Minireview

The draft genome sequence of the nematode *Caenorhabditis briggsae*, a companion to *C. elegans*

Bhagwati P Gupta and Paul W Sternberg

Address: Howard Hughes Medical Institute and Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA.

Correspondence: Paul W Sternberg. E-mail: pws@caltech.edu

Published: 18 November 2003

Genome ***Biology*** 2003, **4:**238

The electronic version of this article is the complete one and can be found online at http://genomebiology.com/2003/4/12/238

Abstract

The publication of the draft genome sequence of *Caenorhabditis briggsae* improves the annotation of the genome of its close relative *Caenorhabditis elegans* and will facilitate comparative genomics and the study of the evolutionary changes during development.

Each genome is fascinating in its own right, but some genomes have been chosen for study because of the added advantages of understanding a closely related genome. *Caenorhabditis briggsae* (Figure 1a) is just such a companion to *Caenorhabditis elegans*, a model system that has been widely used to study the genetic basis of animal development, behavior and physiology. The sequencing of the genome of *C. elegans* [1] has led to rapid advances in our understanding of gene function, particularly through the use of techniques such as RNA interference (RNAi) [2]. Stein *et al.* [3] now report a draft sequence of the *C. briggsae* genome.

C. briggsae is closely related to *C. elegans* (Figure 2) and has almost identical morphology (Figure 1b,c) [4-6]. Given the lack of fossil records, the evolutionary distance between *C. briggsae* and *C. elegans* has been estimated using a molecular clock, which gives a divergence time of between 20 and 120 million years ago (Mya) [7-11]. By analyzing the 338 sets of orthologous genes found in the *C. briggsae, C. elegans, Anopheles* and human genomes and using a molecular clock calibrated by the known date of divergence of nematodes and arthropods, Stein *et al.* [3] now report a much tighter estimate of the divergence time of *C. briggsae* and *C. elegans*, between 80 and 110 Mya. This divergence is slightly greater than the estimate of the human-mouse divergence time (65-75 Mya) [12]. The draft covers 98% of the 104 Mb genome; the slightly larger size of the *C. briggsae* genome compared with that of *C. elegans* (100.3 Mb according to the WS108 release of September 2003 [13]) is primarily due to additional repetitive DNA.

Comparison of genes and non-coding regions

The *C. briggsae* genome was annotated using various gene-finding programs (such as Genefinder [14]) and by comparison with *C. elegans*. As different programs often disagree with each other in predicting genes, Stein *et al.* [3] adopted a 'hybrid' approach by combining the predictions made by multiple gene-finding programs and selecting the consensus. In cases in which a consensus could not be obtained, the authors chose the predictions with best overall similarity with the *C. elegans* genome. This analysis identified 19,507 genes in *C. briggsae*. In general, the 'hybrid' approach was twice as accurate as any single gene-prediction program. Conversely, the *C. briggsae* genome has been extremely useful in the annotation of *C. elegans* genes: 1,275 new genes were predicted in the *C. elegans* genome (6% of the new total of 20,621 genes) on the basis of the *C. briggsae-C. elegans* comparison. Almost 300 of these are confirmed by the open reading-frame sequence tag data of Reboul *et al.* [15]. Gene finding is of supreme importance because essentially all *C. elegans* genes are being studied, for example by RNAi, and this 6% increase in the number of genes will have enormous impact on the intensive analysis of the organism. Further refinements in *C. elegans* gene-structure predictions

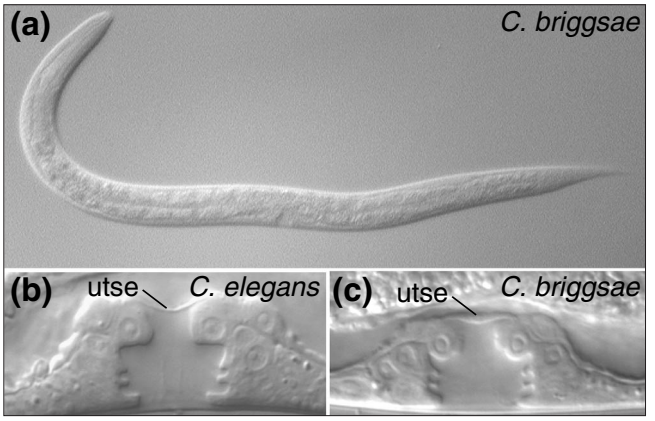

Figure 1
Anatomical comparison of *C. briggsae* and *C. elegans*. **(a)** The *C. briggsae* hermaphrodite. **(b,c)** Vulval invagination in L4 larvae of (b) *C. elegans* and (c) *C. briggsae*. The vulval morphology of *C. briggsae* is almost indistinguishable from that of *C. elegans*; minor differences include a slightly thicker vulval-uterine connection (utse) in *C. briggsae*.

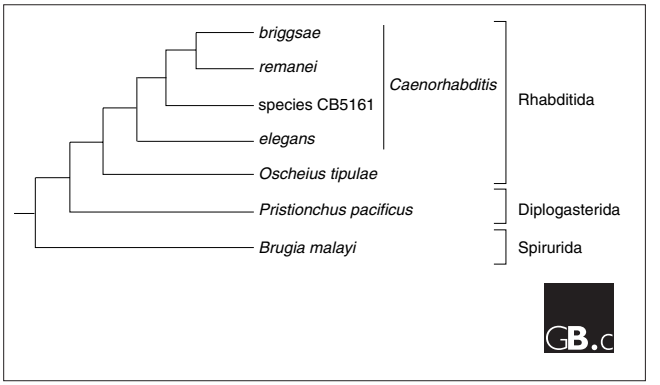

Figure 2
The phylogenetic relationship of *C. briggsae* with other nematodes (modified from [3,6]). *Oscheius tipulae* (family Rhabditida) and *Pristionchus pacificus* (Diplogasterida) are two other models used to study evolutionary changes during development [31].

are likely as many more genes are compared carefully in the two species.

Using two different approaches - best reciprocal BLASTP matches and conserved gene order (synteny) - Stein *et al.* [3] have defined orthologs of *C. elegans* genes in the *C. briggsae* genome (62% of predicted *C. briggsae* genes or 12,155; Figure 3). The identity at the protein level between *C. briggsae-C. elegans* orthologous pairs (mean of 75%) is comparable to mouse-human orthologs (median 78.5%) [3,12]. Comparison of the orthologs reveals that about 11% of introns are species-specific, with *C. elegans* having almost twice as many unique introns (*C. elegans* has about 4,400 and *C. briggsae* about 2,200; see also [16]). On average, half of the genes with orthologs differ in the presence of an intron between the two species. Just 4% (807) of the *C. briggsae* genes do not have significant BLASTP matches in *C. elegans* (Figure 3); these are likely to be highly divergent and novel (species-specific) genes.

The *C. briggsae* genome contains 5,211 genes that have multiple matches in the *C. elegans* genome (Figure 3) and that correspond to various gene families. Although in most cases such families contain comparable number of proteins in the two species (for example, there are 376 protein kinases in *C. elegans* and 399 in *C. briggsae*), the olfactory-type chemosensory receptors and the cyclin-like F-box proteins have significantly more members in *C. elegans* (718 and 243 members, respectively) than in *C. briggsae* (464 and 98, respectively) [3]. For the chemoreceptors, gene expansion is likely to have occurred in the *C. elegans*-specific families. The functional significance of such divergence might indicate physiological or ecological differences between these species.

In addition to finding coding exons, blocks of alignable sequence were found using the WABA algorithm [16]. About 1.3 million such blocks were identified, covering coding exons (32.2%), introns (35.3%) and other regions (untranslated regions, intergenic regions, and so on). These conserved regions are a rich resource for further investigation, but it would be useful to have additional genome sequences in order to help discriminate between regions that are conserved because of selection and those that are conserved because they have not yet faded away during neutral evolution. The WABA algorithm distinguishes different codon positions and thus reveals many of the presumed regions of intergenic conservation to have the signature of coding exons. The *C. briggsae* genome sequence has been a tool to help analyze *cis*-regulatory sequences for some time already. In general, the *C. briggsae-C. elegans* comparison has been helpful in locating functionally critical regions of non-coding DNA [17,18] and may be a good filter for eliminating false positives for some types of bioinformatic searches.

By aligning clear orthologs, Stein *et al.* [3] were able to evaluate the syntenic relationships for much of the two genomes. Alignment of ortholog pairs reveals over 3,000 rearrangements, and these are ten times as common within a chromosome as between chromosomes. A higher-resolution view of the syntenic relationships between the genomes will require additional physical mapping, or genetic mapping, for example using single-nucleotide polymorphisms. Comparison of the genomic rearrangements between *C. briggsae-C. elegans* and *Drosophila* species reveals a breakpoint rate roughly five times higher in nematodes.

One striking feature of nematode genome organization is the existence of *trans*-spliced operons, in which a primary transcript is processed to give multiple protein-coding transcripts by a *trans*-splicing reaction with the SL2 spliced

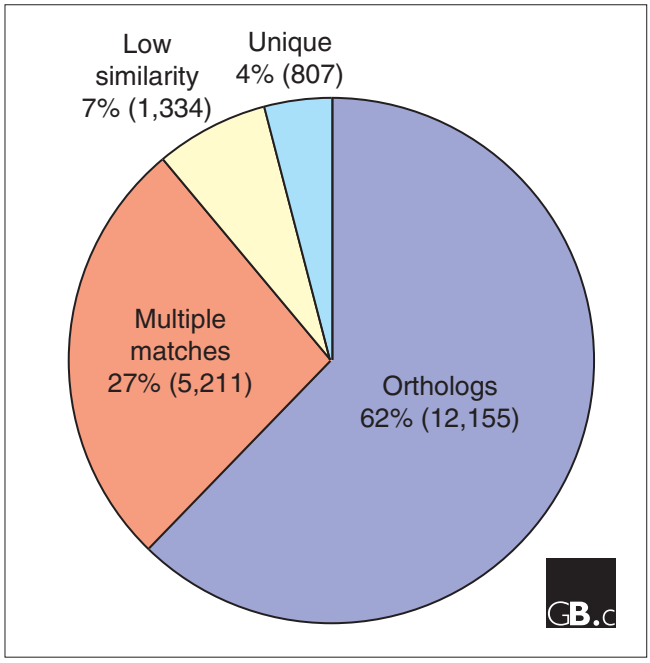

Figure 3
A pie chart showing the relationship between *C. briggsae* and *C. elegans* genes. BLASTP matches and conserved gene order (synteny) have revealed 62% of *C. briggsae* genes to have orthologs in *C. elegans*. About 27% of *C. briggsae* genes have multiple matches in *C. elegans* genome and represent various gene families. Of the remaining 11% of genes, 7% have very weak similarity to *C. elegans* (with a BLASTP *E* value of over 10^{-5}), whereas 4% appear to be unique.

leader sequence. There are estimated to be about 1,000 operons in *C. elegans* (WS108 release) [19]. Of the 800 well-characterized *C. elegans* operons (WS77 release), 32 are broken in *C. briggsae*, either by insertion, transposition or rearrangement.

Developmental differences between *C. briggsae* and *C. elegans*

As we have seen, despite the morphological similarity, there are significant molecular differences between the *C. briggsae* and *C. elegans* genomes. The *C. briggsae* genome contains about 800 genes with no apparent match in *C. elegans* [3]. Together with the divergent genes and gene families (Figure 3), nearly one third of the genome is arguably different from *C. elegans*. Are these changes reflected in significant biological differences? Careful examination has revealed a number of subtle differences between the two species. For example, the excretory system in *C. elegans* plays a critical role in osmoregulation [20], and comprises three cells including a single duct cell [21]. Although the excretory system looks morphologically identical in the two species, *C. briggsae* animals have a more anterior duct opening than *C. elegans* [22]. This difference is the result of the altered expression of *lin-48*, which encodes a member of the Ovo family of transcription factors. Expression of *lin-48* is observed in the excretory duct cell in *C. elegans* but not in *C. briggsae* [22,23]. This is the first example of a morphological difference between the two nematode species with a known molecular basis.

Studies of vulval development in *C. briggsae* have revealed that, although the overall vulval morphology is similar to that in *C. elegans* (Figure 1b,c), there are some differences in the underlying mechanisms. In *C. elegans*, six vulval precursor cells are competent to respond to a gonad-derived inductive signal and to produce the vulval progeny cells [24]. For example, ablation of all vulval precursor cells but the anterior-most one, known as P3.p, allows the latter to generate vulval cells, even though it does not do so in an intact animal. By contrast, P3.p in *C. briggsae* is not competent in this assay [25]. On the other hand, in certain multivulva mutants of *C. briggsae*, P3.p is induced and makes vulval progeny cells (B.P.G., unpublished observations). Thus, there is a subtle difference in the competence of the P3.p cell between the two species. Another example of the differences in vulval development comes from the study of the *glp-1* gene function in three *Caenorhabditis* species: *C. briggsae, C. elegans* and *C. remanei*. In *C. elegans*, GLP-1, a receptor of the LIN-12/Notch family, mediates cell-cell communication during development [26]. Loss-of-function mutations in *glp-1* do not affect *C. elegans* vulval development, but in *C. briggsae*, injection of *glp-1* double-stranded RNA causes a multivulva phenotype, suggesting that *glp-1* inhibits vulval development in *C. briggsae* [27].

Nematodes are morphologically diverse and have two different modes of reproduction (hermaphroditic, such as in *C. briggsae* and *C. elegans*, and male-female, such as in *C. remanei* and *Caenorhabditis* species CB5161); they thus provide an excellent opportunity to study the molecular basis of sexual differentiation and evolution. Several of the known sex-determination genes are remarkably diverged between *C. briggsae* and *C. elegans* [28]; for example, FEM-3 and TRA-2, which physically interact with each other, have coevolved and function strictly in a species-specific manner [29]. RNAi targeting the *fem-1* and *fem-2* genes in *C. briggsae* have revealed functional differences from *C. elegans* ([28] and references therein).

A need for more sequences and functional tests

An immediate reward of the *C. briggsae* genome sequence has been to increase our understanding of *C. elegans* gene structure. The *C. briggsae-C. elegans* comparison will inspire many new projects to study gene function and evolutionary changes. Although sequence comparison has revealed a significant number of gene sequences conserved between *C. briggsae* and *C. elegans* (62% orthologs; see Figure 3), it is not clear whether all these genes are functionally conserved. Furthermore, conservation in the coding

sequences might be misleading if the regulatory elements have evolved differentially (see, for example, [17]). In addition, there are significant differences in some gene families (see above). An understanding of the evolutionary conservation and divergence between the two genomes will require systematic study of the function of *C. briggsae* genes. The recent success of large-scale RNAi screens in *C. elegans* raises the promise of a similar approach in *C. briggsae*. If there is a difference in the RNAi phenotype of a particular gene, however, it will be unclear whether it is a quantitative difference in the use of that gene or whether the gene differs in its susceptibility to RNAi. Thus, classical genetics or targeted gene knockouts are likely to be valuable. To facilitate such experiments, a classical genetic linkage map (B.P.G., R. Johnsen, T. Inoue, A. Mah, G. Jo, D. Baillie and P.W.S., unpublished results) and a single-nucleotide polymorphism map (R. Miller, S. Baird, L. Fulton and R. Waterston, personal communication) are being developed for *C. briggsae*. These maps will help order contigs in the genome and allow researchers to carry out genetic analysis, as well as helping clone *C. briggsae* genes with novel mutant phenotypes and study their biological functions.

The *C. briggsae* genome sequence will probably also help in understanding parasitic nematode genomes such as that of the filarial parasite *Brugia malayi*. Additional genomes could bridge the gap from *C. briggsae* and *C. elegans* to other species (Figure 2). Intensive analysis of transcriptional regulatory networks, in particular *cis*-regulatory elements, has clearly been helped by the availability of *C. briggsae* sequence. There is significant value for computational analysis in having additional close nematode genomes (E. Schwarz, J. DeModena, E. Moon, H. Shizuya, B. Wold and P.W.S., unpublished observations). The biological differences between the other nematodes might make the sequence comparisons slightly less informative, but it is possible to test the function of sequences by reciprocal transformation experiments [30], providing a way to know whether there is indeed conservation in gene function. We look forward to the new insights into nematode biology that more genome sequences and further analysis of the *C. briggsae* genome will bring.

References

1. The *C. elegans* sequencing consortium: **Genome sequence of the nematode *C. elegans*: a platform for investigating biology.** *Science* 1998, **282:**2012-2018.
2. Fire A, Xu S, Montgomery MK, Kostas SA, Driver SE, Mello CC: **Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*.** *Nature* 1998, **391:**806-811.
3. Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent M, Chen NS, Chinwalla A, Clarke L, Clee C, Coghlan A, *et al.*: **The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics.** *PLoS Biology* 2003, **1:**166-192.
4. Nigon V, Dougherty EC: **Reproductive patterns and attempts at receiprocal crossing of *Rhabditis elegans* Maupas, 1900, and *Rhabditis briggsae* Dougherty & Nigon, 1949 (*Nematoda: Rhabditidae*).** *J Exp Zool* 1949, **112:**485-503.
5. Fodor A, Riddle DL, Nelson FK, Golden JW: **Comparison of a new wild-type *Caenorhabditis briggsae* with laboratory strains of *C. briggsae* and *C. elegans*.** *Nematol* 1983, **29:**203-217.
6. Blaxter ML, Ley PD, Garey JR, Liu LX, Scheldeman P, Vierstraete A, Vanfleteren JR, Mackey LY, Dorris M, Frisse LM, *et al.*: **A molecular evolutionary framework for the phylum Nematoda.** *Nature* 1998, **392:**71-75.
7. Coghlan A, Wolfe KH: **Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*.** *Genome Res* 2002, **12:**857-867.
8. Kennedy BP, Aamodt EJ, Allen FL, Chung MA, Heschl MF, McGhee JD: **The gut esterase gene (*ges-1*) from the nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae*.** *J Mol Biol* 1993, **229:**890-908.
9. Lee YH, Huang XY, Hirsh D, Fox GE, Hecht RM: **Conservation of gene organization and *trans*-splicing in the glyceraldehydes-3-phosphate dehydrogenase-encoding genes of *Caenorhabditis briggsae*.** *Gene* 1992, **121:**227-235.
10. Heschl MF, Baillie DL: **Functional elements and domains inferred from sequence comparisons of a heat shock gene in two nematodes.** *J Mol Evol* 1990, **31:**3-9.
11. Prasad SS, Baillie DL: **Evolutionarily conserved coding sequences in the *dpy-20-unc-22* region of *Caenorhabditis elegans*.** *Genomics* 1989, **5:**185-198.
12. Waterston RH, Lindblad-Toh K, Birney E, Rogers J, Abril JF, Agarwal P, Agarwala R, Ainscough R, Alexandersson M, An P, *et al.*: **Mouse genome sequencing consortium. Initial sequencing and comparative analysis of the mouse genome.** *Nature* 2002, **420:**520-562.
13. **Wormbase** [http://www.wormbase.org]
14. **Genefinder** [http://repeatmasker.genome.washington.edu/cgi-bin/Genefinder]
15. Reboul J, Vaglio P, Rual JF, Lamesch P, Martinez M, Armstrong CM, Li S, Jacotot L, Bertin N, Janky R, *et al.*: ***C. elegans* ORFeome version 1.1: experimental verification of the genome annotation and resource for proteome-scale protein expression.** *Nat Genet* 2003, **34:**35-41.
16. Kent WJ, Zahler AM: **Conservation, regulation, synteny, and introns in a large-scale *C. briggsae-C. elegans* genomic alignment.** *Genome Res* 2000, **10:**1115-1125.
17. Kirouac M, Sternberg PW: ***cis*-regulatory control of three cell fate-specific genes in vulval organogenesis of *Caenorhabditis elegans* and *C. briggsae*.** *Dev Biol* 2003, **257:**85-103.
18. Gower NJD, Temple GR, Schein JE, Marra M, Walker DS, Baylis HA: **Dissection of the promoter region of the inositol 1,4,5-trisphosphate receptor gene, *itr-1*, in *C. elegans*: A molecular basis for cell-specific expression of IP_3R isoforms.** *J Mol Biol* 2001, **306:**145-157.
19. Blumenthal T, Gleason KS: ***Caenorhabditis elegans* operons: forms and function.** *Nat Rev Genet* 2003, **4:**112-120.
20. Nelson FK, Riddle DL: **Functional study of the *Caenorhabditis elegans* secretory-excretory system.** *J Exp Zool* 1984, **231:**45-56.
21. Nelson FK, Albert PS, Riddle DL: **Fine structure of the *Caenorhabditis elegans* secretory-excretory system.** *J Ultrastruct Res* 1983, **82:**156-171.
22. Wang X, Chamberlin HM: **Multiple regulatory changes contribute to the evolution of the *Caenorhabditis lin-48 ovo* gene.** *Genes Dev* 2002, **16:**2345-2349.
23. Johnson AD, Fitzsimmons D, Hagman J, Chamberlin HM: **EGL-38 Pax regulates the ovo-related gene *lin-48* during *Caenorhabditis elegans* organ development.** *Development* 2001, **128:**2857-2865.
24. Horvitz HR, Sternberg PW: **Multiple intercellular signalling systems control the development of the *C. elegans* vulva.** *Nature* 1991, **351:**535-541.
25. Delattre M, Felix M-A: **Polymorphism and evolution of vulval precursor cell lineages within two nematode genera, *Caenorhabditis* and *Oscheius*.** *Curr Biol* 2001, **11:**631-643.
26. Kimble J, Simpson P: **The LIN-12/Notch signaling pathway and its regulation.** *Annu Rev Cell Dev Biol* 1997, **13:**333-361.
27. Rudel D, Kimble J: **Conservation of *glp-1* regulation and function in nematodes.** *Genetics* 2001, **157:**639-654.
28. Stothard P, Pilgrim D: **Sex-determination gene and pathway evolution in nematodes.** *BioEssays* 2003, **25:**221-231.
29. Haag ES, Wang S, Kimble J : **Rapid coevolution of the nematode sex-determining genes *fem-3* and *tra-2*.** *Curr Biol* 2002, **12:**2035-2041.
30. Ruvinsky I, Ruvkun G: **Functional tests of enhancer conservation between distantly related species.** *Development* 2003, **130:**5133-5142.
31. Sommer RJ: **Comparative genetics: A third model nematode species.** *Curr Biol* 2000, **10:**R879-R881.