778.t1(redv) redi_g17781.t1(redv) redi_g17826.t1(redv)
redi_g17837.t1(redv) redi_g17839.t1(redv) redi_g17869.t1(redv) redi_g17918.t1(redv) redi_g17931.t1(redv)
redi_g17933.t1(redv) redi_g17934.t1(redv) redi_g17935.t1(redv) redi_g17936.t1(redv) redi_g17943.t1(redv)
redi_g17963.t1(redv) redi_g1800.t1(redv) redi_g18022.t1(redv) redi_g18025.t1(redv) redi_g18038.t1(redv)
redi_g18039.t1(redv) redi_g18040.t1(redv) redi_g1806.t1(redv) redi_g18088.t1(redv) redi_g18089.t1(redv)
redi_g18090.t1(redv) redi_g18092.t1(redv) redi_g18093.t1(redv) redi_g1816.t1(redv) redi_g18430.t1(redv)
redi_g18432.t1(redv) redi_g18511.t1(redv) redi_g18553.t1(redv) redi_g18554.t1(redv) redi_g18652.t1(redv)

redi_g18733.t1(redv) redi_g18740.t1(redv) redi_g18814.t1(redv) redi_g18815.t1(redv) redi_g18819.t1(redv)
redi_g18845.t1(redv) redi_g18860.t1(redv) redi_g18864.t1(redv) redi_g18873.t1(redv) redi_g19006.t1(redv)
redi_g19084.t1(redv) redi_g19095.t1(redv) redi_g19126.t1(redv) redi_g19148.t1(redv) redi_g19151.t1(redv)
redi_g19204.t1(redv) redi_g19331.t1(redv) redi_g19696.t1(redv) redi_g19766.t1(redv) redi_g19767.t1(redv)
redi_g19771.t1(redv) redi_g19814.t1(redv) redi_g19815.t1(redv) redi_g19890.t1(redv) redi_g1992.t1(redv)
redi_g19983.t1(redv) redi_g19984.t1(redv) redi_g20016.t1(redv) redi_g20017.t1(redv) redi_g20019.t1(redv)
redi_g20020.t1(redv) redi_g20021.t1(redv) redi_g20024.t1(redv) redi_g20025.t1(redv) redi_g20026.t1(redv)
redi_g20027.t1(redv) redi_g20034.t2(redv) redi_g20037.t1(redv) redi_g20038.t1(redv) redi_g20045.t1(redv)
redi_g20046.t1(redv) redi_g20086.t1(redv) redi_g20089.t1(redv) redi_g20090.t1(redv) redi_g20223.t1(redv)
redi_g20457.t1(redv) redi_g20458.t1(redv) redi_g20490.t1(redv) redi_g20519.t1(redv) redi_g20524.t1(redv)
redi_g20525.t1(redv) redi_g20531.t1(redv) redi_g20558.t1(redv) redi_g206.t1(redv) redi_g20664.t1(redv) redi_g20890.t1(redv)
redi_g20967.t1(redv) redi_g20987.t1(redv) redi_g20989.t1(redv) redi_g21056.t1(redv) redi_g21057.t1(redv)
redi_g21058.t1(redv) redi_g21066.t1(redv) redi_g21072.t1(redv) redi_g21075.t1(redv) redi_g21076.t1(redv)
redi_g21131.t1(redv) redi_g21132.t1(redv) redi_g21200.t1(redv) redi_g21216.t1(redv) redi_g21242.t1(redv)
redi_g21284.t1(redv) redi_g2132.t1(redv) redi_g215.t1(redv) redi_g2171.t1(redv) redi_g22557.t1(redv) redi_g22587.t1(redv)
redi_g2260.t1(redv) redi_g2271.t1(redv) redi_g2272.t1(redv) redi_g2278.t1(redv) redi_g22800.t1(redv) redi_g22834.t1(redv)
redi_g2285.t1(redv) redi_g22896.t1(redv) redi_g23008.t1(redv) redi_g23009.t1(redv) redi_g23017.t1(redv)
redi_g23163.t1(redv) redi_g23164.t1(redv) redi_g2341.t1(redv) redi_g2371.t1(redv) redi_g23828.t1(redv) redi_g23977.t1(redv)
redi_g24002.t1(redv) redi_g24003.t1(redv) redi_g24231.t1(redv) redi_g24248.t1(redv) redi_g24249.t1(redv)
redi_g24250.t1(redv) redi_g24256.t1(redv) redi_g24279.t1(redv) redi_g24331.t1(redv) redi_g24370.t1(redv)
redi_g24372.t1(redv) redi_g24373.t1(redv) redi_g24374.t1(redv) redi_g24375.t1(redv) redi_g24657.t1(redv)
redi_g24658.t1(redv) redi_g24660.t1(redv) redi_g2474.t1(redv) redi_g248.t1(redv) redi_g24892.t1(redv) redi_g24893.t1(redv)
redi_g249.t1(redv) redi_g24920.t1(redv) redi_g24930.t1(redv) redi_g24940.t1(redv) redi_g24941.t1(redv) redi_g24942.t1(redv)
redi_g24948.t1(redv) redi_g24949.t1(redv) redi_g24950.t1(redv) redi_g24951.t1(redv) redi_g24956.t1(redv)
redi_g25123.t1(redv) redi_g25124.t1(redv) redi_g25302.t1(redv) redi_g25303.t1(redv) redi_g25304.t1(redv)
redi_g25306.t1(redv) redi_g25313.t1(redv) redi_g25314.t1(redv) redi_g25315.t1(redv) redi_g25316.t1(redv)
redi_g25317.t1(redv) redi_g25318.t1(redv) redi_g25329.t1(redv) redi_g25357.t1(redv) redi_g25358.t1(redv)
redi_g25359.t1(redv) redi_g25360.t1(redv) redi_g25361.t1(redv) redi_g25374.t1(redv) redi_g25375.t1(redv)
redi_g25376.t1(redv) redi_g25422.t1(redv) redi_g25423.t1(redv) redi_g25499.t1(redv) redi_g25552.t1(redv)
redi_g25558.t1(redv) redi_g25559.t1(redv) redi_g25638.t1(redv) redi_g25639.t1(redv) redi_g25644.t1(redv)
redi_g25708.t1(redv) redi_g25776.t1(redv) redi_g25777.t1(redv) redi_g25778.t1(redv) redi_g25884.t1(redv)
redi_g25885.t1(redv) redi_g25886.t1(redv) redi_g25910.t1(redv) redi_g25941.t1(redv) redi_g25944.t1(redv)
redi_g2617.t1(redv) redi_g2619.t1(redv) redi_g2693.t1(redv) redi_g2812.t1(redv) redi_g2815.t1(redv) redi_g2834.t1(redv)
redi_g2994.t1(redv) redi_g3045.t1(redv) redi_g3061.t1(redv) redi_g3093.t2(redv) redi_g3101.t1(redv) redi_g3159.t1(redv)
redi_g3160.t1(redv) redi_g3180.t1(redv) redi_g3369.t1(redv) redi_g3370.t1(redv) redi_g3403.t1(redv) redi_g3404.t1(redv)
redi_g3413.t1(redv) redi_g3422.t1(redv) redi_g3548.t1(redv) redi_g3569.t1(redv) redi_g3583.t1(redv) redi_g3584.t1(redv)
redi_g3592.t1(redv) redi_g3593.t1(redv) redi_g3594.t1(redv) redi_g3645.t1(redv) redi_g3721.t1(redv) redi_g3723.t1(redv)
redi_g3776.t1(redv) redi_g3827.t1(redv) redi_g3995.t1(redv) redi_g4019.t1(redv) redi_g4024.t1(redv) redi_g4025.t1(redv)
redi_g4051.t1(redv) redi_g4158.t1(redv) redi_g4160.t1(redv) redi_g4161.t1(redv) redi_g4166.t1(redv) redi_g4190.t1(redv)
redi_g4203.t1(redv) redi_g4252.t1(redv) redi_g4339.t1(redv) redi_g4340.t1(redv) redi_g4357.t1(redv) redi_g438.t1(redv)
redi_g4391.t1(redv) redi_g4393.t1(redv) redi_g4396.t1(redv) redi_g4400.t1(redv) redi_g4498.t1(redv) redi_g4536.t1(redv)
redi_g4558.t1(redv) redi_g4579.t1(redv) redi_g4611.t1(redv) redi_g4673.t1(redv) redi_g47.t1(redv) redi_g4723.t1(redv)
redi_g4724.t1(redv) redi_g4834.t1(redv) redi_g4884.t1(redv) redi_g4987.t1(redv) redi_g4991.t1(redv) redi_g5029.t1(redv)
redi_g5042.t1(redv) redi_g5065.t1(redv) redi_g5066.t1(redv) redi_g5067.t1(redv) redi_g5068.t1(redv) redi_g5085.t1(redv)
redi_g5174.t1(redv) redi_g5218.t1(redv) redi_g5236.t1(redv) redi_g5237.t1(redv) redi_g5238.t1(redv) redi_g5244.t1(redv)
redi_g5259.t1(redv) redi_g5261.t1(redv) redi_g5262.t1(redv) redi_g5432.t1(redv) redi_g5433.t1(redv) redi_g5450.t1(redv)
redi_g5498.t1(redv) redi_g5577.t1(redv) redi_g5581.t1(redv) redi_g5650.t1(redv) redi_g5851.t1(redv) redi_g5863.t1(redv)
redi_g5971.t1(redv) redi_g5983.t1(redv) redi_g6027.t1(redv) redi_g6251.t1(redv) redi_g6255.t1(redv) redi_g6256.t1(redv)
redi_g6259.t1(redv) redi_g6357.t1(redv) redi_g6358.t1(redv) redi_g6379.t1(redv) redi_g6554.t1(redv) redi_g6665.t1(redv)
redi_g6666.t1(redv) redi_g6690.t1(redv) redi_g6760.t1(redv) redi_g6761.t1(redv) redi_g6766.t1(redv) redi_g6767.t1(redv)
redi_g6768.t1(redv) redi_g6802.t1(redv) redi_g6804.t1(redv) redi_g6805.t1(redv) redi_g6886.t2(redv) redi_g6887.t1(redv)

redi_g6890.t2(redv) redi_g6895.t1(redv) redi_g6948.t1(redv) redi_g6949.t1(redv) redi_g6958.t1(redv) redi_g6999.t1(redv) redi_g7016.t1(redv) redi_g7017.t1(redv) redi_g7022.t1(redv) redi_g7024.t1(redv) redi_g7035.t1(redv) redi_g7036.t1(redv) redi_g7043.t1(redv) redi_g7060.t1(redv) redi_g7063.t1(redv) redi_g7077.t1(redv) redi_g7083.t1(redv) redi_g7091.t1(redv) redi_g7117.t1(redv) redi_g7137.t1(redv) redi_g7138.t1(redv) redi_g7144.t1(redv) redi_g7145.t1(redv) redi_g7430.t2(redv) redi_g7502.t1(redv) redi_g7503.t1(redv) redi_g7507.t1(redv) redi_g7509.t1(redv) redi_g7510.t1(redv) redi_g7511.t1(redv) redi_g7515.t1(redv) redi_g7516.t3(redv) redi_g7521.t1(redv) redi_g7522.t1(redv) redi_g7526.t1(redv) redi_g7540.t1(redv) redi_g7662.t1(redv) redi_g7794.t1(redv) redi_g7825.t1(redv) redi_g7843.t1(redv) redi_g7986.t1(redv) redi_g8026.t1(redv) redi_g8115.t1(redv) redi_g8131.t1(redv) redi_g8270.t1(redv) redi_g8271.t1(redv) redi_g8274.t1(redv) redi_g8371.t1(redv) redi_g8392.t1(redv) redi_g8462.t1(redv) redi_g8493.t1(redv) redi_g8513.t1(redv) redi_g8585.t1(redv) redi_g859.t1(redv) redi_g870.t1(redv) redi_g8757.t1(redv) redi_g8763.t1(redv) redi_g8764.t1(redv) redi_g8765.t1(redv) redi_g8771.t1(redv) redi_g8772.t1(redv) redi_g8790.t1(redv) redi_g8793.t1(redv) redi_g8797.t1(redv) redi_g8830.t1(redv) redi_g8835.t1(redv) redi_g8897.t1(redv) redi_g8915.t1(redv) redi_g8983.t1(redv) redi_g9074.t1(redv) redi_g9078.t1(redv) redi_g9079.t2(redv) redi_g91.t1(redv) redi_g9195.t1(redv) redi_g9196.t1(redv) redi_g9324.t1(redv) redi_g9371.t1(redv) redi_g938.t1(redv) redi_g9383.t1(redv) redi_g9384.t1(redv) redi_g9481.t1(redv) redi_g9545.t1(redv) redi_g9646.t1(redv) redi_g9654.t1(redv) redi_g9655.t1(redv) redi_g9709.t1(redv) redi_g9907.t1(redv) redi_g9941.t1(redv)

**ABC transporter-domain-containing protein clusters from *P. redivivus*. Clusters are named by the *C. elegans* orthologous proteins, where present.**
**All ABC transporter-domain-containing proteins are highlighted in yellow and clusters with many *P. redivivus* orthologs where only some are ABC proteins have the number of ABC proteins listed in brackets.**

**ABCE-1**
ORTHOMCL2684(10 genes,9 taxa): BM04839(bmal) GS_14232(asum) PPA10310(ppac) WBGene00012714(cele) bxyl_g0042261(bxyl) mhap_01593(mhap) nvit_g10520(nvit) redi_g3531.t1(zred) tspi_g49731(tspi) tspi_g50509(tspi)

**ABCF-1**
ORTHOMCL5025(8 genes,8 taxa): BM18452(bmal) GS_01526(asum) PPA00336(ppac) WBGene00006512(cele) bxyl_g0012827(bxyl) mhap_01403(mhap) nvit_g11610(nvit) redi_g10273.t2(zred)

**ABCF-2**
ORTHOMCL3979(9 genes,9 taxa): BM04412(bmal) GS_14282(asum) PPA08123(ppac) WBGene00012097(cele) bxyl_g00422287(bxyl) mhap_01245(mhap) nvit_g10139(nvit) redi_g13187.t1(zred) tspi_g57419(tspi)

**ABCF-3**
ORTHOMCL4134(9 genes,9 taxa): BM02573(bmal) GS_17626(asum) PPA28508(ppac) WBGene00018339(cele) bxyl_g000922(bxyl) mhap_00962(mhap) nvit_g12261(nvit) redi_g17665.t1(zred) tspi_g50642(tspi)

**ABCH-1**
ORTHOMCL7541(6 genes,6 taxa): GS_05146(asum) PPA18570(ppac) WBGene00016973(cele) bxyl_g01513292(bxyl) nvit_g10163(nvit) redi_g3293.t1(zred)

**ABCX-1**
ORTHOMCL5422(8 genes,7 taxa): BM04100(bmal) BM11298(bmal) GS_14309(asum) PPA18568(ppac) WBGene00006522(cele) bxyl_g01513293(bxyl) nvit_g10162(nvit) redi_g19728.t1(zred)

**ABT-2**
ORTHOMCL2268(10 genes,7 taxa): BM21025(bmal) GS_05132(asum) PPA07651(ppac) PPA07657(ppac) PPA23016(ppac) WBGene00000020(cele) bxyl_g0114194(bxyl) redi_g18975.t1(zred) tspi_g49537(tspi) tspi_g56075(tspi)

**ABT-4**

ORTHOMCL1659(11 genes,7 taxa): GS_10190(asum) PPA04003(ppac) PPA20763(ppac) WBGene00000022(cele) bxyl_g00351232(bxyl) nvit_g12988(nvit) nvit_g12990(nvit) nvit_g12991(nvit) nvit_g12992(nvit) redi_g11354.t1(zred) tspi_g54183(tspi)

**ABT-5,6 and CED-7**
ORTHOMCL526(19 genes,8 taxa): BM18546(bmal) GS_08484(asum) PPA00756(ppac) PPA01242(ppac) PPA05691(ppac) PPA12235(ppac) WBGene00000023(cele)[ABT-5] WBGene00000421(cele)[CED-7] WBGene00018982(cele)[ABT-6] bxyl_g00351338(bxyl) bxyl_g0064929(bxyl) bxyl_g0064934(bxyl) mhap_00090(mhap) nvit_g11824(nvit) redi_g1215.t1(zred) redi_g8869.t1(zred) redi_g9825.t1(zred) redi_g9828.t1(zred) redi_g9830.t1(zred)

**ABTM-1**
ORTHOMCL5221(8 genes,8 taxa): BM06670(bmal) GS_01865(asum) WBGene00022281(cele) bxyl_g00298237(bxyl) mhap_00906(mhap) nvit_g11654(nvit) redi_g16493.t1(zred) tspi_g54279(tspi)

**HAF Proteins**
ORTHOMCL3254(9 genes,8 taxa): BM21418(bmal) GS_09342(asum) PPA26513(ppac) WBGene00001811(cele) WBGene00001813(cele) bxyl_g011391(bxyl) mhap_00889(mhap) nvit_g16970(nvit) redi_g3308.t1(zred)

**HAF Proteins**
ORTHOMCL442(21 genes,8 taxa): BM03412(bmal) BM16450(bmal) GS_00792(asum) GS_08782(asum) GS_18912(asum) PPA00989(ppac) PPA06384(ppac) WBGene00001812(cele)[HAF-2] WBGene00001814(cele)[HAF-4] WBGene00001819(cele)[HAF-9] bxyl_g01078183(bxyl) bxyl_g0114153(bxyl) bxyl_g01147131(bxyl) bxyl_g01147133(bxyl) mhap_00811(mhap) mhap_02189(mhap) redi_g10515.t1(zred) redi_g18432.t1(zred) redi_g39.t1(zred) redi_g40.t1(zred) tspi_g54367(tspi)

**HAF-6**
ORTHOMCL10050(4 genes,4 taxa): GS_24168(asum) WBGene00001816(cele) bxyl_g005061(bxyl) redi_g20293.t1(zred)

**HMT-1**
ORTHOMCL1652(11 genes,7 taxa): GS_16376(asum) PPA14422(ppac) WBGene00001815(cele) bxyl_g0081362(bxyl) nvit_g13574(nvit) redi_g12612.t1(zred) redi_g21191.t1(zred) redi_g8087.t1(zred) redi_g9675.t1(zred) tspi_g49171(tspi) tspi_g55152(tspi)

**MRP Proteins (MRP-1 is heavy metal resistance)**
ORTHOMCL90(60 genes,9 taxa): BM07007(bmal) BM08831(bmal) BM20194(bmal) BM20195(bmal) GS_06310(asum) GS_07037(asum) GS_08473(asum) GS_08708(asum) GS_20097(asum) PPA06331(ppac) PPA06907(ppac) PPA07998(ppac) PPA17668(ppac) PPA20574(ppac) PPA20782(ppac) PPA24297(ppac) PPA25269(ppac) WBGene00003407(cele)[MRP-1] WBGene00003408(cele)[MRP-2] WBGene00003409(cele)[MRP-3] WBGene00003410(cele)[MRP-4] WBGene00003412(cele)[MRP-6] WBGene00003413(cele)[MRP-7] WBGene00003414(cele)[MRP-8] bxyl_g00116719(bxyl) bxyl_g0033338(bxyl) bxyl_g01109146(bxyl) bxyl_g012112(bxyl) mhap_00078(mhap) nvit_g10688(nvit) nvit_g10689(nvit) nvit_g10690(nvit) nvit_g12592(nvit) nvit_g12659(nvit) nvit_g12660(nvit) nvit_g12724(nvit) nvit_g12725(nvit) nvit_g12726(nvit) nvit_g12728(nvit) nvit_g13439(nvit) nvit_g13919(nvit) nvit_g15158(nvit) nvit_g16516(nvit) nvit_g16518(nvit) nvit_g17018(nvit) nvit_g18185(nvit) nvit_g18247(nvit) nvit_g18755(nvit) nvit_g50086(nvit) nvit_g50115(nvit) redi_g13918.t1(zred) redi_g14752.t1(zred) redi_g17310.t1(zred) redi_g18857.t1(zred) redi_g19999.t1(zred) redi_g2347.t1(zred) redi_g585.t3(zred) redi_g7823.t1(zred) tspi_g53736(tspi) tspi_g53848(tspi)

**MRP-5**
ORTHOMCL3573(9 genes,7 taxa): BM18536(bmal) GS_12380(asum) GS_19177(asum) PPA26346(ppac) PPA26347(ppac) WBGene00003411(cele) bxyl_g00579607(bxyl) mhap_00108(mhap) redi_g2536.t1(zred)

**PGP proteins**

ORTHOMCL77(72 genes,7 taxa): BM06293(bmal) BM17379(bmal) BM20113(bmal) GS_00985(asum) GS_01681(asum) GS_07518(asum) GS_08285(asum) GS_12341(asum) GS_19586(asum) GS_20427(asum) GS_21361(asum) GS_22685(asum) PPA03557(ppac) PPA04690(ppac) PPA07555(ppac) PPA15485(ppac) PPA16243(ppac) PPA17189(ppac) PPA17954(ppac) PPA19458(ppac) PPA24272(ppac) PPA24275(ppac) PPA25898(ppac) WBGene00003995(cele)[PGP-1] WBGene00003996(cele)[PGP-2] WBGene00003997(cele)[PGP-3] WBGene00003998(cele)[PGP-4] WBGene00003999(cele)[PGP-5] WBGene00004000(cele)[PGP-6] WBGene00004001(cele)[PGP-7] WBGene00004002(cele)[PGP-8] WBGene00004003(cele)[PGP-9] WBGene00004005(cele)[PGP-11] WBGene00004006(cele)[PGP-12] WBGene00004007(cele)[PGP-13] WBGene00004008(cele)[PGP-14] bxyl_g00116315(bxyl) bxyl_g00116473(bxyl) bxyl_g00116844(bxyl) bxyl_g0036416(bxyl) bxyl_g0036420(bxyl) bxyl_g0050856(bxyl) bxyl_g0050857(bxyl) bxyl_g00579212(bxyl) bxyl_g01109228(bxyl) bxyl_g01109473(bxyl) nvit_g50599(nvit) redi_g10895.t1(zred) redi_g11074.t1(zred) redi_g12160.t1(zred) redi_g12794.t1(zred) redi_g14521.t1(zred) redi_g17132.t1(zred) redi_g17150.t2(zred) redi_g18108.t1(zred) redi_g18526.t1(zred) redi_g18578.t1(zred) redi_g18582.t1(zred) redi_g19718.t1(zred) redi_g19719.t1(zred) redi_g19721.t1(zred) redi_g19722.t1(zred) redi_g2208.t1(zred) redi_g22296.t1(zred) redi_g22409.t1(zred) redi_g3036.t1(zred) redi_g4627.t1(zred) redi_g5577.t1(zred) redi_g603.t1(zred) redi_g8824.t1(zred) redi_g9667.t1(zred) redi_g9732.t1(zred)

**PGP-10**

ORTHOMCL2008(11 genes,7 taxa): BM02582(bmal) BM18136(bmal) GS_10294(asum) GS_15393(asum) PPA23730(ppac) PPA23731(ppac) WBGene00004004(cele) bxyl_g0066918(bxyl) bxyl_g0066920(bxyl) mhap_00343(mhap) redi_g13159.t1(zred)

**PMP-1,2**

ORTHOMCL4559(8 genes,6 taxa): GS_11334(asum) PPA25170(ppac) PPA25171(ppac) WBGene00004058(cele) WBGene00004059(cele) bxyl_g0013986(bxyl) nvit_g16903(nvit) redi_g18747.t1(zred)

**PMP-3**

ORTHOMCL6833(7 genes,6 taxa): BM02653(bmal) BM16426(bmal) GS_16618(asum) WBGene00004060(cele) mhap_02728(mhap) redi_g19415.t1(zred) tspi_g60592(tspi)

**PMP-4**

ORTHOMCL6559(7 genes,7 taxa): BM14130(bmal) PPA11598(ppac) WBGene00004061(cele) bxyl_g01109477(bxyl) nvit_g10630(nvit) redi_g20447.t1(zred) tspi_g60591(tspi)

**PMP-5**

ORTHOMCL7610(6 genes,4 taxa): GS_00403(asum) GS_09393(asum) PPA02112(ppac) WBGene00004062(cele) redi_g17130.t1(zred) redi_g7104.t1(zred)

**WHT Proteins**

ORTHOMCL371(23 genes,7 taxa): GS_05172(asum) GS_10626(asum) PPA08267(ppac) PPA19948(ppac) WBGene00007513(cele)[WHT-2] WBGene00008950(cele)[WHT-5] WBGene00012925(cele)[WHT-8] WBGene00015479(cele)[WHT-1] WBGene00021535(cele)[WHT-7] bxyl_g003584(bxyl) bxyl_g00579332(bxyl) bxyl_g01109480(bxyl) mhap_01182(mhap) mhap_01826(mhap) nvit_g10033(nvit) nvit_g10034(nvit) nvit_g10274(nvit) nvit_g10275(nvit) nvit_g11347(nvit) nvit_g14844(nvit) nvit_g14845(nvit) nvit_g16927(nvit) redi_g18408.t1(zred) redi_g7599.t1(zred)

**WHT-4**

ORTHOMCL9827(4 genes,4 taxa): PPA28021(ppac) WBGene00017179(cele) bxyl_g01653286(bxyl) redi_g412.t1(zred)

ORTHOMCL714(16 genes,5 taxa): GS_20348(asum) PPA13995(ppac) bxyl_g01143263(bxyl) mhap_01534(mhap) mhap_05610(mhap) mhap_05988(mhap) mhap_06232(mhap) mhap_07324(mhap) mhap_07511(mhap) mhap_08821(mhap) mhap_09531(mhap) mhap_10850(mhap) mhap_11706(mhap) mhap_11769(mhap) redi_g1891.t1(zred) redi_g1893.t1(zred)

ORTHOMCL1824(11 genes,6 taxa): BM08553(bmal) BM08606(bmal) BM13755(bmal) BM20612(bmal) GS_05613(asum) GS_11174(asum) WBGene00012925(cele) bxyl_g0125442(bxyl) bxyl_g01513108(bxyl) nvit_g13209(nvit) redi_g15295.t1(zred)

**P.redivivus lineage-specific ABC transporters:**

[4]ORTHOMCL8241(5 genes,1 taxa): redi_g10477.t1(zred) redi_g10479.t1(zred) redi_g12361.t1(zred) redi_g1270.t1(zred) redi_g6735.t1(zred)

[6]ORTHOMCL4398(8 genes,1 taxa): redi_g11832.t1(zred) redi_g11833.t1(zred) redi_g11984.t1(zred) redi_g1560.t1(zred) redi_g1741.t1(zred) redi_g1742.t1(zred) redi_g1743.t1(zred) redi_g1744.t1(zred)

ORTHOMCL9461(4 genes,1 taxa): redi_g10476.t1(zred) redi_g1271.t1(zred) redi_g6731.t1(zred) redi_g7849.t1(zred)

ORTHOMCL13505(2 genes,1 taxa): redi_g5928.t1(zred) redi_g5934.t1(zred)

ORTHOMCL13566(2 genes,1 taxa): redi_g24041.t1(zred) redi_g3372.t1(zred)

ORTHOMCL13758(2 genes,1 taxa): redi_g16939.t1(zred) redi_g7195.t1(zred)

Orphans:
redi_g9733.t1
redi_g9720.t1
redi_g7109.t1
redi_g7108.t1
redi_g5575.t1
redi_g4346.t1
redi_g2514.t1
redi_g1936.t1
redi_g15828.t1

**File S2**

**Supporting data**

Available for download as an Excel file at http://www.genetics.org/lookup/suppl/doi:10.1534/genetics.112.148809/-/DC1.

**Table S1 Analysis of the transcriptome of *P. redivivus***

| Transcriptome | |
|---|---|
| Assembly size | 21.3 Mb |
| # N bases | 0 Mb |
| GC content | 48.07% |
| N50 | 2.1 kb |
| Total Contig Nr | 18199 |
| Contigs > N50 | 3039 |
| Max Contig size | 15.8 kb |

**Table S2 Bioinformatics workflow of the miRNA-seq data and the number of obtained reads**

| | |
|---|---|
| # reads after base calling | 24161842 |
| # trimmed reads without primer dimers, *E. coli* and tRNA | |
| 10-28 nt reads | 749677 |
| 29-38 nt reads | 23231669 |
| # of 10-28 nt reads with exact map to *P. redivivus* genome | |
| one location | 659159 |
| 2-10 locations | 14505 |
| >10 locations | 450 |
| map to predicted miRNA hairpins | 619026 |
| 22G RNA reads (21-23nt, start with G) | 5795 |
| 21U RNA reads (20-22nt, start with U) | 2086 |
| 26G RNA reads (25-27nt, start with G) | 2137 |
| # of 29-38 nt reads with exact map to *P. redivivus* genome | |
| one location | 21544839 |
| 2-10 locations | 55972 |
| >10 locations | 1727 |
| map to predicted miRNA hairpins | 1829 |

**Table S3 Summary of conservation of miRNAs across different species in the animal kingdom**

| *P.redivivus* | *C. elegans* | *C. briggsae* | *C. remanei* | *P. pacificus* | *B. malayi* | *A. suum* | *D. melanogaster* | *H. sapiens* |
|---|---|---|---|---|---|---|---|---|
| **miRNAs** | **46/223=20%** | **28/140=20%** | **29/109=27%** | **20/124=16%** | **20/32=63%** | **50/97=52%** | **31/240=13%** | **42/1527=3%** |
| pre-let-7 | + | + | + | + | + | + | + | + |
| pre-lin-4 | + | + | + | + | + | + | + | + |
| pre-miR-1 | + | + | + | + | - | + | + | + |
| pre-miR-124 | + | + | + | + | + | + | + | + |
| pre-miR-1834 | + | + | - | - | - | + | + | - |
| pre-miR-2 | + | - | - | - | + | + | - | - |
| pre-miR137 | + | + | - | + | + | + | - | + |
| pre-miR-235 | + | + | + | - | + | + | + | + |
| pre-miR-236 | + | + | + | + | + | + | + | + |
| pre-miR-239 | + | - | + | - | - | + | + | - |
| pre-miR-240 | + | - | - | - | - | - | - | + |
| pre-miR-242 | + | - | - | - | - | - | - | - |
| pre-miR-252 | + | + | + | + | - | + | - | - |
| pre-miR-255 | + | + | - | - | - | - | - | - |
| pre-miR-34 | + | + | + | - | + | + | + | + |
| pre-miR-35 | + | - | - | - | - | + | - | - |
| pre-miR-353# | + | - | - | - | - | - | - | - |
| pre-miR-360 | + | - | - | - | - | - | - | - |
| pre-miR-37 | + | + | + | - | - | + | - | - |
| pre-miR-39 | + | - | - | + | - | - | - | - |
| pre-miR-40 | + | - | - | + | - | + | - | - |
| pre-miR-44 | + | + | + | + | + | + | + | - |
| pre-miR-46 | + | + | + | + | + | + | + | - |
| pre-miR-4809 | + | - | - | - | - | - | - | - |
| pre-miR-4816 | + | - | - | - | - | - | - | + |
| pre-miR-49 | + | + | + | - | - | + | + | + |
| pre-miR-50 | + | + | + | - | + | + | + | + |
| pre-miR-51 | + | - | + | - | + | + | + | + |
| pre-miR-60 | + | + | + | - | - | - | - | - |
| pre-miR-61 | + | + | + | + | + | + | + | - |
| pre-miR-67 | + | + | + | - | - | + | + | - |
| pre-miR-71 | + | + | + | + | + | + | - | - |
| pre-miR-72 | + | - | + | + | + | + | + | + |
| pre-miR-79 | + | + | + | + | + | + | + | + |
| pre-miR-792 | + | - | - | - | - | - | - | - |
| pre-miR-81 | + | - | - | - | - | + | + | - |
| pre-miR-86 | + | + | + | + | - | + | - | - |
| pre-miR-87 | + | + | + | + | + | + | + | - |
| Pred1272_802 | - | - | - | - | - | + | - | - |
| Pred15450_6732 | - | - | - | - | - | + | + | - |
| Pred71781_20019 | - | - | - | - | - | + | + | - |
| Pred133452_28604 | - | - | - | - | + | + | - | - |
| Pred98015_24626 | - | - | - | - | - | + | + | + |
| Pred81850_21914 | - | + | - | - | - | + | - | - |
| Pred69418_19558 | - | + | - | + | - | + | - | - |
| Pred5043_2650 | - | - | - | - | - | + | - | - |
| Pred66491_18964 | - | - | - | - | - | + | - | - |
| Pred17878_7454 | - | - | - | - | - | + | - | - |
| Pred58870_17260 | - | - | - | - | - | + | - | - |
| Pred8772_4189 | - | - | - | - | - | + | + | - |
| Pred40523_13376 | - | - | - | + | - | + | - | - |
| Pred101727_25290 | - | - | + | + | + | + | + | + |
| Pred103329_25632 | - | - | - | - | - | + | - | + |
| Pred105240_25907 | - | - | - | - | - | - | + | - |
| Pred11150_5109 | - | - | - | - | - | - | + | - |
| Pred17878_7460 | - | - | - | - | - | - | - | + |
| Pred5881_3029 | - | - | - | + | - | - | + | + |
| Pred7753_3802 | - | - | - | - | - | - | - | + |
| Pred9058_4290 | - | - | - | - | - | - | - | + |
| Pred81850_21910 | - | + | + | - | - | - | - | - |

**Table S4 Conservation of Argonaute family proteins**

| *C. elegans* protein | *Cele* | *Ppac* | *Pred* | *Bxyl* | *Mhap* | *Bmal* | *Asuu* | *Tspi* | *Nvit* |
|---|---|---|---|---|---|---|---|---|---|
| ALG-1 | 3 | 2 | 3 | 4 | 2 | 4 | 3 | 4 | 3 |
| ALG-2 | 3 | 2 | 3 | 4 | 2 | 4 | 3 | 4 | 3 |
| CSR-1 | 2 | 3 | 5 | 2 | 2 | 1 | 2 | x | x |
| C04F12.1 | 2 | 3 | 5 | 2 | 2 | 1 | 2 | x | x |
| ALG-4* | 1 | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ |
| RDE-1 | 3 | 3 | 2 | 2 | x | 3 | 2 | 2 | 2 |
| PPW-1 | 3 | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ |
| SAGO-1 | 3 | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ |
| SAGO-2 | 3 | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ |
| PPW-2 | 5 | 3 | 2 | 3 | 2 | 2 | 2 | x | x |
| WAGO-4 | 5 | 3 | 2 | 3 | 2 | 2 | 2 | x | x |
| WAGO-1 | 5 | 3 | 2 | 3 | 2 | 2 | 2 | x | x |
| WAGO-2 | 5 | 3 | 2 | 3 | 2 | 2 | 2 | x | x |
| WAGO-5 | 5 | 3 | 2 | 3 | 2 | 2 | 2 | x | x |
| ERGO-1 | 1 | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ |
| PRG-1 | 2 | 1 | x | x | x | x | x | x | 4 |
| PRG-2 | 2 | 1 | x | x | x | x | x | x | 4 |
| NRDE-3 | 5 | 4 | 4 | 1 | 2 | 1 | 1 | x | x |
| WAGO-9 | 5 | 4 | 4 | 1 | 2 | 1 | 1 | x | x |
| WAGO-10 | 5 | 4 | 4 | 1 | 2 | 1 | 1 | x | x |
| WAGO-11 | 5 | 4 | 4 | 1 | 2 | 1 | 1 | x | x |
| C14B1.7 | 5 | 4 | 4 | 1 | 2 | 1 | 1 | x | x |
| ALG-3 | 1 | 1 | x | 1 | x | 1 | 1 | 141 | x |
| HPO-24 | 1 | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ |
| T23B3.2* | 1 | 1 | x | x | x | x | 1 | x | x |
| C06A1.4† | 1 | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ |

**Table S5 Cluster analysis orphan proteins**

| Species | Total proteins in analysis | Proteins clustered | Orphan proteins | % orphans |
|---|---|---|---|---|
| Cluster analysis 1 | | | | |
| *M. hapla* | 13,072 | 8,847 | 4,225 | 32.3 |
| *B. xylophilus* | 18,074 | 12,373 | 5,701 | 31.5 |
| *P. redivivus* | 24,249 | 17,415 | 6,834 | 28.2 |
| *C. elegans* | 20,426 | 15,858 | 4,568 | 22.4 |
| *P. pacificus* | 24,217 | 15,109 | 9,198 | 38.0 |
| *A. suum* | 18,842 | 10,790 | 7,752 | 41.8 |
| *B. malayi* | 21,332 | 16,061 | 5,271 | 24.7 |
| *T. spiralis* | 16,380 | 11,058 | 5,322 | 32.5 |
| *N. vitripennis* | 18,822 | 15,110 | 3,712 | 19.7 |
| Cluster analysis 2 | | | | |
| *H. sapiens* | 32,799 | 30,507 | 2,292 | 7.0 |
| *M. musculus* | 29,617 | 28,021 | 1,596 | 5.4 |
| *T. castaneum* | 16,645 | 11,939 | 4,706 | 28.3 |
| *N. vitripennis* | 18,822 | 16,123 | 2,699 | 14.3 |
| *D. melanogaster* | 24,298 | 21,425 | 2,873 | 11.8 |
| *A. thaliana* | 32,983 | 28,727 | 4,256 | 13.0 |
| *S. cerevisiae* | 5,887 | 668 | 5,219 | 88.7 |

**Table S6 Conservation of the cell death pathway**

| *C. elegans* protein | *Cele* | *Ppac* | *Pred* | *Bxyl* | *Mhap* | *Bmal* | *Asuu* | *Tspi* | *Nvit* |
|---|---|---|---|---|---|---|---|---|---|
| EGL-1* | 1 | x | x | x | x | x | x | x | x |
| Cell death | | | | | | | | | |
| CED-3 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | x | 4 |
| CED-4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| CED-9 | 1 | 1 | x | x | x | 1 | 1 | 1 | 1 |
| CED-8 | 1 | 1 | x | 1 | x | 1 | 1 | x | 1 |
| Engulfment | | | | | | | | | |
| CED-1* | 1 | x | x | x | x | x | x | x | x |
| CED-6 | 1 | x | 1 | 1 | 1 | 1 | 1 | 4 | 1 |
| CED-7 | 3 | 4 | 5 | 3 | 1 | 1 | 1 | 0 | 1 |
| CED-2 | 1 | x | 1 | x | 1 | 1 | 1 | 1 | 1 |
| CED-5 | 1 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 |
| CED-10 | 2 | 1 | 1 | 1 | x | 2 | 1 | 1 | 1 |
| CED-12 | 1 | 1 | 1 | 1 | 1 | x | 1 | 1 | 1 |

**Table S7 Conservation of RNAi pathway**

| *C. elegans* protein | *Cele* | *Ppac* | *Pred* | *Bxyl* | *Mhap* | *Bmal* | *Asuu* | *Tspi* | *Nvit* |
|---|---|---|---|---|---|---|---|---|---|
| Small RNA biosynthetic proteins | | | | | | | | | |
| DRH-3 | 1 | 2 | 3 | 1 | 1 | 1 | 3 | 3 | x |
| DRSH-1 | 1 | x | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| XPO-1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 1 |
| DCR-1 | 2 | 1 | 2 | 2 | x | 1 | 3 | 1 | 3 |
| DRH-1 | 1 | 1 | 1 | 1 | 1 | 3 | 1 | x | x |
| PASH-1 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| RDE-4 | 1 | 1 | x | 1 | x | 2 | 1 | x | 3 |
| XPO-3 | 1 | x | x | x | x | 1 | 1 | x | x |
| Amplification proteins | | | | | | | | | |
| SMG-2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 1 |
| SMG-6 | 1 | 1 | 0 | 1 | 2 | 1 | 2 | 3 | 1 |
| EGO-1 | 5 | 2 | 4 | 2 | 3 | 7 | 4 | 3 | 2 |
| RRF-1 | 5 | 2 | 4 | 2 | 3 | 7 | 4 | 3 | 2 |
| RRF-3 | 1 | 1 | x | 1 | x | 2 | 1 | 1 | x |
| SMG-5 | 1 | x | x | x | x | x | x | x | x |
| RSD-2 | 3 | x | x | x | x | x | x | x | x |
| Spreading proteins | | | | | | | | | |
| RSD-3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| SID-1 | 2 | 1 | x | x | x | 1 | 2 | 1 | 1 |
| SID-2 | 1 | x | x | x | x | x | x | x | x |
| RSD-6 | 1 | x | x | x | x | x | x | x | x |
| RISC proteins | | | | | | | | | |
| TSN-1 | 1 | 1 | 1 | 1 | x | 1 | 1 | 1 | 1 |
| AIN-1 | 1 | x | 1 | 1 | 1 | 4 | 1 | x | x |
| VIG-1 | 1 | x | x | x | x | x | x | x | x |
| AIN-2 | 1 | x | x | x | x | x | x | x | x |
| RNAi inhibitors | | | | | | | | | |
| ERI-1 | 1 | 1 | 2 | x | 1 | 1 | 1 | x | x |
| XRN-2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ADR-2 | 1 | 1 | 1 | 1 | x | x | 1 | x | 1 |
| XRN-1 | 1 | 1 | x | x | x | x | 1 | 1 | 2 |
| ADR-1* | 1 | x | x | x | x | x | x | x | x |
| ERI-5 | 1 | x | x | x | x | x | x | x | x |
| ERI-6 | 3 | x | x | x | x | x | x | x | x |
| ERI-7 | 1 | x | 3 | 1 | x | 2 | 1 | x | 1 |
| Nuclear RNAi effectors | | | | | | | | | |
| ERI-3 | 1 | x | x | x | x | x | x | x | x |
| MUT-7 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 |
| CID-1 | 1 | x | x | 1 | 2 | 1 | 1 | x | 6 |
| EKL-1 | 1 | 1 | 6 | 1 | 1 | 1 | 2 | 1 | x |
| GLF-1 | 1 | 1 | 1 | 1 | x | 2 | 1 | 1 | x |
| MES-2 | 1 | x | 2 | 1 | x | 2 | 1 | 1 | 1 |
| EKL-4 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | x | 1 |
| MES-6 | 1 | x | 2 | 1 | 1 | 1 | 1 | x | 1 |
| RHA-1 | 1 | x | 1 | 1 | x | 1 | 2 | 1 | 2 |
| EKL-6 | 1 | 1 | 1 | 1 | x | x | 2 | x | 1 |
| ZFP-1 | 1 | 2 | x | x | x | 3 | 2 | x | 1 |
| MUT-2 | 1 | 1 | x | x | x | x | x | x | x |
| EKL-5 | 1 | x | x | x | x | x | x | x | x |
| MES-3 | 1 | x | x | x | x | x | x | x | x |
| MUT-16 | 1 | x | x | x | x | x | x | x | x |
| RDE-2 | 1 | x | x | x | x | x | x | x | x |

**Table S8 Putative retroelement Pol genes**

| | |
|---|---|
| Aspartyl protease | 3 |
| Reverse transcriptase | 65 |
| Integrase core domain | 20 |
| Phage integrase | 3 |

**Table S9 Conservation of F-box domain containing proteins**

| *P. redivivus* protein (*C. elegans* ortholog name) | *Cele* | *Ppac* | *Bxyl* | *Mhap* | *Bmal* | *Asuu* | *Tspi* | *Nvit* |
|---|---|---|---|---|---|---|---|---|
| g13760.t1 (T27B4.1) | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| ~~g9130.t1~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ | ~~x~~ |
| g8702.t1 (T07E3.4) | 1 | 1 | 1 | x | x | 1 | x | x |
| g1939.t1 | x | x | x | x | 2 | 1 | x | x |
| g5542.t1 (C02F5.7) | 1 | x | 1 | 1 | 1 | 1 | 4 | 2 |
| g23006.t1 (C14B1.3) | 1 | 2 | 1 | x | 3 | 2 | x | 3 |
| g14039.t1 (C14B1.3) | 1 | 2 | 1 | x | 3 | 2 | x | 3 |

**Table S10 ABC transporter orthologs in *P. redivivus***

| *C. elegans* protein clusters with *P. redivivus* orthologs | *P. redivivus* orthologs |
|---|---|
| ABCE-1 | pred_g3531.t1 |
| ABCF-1 | pred_g10273.t2 |
| ABCF-2 | pred_g13187.t1 |
| ABCF-3 | pred_g17665.t1 |
| ABCH-1 | pred_g3293.t1 |
| ABCX-1 | pred_g19728.t1 |
| ABT-2 | pred_g18975.t1 |
| ABT-4 | pred_g11354.t1 |
| ABT-5, ABT-6, CED-7 (3) | pred_g1215.t1 pred_g8869.t1 pred_g9825.t1 (5) pred_g9828.t1 pred_g9830.t1 |
| ABTM-1 | pred_g16493.t1 |
| HAF-1,3 (2) | pred_g3308.t1 |
| HAF-2,4,9 (3) | pred_g10515.t1 pred_g18432.t1 pred_g39.t1 (4) pred_g40.t1 |
| HAF-6 | pred_g20293.t1 |
| HMT-1 | pred_g12612.t1 pred_g21191.t1 pred_g8087.t1 (4) pred_g9675.t1 |
| MRP-1,2,3,4,6,7,8 (7) | pred_g13918.t1 pred_g14752.t1 pred_g17310.t1 (8) pred_g18857.t1 pred_g19999.t1 pred_g2347.t1 pred_g585.t3 pred_g7823.t1 |
| MRP-5 | pred_g2536.t1 |
| PGP-1,2,3,4,5,6,7,8,9, 11,12,13,14 (13) | pred_g10895.t1 pred_g11074.t1 pred_g12160.t1 (25) pred_g12794.t1 pred_g14521.t1 pred_g17132.t1 pred_g17150.t2 pred_g18108.t1 pred_g18526.t1 pred_g18578.t1 pred_g18582.t1 pred_g19718.t1 pred_g19719.t1 pred_g19721.t1 pred_g19722.t1 pred_g2208.t1 pred_g22296.t1 pred_g22409.t1 pred_g3036.t1 pred_g4627.t1 pred_g5577.t1 pred_g603.t1 pred_g8824.t1 pred_g9667.t1 pred_g9732.t1 |
| PGP-10 | pred_g13159.t1 |
| PMP-1,2 (2) | pred_g18747.t1 |
| PMP-3 | pred_g19415.t1 |
| PMP-4 | pred_g20447.t1 |
| PMP-5 | pred_g17130.t1 pred_g7104.t1 (2) |
| WHT-1,2,5,7 (4) | pred_g18408.t1 pred_g7599.t1 (2) |
| WHT-4 | pred_g412.t1 |
| WHT-8 | pred_g15295.t1 |
| *C. elegans* proteins without *P. redivivus* orthologs | |
| ABT-1, ABT-3, CFT-1, HAF-7, HAF-8, PGP-15, WHT-3, WHT-6, WHT-9 | |