

A Survey of Nematode SmY RNAs

Thomas A. Jones^{a*}, Wolfgang Otto^{b*}, Manja Marz^b, Sean R. Eddy^a, and Peter F. Stadler^{b,c,d,e}

^aHHMI Janelia Farm Research Campus
Ashburn VA 20147, USA;

^bBioinformatics Group, Department of Computer Science;
and Interdisciplinary Center for Bioinformatics,
University of Leipzig, Härtelstrasse 16-18, D-01407,
Leipzig, Germany;

^cInstitute for Theoretical Chemistry, University of Vienna,
Währingerstraße 17, A-1090 Wien, Austria;

^dFraunhofer Institute for Cell Therapy und Immunology,
Perlickstr.1, D-04103 Leipzig, Germany;

^eSanta Fe Institute, 1399 Hyde Park Rd.,
Santa Fe, NM 87501, USA.

*These authors contributed equally.

November 20, 2008

Abstract

SmY RNAs are a family of ~70-90 nt small nuclear RNAs found in nematodes. In *C. elegans*, SmY RNAs copurify in a small ribonucleoprotein (snRNP) complex related to the SL1 and SL2 snRNPs that are involved in nematode mRNA trans-splicing. Here we describe a comprehensive computational analysis of SmY RNA homologs found in the currently available genome sequences. We identify homologs

in all sequenced nematode genomes in class *Chromadorea*. We are unable to identify homologs in a more distantly related nematode species, *Trichinella spiralis* (class: *Dorylaimia*), and in representatives of non-nematode phyla that use trans-splicing. Using comparative RNA sequence analysis, we infer a conserved consensus SmY RNA secondary structure consisting of two stems flanking a consensus Sm protein binding site. A representative seed alignment of the SmY RNA family, annotated with the inferred consensus secondary structure, has been deposited with the Rfam RNA families database.

Introduction

The SmY RNAs are a family of small nuclear RNAs found in Nematoda species. The first SmY RNA was discovered in purified *Ascaris lumbricoides* spliceosome preparations, as well as a second RNA called SmX that is not detectably homologous to SmY [1]. Twelve SmY homologs were identified computationally in *Caenorhabditis elegans*, and ten in *Caenorhabditis briggsae* [2]. Several transcripts from these SmY genes were cloned and sequenced in a systematic survey of small noncoding RNA transcripts in *C. elegans* [3]. SmY RNAs are about 70-90 nucleotides long, with a conserved consensus binding site for the Sm protein, a shared component of spliceosomal snRNPs [1, 2]. In *C. elegans*, SmY RNAs copurify in a complex with Sm, SL75p, and SL26p proteins, while the better-characterized *C. elegans* SL1 trans-splicing snRNA copurifies in a complex with Sm, SL75p and SL21p (a paralog of SL26p) [2]. Loss of function of either SL21p or SL26p individually causes only a weak cold-sensitive phenotype, whereas knockdown of both is lethal, as is a SL75p knockdown. Based on these results, the SmY RNAs are believed to have a function in trans-splicing.

To date, SmY RNAs have been described in *C. elegans*, *C. briggsae*, and *A. lumbricoides*. The range of species possessing SmY RNAs has not been well characterized. Here we report the results of a comprehensive computational characterization of SmY RNA genes in available genome sequences.

Results

Initial SmY sequences. Thirteen identified SmY sequences are in public DNA databases: *Ascaris lumbricoides* SmY RNA [1] and twelve SmY RNAs from *Caenorhabditis elegans* [2] (Table 1). Full length 5' and 3' ends for all these sequences are experimentally determined [1–3, 5], with three exceptions. SmY-12

Table 1: Previously published SmY RNA sequences.

Name	Alternative Names	Accession numbers	length (nt)	References
<i>Ascaris lumbricoides</i>				
SmY		U52372.1	72	[1]
<i>Caenorhabditis elegans</i>				
SmY-1	CeN32, C33A12.22	AY948626.1, NR_003443.1	77	[3]
SmY-2	CeN25-1, C33A12.21 ^a , Ce135 ^a	AY948618.1, NR_003442.1 ^a , DQ789540.1 ^a	77 ^b	[3, 4]
SmY-3	CeN25-2, D1086.14 ^a , Ce96 ^a	AY948619.1, NR_003469.1 ^a , DQ789534.1 ^a	82 ^b	[3, 4]
SmY-4	D1086.16	NR_003471.1	81	CESC
SmY-5	CeN25-3, D1086.15	NR_003470.1, AY948620.1	77 ^c	CESC
SmY-6	CeN25-5	AM286190.1	83	[5]
SmY-7	Y73B6BL.46	NR_003463.1	82 ^d	CESC
SmY-8	CeN31, Y45F10B.19	AY948625.1, NR_003460.1	79	[3]
SmY-9	CeN25-7	AM286192.1	77	[5]
SmY-10	CeN112, Y45F10B.20	AY948610.1, NR_003461.1	90	[3]
SmY-11	CeN25-4, Y57G11C.55	AY948621.1, NR_003462.1	78	[3]
SmY-12	CeN25-6	AM286191.1	81 ^e	[5]

^a misclassified as small nucleolar RNA;

^b accessions conflict on exact size/sequence, used sequence reported by [3];

^c experimentally determined 78nt sequence from [3] includes 5' G not encoded by WS150 genome; used CESC 77nt version;

^d reported database sequence is on incorrect strand;

^e accession reports partial 73nt 3'-truncated sequence, we inferred an additional 3' 8nt from genomic sequence.

CESC: The *C. elegans* Sequencing Consortium.

was obtained as a partial 3'-truncated sequence [5], and SmY-4 and SmY-7 are predicted from sequence similarity [2].

SmY-2 and SmY-3 have also been identified with slightly different transcript sizes and called C/D small nucleolar RNAs Ce135 (72nt) and Ce96 (98nt) by [4], who criticized Deng *et al.* [3] for classifying these sequences as "small nuclear RNA like". Our analysis agrees with MacMorris *et al.* [2] in assigning these as SmY small nuclear RNA homologs, and we have used the transcript sequences deposited by Deng *et al.* [3].

In two cases, we modified a sequence from the accessioned version. We added 8 nt of genomic sequence to the 3' end of the truncated SmY-12 sequence to make it conform to our full-length consensus model. We reversed the orientation of SmY-7, because the accessioned version is in the incorrect (antisense) orientation.

Homology searches and a representative seed alignment. Starting from the sequences in Table 1, we conducted a number of different iterative searches, using a combination of BLAST [6] and Infernal 1.0 [7] [<http://infernal.janelia.org>] to identify SmY RNA homologs in a variety of genome sequences. Putative homologs were identified in the following 13 nematode genome sequence assemblies:

Caenorhabditis elegans (WormBase, WS150 [8, 9]); *Caenorhabditis briggsae* (WashU GSC, v1.0cb3, Jan 2007 [10]); *Caenorhabditis remanei* (WashU GSC, v15.0.1, May 2007); *Caenorhabditis japonica* (WashU GSC, v3.0.2, March 2008); *Caenorhabditis brenneri* (WashU GSC, v4.0, Jan 2007); *Pristionchus pacificus* (WashU GSC, v5.0.1, Jan 2008 [11]); *Haemonchus contortus* (Sanger Institute, unversioned, Jan 27 2006); *Meloidogyne incognita* (PNGG, unversioned, Nov 2007); *Meloidogyne hapla* (EBI, unversioned, Sept 2008 [12]); *Heterodera glycines* (Monsanto, Genbank ABLA00000000.1, Apr 2008); *Brugia malayi* (TIGR, unversioned, Sept 2007 [13]); *Ascaris suum* (www.nematode.net, unversioned, Nov 2006); and *Trichinella spiralis* (WashU GSC, v1.0, Mar 2006).

68 sequences were selected to be representative of the family. Starting from automated Infernal alignments, a multiple alignment was assembled and manually refined by structure and sequence conservation to form a curated seed alignment suitable for the Rfam database [14]. A Stockholm format text file of this alignment is provided in the Supplementary Material (`SmY_seed.stk`).

We used manual comparative sequence analysis to deduce a consensus secondary structure, and also independently predicted a consensus structure using the program locARNATE [15]. The two structure predictions largely agree with each other, and with a consensus structure previously published by [2]. The manual comparative analysis was favored where details differed. Figure 1 shows the predicted consensus secondary structure, together with a summary of the extensive base-pair covariation evidence in the seed alignment that supports it.

We did a retrospective analysis to establish the support for each individual sequence's probable homology to the rest of the family, which confirmed that each sequence is supported by significant ($< 1 \times 10^{-4}$) BLASTN or Infernal E-values when searched against phylogenetically independent subsets of the seed alignment (using the `-Z` option of both programs to calculate E-values for an effective search space size of 200 MB), with four exceptions. Four distantly related SmY sequences are predicted in Tylenchid nematode species – two paralogs in *Heterodera glycines*, and one SmY each in the related species *Meloidogyne hapla* and *M. incognita*. The assignment of these sequences as SmY homologs is supported by borderline Infernal E-values (0.01-0.001) to more than one Infernal model built of other independent SmY sequence subsets, and by the fact that they share the expected pattern of conservation, including base pair covariations consistent with stem 2.

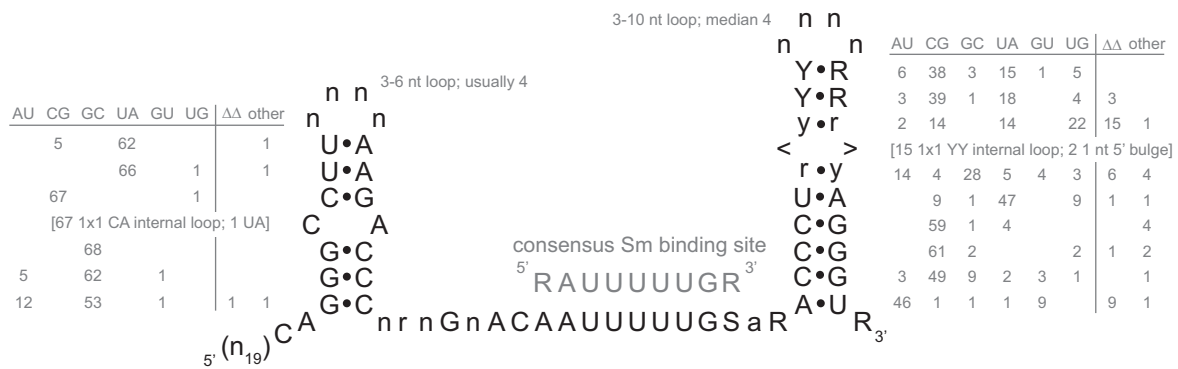


Figure 1: Consensus SmY RNA structure, with tables annotating the number of compensatory base pair substitutions, base pair deletions ($\Delta\Delta$ column), and noncompensatory substitutions (“other” column) observed in the 68 sequences of the representative seed alignment, in support of the structure prediction. The most highly conserved residues are shown as upper case letters in the structure. The sequence at the 5’ end of SmY RNAs is highly variable; the consensus is shown here as n_{19} , but it varies in both length and sequence in individual SmY sequences.

Phylogenetic diversity of SmY RNAs. In the system used by the Rfam RNA database, a consensus Infernal statistical model is built from a stable, curated seed alignment, and this consensus model is used to automatically identify and annotate homologs in genome sequences. The seed should be sufficiently representative that this single model identifies all known homologs. We used an Infernal model of the 68-sequence seed alignment to search the 13 nematode genomes. This search identifies 155 loci with E-values < 0.001 , and these loci include all the sequences we gathered in our initial iterative searches. An annotated table of all these loci is provided in the Supplementary Material (`SmY_all.tbl`).

Each of these loci was examined in detail. All appear to be plausible SmY homologs based on their overall pattern of conservation. Eight candidates appear to be artifacts of underassembled contigs in draft genomes. We annotated 26 as putative pseudogenes based on significant local deviations from the expected consensus (such as disruption of one of the stems) and/or the lack of an upstream proximal sequence element (PSE), a conserved transcriptional control motif generally found upstream of small nuclear RNAs [16], including SmY RNAs [3]. We annotated the remaining 121 loci as putative SmY RNA genes. Our gene/pseudogene labeling is only a best guess; for noncoding RNAs, it is generally not possible to unam-

biguously distinguish pseudogenes from genes by computational analysis.

In *C. briggsae*, we assigned 11 SmY genes and 1 pseudogene. Nine of these eleven genes were previously identified and named *cbSmY-1* through *cbSmY-9* [2]; we retained these names, though our analysis revises the predicted 5' and 3' ends of the genes. An additional locus named *cbSmY-10* by [2] does not appear to us to be an SmY RNA homolog. We detected two additional *C. briggsae* SmY genes, which we named *cbSmY-11* and *cbSmY-12* to be consistent with [2]. In all other species, we have not assigned gene names, but rather have identified putative SmY loci by their assembly contig name and sequence coordinates.

Figure 2 shows the phylogenetic distribution of SmY RNA genes and pseudogenes. The SmY family has undergone a large paralogous expansion in *Caenorhabditis* and *Pristionchus* species, with copy numbers of 10-26, compared to 1-4 copies in other nematode genomes. Many of these paralogs within a species are more related to each other than to any homolog in another species, suggesting independent paralogous expansions and/or evolutionary turnover (balancing gene loss and paralogous duplication) in these lineages. An extreme case of apparently recent expansion is *Pristionchus*, where most SmY RNAs have 100% identical paralogs. Only one SmY locus appears to be syntenically conserved among the five *Caenorhabditis* species, with a single copy in *C. japonica* and two copies in the other four species corresponding to *C. elegans* SmY-1 and SmY-2. Rapid turnover of paralogs is a common feature of multicopy structural RNA genes; similar features are seen for tRNA gene families, for example [17].

We also used this model to search for SmY homologs in six non-nematode genomes representing other phyla. We chose genome sequence assemblies of the trematode *Schistosoma mansoni* (TIGR, unversioned, 15 May 2007) [18] and the urochordates *Ciona intestinalis* (JGI, v2.0, Oct 2002) [19] and *Oikopleura dioica* (Genoscope, v3.0, Sept 2006) because these metazoans employ spliced leader trans-splicing [20]. The leech *Helobdella robusta* (JGI, v1.0, July 2007), the snail *Lottia gigantea* (JGI, v1.0, August 2006), and the fruit fly *Drosophila melanogaster* (BDGP, v5.10, July 2008) were chosen as additional representative outgroups to the phylum *Nematoda*. No Infernal hit with an E-value better than 0.01 was identified.

Discussion

SmY RNA appears to be associated with trans-splicing and spliceosome proteins in *Caenorhabditis elegans* and *Ascaris*, but unlike the trans-spliced leader RNAs SL1 and SL2, it apparently does not contribute

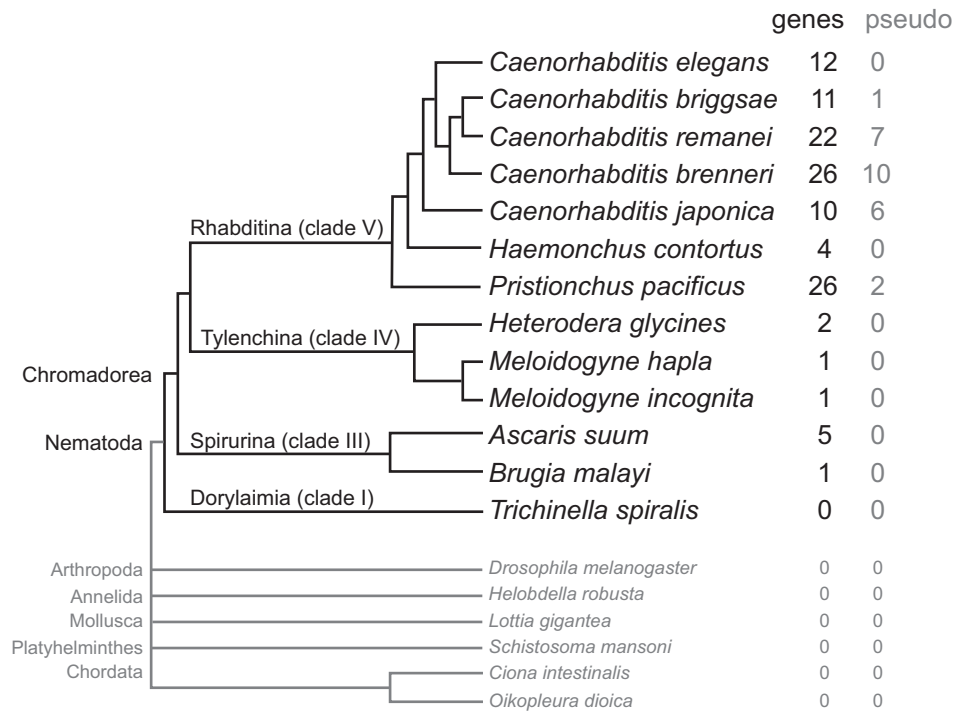


Figure 2: Phylogenetic distribution of 147 identified SmY RNA homologs (not shown are another 8 hits that were attributed to underassembled contigs in draft genomes). The species phylogeny is represented as a cladogram (branch lengths are arbitrary), combining the *Caenorhabditis* species phylogeny from [21] with the phylogeny of phylum Nematoda from [22, 23]. To our knowledge there is no sequenced genome yet from a representative of clade II (Enoplia) in the Nematoda. At the root, the relationship of Nematoda to other metazoan phyla is shown as a multifurcation, because most of these relationships remain in some doubt.

a spliced leader sequence to mRNAs. What does SmY RNA do, then? MacMorris *et al.* hypothesized that the role of SmY RNA may be in recycling spliceosome proteins after SL RNAs are consumed in the trans-splicing reaction [2]. They proposed a specific model in *C. elegans* in which the stem-loop 2 sequence of one SmY RNA, SmY-10, base-pairs to SL1 RNAs (which are encoded by a tandem array of about 110 near-identical genes), while stem-loop 2 of the other SmY RNAs base pairs to SL2 RNAs (which are encoded by 18-20 dispersed genes with significant sequence variation). MacMorris' model suggests that the diversification of SmY RNA gene copies (accompanied by sequence variations in stem-loop 2, the more variable stem) may be driven by the diversification of SL2 RNA genes. Although we have not conducted a detailed joint comparative analysis of SL RNAs and SmY RNAs, the results of our SmY RNA survey are broadly in accordance with this model's expectations. SL2 RNAs have as yet only been identified in Rhabditina species, whereas SL1 RNAs have been found throughout the other species of Chromadorea [24]. We find the largest proliferation of paralogous SmY RNA genes in species that have SL2 genes, and smaller numbers of SmY RNAs in species that only have SL1.

We were not able to identify any SmY RNA homologs in the more distantly related Dorylaimid species *Trichinella spiralis*, which has a noncanonical and polymorphic family of SL1-like trans-spliced RNAs [25], but this negative result is inconclusive. The SmY homologs we identified in clade IV Tylenchid nematodes are at the detection limit of the Infernal software (and well beyond BLAST's limits), so it may be that *Trichinella* SmY homologs exist but are too diverged to be detectable with our methods. The same caution applies to our inability to identify SmY RNAs outside the nematode phylum.

By eye, we do note one suggestive similarity outside nematodes. The SmY RNA structure strongly resembles the proposed structure of a herpesvirus HSUR3 RNA, one of five U snRNAs expressed by herpesvirus saimiri [26]. Like SmY RNA, HSUR3 is a small (75 nt) RNA proposed to have a consensus Sm binding site flanked by two stem-loops of similar length and loop size as the SmY stem-loops, including a C:A mismatch in stem 1. However, an Infernal model of SmY does not assign a significant homology score to HSUR3. We note this suggestive visual similarity because the function of the herpesvirus U RNAs remains unknown, and perhaps there is a useful link to the role of the SmY RNAs.

Supplementary Material

Three ASCII text files are provided as supplementary material. `SmY_seed.stk` is the seed alignment of 68 representative sequences in Stockholm format [14], with consensus secondary structure annotation. `SmY_all.tbl` is a tabular file listing all 155 SmY-related loci we identified, sorted by species and by Infernal E-value, and annotated by whether we believe the locus to be a true gene, a pseudogene, or a draft assembly artifact. `SmY_all.fa` is a FASTA file containing the 147 gene and pseudogene sequences (excluding the 8 sequences that we believe are assembly artifacts).

Acknowledgements

This work was supported in part by the *Graduierten-Kolleg Wissensrepräsentation*, the Bioinformatics Initiative of the *Deutsche Forschungs-Gemeinschaft* and by the *Konrad-Adenauer-Stiftung*. The Leipzig group (WO, MM, PFS) thanks Petra Pregel, Jens Steuck and Alexander Donath for making work much easier. The Janelia group (TAJ, SRE) thanks Diana Kolbe and Eric Nawrocki for their work on the Infernal software, and Goran Ceric for managing Janelia’s computing resources.

References

- [1] P. A. Maroney, Y. T. Yu, M. Jankowska, and T. W. Nilsen. Direct analysis of nematode cis and trans-spliceosomes: a functional role for U5 snRNA in spliced leader addition trans-splicing and the identification of novel Sm snRNPs. *RNA*, 2:735–745, 1996.
- [2] M. MacMorris, M. Kumar, E. Lasda, A. Larsen, B. Kraemer, and T. Blumenthal. A novel family of *C. elegans* snRNPs contains proteins associated with trans-splicing. *RNA*, 13:511–520, 2007.
- [3] W. Deng, X. Zhu, G. Skogerbø, Y. Zhao, Z. Fu, Y. Wang, H. He, L. Cai, H. Sun, C. Liu, B. Li, B. Bai, J. Wang, D. Jia, S. Sun, H. He, Y. Cui, Y. Wang, D. Bu, and R. Chen. Organization of the *Caenorhabditis elegans* small non-coding transcriptome: Genomic features, biogenesis, and expression. *Genome Res.*, 16:20–29, 2006.

- [4] A. Zemann, A. op de Bekke, M. Kiefmann, J. Brosius, and J. Schmitz. Evolution of small nucleolar RNAs in nematodes. *Nucl. Acids Res.*, 34:2676–2685, 2006.
- [5] H. He, J. Wang, T. Liu, X. S. Liu, T. Li, Y. Wang, Z. Qian, H. Zheng, X. Zhu, T. Wu, B. Shi, W. Deng, W. Zhou, G. Skogerbø, and R. Chen. Mapping the *C. elegans* noncoding transcriptome with a whole-genome tiling microarray. *Genome Res.*, 17:1471–1477, 2007.
- [6] S. F. Altschul, T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl. Acids Res.*, 25:3389–3402, 1997.
- [7] E. P. Nawrocki and S. R. Eddy. Query-dependent banding (QDB) for faster RNA similarity searches. *PLoS Comput. Biol.*, 3:e56, 2007.
- [8] The *C. elegans* Genome Sequencing Consortium. Genome sequence of the nematode *C. elegans*: A platform for investigating biology. *Science*, 282:2012–2018, 1998.
- [9] A. Rogers, I. Antoshechkin, T. Bieri, D. Blasiar, C. Bastiani, P. Canaran, J. Chan, W. J. Chen, P. Davis, J. Fernandes, T. J. Fiedler, M. Han, T. W. Harris, R. Kishore, R. Lee, S. McKay, H. M. Müller, C. Nakamura, P. Ozersky, A. Petcherski, G. Schindelman, E. M. Schwarz, W. Spooner, M. A. Tuli, K. Van Auken, D. Wang, X. Wang, G. Williams, K. Yook, R. Durbin, L. D. Stein, J. Spieth, and P. W. Sternberg. WormBase 2007. *Nucl. Acids Res.*, 36:D612–D617, 2008.
- [10] L. D. Stein, Z. Bao, D. Blasiar, T. Blumenthal, M. R. Brent, N. Chen, A. Chinwalla, L. Clark, C. Clee, A. Coghlan, A. Coulson, P. D’Eustachio, D. H. A. Fitch, L. A. Fulton, R. E. Fulton, S. Griffiths-Jones, T. W. Harris, L. W. Hillier, R. Kamath, P. E. Kuwabara, E. R. Mardis, M. A. Marra, T. L. Miner, P. Minx, J. C. Mullkin, R. W. Plumb, J. Rogers, J. E. Schein, M. Sohrmann, J. Spieth, J. E. Stajich, C. Wie, D. Willey, R. K. Wilson, R. Durbin, and R. H. Waterston. The genome sequence of *Caenorhabditis briggsae*: A platform for comparative genomics. *PLoS Biol.*, 1:166–192, 2003.
- [11] C. Dieterich, S. W. Clifton, L. N. Schuster, A. Chinwalla, K. Delehaunty, I. Dinkelacker, L. Fulton, R. Fulton, J. Godfrey, P. Minx, M. Mitreva, W. Roeseler, H. Tian, H. Witte, S. P. Yang, R. K. Wilson,

- and R. J. Sommer. The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism. *Nat. Genet.*, 40:1193–1198, 2008.
- [12] C. H. Opperman, D. M. Bird, V. M. Williamson, D. S. Rokhsar, M. Burke, J. Cohn, J. Cromer, S. Diener, J. Gajan, S. Graham, T. D. Houfek, Q. Liu, T. Mitros, J. Schaff, R. Schaffer, E. Scholl, B. R. Sosinski, V. P. Thomas, and E. Windham. Sequence and genetic map of *Meloidogyne hapla*: A compact nematode genome for plant parasitism. *Proc. Natl. Acad. Sci. USA*, 105:14802–14807, 2008.
- [13] E. Ghedin, S. Wang, D. Spiro, E. Caler, Q. Zhao, J. Crabtree, J. E. Allen, A. L. Delcher, D. B. Guiliano, D. Miranda-Saavedra, S. V. Angiuoli, T. Creasy, P. Amedeo, B. Haas, N. M. El-Sayed, J. R. Wortman, T. Feldblyumand L. Tallon, M. Schatz, M. Shumway, H. Koo, S. L. Salzberg, S. Schobel, M. Pertea, M. Pop, O. White, G. J. Barton, C. K. Carlow, M. J. Crawford, J. Daub, M. W. Dimmic, C. F. Estes, J. M. Foster, M. Ganatra, W. F. Gregory, N. M. Johnson, J. Jin, R. Komuniecki, I. Korf, S. Kumar, S. Laney, B. W. Li, W. Li, T. H. Lindblom, S. Lustigman, D. Ma, C. V. Maina, D. M. Martin, J. P. McCarter, L. McReynolds, M. Mitreva, T. B. Nutman, J. Parkinson, J. M. Peregrín-Alvarez, C. Poole, Q. Ren, L. Saunders, A. E. Sluder, K. Smith, M. Stanke, T. R. Unnasch, J. Ware, A. D. Wei, G. Weil, D. J. Williams, Y. Zhang, S. A. Williams, C. Fraser-Liggett, B. Slatko, M. L. Blaxter, and A. L. Scott. Draft genome of the filarial nematode parasite *Brugia malayi*. *Science*, 317:1756–1760, 2007.
- [14] P. P. Gardner, J. Daub, J. G. Tate, E. P. Nawrocki, D. L. Kolbe, S. Lindgreen, A. C. Wilkinson, R. D. Finn, S. Griffiths-Jones, S. R. Eddy, and A. Bateman. Rfam: Updates to the RNA families database. NAR, in press, 2009.
- [15] W. Otto, S. Will, and R. Backofen. Structural local multiple alignment of RNA. In M. Schroeder, editor, *GCB '08: German Conference on Bioinformatics*. GI, 2008. accepted.
- [16] J. Thomas, K. Lea, E. Zucker-Aprison, and T. Blumenthal. The spliceosomal snRNAs of *Caenorhabditis elegans*. *Nucl. Acids Res.*, 18:2633–2642, 1990.
- [17] International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409:860–921, 2001.

- [18] B. J. Haas, M. Berriman, H. Hirai, G. G. Cerqueira, P. T. LoVerde, and N. M. El-Sayed. *Schistosoma mansoni* genome: Closing in on a final gene set. *Exp. Parasitol.*, 117:225–228, 2007.
- [19] P. Dehal, Y. Satou, R. K. Campbell, J. Chapman, B. Degnan, A. De Tomaso, B. Davidson, A. Di Gregorio, M. Gelpke, D. M. Goodstein, N. Harafuji, K. E. Hastings, I. Ho, K. Hotta, W. Huang, T. Kawashima, P. Lemaire, D. Martinez, I. A. Meinertzhagen, S. Necula, M. Nonaka, N. Putnam, S. Rash, H. Saiga, M. Satake, A. Terry, L. Yamada, H. G. Wang, S. Awazu, K. Azumi, J. Boore, M. Branno, S. Chin-Bow, R. DeSantis, S. Doyle, P. Francino, D. N. Keys, S. Haga, H. Hayashi, K. Hino, K.S. Imai, K. Inaba, S. Kano, K. Kobayashi, M. Kobayashi, B. I. Lee, K. W. Makabe, C. Manohar, G. Matassi, M. Medina, Y. Mochizuki, S. Mount, T. Morishita, S. Miura, A. Nakayama, S. Nishizaka, H. Nomoto, F. Ohta, K. Oishi, I. Rigoutsos, M. Sano, A. Sasaki, Y. Sasakura, E. Shoguchi, T. Shin-i, A. Spagnuolo, D. Stainier, M. M. Suzuki, O. Tassy, N. Takatori, M. Tokuoka, K. Yagi, F. Yoshizaki, S. Wada, C. Zhang, P. D. Hyatt, F. Larimer, C. Detter, N. Doggett, T. Glavina, T. Hawkins, P. Richardson, S. Lucas, Y. Kohara, M. Levine, N. Satoh, and D. S. Rokhsar. The draft genome of *Ciona intestinalis*: Insights into chordate and vertebrate origins. *Science*, 298:2157–2167, 2002.
- [20] K. E. Hastings. SL trans-splicing: Easy come or easy go? *Trends Genet.*, 21:240–247, 2005.
- [21] W. Sudhaus and K. Kiontke. Comparison of the cryptic nematode species *Caenorhabditis brenneri* sp. n. and *C. remanei* (Nematoda: Rhabditidae) with the stem species pattern of the *Caenorhabditis Elegans* group. *Zootaxa*, 1456:45–62, 2007.
- [22] M. L. Blaxter, P. De Ley, J. R. Garey, L. X. Liu, P. Scheldeman, A. Vierstraete, J. R. Vanfleteren, L. Y. Mackey, M. Dorris, L. M. Frisse, J. T. Vida, and W. K. Thomas. A molecular evolutionary framework for the phylum Nematoda. *Nature*, 392:71–75, 1998.
- [23] M. Mitreva, M. L. Blaxter, D. M. Bird, and J. P. McCarter. Comparative genomics of nematodes. *Trends Genet.*, 21:573–581, 2005.
- [24] T. Blumenthal. Trans-splicing and operons. In The *C. elegans* Research Community, editor, *Worm-Book*. doi/10.1895/wormbook.1.5.1, <http://www.wormbook.org>, 2005.

- [25] J. Pettitt, B. Müller, I. Stansfield, and B. Connolly. Spliced leader *trans*-splicing in the nematode *Trichinella spiralis* uses highly polymorphic, noncanonical spliced leaders. *RNA*, 14:760–770, 2008.
- [26] S. I. Lee and J. A. Steitz. Herpesvirus saimiri U RNAs are expressed and assembled into ribonucleo-protein particles in the absence of other viral genes. *J. Virol.*, 64:3905–3915, 1990.