Comparative Genomics of Gene Loss and Gain in *Caenorhabditis* and Other Nematodes

Christian Rödelsperger

Department for Evolutionary Biology

Max Planck Institute for Developmental Biology

Spemannstr. 35, 72076 Tübingen, Germany

christian.roedelsperger@tuebingen.mpg.de

Cite this protocol as:

Rödelsperger C. (2018) Comparative Genomics of Gene Loss and Gain in *Caenorhabditis* and Other Nematodes. In: Setubal J., Stoye J., Stadler P. (eds) Comparative Genomics. Methods in Molecular Biology, vol 1704. Humana Press, New York, NY

Abstract

Nematodes, such as *Caenorhabditis elegans*, form one of the most species-rich animal phyla. By now more than 30 nematode genomes have been published allowing for comparative genomic analyses at various different time-scales. The majority of a nematode's gene repertoire is represented by either duplicated or so called orphan genes of unknown origin. This indicates the importance of mechanisms that generate new genes during the course of evolution. While it is certain, that nematodes have acquired genes by horizontal gene transfer from various donors, this process only explains a small portion of the nematode gene content. As evolutionary genomic analyses strongly support that most orphan genes are indeed protein-coding, future studies will have to decide, whether they are result from extreme divergence or evolved *de novo* from previously non-coding sequences. In this contribution, I summarize several studies investigating gene loss and gain in nematodes and discuss the strengths and weaknesses of individual approaches and data sets. These approaches can be used to ask nematode-specific questions such as associated with the evolution of parasitism or with switches in mating systems, but also can complement studies in other animal phyla like vertebrates and insects to broaden our general view on genome evolution.

Keywords: duplication, lateral gene transfer, orphan genes, parasite, genome evolution

1. Introduction

Together with insects, nematodes are the most successful animal phyla. They have invaded almost all ecological niches and have developed parasitism multiple times independently [1].

The nematode *Caenorhabditis elegans* is one of the most widely studied model organisms and its genome was the first to be sequenced from a multicellular organism [2]. The sequencing of the related nematode *Caenorhabditis briggsae* initiated various comparative genomic studies on nematodes [3,4,5,6]. But also outside the *Caenorhabditis* genus, there is great potential for further comparisons such as with the necromenic species *Pristionchus pacificus*, which has been established as a satellite model for comparative studies in various aspects of biology involving developmental plasticity [7,8], immunity [6,9] but also population and comparative genomics [10,11]. By now, more than thirty nematode genomes have been published (Fig. 1), which form a robust phylogenetic framework for evolutionary studies at many different levels and time-scales [5,10,12,13,14,15]. Their relatively small genome size [16] and species richness makes them a good system to ask various biological questions and enables sequencing of large numbers of individual strains for population genomic analysis [10,13,17,18]. Furthermore, the establishment of various reverse genetic tools in multiple species [19,20,21] allows detailed functional characterizations of individual genes and gene families [4,22]. Recent comparative and experimental studies in nematodes have shown that novel genes that have arisen either by duplication or are of unknown origin can control ecologically relevant traits in nematodes and play important roles in development [7,8,23]. Novel genes can be introduced into a genome by basically three mechanisms: duplication, horizontal gene transfer, and *de novo* gene formation. In this contribution I will discuss several approaches to investigating patterns of gene gain and loss in nematodes. Please note, that none of the described methods is nematode-specific, but I give numerous examples of their applications in nematode contexts.

2. Gene Duplications

Duplicated genes represent more than a third of typical multicellular genome [24] supporting a major role of duplication events during genome evolution. In his visionary book, Susumu Ohno outlined several alternative scenarios that could determine the fate of genes following duplication. Most prominent is the idea that duplication relaxes the sequence constraint on the ancestral gene and allows the two new copies to functionally diverge either to separate the ancestral functions (subfunctionalization) or even to develop novel functions (neofunctionalization) [25]. While empirical studies have shown accelerated evolution following duplication [26,27], the most likely scenario for the largest part of duplicate genes is that one of the copies is lost [26,28].

The long-term retention of duplicates suggests that there is a fitness cost associated with the loss of one of the duplicates [29] and multiple scenarios have been discussed in the literature that may explain the retention of duplicates with (e.g., subfunctionalization, neofunctionalization) or without (e.g., selection for higher gene dosage) diversification. There have been numerous studies revealing evidence for one or the other scenario [8,22,23,26,27,28], but it is difficult to assess which one has the largest contribution, because different scenarios often result in similar predictions [29]. For example, asymmetry [30] in the rate of evolution between the two duplicates is consistent with neofunctionalization but also with subfunctionalization. One additional complication is that duplicate genes may diversify at many different levels. Thus, while selection for higher gene dosage might have maintained a set of initial duplicates, these duplicates may later specialize or gain new functions in certain tissues or stages. Thus, the most robust characterization of gene duplications can only be made by detailed functional studies on individual gene families [22]. However, comparative genomics approaches are still very useful in describing the general patterns of sequence and expression divergence without making any strong inferences about what processes have generated these patterns.

2.1 Detecting Gene Family Expansions by Comparing Protein Domain Counts

One simple way to screen for gene family expansions is to compare the distribution of protein domain counts (e.g., PFAM) for two genomes [5,14,31]. This may reveal very drastic changes in gene family sizes such as the expansion of BTB domain-containing proteins and of the cullin scaffold proteins in *Panagrellus redivivus* [5]. However it has been shown that independent expansions and losses in both lineages can lead to an overall similar count and let a gene family appear as conserved, while in reality it undergoes substantial turnover. One very extreme case that demonstrates the missing resolution in the detection of duplications based on protein domain counts is the finding that the GST gene family has the same size (N=59) in *C. elegans* and *P. pacificus* but only one of the genes is actually preserved as one-to-one ortholog between the two nematodes [32]. Furthermore this approach has the limitation that visualization and interpretation of the data becomes increasingly difficult as soon as the comparison includes three or more species. In addition, at least three species are needed to distinguish expansions in one lineage from the deletions in the other lineage. Another disadvantage of using protein domains is that many nematode-specific gene families will be ignored if they have not yet been annotated in the PFAM database. As multiple methods for the detection of orthologs exist [33,34,35,36], similar comparisons can be done using gene family sizes as derived from orthologous clusters. This would overcome also the limitation that gene families defined by PFAM are often partially overlapping.

2.2 Phylogenetic Analysis

Reconstructing phylogenetic trees of individual gene families so far constitutes the gold standard for the detection of duplications and assigning orthology relationships. Such phylogenetic analyses usually start with homology searches by programs like BLAST to define the given gene family. This is followed by multiple sequence alignment [37,38] with simultaneous hand curation of the obtained data sets taking into account all possible sources of errors in the data (see below). The next

step is to find the best substitution model that explains the observed alignment [39], and finally the gene tree can be reconstructed [40]. Visual inspection of the phylogenetic trees might give some hints for which genes the initial BLAST searches might have missed orthologous sequences. This means that the whole process can be reiterated infinitely to continuously improve the data set and final gene tree. This kind of work depends to a large extent on somewhat arbitrary decisions that have to be made by an experienced researcher. Thus, it is invaluable to talk to experienced people to check the data. For certain gene families such as Cytochrome P450 there are even researchers such as David Nelson [41] who are willing to clean the data and to annotate Cytochrome P450 sequences for individual genome projects. Despite its labor-intensive nature and the need to deposit final data in public repositories to ensure reproducibility, detailed phylogenetic analysis on hand-curated data sets is still superior to any automated approach for ortholog detection and gene tree reconstruction. This is because the amount of error in current gene annotations is still substantial [23,42] and actually multiplies with the number of investigated species.

2.3 Assembly Artifacts and Need for Manual Curation

The large number of published nematode genomes comes at the cost that the quality of published genomes is quite heterogeneous. While the genome of *C. elegans* is probably the best available assembly among all animals (it does not contain any gaps), a number of highly fragmented draft assemblies were recently published with N50 values of only around 10 kb [43,44,45]. For clarification, the N50 value is a measure of contiguity of a genome assembly and is typically defined as the size of the smallest contig in the set of largest contigs that together make up half of the total genome size. Furthermore, only few genomes are constantly improved by incorporating new sequencing data or novel annotation pipelines. Thus, it is very important to bear in mind that genomes and consequently also gene annotations may be incomplete [42]. Especially gene absence calls are difficult because based on protein homology searches alone it is not clear whether the gene

was not annotated, not assembled, or is truly not present in a given genome. Therefore one should at least confirm by TBLASTN searches against the corresponding genome assembly and ideally also check raw sequence reads to confirm that there is really no evidence for homologous sequences in a given genome. Finally, additional support can be gained from closely related genomes of the same genus. By now, multiple nematode genera such as *Caenorhabditis*, *Pristionchus*, *Strongyloides*, and *Steinernema* have several published genomes and missing evidence in multiple independently assembled nematode genomes gives much more confidence that a gene is truly absent [46]. Furthermore it has been shown that assembly artifacts and misannotation can generate inflated gene counts and lead to errors in automated orthology detection methods [23,47]. Thus, I strongly recommend for studies involving just one or only a small number of gene families to hand curate the sequences, e.g. use conserved sequences from related nematodes to close gaps in multiple alignments [22,32]. While evidence for large gene duplications is typically quite robust, for many nematode genomes, there is still the chance that heterozygosity can lead to independently assembled haplotypes and thus could potentially lead to false signals of duplications [48]. It is important to note that the importance of this problem is quite different across nematode genomes. While most nematodes are too small to get enough DNA material for sequencing from a single individual, with exception of a few studies [45,49,50], typically populations of inbred worms are sequenced. As some species are hermaphroditic and do not suffer from inbreeding depression, inbreeding can be much more effective in these species resulting in overall better assemblies. Thus, differences in genetic diversity together with different inbreeding schemes create another layer of heterogeneity across nematode assemblies. One way to investigate whether heterozygosity can lead to additional duplications is to examine coverage profiles and the general performance of the assembly program [31]. Most assemblers employ different heuristic strategies to generate haploid assembly from diploid genomes. These strategies can lead to overcompression of repeats and also of recently duplicated sequences. If this were the trend, then this would indirectly support that the identified duplications are indeed true.

2.4 Microevolutionary Studies of Gene Loss and Gain

One advantage of nematodes as compared to vertebrates is their small genome size [16] which allows population genomic studies involving whole genome sequencing of large numbers of natural isolates [10,13,18]. As gene loss and gains are usually introduced in individual worms that are members of particular populations, it will take some time until they either become fixated or will be lost in a given population. As selection needs time to purge deleterious alleles, genome-wide patterns might differ a lot between intra-species comparisons and inter-species comparisons. Thus, we have recently seen that comparisons between the nematodes *C. elegans* and *P. pacificus* which shared a common ancestor in the order of 10^8 years ago [51], reveal strong evidence for the selection for higher gene dosage driving duplications of developmentally regulated genes [23]. In contrast, intra-species comparisons of different *P. pacificus* strains with divergence times in the order of 10^6 generations [12] showed that most duplication events do not seem to be functional in the sense that the cumulative gene dosage in strains with a given duplication is not higher than in strains carrying a single copy [28]. This was in strong contrast to deletions, which showed a strong impact on gene expression levels. In the mentioned study, we identified duplications and deletions based on differences in read coverage for Illumina resequencing data [52]. This is quite comparable to the previously employed microarray hybridization techniques [53,54,55]. Both the approaches have the disadvantages that the resolution is in the order of kilobases and breakpoints are just loosely defined, but in the case of resequencing data, other alignment-based approaches can be applied that identify breakpoints at nucleotide level resolution [56,57]. As is the case in any pairwise comparison, in the absence of an outgroup species, it is not possible to correctly interpret a deletion or duplication event relative to a reference strain as a derived event. It may well be that a recent duplication in the reference strain will be detected as a deletion in the strain of interest [7]. Thus, exploiting the species richness of nematodes, it is possible to polarize a comparison between strains of the same species with the help of genomic data for a closely related outgroup [28]. However, even without a polarized comparison, it is possible to detect hotspots of structural

variations in a given genome and to identify gene classes that are preferentially affected by such events [28,54,55].

3. Horizontal Gene Transfer

One particular specific feature of nematode genomes as opposed to most insects and vertebrates is the presence of various well-characterized cases of horizontal gene transfers (HGTs) that is the transmission of genes across species borders. The most prominent cases consist in cell wall degrading enzymes (cellulases) that have been acquired at least three times independently in nematodes. While cellulases in plant parasitic nematodes of the *Meloidogyne* lineage and in the necromenic *P. pacificus* belong to the GHF5 class [51,58], other plant parasitic nematodes of the *Bursaphelenchelus* lineage have acquired cellulases of a different class (GHF45) [59]. However, even though *Pristionchus* and *Meloidogyne* cellulases are members of the same class, they are so distantly related that it is most parsimonious to assume that they have been acquired in both lineages independently [60]. More precisely, while cellulases in the *Meloidogyne* lineage have most likely been acquired from bacteria, the most similar sequences to the *Pristionchus* cellulases are found in algae and slime molds [11,51,60]. Furthermore, the GHF45 cellulases in the *Bursaphelenchelus* lineage have been shown to be of fungal origin [59]. While horizontally transferred genes are typically identified as genes that do not show homology in any other nematode but have database hits outside nematodes [51,61], the first question to be answered is whether the identified candidates represent contamination during the sequencing process. Strong evidence for contamination is if the sequences resemble known bacterial at very high sequence identity (e.g., >98%). To further support, that horizontally transferred genes are actually part of the nematode genomes, the assembly and raw reads can be checked to show that horizontally transferred and nematode DNA are actually physically linked and are sequenced at similar depth. This can also be validated by inverse PCR. Furthermore expression evidence in the form of RNA-seq [60], qPCR [62], in situ hybridization [51], or reporter constructs and in the case of cellulases, further functional

assays [51,60] are very helpful to show that the gene has been integrated into the biology of the host organism. In addition to cellulases, two other gene families, antimicrobial peptides most likely from beetles and insect transposons were identified as horizontally acquired in nematodes of the *Pristionchus* lineage [51,61]. Furthermore, analysis of codon usage suggested that several unknown genes in *P. pacificus* could be of insect origin [61]. At that time, we hypothesized that the lack of homology was due to the absence of genomic data of a host for *Pristionchus* nematodes and that sequencing of a scarab beetle genome could potentially reveal further evidence for HGTs. However, sequencing the genome of the scarab beetle *Oryctes borbonicus* that is host of *P. pacificus* on la Réunion Island in the Indian Ocean did not reveal any additional HGT events [31]. Together with the relative small number of confirmed HGTs, this finding supports that HGT only explains a tiny portion of the nematode gene content.

4. Orphan Genes

Orphan genes have been generally defined as genes that are restricted to certain taxonomic groups [63]. Such taxonomically restricted genes have been frequently called orphan, pioneer, novel, or young genes. Depending on the definition and phylogenetic sampling up to one third of an animal's gene repertoire are frequently identified as orphan genes [63]. Three mechanisms have been suggested to explain the emergence of orphan genes, strong divergence beyond the level of identification as homologous sequence, *de novo* gene formation, and horizontal gene transfer [16,63]. Based on extensive screens for horizontal gene transfers in nematodes which only confirmed a handful of cases [14,16,31,51,58,59,63], it seems unlikely that this explains a substantial part of orphan genes. While it is very difficult to distinguish the two remaining scenarios on a genome-wide scale, another alternative explanation results from the fact that fragmented assemblies can result in extensive error in gene numbers and consequently could result in artificial orphan genes [47]. In a recent study, we addressed the question whether orphan genes are really protein-coding, artifacts, or even represent non-coding RNAs [64]. We used a simple set of

assumptions to test, which fraction of orphan genes belongs to each of these classes. First, considering every sequence that is supported by expression evidence as a real biological entity allowed us to get an upper bound for genomic artifacts. Second, every sequence that either shows direct evidence for translation based on available proteomic data or shows evidence of selection against nonsynonymous mutations can be considered a truly protein-coding gene. In the example of *P. pacificus* where 9885 orphan genes were defined based on the absence of homologs in any available genome outside the family Diplogastridae (for comparison, *C. elegans* has ~6000 orphans when defined at the genus level [16]), we could show that depending on the thresholds, 42-81% of orphan genes are transcribed and 40-77% of them are truly protein-coding. While *P. pacificus* has more than 14 published RNA-seq data sets [6,8,23], proteome evidence for *P. pacificus* is restricted to only two studies [65,66] and directly supports only 4% of orphan genes. Thus, the strongest contribution towards assigning a gene as truly protein-coding came from comparative genomic analyses. We therefore exploited the genome assembly of a closely related sister species *P. exspectatus* [18] to calculate estimates of selection on protein-coding genes for orthologous clusters [33]. Similarly, we calculated the strength of selection on individual genes between two divergent lineages within *P. pacificus [18]*, which allowed us to make a statement even for species-specific orphans. Finally, we used the idea that selection estimates can be calculated even from data of a single genome alone, if paralogous sequences are available [67]. This demonstrated that a single genome alone is enough to show negative selection in 8-10% of orphan genes, but adding population and comparative genomics data was able to increase this estimate to 40-77%. Once, the validity of orphan genes has been demonstrated, further studies can ask how old are they and how they evolve.

5. Gene Loss

The large amount of duplicated genes in present genomes [23,24] and high rates of observed duplications in experimental evolution studies [53,68] are equalized by continuous gene loss. Thus,

in a pioneering work of comparative genomics, Lynch and Connery have demonstrated that the number of duplicated genes in *C. elegans* and other genomes decreases with the age of genes as approximated by synonymous site divergence [26]. Interestingly, while duplications have frequently been assumed to be associated with adaptations to novel environments [5,51,69], there seems to be much less research on gene loss. However, one context, in which deletions are studied in nematodes, is the evolution of mating systems. In the genus *Caenorhabditis*, hermaphroditism has arisen three times independently from outcrossing ancestral species [70]. Similarly, in the *Pristionchus* genus there have been most likely five independent mating type switches [71]. With the limited genome sequencing data available, it seems that hermaphroditic genomes tend to be 20-40% smaller than the genomes of closely related outcrossing species [18,70] and it has been found that an observed segregation bias favoring smaller chromosomes in hermaphroditic species predicts this trend [72]. This would also be consistent with the finding that while in plants different genome sizes can be explained by deletions in non-coding regions and transposons, in *Caenorhabditis* all classes of genomic segments have been reduced in similar proportions [70]. In addition to segregation bias favoring smaller chromosomes [71], mating type transitions to hermaphroditism predict a relaxed constraint and potential loss of certain sex-specific genes [73,74,75]. However, empirical studies on genome size evolution in nematodes are still in their infancy, because similar to the studies on independent evolution of parasitism, studying the impact of mating system transitions on genome evolution requires sequencing sets of genomes and transcriptomes in different parts of the phylogeny.

6. Conclusion

Nematode genomes offer a range of very interesting and nematode-specific questions to ask such as “What are the genomic footprints of evolution of parasitism or mating system transitions?”. But even for more general questions, such as “What is the origin of orphan genes?” and “What drives the retention of duplicate genes?” studies in nematodes offer a complementary animal system to

support genomic trends identified in vertebrates or insects. For example, similar to recent work in beetles [31], one could ask whether certain nematode lineages (e.g., parasites) exhibit accelerated rates of gene turn over. While most nematode studies did such analysis on specific data sets such as individual gene families [32] or a specific clade of nematode species [68], a comprehensive genome-wide study of the rates of gene gain and loss is still lacking. In addition, established genetic tools and short generation times make some species excellent models for functional investigation of selected candidate genes [4,22] and experimental evolution studies [12,15,53]. Their species-richness together with their small genome size facilitates easy generation of novel comparative genomic data sets that are particularly suited to ask a specific question [69,76]. However, working with nematode genomes comes with certain costs, for example that so far there are no reliable estimates of divergence times between species, as informative fossil records are extremely scarce and approaches based on extrapolating divergence times from the per generation mutation rates suffer from the fact that the number of generations per year might vary by two orders of magnitude [12,77,78]. Further, the small size of most nematode species makes tissue-specific transcriptome profiling extremely difficult. Thus, the number of tissue-specific transcriptome studies [6,79,80] is far less than in vertebrates. Together with the fact that strong functional divergence between species [4,81,82] makes Gene Ontology term assignments for biological processes relatively unreliable, this makes the interpretation of gene sets, identified by comparative genomics approaches quite difficult. Nevertheless, I am convinced that existing and future nematode data sets form a powerful phylogenetic framework to extend our views on genome evolution.

Acknowledgments

I would like to thank Praveen Baskaran, Neel Prabh, and Gabriel Markov for helpful comments on the manuscript.

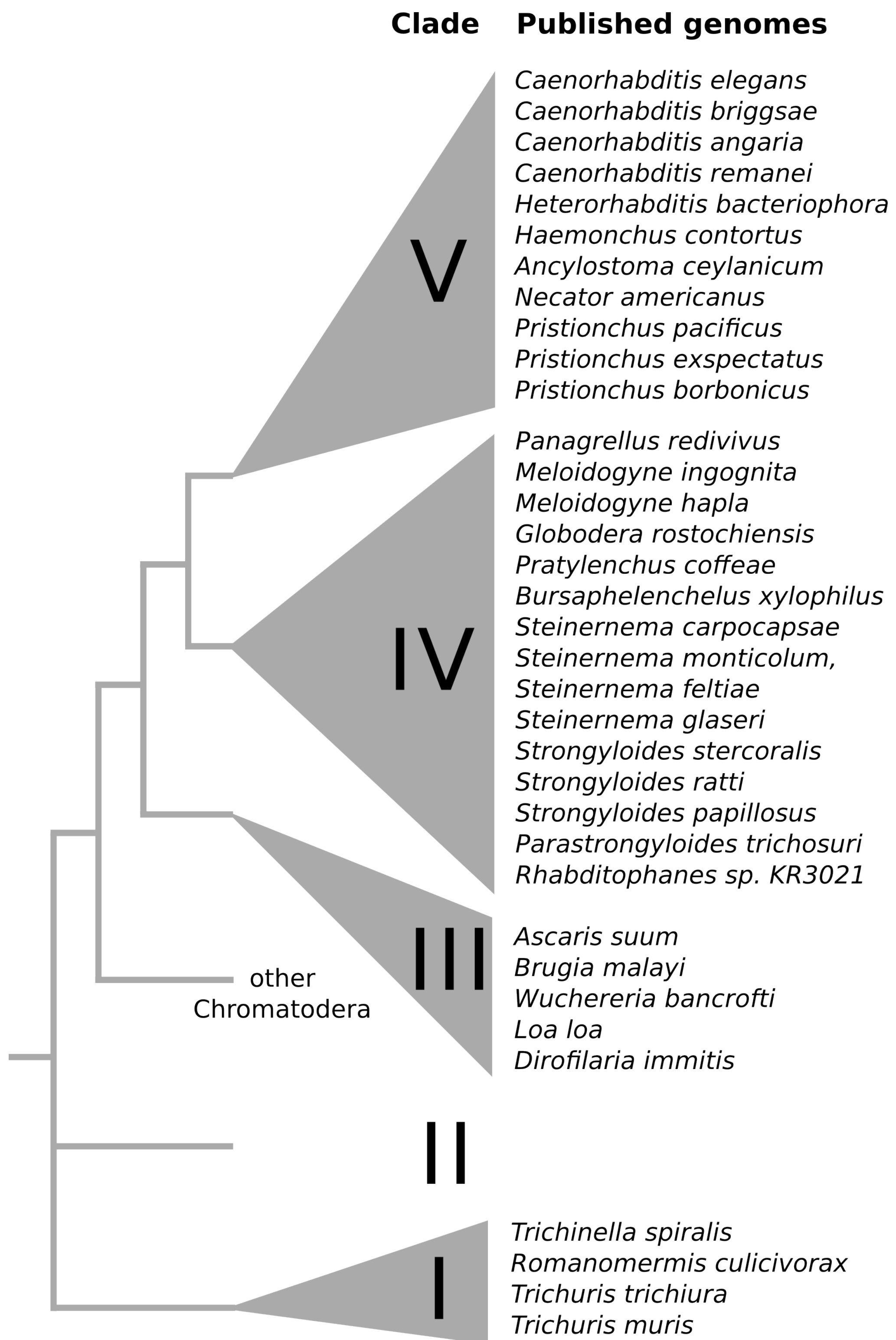

Figure 1 - Nematode phylogeny

The tree shows a schematic phylogeny of nematode clades [1] with names of species for which a draft genome assembly has been published.

References

1. Blaxter ML, De Ley P, Garey JR, Liu LX, Scheldeman P, et al. (1998) A molecular evolutionary framework for the phylum Nematoda. Nature 392: 71-75.
2. C. elegans Sequencing Consortium (1998) Genome sequence of the nematode C. elegans: a platform for investigating biology. Science 282: 2012-2018.
3. Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, et al. (2003) The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. PLoS Biol 1: E45.
4. Verster AJ, Ramani AK, McKay SJ, Fraser AG (2014) Comparative RNAi screens in C. elegans and C. briggsae reveal the impact of developmental system drift on gene function. PLoS Genet 10: e1004077.
5. Srinivasan J, Dillman AR, Macchietto MG, Heikkinen L, Lakso M, et al. (2013) The draft genome and transcriptome of Panagrellus redivivus are shaped by the harsh demands of a free-living lifestyle. Genetics 193: 1279-1295.
6. Lightfoot JW, Chauhan VM, Aylott JW, Rödelsperger C (2016) Comparative transcriptomics of the nematode gut identifies global shifts in feeding mode and pathogen susceptibility. BMC Res Notes 9: 142.
7. Mayer MG, Rödelsperger C, Witte H, Riebesell M, Sommer RJ (2015) The Orphan Gene dauerless Regulates Dauer Development and Intraspecific Competition in Nematodes by Copy Number Variation. PLoS Genet 11: e1005146.
8. Ragsdale EJ, Muller MR, Rödelsperger C, Sommer RJ (2013) A developmental switch coupled to the evolution of plasticity acts through a sulfatase. Cell 155: 922-933.
9. Rae R, Witte H, Rödelsperger C, Sommer RJ (2012) The importance of being regular: Caenorhabditis elegans and Pristionchus pacificus defecation mutants are hypersusceptible to bacterial pathogens. Int J Parasitol 42: 747-753.
10. McGaughran A, Rödelsperger C, Grimm DG, Meyer JM, Moreno E, et al. (2016) Genomic Profiles of Diversification and Genotype-Phenotype Association in Island Nematode Lineages. Mol Biol Evol.
11. Rödelsperger C, Dieterich C (2015) Comparative and functional genomics. In: Sommer RJ, editor. Pristionchus pacificus A Nematode Model for Comparative and Evolutionary Biology: Brill. pp. 141-165.
12. Weller AM, Rödelsperger C, Eberhardt G, Molnar RI, Sommer RJ (2014) Opposing forces of A/T-biased mutations and G/C-biased gene conversions shape the genome of the nematode Pristionchus pacificus. Genetics 196: 1145-1152.
13. Thomas CG, Wang W, Jovelin R, Ghosh R, Lomasko T, et al. (2015) Full-genome evolutionary histories of selfing, splitting, and selection in Caenorhabditis. Genome Res 25: 667-678.
14. Kikuchi T, Cotton JA, Dalzell JJ, Hasegawa K, Kanzaki N, et al. (2011) Genomic insights into the origin of parasitism in the emerging plant pathogen Bursaphelenchus xylophilus. PLoS Pathog 7: e1002219.
15. Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledo JI, et al. (2009) A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. Proc Natl Acad Sci U S A 106: 16310-16314.
16. Rödelsperger C, Streit A, Sommer RJ (2013) Structure, Function and Evolution of the Nematode genome. eLS. Chichester: John Wiley & Sons, Ltd.
17. Palomares-Rius JE, Tsai IJ, Karim N, Akiba M, Kato T, et al. (2015) Genome-wide variation in the pinewood nematode Bursaphelenchus xylophilus and its relationship with pathogenic

traits. BMC Genomics 16: 845.
18. Rödelsperger C, Neher RA, Weller AM, Eberhardt G, Witte H, et al. (2014) Characterization of genetic diversity in the nematode Pristionchus pacificus from population-scale resequencing data. Genetics 196: 1153-1165.
19. Lo TW, Pickle CS, Lin S, Ralston EJ, Gurling M, et al. (2013) Precise and heritable genome editing in evolutionarily diverse nematodes using TALENs and CRISPR/Cas9 to engineer insertions and deletions. Genetics 195: 331-348.
20. Tabara H, Grishok A, Mello CC (1998) RNAi in C. elegans: soaking in the genome sequence. Science 282: 430-431.
21. Witte H, Moreno E, Rödelsperger C, Kim J, Kim JS, et al. (2015) Gene inactivation using the CRISPR/Cas9 system in the nematode Pristionchus pacificus. Dev Genes Evol 225: 55-62.
22. Markov GV, Meyer JM, Panda O, Artyukhin AB, Claassen M, et al. (2016) Functional Conservation and Divergence of daf-22 Paralogs in Pristionchus pacificus Dauer Development. Mol Biol Evol.
23. Baskaran P, Rödelsperger C, Prabh N, Serobyan V, Markov GV, et al. (2015) Ancient gene duplications have shaped developmental stage-specific expression in Pristionchus pacificus. BMC Evol Biol 15: 185.
24. Wagner A (2001) Birth and death of duplicated genes in completely sequenced eukaryotes. Trends Genet 17: 237-239.
25. Ohno S (1970) Evolution by Gene Duplication. Springer.
26. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. Science 290: 1151-1155.
27. Pegueroles C, Laurie S, Alba MM (2013) Accelerated evolution after gene duplication: a time-dependent process affecting just one copy. Mol Biol Evol 30: 1830-1842.
28. Baskaran P, Rödelsperger C (2015) Microevolution of Duplications and Deletions and Their Impact on Gene Expression in the Nematode Pristionchus pacificus. PLoS One 10: e0131136.
29. Hahn MW (2009) Distinguishing among evolutionary models for the maintenance of gene duplicates. J Hered 100: 605-617.
30. Katju V (2013) To the beat of a different drum: determinants implicated in the asymmetric sequence divergence of Caenorhabditis elegans paralogs. BMC Evol Biol 13: 73.
31. Meyer JM, Markov GV, Baskaran P, Herrmann M, Sommer RJ, et al. (2016) Draft genome of the scarab beetle Oryctes borbonicus on La Reunion Island. Genome Biol Evol.
32. Markov GV, Baskaran P, Sommer RJ (2015) The same or not the same: lineage-specific gene expansions and homology relationships in multigene families in nematodes. J Mol Evol 80: 18-36.
33. Li L, Stoeckert CJ, Jr., Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res 13: 2178-2189.
34. Remm M, Storm CE, Sonnhammer EL (2001) Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol 314: 1041-1052.
35. Rödelsperger C, Dieterich C (2008) Syntenator: multiple gene order alignments with a gene-specific scoring function. Algorithms Mol Biol 3: 14.
36. Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. Science 278: 631-637.
37. Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27: 221-224.
38. Edgar RC (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5: 113.
39. Darriba D, Taboada GL, Doallo R, Posada D (2011) ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27: 1164-1165.

40. Schliep KP (2011) phangorn: phylogenetic analysis in R. Bioinformatics 27: 592-593.
41. Nelson DR (2009) The cytochrome p450 homepage. Hum Genomics 4: 59-65.
42. Gilabert A, Curran DM, Harvey SC, Wasmuth JD (2016) Expanding the view on the evolution of the nematode dauer signalling pathways: refinement through gene gain and pathway co-option. BMC Genomics 17: 476.
43. Godel C, Kumar S, Koutsovoulos G, Ludin P, Nilsson D, et al. (2012) The genome of the heartworm, Dirofilaria immitis, reveals drug and vaccine targets. FASEB J 26: 4650-4661.
44. Mortazavi A, Schwarz EM, Williams B, Schaeffer L, Antoshechkin I, et al. (2010) Scaffolding a Caenorhabditis nematode genome with RNA-seq. Genome Res 20: 1740-1747.
45. Susoy V, Herrmann M, Kanzaki N, Kruger M, Nguyen CN, et al. (2016) Large-scale diversification without genetic isolation in nematode symbionts of figs. Sci Adv 2: e1501031.
46. Rödelsperger C, Menden K, Serobyan V, Witte H, Baskaran P (2016) First insights into the nature and evolution of antisense transcription in nematodes. BMC Evol Biol 16: 165.
47. Denton JF, Lugo-Martinez J, Tucker AE, Schrider DR, Warren WC, et al. (2014) Extensive error in the number of genes inferred from draft genome assemblies. PLoS Comput Biol 10: e1003998.
48. Barriere A, Yang SP, Pekarek E, Thomas CG, Haag ES, et al. (2009) Detecting heterozygosity in shotgun genome assemblies: Lessons from obligately outcrossing nematodes. Genome Res 19: 470-480.
49. Jex AR, Liu S, Li B, Young ND, Hall RS, et al. (2011) Ascaris suum draft genome. Nature 479: 529-533.
50. Tallon LJ, Liu X, Bennuru S, Chibucos MC, Godinez A, et al. (2014) Single molecule sequencing and genome assembly of a clinical specimen of Loa loa, the causative agent of loiasis. BMC Genomics 15: 788.
51. Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, et al. (2008) The Pristionchus pacificus genome provides a unique perspective on nematode lifestyle and parasitism. Nat Genet 40: 1193-1198.
52. Xie C, Tammi MT (2009) CNV-seq, a new method to detect copy number variation using high-throughput sequencing. BMC Bioinformatics 10: 80.
53. Farslow JC, Lipinski KJ, Packard LB, Edgley ML, Taylor J, et al. (2015) Rapid Increase in frequency of gene copy-number variants during experimental evolution in Caenorhabditis elegans. BMC Genomics 16: 1044.
54. Maydan JS, Flibotte S, Edgley ML, Lau J, Selzer RR, et al. (2007) Efficient high-resolution deletion discovery in Caenorhabditis elegans by array comparative genomic hybridization. Genome Res 17: 337-347.
55. Maydan JS, Lorch A, Edgley ML, Flibotte S, Moerman DG (2010) Copy number variation in the genomes of twelve natural isolates of Caenorhabditis elegans. BMC Genomics 11: 62.
56. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z (2009) Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25: 2865-2871.
57. Emde AK, Schulz MH, Weese D, Sun R, Vingron M, et al. (2012) Detecting genomic indel variants with exact breakpoints in single- and paired-end sequencing data using SplazerS. Bioinformatics 28: 619-627.
58. Kyndt T, Haegeman A, Gheysen G (2008) Evolution of GHF5 endoglucanase gene structure in plant-parasitic nematodes: no evidence for an early domain shuffling event. BMC Evol Biol 8: 305.
59. Kikuchi T, Jones JT, Aikawa T, Kosaka H, Ogura N (2004) A family of glycosyl hydrolase family 45 cellulases from the pine wood nematode Bursaphelenchus xylophilus. FEBS Lett 572: 201-205.

60. Mayer WE, Schuster LN, Bartelmes G, Dieterich C, Sommer RJ (2011) Horizontal gene transfer of microbial cellulases into nematode genomes is associated with functional assimilation and gene turnover. BMC Evol Biol 11: 13.
61. Rödelsperger C, Sommer RJ (2011) Computational archaeology of the Pristionchus pacificus genome reveals evidence of horizontal gene transfers from insects. BMC Evol Biol 11: 239.
62. Schuster LN, Sommer RJ (2012) Expressional and functional variation of horizontally acquired cellulases in the nematode Pristionchus pacificus. Gene 506: 274-282.
63. Tautz D, Domazet-Loso T (2011) The evolutionary origin of orphan genes. Nat Rev Genet 12: 692-702.
64. Prabh N, Rödelsperger C (2016) Are orphan genes protein-coding, prediction artifacts, or non-coding RNAs? BMC Bioinformatics 17: 226.
65. Borchert N, Dieterich C, Krug K, Schutz W, Jung S, et al. (2010) Proteogenomics of Pristionchus pacificus reveals distinct proteome structure of nematode models. Genome Res 20: 837-846.
66. Borchert N, Krug K, Gnad F, Sinha A, Sommer RJ, et al. (2012) Phosphoproteome of Pristionchus pacificus provides insights into architecture of signaling networks in nematode models. Mol Cell Proteomics 11: 1631-1639.
67. Katju V, Lynch M (2003) The structure and early evolution of recently arisen gene duplicates in the Caenorhabditis elegans genome. Genetics 165: 1793-1803.
68. Lipinski KJ, Farslow JC, Fitzpatrick KA, Lynch M, Katju V, et al. (2011) High spontaneous rate of gene duplication in Caenorhabditis elegans. Curr Biol 21: 306-310.
69. Hunt VL, Tsai IJ, Coghlan A, Reid AJ, Holroyd N, et al. (2016) The genomic basis of parasitism in the Strongyloides clade of nematodes. Nat Genet 48: 299-307.
70. Fierst JL, Willis JH, Thomas CG, Wang W, Reynolds RM, et al. (2015) Reproductive Mode and the Evolution of Genome Size and Structure in Caenorhabditis Nematodes. PLoS Genet 11: e1005323.
71. Weadick CJ, Sommer RJ (2016) Mating System Transitions Drive Life Span Evolution in Pristionchus Nematodes. Am Nat 187: 517-531.
72. Wang J, Chen PJ, Wang GJ, Keller L (2010) Chromosome size differences may affect meiosis and genome size. Science 329: 293.
73. Wright AE, Mank JE (2013) The scope and strength of sex-specific selection in genome evolution. J Evol Biol 26: 1841-1853.
74. Cutter AD (2008) Reproductive evolution: symptom of a selfing syndrome. Curr Biol 18: R1056-1058.
75. Thomas CG, Woodruff GC, Haag ES (2012) Causes and consequences of the evolution of reproductive mode in Caenorhabditis nematodes. Trends Genet 28: 213-220.
76. Dillman AR, Macchietto M, Porter CF, Rogers A, Williams B, et al. (2015) Comparative genomics of Steinernema reveals deeply conserved gene regulatory networks. Genome Biol 16: 200.
77. Cutter AD (2008) Divergence times in Caenorhabditis and Drosophila inferred from direct estimates of the neutral mutation rate. Mol Biol Evol 25: 778-786.
78. Mayer MG, Sommer RJ (2011) Natural variation in Pristionchus pacificus dauer formation reveals cross-preference rather than self-preference of nematode dauer pheromones. Proc Biol Sci 278: 2784-2790.
79. Spencer WC, Zeller G, Watson JD, Henz SR, Watkins KL, et al. (2011) A spatial and temporal map of C. elegans gene expression. Genome Res 21: 325-341.
80. McGhee JD, Sleumer MC, Bilenky M, Wong K, McKay SJ, et al. (2007) The ELT-2 GATA-factor and the global regulation of transcription in the C. elegans intestine. Dev Biol 302: 627-645.
81. Sommer RJ (2012) Evolution of regulatory networks: nematode vulva induction as an example

of developmental systems drift. Adv Exp Med Biol 751: 79-91.
82. Sinha A, Sommer RJ, Dieterich C (2012) Divergent gene expression in the conserved dauer stage of the nematodes Pristionchus pacificus and Caenorhabditis elegans. BMC Genomics 13: 254.