

RESEARCH ARTICLE

Open Access

# The genome of *Romanormis culicivorax*: revealing fundamental changes in the core developmental genetic toolkit in Nematoda

Philipp H Schiffer<sup>1\*</sup>, Michael Kroiher<sup>1</sup>, Christopher Kraus<sup>1</sup>, Georgios D Koutsovoulos<sup>2</sup>, Sujai Kumar<sup>2</sup>, Julia I R Camps<sup>1</sup>, Ndifon A Nsah<sup>1</sup>, Dominik Stappert<sup>3</sup>, Krystalynne Morris<sup>4</sup>, Peter Heger<sup>1</sup>, Janine Altmüller<sup>5</sup>, Peter Frommolt<sup>5</sup>, Peter Nürnberg<sup>5</sup>, W Kelley Thomas<sup>4</sup>, Mark L Blaxter<sup>2</sup> and Einhard Schierenberg<sup>1</sup>

## Abstract

**Background:** The genetics of development in the nematode *Caenorhabditis elegans* has been described in exquisite detail. The phylum Nematoda has two classes: Chromadorea (which includes *C. elegans*) and the Enoplea. While the development of many chromadorean species resembles closely that of *C. elegans*, enoplean nematodes show markedly different patterns of early cell division and cell fate assignment. Embryogenesis of the enoplean *Romanormis culicivorax* has been studied in detail, but the genetic circuitry underpinning development in this species has not been explored.

**Results:** We generated a draft genome for *R. culicivorax* and compared its gene content with that of *C. elegans*, a second enoplean, the vertebrate parasite *Trichinella spiralis*, and a representative arthropod, *Tribolium castaneum*. This comparison revealed that *R. culicivorax* has retained components of the conserved ecdysozoan developmental gene toolkit lost in *C. elegans*. *T. spiralis* has independently lost even more of this toolkit than has *C. elegans*. However, the *C. elegans* toolkit is not simply depauperate, as many novel genes essential for embryogenesis in *C. elegans* are not found in, or have only extremely divergent homologues in *R. culicivorax* and *T. spiralis*. Our data imply fundamental differences in the genetic programmes not only for early cell specification but also others such as vulva formation and sex determination.

**Conclusions:** Despite the apparent morphological conservatism, major differences in the molecular logic of development have evolved within the phylum Nematoda. *R. culicivorax* serves as a tractable system to contrast *C. elegans* and understand how divergent genomic and thus regulatory backgrounds nevertheless generate a conserved phenotype. The *R. culicivorax* draft genome will promote use of this species as a research model.

**Keywords:** Nematode, Genome, Evolution, Development, *Caenorhabditis*, Mermithida, *Romanormis*

## Background

Nematodes have a generally conserved body plan. Their typical form is dictated by the presence of a single-chamber hydroskeleton, where longitudinal muscles act against an inextensible extracellular cuticle. The conservation of organ systems between nematode species is even

more striking, with, for example, the nervous system, the somatic gonad and the vulva having very similar general organisations and cellular morphologies across the phylum. It might be thought that these similarities arise from highly stereotypical developmental programmes, but comparative studies challenge this “all nematodes are equal” view.

Embryonic development of the nematode *Caenorhabditis elegans* has become a paradigmatic model for studying developmental processes in animals, including early soma-germline separation, fate specification including

\*Correspondence: philipp.schiffer@googlemail.com

<sup>1</sup>Zoologisches Institut, Universität zu Köln, Cologne, NRW, Germany, ORCID:0000-0001-6776-0934

Full list of author information is available at the end of the article

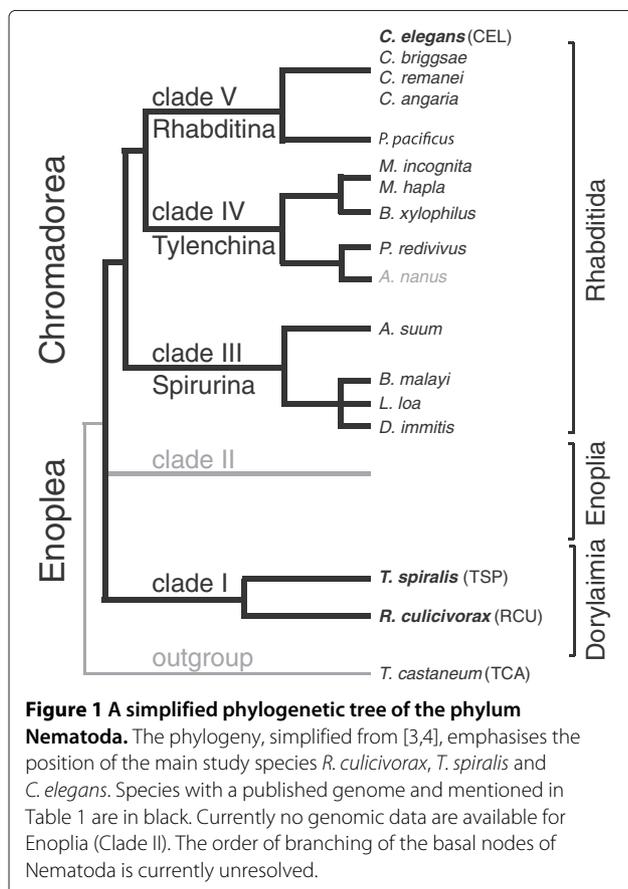
inductive interactions, and tissue-specific differentiation. The particular mode of development of *C. elegans* is distinct within the major metazoan model organisms, but much of the regulatory logic of its development is comparable to that observed in other phyla. One key aspect in which *C. elegans* differs from vertebrate and arthropod models is that *C. elegans* shows a strictly determined development [1], with a largely invariant cell-lineage giving rise to predictable sets of differentiated cells [2]. Inductive cell-cell interactions are, nevertheless, essential for its correct development [1]. *C. elegans* is a rhabditid nematode, one of approximately 23,000 described and 1 million estimated nematode species. Molecular and morphological systematics of the phylum Nematoda identify two classes: Chromadorea (including Rhabditida, and thus *C. elegans*), and Enoplea (subdivided into Dorylaimia and Enoplia) [3,4] (Figure 1). *C. elegans* is a chromadorean, and most investigation of developmental biology of nematodes has been carried out on Chromadorean species. The first description of the early embryonic cell-lineage of a nematode, that of *Ascaris* (Spirurina within Chromadorea) in the 1880's [5,6], conforms to the *C. elegans* model. Early development across all suborders of the Rhabditida is very similar [7,8]. In general, only minor variations of the division pattern observed in *C. elegans*

have been described in these nematodes [9,10], including heterochrony in the timing of cell divisions, and restrictions in cell-cell interaction due to different placement of blastomeres in the developing embryo. From these observations it might be assumed that all nematodes follow a *C. elegans*-like pattern of development. However, deviations from the *C. elegans* pattern observed in other rhabditid nematodes show that the strictly determined mode of development is subject to evolutionary change, making it particularly attractive for the study of underpinning regulatory logic of developmental mechanisms. Indeed, a greater role for regulative interactions in early development has been demonstrated in another rhabditid, *Acrobeloides nanus* (Tylenchina) [11,12].

Regulative development is common among Metazoa, and is also observed in other Ecdysozoa, including Arthropoda. Indeed, in several enoplean species, early embryos have been found to not display polarised early divisions, arguing against a strongly determined mode of development in this group [13,14]. The determined mode found in *C. elegans* is thus likely to be derived even within Nematoda [15], implying that the core developmental system in Nematoda has changed, while maintaining a very similar organismal output. This phenomenon, termed "developmental system drift" [16], reveals independent selection on the developmental mechanism and the final form produced.

To explore the genetics of development of enoplean and other non-rhabditid nematodes requires tractable experimental systems with a suitable set of methodological tools and extensive genomic data. While *C. elegans* and its embryos are relatively easily manipulated and observed, and the *C. elegans* genome has been fully sequenced [17], embryos from the Enoplia and Dorylaimia are much harder to culture and manipulate. Few viable laboratory cultures exist and obtaining large numbers of embryos from wild material is difficult. Functional molecular analysis of most nematodes, in particular Enoplea, is further hindered by the lack of genetic tools such as mutant analysis or gene-knockdown via RNAi. Performing detailed comparative experimental embryology on a phylogenetically representative set of species across the phylum Nematoda thus remains a distant goal.

The genetic toolkit utilised by a species is represented in its genome, and direct assessment of the genetic capabilities of an organism can thus be assessed through analysis of genome data. Using the background knowledge of pathways and modules used in other taxa, the underpinning logic of a species' developmental system can be inferred from its genome, and the developmental toolkits of different species can be compared. These comparisons can reveal changes in developmental logic between taxa by identifying gene losses during evolution that must result in changed pathway functioning, and similarly identify



genes recruited to developmental regulatory roles in particular lineages.

Efficient generation of genomic resources for non-model species, and the inference of developmental regulatory pathways from the encoded gene sets, is now possible. The majority of the fifteen nematode genomes published to date have been from Rhabditida (Figure 1) [18-26]. The single enoplean genome sequences is from the mammalian parasite *Trichinella spiralis* (Dorylaimia; order Trichocephalida) [27]. *T. spiralis* is ovoviviparous, proper development requires intrauterine environment, and early blastomeres are extremely transparent [28] such that individual nuclei are hard to identify (E.S., unpublished observations). Hence, this species is of very limited value for light microscopical image analysis and experimental investigation correlating cell dynamics with the molecular circuitry regulating early development.

Although the genomes of many additional nematode species are being sequenced [29,30], even in this wider sampling of the phylum, Enoplea remains neglected. The enoplean *Romanomermis culicivorax* (order Mermithida within Dorylaimia) has been established in culture for decades. It infects and kills the larvae of many different mosquito species [31], and is being investigated for its potential as a biocontrol agent of malaria and other disease vectors [31,32]. *R. culicivorax* and *T. spiralis* differ fundamentally in many life-cycle and phenotypic characters. *R. culicivorax* reproduces sexually. A single female can produce more than a thousand eggs, and embryos are easily studied under laboratory conditions. They display a developmental pattern that differs markedly from *C. elegans*. As in other Enoplea [14,33] the first division is equal, and not asymmetric as in *C. elegans*. *R. culicivorax* also shows an inversion of dorso-ventral axis polarity compared to *C. elegans*, while a predominantly monoclonal fate distribution indicates fewer modifying inductions between blastomeres [33,34]. Generation of the hypodermis involves repetitive cell elements extending from posterior to anterior over the remainder of the embryo, a process distinct from that observed in *C. elegans* [34].

We here catalogue the *R. culicivorax* developmental toolkit derived from annotation of a draft genome sequence. We contrast genes and proteins identified in *R. culicivorax* and *T. spiralis* with those of *C. elegans*, and other Ecdysozoa, represented by the arthropod *Tribolium castaneum*. We conclude that major changes in the regulatory logic of development have taken place during nematode evolution, possibly as a consequence of developmental system drift, and that the model species *C. elegans* is considerably derived compared to an ecdysozoan (and possibly metazoan) ground system. However, we are still able to define conserved gene sets that may act in “phylotypic” developmental stages.

## Results and discussion

### *Romanomermis culicivorax* has a large and repetitive genome

A draft genome assembly for *R. culicivorax* was generated from 26.9 gigabases (Gb) of raw data (filtered from a total of 41 Gb sequenced; Additional file 1: Table S1). The assembly has a contig span of 267 million base pairs (Mb) and a scaffold span of 323 Mb. The 52 Mb of spanned gaps are likely inflated estimates derived from use of the SSPACE scaffolder. We do not currently have a validated independent estimate of genome size for *R. culicivorax*, but preliminary measurements with Feulgen densitometry suggest a size greater than 320 Mb (Elizabeth Martínez Salazar pers. comm.). The *R. culicivorax* genome is thus three times bigger than that of *C. elegans*, and five times that of *T. spiralis* (Table 1). The assembly is currently in 62,537 scaffolds and contigs larger than 500 bp, with an N50 of 17.6 kilobases (kb). The N50 for scaffolds larger than 10 kb is 29.9 kb, and the largest scaffold is over 200 kb. The GC content is 36%, comparable to 38% of *C. elegans* and 34% in *T. spiralis*. We identified 47% of the *R. culicivorax* genome as repetitive. To validate this estimate we applied our repeat-finding approach to previously published genomes and achieved good accordance with these data (Table 1). The non-repetitive content of the *R. culicivorax* genome is thus approximately twice that of *C. elegans* and three times that of *T. spiralis*. *T. spiralis* thus stands out as having the least complex nematode genome sequenced so far, and the contrast with *R. culicivorax* indicates that small genomes are not characteristic of Dorylaimia.

We generated 454 Sequencing transcriptome data from mixed adults, and assembled 29,095 isotigs in 22,418 isogroups, spanning 23 Mb. These are likely to be a reasonable estimate of the *R. culicivorax* transcriptome. Using BLAT [35], 21,204 of the isotigs were found to be present (with matches covering >80% of the isotig) in single contigs or scaffolds of the genome assembly, suggesting reasonable biological completeness and contiguity of the genome. We also used the CEGMA [36] approach to assess quality of the genome assembly, and found a high representation (90% partial, 75% complete) and a low proportion of duplicates (1.1 fold) (Table 2). Automated gene prediction with iterative rounds of the MAKER pipeline [37], using the transcriptome data as evidence both directly and through GenomeThreader-derived mapping, yielded a total of nearly 50,000 gene models. These were reduced to 48,171 by merging those with identities >99% using Cd-hit [38]. Within the 48,171 models, 12,026 were derived from the AUGUSTUS modeller [39] and 36,145 from SNAP. Because AUGUSTUS predictions conservatively require some external evidence (transcript mapping and/or sequence similarity to other known proteins), we regarded these as the most reliable

**Table 1 Genome statistics**

Species	Approximate <sup>#</sup> genome size	Estimated repeat content	Median <sup>†</sup> exon length	Median <sup>†</sup> intron length	GC content	Source
<i>C.elegans</i>	100Mb	17% (16.5%)	145bp	69bp	38%	[17,18]
<i>P.pacificus</i>	165Mb	15.3% (17%)	85bp	141bp	42%	[20,25]
<i>A.suum</i>	334Mb	4.4%	144bp	907bp	37.9%	[21,40]
<i>B.malayi</i>	95Mb	16.5% (15%)	140bp	219bp	30%	[22]
<i>B.xylophilus</i>	69Mb	22.5%	183bp	69bp	40%	[25]
<i>M.incognita</i>	~200Mb	36.7%	136bp	82bp	31%	[24]
<i>T.spiralis</i>	63Mb	19.8% (18%)	128bp	283bp	34%	[27]
<i>R.culicivora</i>	>270Mb	48.2%	161bp	405bp	36%	this work

Repeat content of different nematode genomes appears not to be directly correlated with genome size. Re-calculation in selected genomes shows little deviance from published data (in parentheses)\* and thus indicates the validity of our inference for *R. culicivora*.

\*For *B. xylophilus* and *M. incognita* only reference data is given as the same programs were used for initial inference (see references); *A. suum* not re-calculated.

#*M. incognita* genome size given as 86Mbp in [24] has been re-estimated to about 150Mbp (E. Danchin pers. comm.).

†Median lengths for *A. suum* and *T. spiralis* were calculated in this work as these data are not given in the cited publications.

and biologically complete. In comparison *C. elegans* has ~22,000 genes, and *T. spiralis* has ~16,000. The satellite model nematode *Pristionchus pacificus* has ~27,000 genes [20]. Exons of the AUGUSTUS-predicted genes in *R. culicivora* had a median length of 161 bp, slightly larger than those in *C. elegans* (137bp) and *T. spiralis* (128bp). Introns of the *R. culicivora* AUGUSTUS models, with a median of 405 bp, were much larger than those of *C. elegans*

(69 bp) or *T. spiralis* (283bp). The small introns observed in *C. elegans* and other rhabditid nematodes (Table 2) are thus likely to be a derived feature.

We annotated 1,443 tRNAs in the *R. culicivora* genome using INFERNAL [41] and tRNAscan-SE [42], of which 382 were pseudogenes (see Additional file 1: Table S2 for details). In comparison, *T. spiralis* has 134 tRNAs of which 7 are pseudogenes, while *C. elegans* has 606 tRNAs with 36 pseudogenes [43]. Threonine (Thr) tRNAs were particularly overrepresented (676 copies), a finding echoed in the genomes of *Meloidogyne incognita* and *Meloidogyne floridensis* [24,43] and in *P. pacificus* [20]. The latter has also an overrepresentation of Arginine tRNAs [43].

We have made available the annotated *R. culicivora* genome, with functional categorisations of predicted genes and proteins and annotation features, in a dedicated genome browser at <http://romanomermis.nematod.es>.

#### The *R.culicivora* gene set is more representative of Dorylaimia than *T. spiralis*

The phylogenetic placement of *R. culicivora* makes its genome attractive for exploring the likely genetic complexity of an ancestral nematode. With *T. spiralis*, it can be used to reveal the idiosyncrasies of the several genomes available for Rhabditida. To polarise this comparison, we used the arthropod *Tribolium castaneum*, for which a high quality genome sequence is available [44]. *T. castaneum* development is considered less derived than that of the major arthropod model *Drosophila melanogaster* [45]. The OrthoMCL pipeline accurately clusters orthologous proteins, facilitating the complex task of grouping proteins that are likely to share biological function in divergent organisms [46], and performs better than approaches that simply use domain presence information or aggregative approaches such as psiBLAST [47]. We used the

**Table 2 Assembly and annotation statistics**

Metric	Result
Contigs > 100bp span	267,342,457bp
Scaffolds > 500bp span	322,765,761bp
Num. contigs/scaffolds	62,537
N50 contigs/scaffolds > 500bp	17,632 bp
N50 scaffolds > 500bp	29,995bp
Max contig length	28,847bp
Max scaffold length	201,054bp
Mean transcript length	593bp
Mean protein length	190aa
MAKER AUGUSTUS predictions	12,026 proteins
MAKER SNAP predictions	36,145 proteins
Num. ESTs (isogroups)	22,418 ESTs
Mean EST length	330bp
80% BLAT sequence coverage	21,204 ESTs
CEGMA compl. completeness	75.40%
CEGMA Group 1 part. compl.	81.82%
CEGMA Group 2 part. compl.	91.07%
CEGMA Group 3 part. compl.	91.80%
CEGMA Group 4 part. compl.	95.38%

OrthoMCL pipeline to generate a set of protein clusters for the four species (*R. culicivora*x, *T. spiralis*, *C. elegans* and *T. castaneum*). While the large divergence between these species may obscure relationships between protein sequences, making inference of orthology problematic [48-50], the parameters used were most inclusive [50-52]. Additionally, as the *R. culicivora*x genome assembly may not be complete we based inference of absence on shared loss in both *R. culicivora*x and *T. spiralis*. Additionally, we validated inferences of absence from the OrthoMCL analyses by performing detailed sequence comparisons using BLAST+ [53] (Additional file 2).

We identified 3,274 clusters that contained protein representatives from all three nematodes, and 2,833 of these also contained at least one *T. castaneum* representative (Figure 2). These 2,833 clusters represent a conserved ecdysozoan (and possibly metazoan) core proteome. Many clusters had *T. castaneum* members, and members from some but not all of the three nematodes, representing candidate examples of loss in one or more nematode lineages of ancient proteins. For example, we identified clusters containing proteins from only one of the nematode species. *T. spiralis* had the lowest number of these (975), while *C. elegans* and *R. culicivora*x each had over two thousand. Interestingly, of the 2,747 clusters with only *R. culicivora*x proteins from Nematoda, 324 included *T. castaneum* orthologues, whereas *C. elegans* only shared 283 clusters uniquely with the beetle. *T. spiralis* has lost more of these phylogenetically ancient genes than has either *R. culicivora*x or *C. elegans*. *T. spiralis* and *C. elegans* shared only 412 clusters exclusive of *R. culicivora*x members, while *R. culicivora*x and *C. elegans* shared about 1300 clusters exclusive of *T. spiralis*. Despite their phylogenetic affinity, *R. culicivora*x

and *T. spiralis* only shared 600 clusters exclusive of *C. elegans* (Figure 2). We suggest that *T. spiralis* genome is not typical of dorylaims. In comparison to other nematodes it is smaller, has fewer genes overall, and has fewer phylogenetically ancient genes. This is congruent with the previously reported loss of proteins with metabolic function in *T. spiralis* [27]. This reduction in genetic complexity could be due to evolutionary pressures following acquisition of a lifestyle that lacks a free-living stage. Many parasitic and endosymbiotic prokaryotes and eukaryotes have reduced genome sizes, though this is not an absolute rule [54].

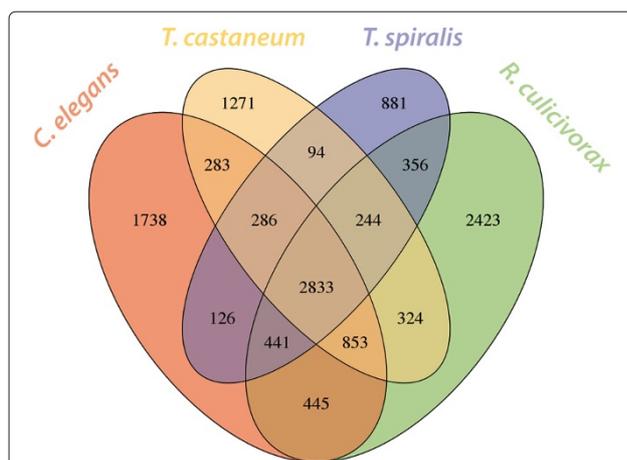
Clusters containing only *R. culicivora*x and *T. spiralis* proteins might identify functions distinct to these dorylaim nematodes. In the 461 *T. spiralis* and 806 *R. culicivora*x proteins in these clusters, a total of 65 GO terms were found to be overrepresented (single test  $p < 0.05$  by Fisher's exact test). While *C. elegans* has a reduced ability to methylate DNA [55], we found four methylation-associated GO terms among the 64 overrepresented. We also detected significant enrichment (single test  $p < 0.05$ ) for GO terms describing chromatin and DNA methylation functions in the set of *R. culicivora*x proteins that lacked homologues in *C. elegans* (see Additional file 3). Important roles for methylation and changes in methylation patterns during *T. spiralis* development have been inferred from transcriptional profiling [56]. Methylation is important for the silencing of transposable elements [57,58] and could play a crucial role in the highly repetitive *R. culicivora*x genome.

The clusters that contained *R. culicivora*x, *T. spiralis* and *T. castaneum* proteins but no *C. elegans* orthologues might contain proteins involved in core ecdysozoan processes lost in *C. elegans*. In these clusters we identified 40 GO terms overrepresented (single test  $p < 0.05$ ) compared to the *C. elegans* proteome (see Additional file 3). Some of these GO terms were linked to chromatin remodelling and methylation (e.g. Ino80 complex, histone arginine methylation). Other overrepresented GO terms were related to cell signalling (the Wnt receptor pathway; the *C. elegans* Wnt signalling system is distinct from other metazoa [59]), and ecdysone receptor holocomplex (potentially a basic ecdysozoan function [60]).

### The genetic background of development in *R. culicivora*x and *T. spiralis* differs markedly from that of *C. elegans*

In a recent multi-species developmental time course expression analysis within several *Caenorhabditis* species, conserved sets of genes were found to have conserved patterns of differential expression in discrete phases in the timeline from zygote to the hatching larva [61].

Nearly half (845) of these 1725 conserved, differentially expressed *C. elegans* proteins were not clustered with *R. culicivora*x or *T. spiralis* proteins using OrthoMCL. We



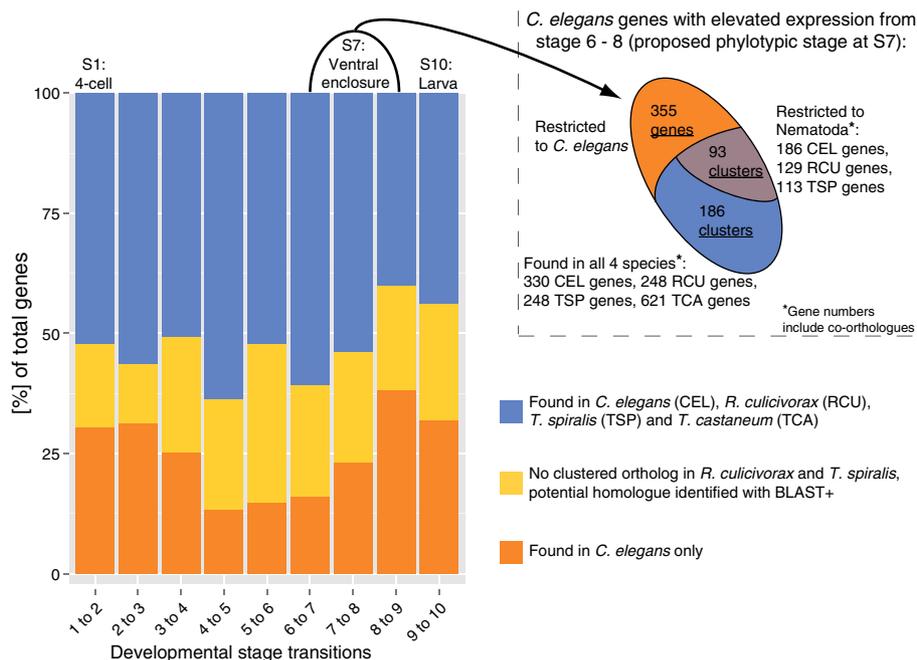
**Figure 2 Clusters of homologous proteins.** Shared and species-unique clusters of homologous proteins from a comparison of the proteomes of *Romanomeris culicivora*x, *Trichinella spiralis*, *Caenorhabditis elegans* and *Tribolium castaneum* using OrthoMCL.

were unable to identify any sequence similarity for 450 of these *C. elegans* proteins, while 395 had only marginal similarities insufficient for OrthoMCL clustering. Eighteen of these 395 are members of *C. elegans* nuclear hormone receptor subfamilies, 5 are innexin type gap-junction proteins, 6 are TWiK potassium channel proteins and 5 are acetylcholine receptor proteins. These protein families are particularly diverse and expanded in *C. elegans* [62-65] and we suggest that they represent rapidly evolved, divergent duplications within the lineage leading to *C. elegans*. The proportion of *Caenorhabditis*-restricted genes across the developmental time course examined by Levin et al. [61] varied from 36% to 60% (Figure 3 and Additional file 4). Thus a surprisingly high proportion of *Caenorhabditis* genes with conserved expression during embryogenesis appear to be unique to the genus or are so divergent that we could not detect possible orthologues in the dorylaids. The pattern of higher retention of conserved genes in *R. culicivora* compared to *T. spiralis* was also evident in these conserved-expression developmental genes, as 238 had *R. culicivora* orthologues but lacked a *T. spiralis* orthologues. Given the conservatism of body plan evolution in nematodes, these dramatic genetic differences suggest extensive, largely phenotypically “silent” changes in the genetic programmes orchestrating nematode development.

### Core developmental pathways differ between nematodes

There are important differences in cell behaviour during early embryogenesis between *R. culicivora* and *C. elegans* [33,34]. We used the genomic data to follow up on some of the striking contrasts between the dorylaids and the rhabditid patterns of development: establishment of primary axis polarity, segregation of maternal message within the early embryo, hypodermis formation, the vulval specification pathways, epigenetic pathways (especially DNA methylation), sex determination and light sensing (see Additional file 1).

The mechanisms of sex determination differ considerably among animals and it has been claimed to be one of the developmental programs most influenced by developmental system drift [16]. Divergence in sex determination pathways is thus not unexpected. While sex is determined by X to autosome ratio in *C. elegans* [66], sex ratios in *R. culicivora* are likely to be environmentally determined through in-host nematode density [67]. Environmental sex determination is found in many nematode taxa, including Strongyloididae and Meloidogyninae (both Tylenchina), taxa more closely related to *C. elegans*. In *C. elegans*, the X to autosome ratio is read by the master switch XOL-1 [68], which acts through the three *sdc* genes [69-71] to regulate the secretion of HER-1, a ligand for the TRA-2 receptor [72-74]. TRA-2 in turn



**Figure 3** Many genes that are developmentally important in *Caenorhabditis elegans* were not present in *Romanormis culicivora* or *Trichinella spiralis*. *R. culicivora* and *T. spiralis* orthologues of the 1,725 genes identified as important in embryogenesis in an analysis of gene expression in *Caenorhabditis* species [61] were sought. For each embryonic stage (1-10) in *C. elegans* we calculated the proportion of these genes that were apparently unique to the genus *Caenorhabditis*. The inset depicts numbers of two sets of genes and corresponding clusters that could play a role in a phylotypic stage of Nematoda or all Ecdysozoa, respectively.

negatively regulates a complex of *fem* genes, which regulates nuclear translocation of TRA-1, the final shared step in the pathway that switches between male and hermaphrodite systems. No credible homologues of XOL-1, SDC-1, SDC-2, SDC-3, HER-1 or TRA-2 in either *T. spiralis* or *R. culicivora*x were detected through OrthoMCL and re-confirmation with BLAST+ (Table 3; Additional file 2), and thus these species are unlikely to use the HER-1 – TRA-2 ligand-receptor system to coordinate sexual differentiation.

Other developmental processes are however more conserved between metazoan taxa. In *C. elegans* and many other animals *par* genes are essential for cell polarisation [75]. Polarised distribution of PAR proteins results in the restriction of mitotic spindle rotation to the germline cell in the *C. elegans* two-cell stage [76-78]. This rotation is not observed in *R. culicivora*x [33]. The division pattern of *C. elegans* mutants lacking *par-2* and *par-3* genes resembles that of the early *R. culicivora*x embryo [33,79]. The *par-2* gene was absent from both *R. culicivora*x and *T. spiralis* (Figure 4; Table 3). Additionally, no orthologues for the *par-2*-interacting genes *let-99*, *gpr-1* or *gpr-2*, required for proper embryonic spindle orientation in *C. elegans* [80], were identified in the dorylaims using OrthoMCL clustering or sensitive BLAST+ searches. Although we identified a protein with weak similarity to *par-3* in *R. culicivora*x, this was so divergent from *C. elegans*, *T. castaneum* and *T. spiralis* *par-3* that it was not clustered in our analysis. In *D. melanogaster* a *par-3* orthologue, *bazooka*, functions in anterior-posterior axis formation [81], but *par-2* is absent from the fly. Thus, we hypothesise that the PAR-3/PAR-2 system for regulating spindle positioning evolved within the lineage leading to the genus *Caenorhabditis*. If the divergent *par-3*-like gene in *R. culicivora*x is involved in axis formation, it probably interacts with different partner proteins.

Once polarity has been established in the early *C. elegans* embryo, many maternal messages are differentially segregated into anterior or posterior blastomeres [78,82]. MEX-3 is an RNA-binding protein translated from maternally-provisioned mRNAs found predominantly in early anterior blastomeres [83,84]. We identified a highly divergent MEX-3 orthologue in *R. culicivora*x, but no orthologue in *T. spiralis*. We explored embryonic expression of *mex-3* in *R. culicivora*x embryos using *in situ* hybridisation (Figure 5). In the fertilized egg the *mex-3* mRNA is initially equally distributed. Prior to first cleavage it is segregated to the anterior pole and thus becomes essentially restricted to the somatic S1 blastomere (for nomenclature, see [14]). With the division of S1 it is localized to both daughter cells. After the 4-cell stage the signal disappears gradually. This expression pattern is similar to that of *C. elegans* *mex-3*, affirming that the *R. culicivora*x gene is likely to be an orthologue retaining

**Table 3 Presence and absence of selected\* *C. elegans* proteins in Dorylaimia**

Protein	<i>T. spiralis</i>	<i>R. culicivora</i> x
<b>Early asymmetry</b>		
CDC-42	+	+
PKC-3	+	+
GPR-1	-	-
GPR-2	-	-
PAR-2	-	-
PAR-6	+	+
MES-6	+	+
MES-3	-	-
MES-4	-	-
GFL-1	+	+
LET-70	+	+
<b>Axis formation</b>		
NUM-1	+	+
ZIM-1	-	-
MES-2	-	-
POS-1	-	-
SMA-6	+	+
SET-2	-	-
UBC-18	+	+
LET-99	-	-
OOC-3	-	-
OOC-5	+	+
GPA-16	+	+
PAR-5	-	-
ATX-2	-	-
MEX-5	-	-
MEX-6	-	-
UNC-120	-	-
NOS-2	-	-
OMA-1	-	-
RME-2	+	+
SPN-4	-	-
<b>Sex determination</b>		
XOL-1	-	-
HER-1	-	-
SEX-1	+	+
FOX-1	+	+
SDC-1	-	-
SDC-2	-	-
SDC-3	-	-
TRA-2	-	-
FEM-1	+	+
FEM-2	+	+

**Table 3 Presence and absence of selected\* *C. elegans* proteins in Dorylaimia (Continued)**

Protein	<i>T. spiralis</i>		<i>R. culicivora</i> x
	Hypodermis and vulva formation		
AFF-1	-	-	-
BAR-1	-	-	-
CEH-2	-	-	-
CEH-27	-	-	-
GRL-15	-	-	-
INX-5	-	-	-
LIN-1	-	-	-
PEB-1	-	-	-
ELT-3	-	-	-
ELT-1	+	-	+
SMA-3	-	-	-
SMA-5	-	-	-

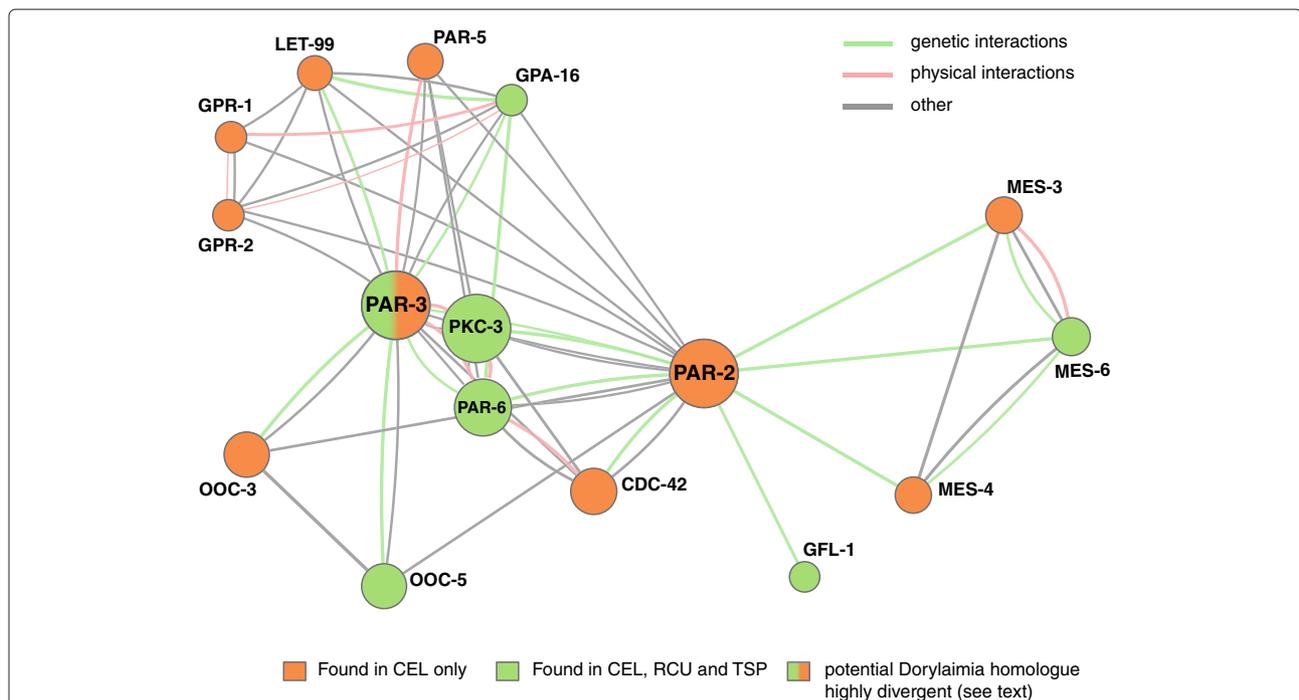
\*For additional proteins see Additional files 2 and 4.

similar functions. However, despite the presence, and apparent conservation of the *mex-3* expression pattern, we were unable to identify other interacting partners of the *C. elegans* MEX-3 protein, such as MEX-5, MEX-6 and SPN-4 in either dorylaim species. While MEX-5

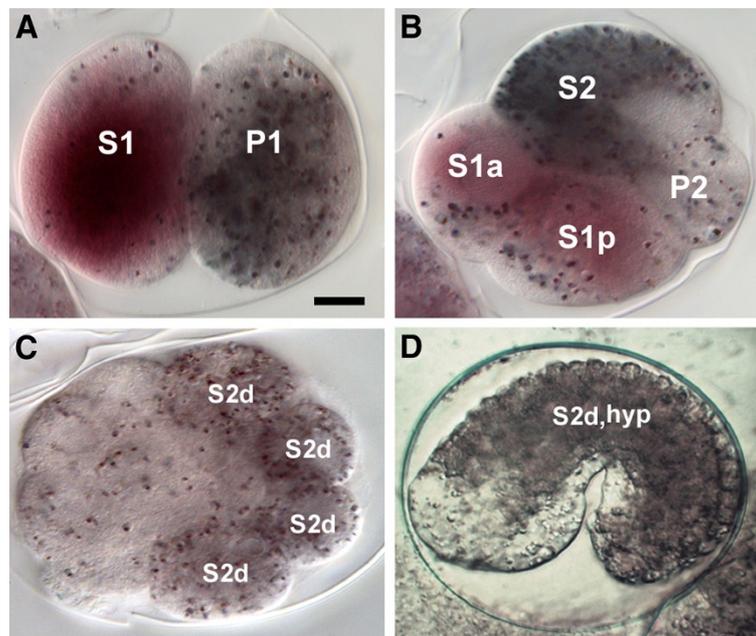
and MEX-6 are important for controlled MEX-3 expression in *C. elegans* [85], the apparent absence of SPN-4 in *R. culicivora*x and *T. spiralis* is particularly intriguing. SPN-4 links embryonic polarity conferred by the *par* genes and partners to cell fate specification through maternally deposited mRNAs and proteins [86,87]. Our findings suggest that the core regulatory logic of the early control of axis formation and cell fate specification must differ significantly between the dorylaim species and *C. elegans*.

The hypodermis in *C. elegans* is derived from specific descendants of the anterior and posterior founder cells [88]. In contrast, in *R. culicivora*x hypodermis is derived from descendants of a single cell [34]. Several *C. elegans* genes expressed in the hypodermis or associated with hypodermal development were absent from *R. culicivora*x and *T. spiralis* (see Table 3 and Additional file 3). For example the GATA-like transcription factors ELT-1 and ELT-3 act redundantly in *C. elegans* [89]. ELT-3 was absent from the dorylaim species, but ELT-1 was conserved in *R. culicivora*x, *T. spiralis* and *T. castaneum*. Thus, ELT-3 appears to be an innovation in the rhabditid lineage, suggesting changes of interaction complexity during nematode evolution.

In *C. elegans*, vulva formation is highly dependent on initial inductive signals from the anchor cell that activate



**Figure 4 The network of proteins interacting with PAR-2 and PAR-3 in *Caenorhabditis elegans* and their orthologues in *Romanomeris culicivora*x and *Trichinella spiralis*.** The network cartoon is based on the core polarity pathway extracted from WormBase, derived from both genetic and physical interactions. PAR-2 was missing from the dorylaim nematodes, as were the directly connected MES-3 and MES-4 genes. The *R. culicivora*x PAR-3-like protein was not retrieved as an orthologue of *C. elegans* and *T. spiralis* PAR-3 proteins, but was identified employing sensitive sequence similarity search. See Table 3 for additional proteins interacting with PAR proteins and their presence-absence patterns.



**Figure 5** *In situ* hybridisation mapping of *mex-3* mRNA distribution in early embryos of *Romanormermis culicivorax*. We used the *R. culicivorax mex-3* gene to prove application of the *in situ* technique in this species and investigated the segregation patterns of this maternal RNA in early development. The *R. culicivorax mex-3* expression pattern is similar to that of *C. elegans* [83]. *R. culicivorax* embryos contain dark pigment granules that are asymmetrically segregated in development. **(A)** At the 2-cell stage, maternal *mex-3* mRNA is detected in the S1 blastomere. The cytoplasmic pigment granules are predominantly in the P1 blastomere. **(B)** At the 4-cell stage, *mex-3* mRNA is detected in daughters of the anterior S1 cell. Cytoplasmic pigment granules are predominantly in the S2 blastomere. **(C)** At a later stage (>20 cells), *mex-3* mRNA is absent. The pigment granules are found in descendants of S2 (S2d). **(D)** During early morphogenesis, the pigment granules are found in S2 descendants forming hypodermis, (S2d, hyp). **(A-C)** fixed embryos; **(D)** live embryo. Scale bar 10  $\mu$ m. Orientation: anterior left.

a complex gene regulatory network, which drives tissue specific cell division and differentiation. The evolutionary plasticity of this system has been explored in rhabditid nematodes, revealing the changing relative importance of cell-cell interactions, inductions, and lineage-autonomous specifications [90,91]. The signal transduction pathways include a RTK/RAS/MAPK cascade, activated by EGF- and wnt-signalling [92]. Among the downstream targets in *C. elegans* are for example LIN-1 and the  $\beta$ -catenin BAR-1, which in turn regulates the HOX-5 orthologue LIN-39 [93-95]. These important regulators of vulva development are completely absent from the genomes of *R. culicivorax* and *T. spiralis* (Table 3 and Additional file 2). We identified a *R. culicivorax* protein with low similarity to *C. elegans* BAR-1 (24% sequence identity). However, this protein is not clustered with other dorylaim proteins, and appears to be either a duplication of the  $\beta$ -catenin ortholog HMP-2 or another armadillo repeat-containing protein rather than an orthologue of BAR-1 (see Additional file 5). These shared patterns of absence again suggest that similar morphological structures can be generated with very different genetic underpinnings. Vulva formation in the dorylaims may be regulated

without the BAR-1 – LIN-39 interaction, as observed in *P. pacificus* [96]. In *C. elegans* Hox gene expression is cell-lineage dependent [97,98], organised so that the cells that express specific Hox genes are clustered along the anterior-posterior axis (see e.g. [99]). It will be informative to test whether in *R. culicivorax* and other non-rhabditid nematodes Hox genes act in an axis position-dependent, but cell lineage-independent manner, as observed in many other animals, notably arthropods [100,101]. Epigenetic regulation is key to developmental processes in many animals, but its roles in *C. elegans* are more muted (see above). Notably *C. elegans* is depleted for chromatin remodelling genes of the Polycomb and Trithorax groups [102]. It is intriguing that we found orthologues of *T. castaneum pleiohomeotic* in *R. culicivorax* and *T. spiralis*, and orthologs of *T. castaneum trithorax* and *Sex comb on midleg (Scm)* in *R. culicivorax*. This suggests that dorylaim chromatin restructuring mechanisms may be more arthropod-like than in *C. elegans*. The presence of an intact methylation machinery and conserved chromatin re-modelling factors opens the prospects for a role for epigenetic modification in developmental regulation of dorylaim nematodes.

### Defining a set of potential phylotypic stage genes

While the examples above demonstrate considerable developmental system drift in Nematoda, we also identified many sets of orthologous proteins conserved between *Dorylaimia* and *C. elegans*. We asked if these could be correlated with functions in distinct developmental phases with a conserved phenotype. Shortly before the start of morphogenesis, at the point of ventral enclosure, nematode embryos from Chromadorea and Enoplea share a similar morphology [14]. Levin et al. [61] found that in five *Caenorhabditis* species a distinct set of genes had elevated expression around the ventral enclosure stage (their stage 7) (Figure 3) and proposed that this constitutes a "phylotypic stage" for nematodes. We used *T. spiralis* and *R. culicivora*x gene sets to refine and restrict this set of phylotypic stage genes. Of the 834 *C. elegans* genes with elevated expression between stages 6 to 8 [61], 355 had no orthologue in *R. culicivora*x, *T. spiralis* or *T. castaneum*. The remaining 479 phylotypic stage candidates from *C. elegans* were present in 279 of our OrthoMCL clusters. Of these clusters 93 were nematode-restricted containing 186 *C. elegans* proteins grouped with 129 *R. culicivora*x and 113 *T. spiralis* homologues. The remaining 186 clusters were part of the conserved ecdysozoan core proteome (see above) and contained 330 *C. elegans* proteins together with 248 *R. culicivora*x, 248 *T. spiralis* and 621 *T. castaneum* proteins (Figure 3; The total number of *C. elegans* candidates is larger than 479 due to the inclusion of co-orthologues in this species). In the set of phylotypic stage genes identified by Levin et al. [61] are proteins functioning in processes such as muscle and neuron formation, signalling between cells, and morphogenesis. This pattern was retained in the conserved clusters (see Additional file 5). Although time-resolved expression data will be needed to confirm the activity of these genes in developmental stages of *R. culicivora*x, their retention in the *Dorylaimia* supports their general importance. We can now sub-classify the set of conserved proteins expressed at the potential nematode phylotypic stage. A first, nematode-restricted set includes many proteins that are important for cuticle formation (e.g. collagen proteins) and some hedgehog-like proteins, expressed in the *C. elegans* hypodermis [103]. As cuticle formation follows ventral enclosure in nematodes, these proteins may be involved in this nematode-specific function. The second set, comprising clusters conserved between the nematodes and *T. castaneum*, contains many important developmental transcription factors, such as the Hox gene *mab-5*, other homeobox genes, and helix-loop-helix and C2H2-type zinc finger transcription factors. This second set may represent a genetic backbone driving formation of phylotypic stage in diverse animal taxa, in accordance with the recent extension of the concept to Metazoa [104-106].

### Conclusions

To be useful as a contrasting system to the canonical *C. elegans* model, any nematode species must be accessible to both descriptive and manipulative investigation. The reference genome for *R. culicivora*x lays bare the core machinery available for developmental regulation, and we have demonstrated that *in situ* hybridisation approaches are feasible for this species. Along with the long established, robust laboratory cultures, this makes *R. culicivora*x an attractive and tractable alternative model for understanding the evolutionary dynamics of nematode development. By combining the *R. culicivora*x genome with that of *T. spiralis*, we have been able to explore the molecular diversity of *Dorylaimia*, and provide robust contrasts with the intensively studied Rhabditida. Particularly surprising are the differences between *R. culicivora*x and *T. spiralis*. The *R. culicivora*x genome is much larger than that of *T. spiralis*, and contained a high proportion of repetitive sequence, including many transposable elements. Despite the phylogenetic and lifestyle affinities between the two dorylaims compared to *C. elegans*, the *R. culicivora*x genome retained many more genes in common with *C. elegans* than did *T. spiralis*. We suggest that *T. spiralis* may be an atypical representative of dorylaim nematodes, perhaps due to its highly derived life cycle.

Our analyses identified many genes apparently absent from the dorylaim genomes, despite relaxed analysis parameters. In particular, for genes identified as critical to *C. elegans* development but apparently absent from the dorylaims, we were unable to identify credible orthologues using sensitive search strategies. In this phylum-spanning comparison, inferences of gene orthology can be obscured by levels of divergence. In addition, the gene family birth rate in the chromadorean lineage leading to *C. elegans* is high [25,27], and therefore *C. elegans* was expected to have many genes absent from the dorylaim species. Thus, we might not have found a *R. culicivora*x orthologue for a specific gene for three reasons: it may have arisen in the branch leading to *C. elegans*; its sequence divergence may be too great to permit clustering with potential homologs; or it was not assembled in the draft dorylaim genomes. The case of *C. elegans* PAR-3 and *D. melanogaster bazooka* illustrate some of these difficulties: the possible *R. culicivora*x orthologue was highly divergent. Whether or not we have been able to identify all the orthologues of the key *C. elegans* genes present in the *R. culicivora*x and *T. spiralis* genomes, the absence of an identified orthologue maximally implies loss from the genome, and minimally implies significant sequence divergence. In the latter case this would most likely cause changes in the networks and pathways in which genes interact to deliver biological function.

Between the model nematode *C. elegans* and arthropod models such as *T. castaneum* many key mechanisms

governing early cell patterning are divergent [76]. Our data strongly support the view that major variation also exists within Nematoda. *T. spiralis* and *R. culicivora*x both lack orthologues of genes involved in core developmental processes in *C. elegans*, and many of these *C. elegans* genes appear to be restricted to the Rhabditida. It is thus doubtful that these processes are regulated by same molecular interactions across the phylum. We suggest that developmental system drift has played a major role in nematode evolution. The phenotypic conservatism associated with the vermiform morphology of nematodes [107] has fostered unjustified expectations concerning the conservation of genetic programmes that determine these morphologies. Despite this divergence in developmental systems, we were able to define two sets of conserved genes possibly active in a taxon-specific phase of ventral enclosure and cuticle formation in Nematoda, and in a potential phylotypic stage of Ecdysozoa. The advent of robust, affordable and rapid genome sequencing also opens the vista of large-scale comparative genomics of development across the phylum Nematoda [29] to better understand the diversity of the phylum and also place the remarkable *C. elegans* model in context of its peers. It will next be necessary to extend these studies to a broader sampling of developmental pathway genes from a wider and representative sampling of nematode genomes across the full diversity of the phylum. We have highlighted a few of the possible avenues a research programme could follow: early axis formation and polarisation, the specification of hypodermis, sex determination, vulva formation, the roles of epigenetic processes in developmental regulation and the confirmation of potential “phylotypic stage genes” with expression analysis in *R. culicivora*x.

## Methods

### Sequencing and genome assembly

Genomic DNA was extracted from several hundred, mixed-sex, adult *R. culicivora*x specimens from a culture first established in Ed Platzer’s laboratory in Riverside, California. Illumina paired end and mate pair sequencing with libraries of varying insert sizes, and Roche 454 single end sequencing, was performed at the Cologne Center for Genomics (CCG: <http://www.ccg.uni-koeln.de>). A Roche 454 dataset of transcriptome reads from cDNA synthesised from mixed developmental stages and sexes was also generated (see Additional file 1: Table S1 for details of data generated).

The quality of the raw data was assessed with FastQC (v.0.9; <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter sequences and low quality data were trimmed from the Illumina paired end data with custom scripts (see <http://github.com/sujaikumar/semblage>) and from the mate pair libraries with Cutadapt (v.1.0) [108]. We constructed a preliminary

genome assembly, with relaxed insert size parameters, from the paired end Illumina libraries with the de-novo-assemble option of the *clcAssemblyCell* (v.4.03b) [109]. We validated the actual insert sizes of our libraries by mapping back the reads to this preliminary assembly using *clcAssemblyCell*. The preliminary assembly was also used to screen out bacterial and other contaminant data [110]. The transcriptome data were assembled with Roche *GSAssembler* (Newbler; version 2.5). For the production assembly, we explored assembly parameters using different mixes of our data, evaluating each for total span, maximal contig lengths, N50, number of contigs, representation of the transcriptome, and conserved eukaryotic gene content (using the CEGMA pipeline v.2.1 [36]). The most promising assembly was scaffolded with the filtered Illumina mate pair read sets using *SSPACE* (v.1.2) [111]. As our genomic DNA derived from a population of nematodes of unknown genetic diversity, we removed short contigs that mapped entirely within larger ones using *Cd-hit* (v.4.5.7) [38] at a 95% cutoff. A final round of superscaffolding was performed, linking scaffolds that had logically consistent matches to the transcriptome data based on BLAT [35] hits and processed with *SCUBAT* (B. Elsworth, pers. comm.; <http://github.com/elswob/SCUBAT>). The final genome assembly was again assessed for completeness by assessing the mapping of the transcriptome contigs and with the CEGMA pipeline [36].

### Genome annotation

*RepeatMasker* (v.3.3.0) [112,113], *RepeatFinder* [114] and *RepeatModeler* (v.1.0.5; <http://www.repeatmasker.org/RepeatModeler.html>; combining *RECON* (v.1.07) [115] and *RepeatScout* (v.1.05) [116]), were used to identify known and novel repetitive elements in the *R. culicivora*x genome. We employed the *MAKER* pipeline [37] to find genes in the *R. culicivora*x genome assembly. In a first pass, the *SNAP* gene predictor included in *MAKER* was trained with a CEGMA [36] derived output of predicted highly conserved genes. As additional evidence we included the transcriptome assembly and a set of approximately 15,000 conserved nematode proteins derived from the NEMBASE4 database [117] (recalculated by J. Parkinson; pers. comm.). In the second, definitive, pass we used the gene set derived from this first *MAKER* iteration to train *AUGUSTUS* [39] inside the *MAKER* pipeline for a second run, also including evidence from transcriptome to genome mapping obtained with *GenomeThreader* [118]. Codon usage in *R. culicivora*x, *T. spiralis*, and *C. elegans* was calculated using *INCA* (v2.1) [119]. Results were then compared to data from [120] (see Additional files 1 and 6).

We used *Blast2GO* (*Blast2GO4Pipe*, v.2.5, January 2012 database issue) [121] to annotate the gene set with Gene

Ontology terms [122], based on BLAST matches with expect values less than  $1e^{-5}$  to the UniProt/SwissProt database (March 2012 snapshot), and domain annotations derived from the InterPro database [123]. Comparison of annotations between three nematode species (*R. culicivora*, *C. elegans*, and *T. spiralis*) and, as a reference outgroup, the holometabolous coleopteran arthropod *Tribolium castaneum*, was based on GO Slim data retrieved with Blast2GO. RNA genes were predicted using INFERNAL (v.1.0.2)[41] and the Rfam database [124], and tRNAscan-SE (v.1.3.1) [42].

### Orthology screen

We inferred clusters of orthologous proteins between *R. culicivora*, *T. spiralis*, and *C. elegans*, and the beetle *T. castaneum* using OrthoMCL (v.2.0.3) [125]. *T. spiralis*, *C. elegans* and *T. castaneum* protein sets were downloaded from NCBI and WormBase (see Additional file 1: Table S3) and redundancy screened with Cd-hit at the 99% threshold. We selected an inflation parameter of 1.5 for MCL clustering (based on [126,127]) within OrthoMCL to generate an inclusive clusterings in our analysis likely to contain even highly diverged representatives from the four species. In analyses of selected developmental genes, clusters were manually validated using NCBI-BLAST+ [53]. We affirmed the uniqueness of *C. elegans* proteins identified as lacking homologues in the enoplean nematodes by comparing them to the *R. culicivora* proteome using BLAST. Those with no significant matches at all (all matches with E-values  $> 1e^{-5}$ ) were classified as confirmed absent. Those having matches with E-values  $< 1e^{-5}$  were investigated further by surveying the cluster memberships of the *R. culicivora* matches. If the *R. culicivora* protein was found to cluster with a different *C. elegans* protein, the uniqueness to *C. elegans* was again confirmed. If the *R. culicivora* protein did not cluster with an alternative *C. elegans* protein, we reviewed the BLAST statistics (E-value, identity and sequence coverage) of the match and searched the GenBank non redundant protein database for additional evidence of possible orthology. Only if these tests yielded no indication of direct orthology was the *C. elegans* protein designated absent from the enoplean set. Further details of the process are given in Additional file 5.

We identified the protein sequences of 1,725 genes differentially expressed in *C. elegans* developmental stages [61] and selected, using our OrthoMCL clustering, those apparently lacking orthologues in *R. culicivora* and *T. spiralis* (verified as above). Using Wormbase (<http://www.wormbase.org>, release WS233) we surveyed the *C. elegans*-restricted genes for their experimentally-defined roles in development.

Custom Perl scripts were used to group orthoMCL clusters on the basis of species membership patterns. The

sets of clusters that contained (i) both *T. spiralis* and *R. culicivora* members but no *C. elegans* members and (ii) *T. spiralis* and *R. culicivora* and *T. castaneum* members but no *C. elegans* members were surveyed for GO annotations enriched in comparison to the whole *C. elegans* proteome (sets i and ii) and the *T. castaneum* proteome (set i), conducting Fisher's exact test as implemented in Blast2GO. Due to the small size of both sets compared to the large reference set, p-values could not be corrected for multiple testing. To improve annotation reliability, these proteins were recompared (using BLAST) to the UniProt/SwissProt database and run through the Blast2GO pipeline as described above.

### Whole-mount in situ hybridization

For in situ hybridisation we modified the freeze-crack procedure described previously for *C. elegans* [128] and revised by Maduro et al. (2007; <http://www.faculty.ucr.edu/~mmaduro/resources.htm>). In particular, to achieve reliable penetration of the durable *R. culicivora* egg envelopes we initially partly removed the protective layer by incubation in alkaline bleach solution (see [33]). Digoxigenine-labeled sense and antisense RNA probes were generated from linearized pBs vectors (Stratagene, La Jolla, USA) containing a 400 bp fragment of *R. culicivora mex-3* via run off *in vitro* transcription with T7 or T3 RNA-polymerase according to the manufacturer's protocol (Roche, Mannheim, Germany). The concentration of the labelled probes was about  $300 \text{ ng} \times \text{ml}^{-1}$ .

### Additional files

**Additional file 1: Supplementary data figures and tables.**

**Additional file 2: Analysis of OrthoMCL output by BLAST+.** BLAST+ results for specific *C. elegans* proteins not found in a cluster with Dorylaimia proteins.

**Additional file 3: Fisher's exact test data.** GO terms enriched in a set of protein clusters shared between Dorylaimia in comparison to (i) *C. elegans* and (ii) *T. castaneum* proteomes.

**Additional file 4: Levin data.** Genes identified as being differentially expressed in *Caenorhabditis* development by Levin et al. [61].

**Additional file 5: Analysis of Phylotypic stage genes.** *C. elegans* orthologues of genes possibly acting in (i) a potential nematode specific phylotypic stage and (ii) a metazoan phylotypic stage.

**Additional file 6: Codon usage in *R. culicivora*.** Codon usage data.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

PHS conceived study, assembled and annotated the genome, conducted analyses and wrote paper; MK conceived study, conducted analyses and wrote part of the paper; CK conceived part of the study, conducted analyses on developmental expression set and wrote part of the paper; GDK helped with genome assembly and annotation; SK helped with genome assembly and wrote/provided Perl scripts; JIRC analysed MEX-3 dataset; NAN analysed PAR dataset; DS analysed SEX determination dataset; KM conducted RNA sequencing and initial EST assembly; PH performed preparative laboratory

experiments and conceived sequencing strategy; JA conceived sequencing strategy and conducted genome sequencing; PF helped with initial genome pre-assembly; PN initiated study and conceived sequencing strategy; WKT conceived parts of study; MLB conceived study and wrote paper; ES initiated and conceived study and wrote paper. All authors read and approved the final manuscript.

#### Acknowledgements

We are indebted to E. Platzer, Riverside, for the continuous supply with *R. culicivora* nematodes. We thank J. Schulze, Cologne, for advice on nematode cultivation and C. Becker and K. Konrad for expert technical assistance in the genome sequencing experiments. We are also grateful to H. Oezden, Cologne for assistance with In-situ hybridisations. We thank J. Parkinson, Toronto, for providing a conserved NEMBASE4 protein set, Elizabeth Martínez Salazar, Zacatecas, Mexico, for Feulgen C-value data and A. H. Jay Burr, Vancouver, Canada for sharing preliminary results on phototaxis in *R. culicivora*. Assemblies and other computations were conducted on the HPC cluster "CHEOPS" at the University of Cologne (<http://rrzk.uni-koeln.de/cheops.html>).

#### Funding

This work was partly funded through the SFB 680: "Molecular Basis of Evolutionary Innovations". Philipp H. Schiffer is funded by the VolkswagenStiftung in the "Förderinitiative Evolutionsbiologie". Gerogios D. Koutsovoulos is funded by a UK BBSRC Research Studentship and an Overseas Research Studentship from the University of Edinburgh. Additional Funding came through the BMBF-Projekt "NGSgoesHPC".

#### Accession numbers

Raw genome and transcriptome sequence data reported in this manuscript have been deposited in the ERA under accession ERP002111 (<http://www.ebi.ac.uk/ena/data/view/ERP002111>), and assembled genomic contigs deposited in the ENA INSDC database under accession numbers CAQS01000001-CAQS01062537. Annotation information and additional data are available through (<http://romanomermis.nematod.es>).

#### Author details

<sup>1</sup>Zoologisches Institut, Universität zu Köln, Cologne, NRW, Germany, ORCID:0000-0001-6776-0934. <sup>2</sup>Institute of Evolutionary Biology, School of Biological Sciences, The University of Edinburgh, Edinburgh, Scotland, UK. <sup>3</sup>Institute für Entwicklungsbiologie, Universität zu Köln, Cologne, NRW, Germany. <sup>4</sup>Hubbard Center for Genome Studies, University of New Hampshire, Durham, NH, USA. <sup>5</sup>Cologne Center for Genomics, Universität zu Köln, Cologne, NRW, Germany.

Received: 7 May 2013 Accepted: 17 December 2013

Published: 27 December 2013

#### References

1. Maduro MF: **Cell fate specification in the *C. elegans* embryo.** *Dev Dyn* 2010, **239**:1315–1329.
2. Sulston JE, Schierenberg E, White JG, Thomson JN: **The embryonic cell lineage of the nematode *Caenorhabditis elegans*.** *Dev Biol* 1983, **100**:64–119.
3. Blaxter M, de Ley P, Garey J, Liu L, Scheldeman P, Vierstraete A, Vanfleteren J, Mackey L, Dorris M, Frisse L: **A molecular evolutionary framework for the phylum Nematoda.** *Nature* 1998, **392**(6671):71–75.
4. Meldal B, Debenham N, de Ley P, de Ley I, Vanfleteren J, Vierstraete A, Bert W, Borgonie G, Moens T, Tyler P, Austen M, Blaxter M, Rodgers A, Lamshead P: **An improved molecular phylogeny of the Nematoda with special emphasis on marine taxa.** *Mol Phylogenet Evol* 2007, **42**(3):622–636.
5. Boveri T: *Die Entwicklung von Ascaris megalocephala mit besonderer Rücksicht auf die Kernverhältnisse.* Fischer: Festschrift fuer Carl von Kupffer; Jena; 1899.
6. Müller H: **Beitrag zur Embryonalentwicklung von Ascaris megalocephala.** *Zoologica* 1903, **41**:60.
7. Vangestel S, Houthoofd W, Bert W, Borgonie G: **The early embryonic development of the satellite organism *Pristionchus pacificus*: differences and similarities with *Caenorhabditis elegans*.** *Nematology* 2008, **10**:301–312.
8. Skiba F, Schierenberg E: **Cell lineages, developmental timing, and spatial pattern formation in embryos of free-living soil nematodes.** *Dev Biol* 1992, **151**(2):597–610.
9. Lahl V, Schulze J, Schierenberg E: **Differences in embryonic pattern formation between *Caenorhabditis elegans* and its close parthenogenetic relative *Diploscapter coronatus*.** *Int J Dev Biol* 2009, **53**(4):507–515.
10. Brauchle M, Kiontke K, Macmenamin P, Fitch DHA, Piano F: **Evolution of early embryogenesis in rhabditid nematodes.** *Dev Biol* 2009, **335**:253–262.
11. Wiegner O, Schierenberg E: **Specification of gut cell fate differs significantly between the Nematodes *Acroboloides nanus* and *Caenorhabditis elegans*.** *Dev Biol* 1998, **204**:3–14.
12. Wiegner O, Schierenberg E: **Regulative development in a nematode embryo: a hierarchy of cell fate transformations.** *Dev Biol* 1999, **215**:1–12.
13. Voronov DA, Panchin YV: **Cell lineage in marine nematode *Enoplus brevis*.** *Dev* 1998, **125**:143–150.
14. Schulze J, Schierenberg E: **Evolution of embryonic development in nematodes.** *EvoDevo* 2011, **2**:18.
15. Schulze J, Houthoofd W, Uenk J, Vangestel S, Schierenberg E: **Plectus - a stepping stone in embryonic cell lineage evolution of nematodes.** *EvoDevo* 2012, **3**:13.
16. True JR, Haag ES: **Developmental system drift and flexibility in evolutionary trajectories.** *Evol Dev* 2001, **3**(2):109–119.
17. *C. elegans* Sequencing Consortium: **Genome sequence of the nematode *C. elegans*: a platform for investigating biology.** *Science* 1998, **282**(5396):2012–2018.
18. Stein LD, Bao Z, Blasiar D, Blumenthal T, Brent MR, Chen N, Chinwalla A, Clarke L, Clee C, Coghlan A, Coulson A, D'eustachio P, Fitch DHA, Fulton LA, Fulton RE, Griffiths-Jones S, Harris TW, Hillier LW, Kamath R, Kuwabara PE, Mardis ER, Marra MA, Miner TL, Minx P, Mullikin JC, Plumb RW, Rogers J, Schein JE, Sohrmann M, Spieth J, et al.: **The genome sequence of, *Caenorhabditis briggsae*: a platform for comparative genomics.** *PLoS Biol* 2003, **1**(2):E45.
19. Mortazavi A, Schwarz EM, Williams B, Schaeffer L, Antoshechkin I, Wold BJ, Sternberg PW: **Scaffolding a *Caenorhabditis* nematode genome with RNA-seq.** *Genome Res* 2010, **20**(12):1740–1747.
20. Dieterich C, Clifton SW, Schuster LN, Chinwalla A, Delehaunty K, Dinkelacker I, Fulton L, Fulton R, Godfrey J, Minx P, Mitreva M, Roeseler W, Tian H, Witte H, Yang SP, Wilson RK, Sommer RJ: **The *Pristionchus pacificus* genome provides a unique perspective on nematode lifestyle and parasitism.** *Nature Genet* 2008, **40**(10):1193–1198.
21. Jex AR, Liu S, Li B, Young ND, Hall RS, Li Y, Yang L, Zeng N, Xu X, Xiong Z, Chen F, Wu X, Zhang G, Fang X, Kang Y, Anderson GA, Harris TW, Campbell BE, Vlaminck J, Wang T, Cantacessi C, Schwarz EM, Ranganathan S, Geldhof P, Nejsum P, Sternberg PW, Yang H, Wang J, Wang J, Gasser RB: ***Ascaris suum* draft genome.** *Nature* 2011, **479**(7374):529–533.
22. Ghedin E, Wang S, Spiro D, Caler E, Zhao Q, Crabtree J, Allen JE, Delcher AL, Guilianio DB, Miranda-Saavedra D, Angiuoli SV, Creasy T, Amedeo P, Haas B, El-Sayed NM, Wortman JR, Feldblyum T, Tallon L, Schatz M, Shumway M, Koo H, Salzberg SL, Schobel S, Pertea M, Pop M, White O, Barton GJ, Carlow KKS, Crawford MJ, Daub J, et al.: **Draft genome of the filarial nematode parasite *Brugia malayi*.** *Science* 2007, **317**(5845):1756–1760.
23. Godel C, Kumar S, Koutsovoulos G, Ludin P, Nilsson D, Comandatore F, Wrobel N, Thompson M, Schmid CD, Goto S, Bringaud F, Wolstenholme A, Bandi C, Epe C, Kaminsky R, Blaxter M, Mäser P: **The genome of the heartworm, *Dirofilaria immitis*, reveals drug and vaccine targets.** *FASEB J* 2012, **26**(11):4650–4661.
24. Abad P, Gouzy J, Aury JM, Castagnone-Sereno P, Danchin EGJ, Deleury E, Perfus-Barbeoch L, Anthouard V, Artiguenave F, Blok VC, Caillaud MC, Coutinho PM, Dasilva C, De Luca F, Deau F, Esquibet M, Flutre T, Goldstone JV, Hamamouch N, Hewezi T, Jaillon O, Jubin C, Leonetti P, Magliano M, Maier TR, Markov GV, Mcveigh P, Pesole G, Poulain J, Robinson-Rechavi M, et al.: **Genome sequence of the metazoan plant-parasitic nematode, *Meloidogyne incognita*.** *Nature Biotechnol* 2008, **26**(8):909–915.
25. Kikuchi T, Cotton JA, Dalzell JJ, Hasegawa K, Kanzaki N, Mcveigh P, Takanashi T, Tsai IJ, Assefa SA, Cock PJA, Otto TD, Hunt M, Reid AJ, Sanchez-Flores A, Tsuchihara K, Yokoi T, Larsson MC, Miwa J, Maule AG, Sahashi N, Jones JT, Berriman M: **Genomic insights into the origin of**

- parasitism in the emerging plant pathogen *Bursaphelenchus xylophilus*. *PLoS Pathogens* 2011, **7**(9):e1002219.
26. Srinivasan J, Dillman AR, Macchietto MG, Heikkinen L, Lakso M, Fracchia KM, Antoshechkin I, Mortazavi A, Wong G, Sternberg PW: **The draft genome and transcriptome of Panagrellus redivivus are shaped by the harsh demands of a free-living lifestyle.** *Genetics* 2013, **193**(4):1279–1295.
  27. Mitreva M, Jasmer DP, Zarlenga DS, Wang Z, Abubucker S, Martin J, Taylor CM, Yin Y, Fulton LA, Minx P, Yang SP, Warren WC, Fulton RS, Bhonagiri V, Zhang X, Hallsworth-Pepin K, Clifton SW, Mccarter JP, Appleton J, Mardis ER, Wilson RK: **The draft genome of the parasitic nematode *Trichinella spiralis*.** *Nature Genet* 2011, **43**(3):228–235.
  28. Hope IA: **Embryology, Developmental Biology and the Genome.** In *The Biology of Nematodes*. 1st edition. Edited by Lee DL. New York: Taylor & Francis; 2002:121–145.
  29. Kumar S, Schiffer PH, Blaxter M: **959 Nematode Genomes: a semantic wiki for coordinating sequencing projects.** *Nucl Acids Res* 2012, **40**(D1):D1295–D1300.
  30. Kumar S, Koutsovoulos G, Kaur G, Blaxter M: **Toward 959 nematode genomes.** *Worm* 2012, **1**:0–8.
  31. Petersen JJ: **Nematodes as biological control agents: Part I. Mermithidae.** In *Advances in Parasitology*. Edited by Baker JR, Muller R. London: Academic Press; 1985:307–346.
  32. Petersen JJ, Chapman HC, Willis OR, Fukuda T: **Release of Romanomermis culicivorax for the control of Anopheles albimanus in El Salvador II. Application of the nematode.** *Ame J Trop Med Hyg* 1978, **27**(6):1268–1273.
  33. Schulze J, Schierenberg E: **Cellular pattern formation, establishment of polarity and segregation of colored cytoplasm in embryos of the nematode Romanomermis culicivorax.** *Dev Biol* 2008, **315**(2):426–436.
  34. Schulze J, Schierenberg E: **Embryogenesis of Romanomermis culicivorax: an alternative way to construct a nematode.** *Dev Biol* 2009, **334**:10–21.
  35. Kent W: **BLAT—the BLAST-like alignment tool.** *Genome Res* 2002, **12**(4):656.
  36. Parra G, Bradnam K, Korf I: **CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes.** *Bioinformatics* 2007, **23**(9):1061–1067.
  37. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, Yandell M: **MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes.** *Genome Res* 2008, **18**:188–196.
  38. Li W, Godzik A: **Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences.** *Bioinformatics* 2006, **22**(13):1658–1659.
  39. Stanke M, Waack S: **Gene prediction with a hidden Markov model and a new intron submodel.** *Bioinformatics* 2003, **19**(2):ii215–ii225.
  40. Wang J, Mitreva M, Berriman M, Thorne A, Magrini V, Koutsovoulos G, Kumar S, Blaxter ML, Davis RE: **Silencing of germline-expressed genes by DNA elimination in somatic cells.** *Dev Cell* 2012, **23**(5):1072–1080.
  41. Nawrocki EP, Kolbe DL, Eddy SR: **Infernal 1.0: inference of RNA alignments.** *Bioinformatics* 2009, **25**(10):1335–1337.
  42. Lowe TM, Eddy SR: **tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence.** *Nucl Acids Res* 1997, **25**(5):0955–0964.
  43. Kumar S: **Data for PhD thesis on next generation nematode genomes.** 2012. <http://dx.doi.org/10.6084/m9.figshare.96089>.
  44. Richards S, The *T. castaneum* Genome Consortium: **The genome of the model beetle and pest *Tribolium castaneum*.** *Nature* 2008, **452**(7190):949–955.
  45. Schröder R, Beermann A, Wittkopp N, Lutz R: **From development to biodiversity—*Tribolium castaneum*, an insect model organism for short germband development.** *Dev Genes Evol* 2008, **218**(3–4):119–126.
  46. Chen F, Mackey AJ, Vermunt JK, Roos DS: **Assessing performance of orthology detection strategies applied to eukaryotic genomes.** *PLoS ONE* 2007, **2**(4):e383.
  47. Altschul S, Madden T, Schäffer A: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucl Acids Res* 1997. [<http://nar.oxfordjournals.org/content/25/17/3389.short>]
  48. Jensen RA: **Orthologs and paralogs - we need to get it right.** *Genome Biol* 2001, **2**(8):interactions1002.1–1002.3.
  49. Koonin E: **Orthologs, paralogs, and evolutionary genomics.** *Ann Rev Genet* 2005, **39**:309–338.
  50. Moreno-Hagelsieb G, Latimer K: **Choosing BLAST options for better detection of orthologs as reciprocal best hits.** *Bioinformatics* 2008, **24**(3):319–324.
  51. Shaye DS, Greenwald I: **OrthoList: A compendium of *C. elegans* Genes with human Orthologs.** *PLoS ONE* 2011, **6**(5):e20085.
  52. Tautz D, Domazet-Lošo T: **The evolutionary origin of orphan genes.** *Nat Rev Genet* 2011, **12**(10):692–702.
  53. Altschul SF, Gish W, Miller W, Myers EW: **Basic local alignment search tool.** *J Mol Biol* 1990, **215**:403–410.
  54. Keeling PJ, Corradi N, Morrison HG, Haag KL, Ebert D, Weiss LM, Akiyoshi DE, Tzipori S: **The reduced genome of the parasitic microsporidian *Enteroctyozoon bienersi* lacks genes for core carbon metabolism.** *Genome Biol Evol* 2010, **2**(0):304–309.
  55. Bird A: **DNA methylation patterns and epigenetic memory.** *Genes & Dev* 2002, **16**:6–21.
  56. Gao F, Liu X, Wu XP, Wang XL, Gong D, Lu H, Xia Y, Song Y, Wang J, Du J, Liu S, Han X, Tang Y, Yang H, Jin Q, Zhang X, Liu M: **Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*.** *Genome Biol* 2012, **13**(10):R100.
  57. Tran RK, Zilberman D, de Bustos C, Ditt RF, Henikoff JG, Lindroth AM, Delrow J, Boyle T, Kwong S, Bryson TD, Jacobsen SE, Henikoff S: **Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in *Arabidopsis*.** *Genome Biol* 2005, **6**(11):R90.
  58. Martienssen RA, Colot V: **DNA methylation and epigenetic inheritance in plants and filamentous fungi.** *Science* 2001, **293**(5532):1070–1074.
  59. Eisenmann DM: **Wnt signaling.** *WormBook* 2005. [[http://www.wormbook.org/chapters/www\\_wntsignaling/wntsignaling.html](http://www.wormbook.org/chapters/www_wntsignaling/wntsignaling.html)]
  60. Graham LD, Kotze AC, Fernley RT, Hill RJ: **Molecular & biochemical parasitology.** *Mol Biochem Parasitol* 2010, **171**(2):104–107. [<http://dx.doi.org/10.1016/j.molbiopara.2010.03.003>]
  61. Levin M, Hashimshony T, Wagner F, Yanai I: **Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo.** *Dev Cell* 2012, **22**(5):1101–1108.
  62. Phelan P: **Innexins: members of an evolutionarily conserved family of gap-junction proteins.** *Biochimica et Biophysica Acta (BBA) - Biomembranes* 2005, **1711**(2):225–245.
  63. Jones AK, Sattelle DB: **Functional genomics of the nicotinic acetylcholine receptor gene family of the nematode, *Caenorhabditis elegans*.** *BioEssays* 2003, **26**:39–49.
  64. Antebi A: **Nuclear hormone receptors in *C. elegans*.** In *Wormbook*. Edited by The C. elegans Research Community. WormBook; 2006. <http://www.wormbook.org>.
  65. Altun ZF, Chen B, Wang ZW, Hall DH: **High resolution map of *Caenorhabditis elegans* gap junction proteins.** *Dev Dyn* 2009, **238**(8):1936–1950.
  66. Haag E: **The evolution of nematode sex determination: *C. elegans* as a reference point for comparative biology.** In *WormBook*. Edited by The C. elegans Research Community. WormBook; 2005. <http://www.wormbook.org>.
  67. Tingley GA, Anderson RM: **Environmental sex determination and density-dependent population regulation in the entomogenous nematode *Romanomermis culicivorax*.** *Parasitol* 1986, **92**:431–449.
  68. Powell JR, Jow MM, Meyer BJ: **The T-box transcription factor SEA-1 is an autosomal element of the X:A signal that determines *C. elegans* sex.** *Dev Cell* 2005, **9**(3):339–349.
  69. Chu DS, Dawes HE, Lieb JD, Chan RC, Kuo AF, Meyer BJ: **A molecular link between gene-specific and chromosome-wide transcriptional repression.** *Genes & Dev* 2002, **16**(7):796–805.
  70. Meyer B: **X-Chromosome dosage compensation.** In *WormBook*. Edited by The C. elegans Research Community. WormBook; 2005. <http://www.wormbook.org>.
  71. Zarkower D: **Somatic sex determination.** In *WormBook*. Edited by The C. elegans Research Community. WormBook; 2006. <http://www.wormbook.org>.
  72. Kuwabara PE, Okkema PG, Kimble J: **tra-2 encodes a membrane protein and may mediate cell communication in the**

- Caenorhabditis elegans sex determination pathway.** *Mol Biol Cell* 1992, **3**(4):461–473.
73. Goodwin EB, Ellis RE: **Turning clustering loops: sex determination in *Caenorhabditis elegans*.** *Curr Biol* 2002, **12**(3):R111–R120.
74. Baldi C, Cho S, Ellis RE: **Mutations in two independent pathways are sufficient to create hermaphroditic nematodes.** *Science* 2009, **326**(5955):1002–1005.
75. Goldstein B, Macara IG: **The PAR proteins: fundamental players in animal cell polarization.** *Dev Cell* 2007, **13**(5):609–622.
76. Bowerman B: **Embryonic polarity: Protein stability in asymmetric cell division.** *Curr Biol* 2000, **10**(17):R637–R641.
77. Severson AF, Bowerman B. *J Cell Biol* 2003, **161**:21–26.
78. Gönczy P, Rose LS: **Asymmetric cell division and axis formation in the embryo.** In *WormBook*. Edited by The C. elegans Research Community. WormBook:2005. <http://www.wormbook.org>.
79. Cheng NN, Kirby CM, Kempthues KJ: **Control of cleavage spindle orientation in *Caenorhabditis elegans*: the role of the genes par-2 and par-3.** *Genetics* 1995, **139**(2):549–559.
80. Wu JC, Rose LS: **PAR-3 and PAR-1 inhibit LET-99 localization to generate a cortical band important for spindle positioning in *Caenorhabditis elegans* embryos.** *Mol Biol Cell* 2007, **18**(11):4470–4482.
81. Doerflinger H, Vogt N, Torres IL, Mirouse V, Koch I, Nusslein-Volhard C, St Johnston D: **Bazooka is required for polarisation of the *Drosophila* anterior-posterior axis.** *Development* 2010, **137**(10):1765–1773.
82. Goldstein B, Frisse L, Thomas W: **Embryonic axis specification in nematodes: evolution of the first step in development.** *Curr Biol* 1998, **8**(3):157–160.
83. Draper BW, Mello CC, Bowerman B, Hardin J, Priess JR: **MEX-3 is a KH domain protein that regulates blastomere identity in early *C. elegans* embryos.** *Cell* 1996, **87**(2):205–216.
84. Huang N, Mootz D, Walhout A, Vidal M, Hunter CP: **MEX-3 interacting proteins link cell polarity to asymmetric gene expression in *Caenorhabditis elegans*.** *Development* 2002, **129**(3):747–759.
85. Evans TC, Hunter CP: **Translational control of maternal RNAs.** In *WormBook*. Edited by The C. elegans Research Community. WormBook:2005. <http://www.wormbook.org>.
86. Gomes JE, Encalada SE, Swan KA, Shelton CA, Carter JC, Bowerman B: **The maternal gene spn-4 encodes a predicted RRM protein required for mitotic spindle orientation and cell fate patterning in early *C. elegans* embryos.** *Dev* 2001, **128**(21):4301–4314.
87. Labbé JC, Goldstein B: **Embryonic development: A New SPN on cell fate specification.** *Curr Biol* 2002, **12**(11):R396–R398.
88. Simske JS, Hardin J: **Getting into shape: epidermal morphogenesis in *Caenorhabditis elegans* embryos.** *BioEssays* 2001, **23**:12–23.
89. Gilleard JS, McGhee JD: **Activation of hypodermal differentiation in the *Caenorhabditis elegans* embryo by GATA transcription factors ELT-1 and ELT-3.** *Mol Cell Biol* 2001, **21**(7):2533–2544.
90. Sommer R: **As good as they get: cells in nematode vulva development and evolution.** *Curr Opin Cell Biol* 2001, **13**(6):715–720.
91. Kiontke K, Barriere A, Kolotuev I, Podbilewicz B, Sommer R, Fitch DHA, Félix MA: **Trends, stasis, and drift in the evolution of nematode vulva development.** *Curr Biol: CB* 2007, **17**(22):1925–1937.
92. Sternberg PW: **Vulval development.** In *WormBook*. Edited by The C. elegans Research Community. WormBook; 2005. <http://www.wormbook.org>.
93. Salsler SJ, Loer CM, Kenyon C: **Multiple HOM-C gene interactions specify cell fates in the nematode central nervous system.** *Genes & Dev* 1993, **7**(9):1714–1724.
94. Eisenmann DM, Kim SK: **Protruding vulva mutants identify novel loci and Wnt signaling factors that function during *Caenorhabditis elegans* vulva development.** *Genetics* 2000, **156**(3):1097–1116.
95. Shemer G, Podbilewicz B: **LIN-39/Hox triggers cell division and represses EFF-1/fusogen-dependent vulval cell fusion.** *Genes & Dev* 2002, **16**(24):3136–3141.
96. Tian H, Schlager B, Xiao H, Sommer RJ: **Wnt signaling induces vulva development in the nematode *Pristionchus pacificus*.** *Curr Biol* 2008, **18**(2):142–146.
97. Streit A, Kohler R, Marty T, Belfiore M, Takacs-Vellai K, Vigano MA, Schnabel R, Affolter M, Müller F: **Conserved Regulation of the *Caenorhabditis elegans* labial/Hox1 Gene ceh-13.** *Dev Biol* 2002, **242**(2):96–108.
98. Aboobaker AA, Blaxter ML: **Hox Gene Loss during dynamic evolution of the nematode cluster.** *Curr Biol* 2003, **13**:37–40.
99. Chisholm A: **Control of cell fate in the tail region of *C. elegans* by the gene egl-5.** *Deve* 1991, **111**(4):921–932.
100. Aboobaker A, Blaxter M: **Hox gene evolution in nematodes: novelty conserved.** *Current Opinion in Genet & Dev* 2003, **13**:593–598.
101. Lemons D, McGinnis W: **Genomic evolution of Hox gene clusters.** *Science* 2006, **313**(5795):1918–1922.
102. Chamberlin HM, Thomas JH: **The bromodomain protein LIN-49 and trithorax-related protein LIN-59 affect development and gene expression in *Caenorhabditis elegans*.** *Development* 2000, **127**(4):713–723.
103. Aspöck G, Kagoshima H, Niklaus G, Bürglin TR: ***Caenorhabditis elegans* Has scores of hedgehog related genes: sequence and expression analysis.** *Genome Res* 1999, **9**:909–923. [genome.cshlp.org](http://genome.cshlp.org).
104. Domazet-Lošo T, Tautz D: **A phylogenetically based transcriptome age index mirrors ontogenetic divergence patterns.** *Nature* 2010, **468**(7325):815–818.
105. Richardson MK: **A phylotypic stage for all animals?** *Dev Cell* 2012, **22**(5):903–904.
106. Kalinka AT, Tomancak P: **The evolution of early animal embryos: conservation or divergence?** *Trends In Ecology & Evolution* 2012, **27**(7):385–393.
107. De Ley P: **A quick tour of nematode diversity and the backbone of nematode phylogeny.** In *WormBook*. Edited by The C. elegans Research Community. WormBook; 2006. <http://www.wormbook.org>.
108. Martin M: **Cutadapt removes adapter sequences from high-throughput sequencing reads.** *EMBnet.Journal* 2011, **17**(1).
109. CLCbio: *White Paper on de Novo Assembly in CLC Assembly Cell*. CLC Bio: Whitepaper; 2010.
110. Kumar S, Blaxter ML: **Simultaneous genome sequencing of symbionts and their hosts.** *Symbiosis* 2012, **55**(3):119–126.
111. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W: **Scaffolding pre-assembled contigs using SSPACE.** *Bioinformatics* 2011, **27**(4):578–579.
112. Smit AFA, Hubley R, Green P: **RepeatMasker Open-3.0. 1996-2010.** <http://www.repeatmasker.org>.
113. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J: **Repbase Update, a database of eukaryotic repetitive elements.** *Cytogenetic and Genome Res* 2005, **110**(1-4):462–467.
114. Volfovsky N, Haas BJ, Salzberg SL: **A clustering method for repeat analysis in DNA sequences.** *Genome Biol* 2001, **2**(8):research0027.1–research0027.11.
115. Bao Z, Eddy SR: **Automated de novo identification of repeat sequence families in sequenced genomes.** *Genome Res* 2002, **12**(8):1269–1276.
116. Price AL, Jones NC, Pevzner PA: **De novo identification of repeat families in large genomes.** *Bioinformatics* 2005, **21**(Suppl 1):i351–i358.
117. Parkinson J, Mitreva M, Whitton C, Thomson M: **A transcriptomic analysis of the phylum Nematoda.** *Nature Genet* 2004, **36**:1259–1267.
118. Gremme G, Brendel V, Sparks ME, Kurtz S: **Engineering a software tool for gene structure prediction in higher organisms.** *Information and Software Technol* 2005, **47**(15):965–978.
119. Supek F, Vlahovicek K: **INCA: synonymous codon usage analysis and clustering by means of self-organizing map.** *Bioinformatics* 2004, **20**(14):2329–2330.
120. Cutter AD, Wasmuth JD, Blaxter ML: **The evolution of biased codon and amino acid usage in nematode genomes.** *Mol Biol Evol* 2006, **23**(12):2303–2315.
121. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M: **Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.** *Bioinformatics* 2005, **21**(18):3674–3676.
122. The Gene Ontology Consortium: **Gene Ontology: tool for the unification of biology.** *Nature Genetics* 2000, **25**:25–29.
123. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R: **InterProScan: protein domains identifier.** *Nucl Acids Res* 2005, **33**(Web Server):W116–W120.

124. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy S, Bateman A: **Rfam: annotating non-coding RNAs in complete genomes.** *Nucl Acids Res* 2004, **33**(Database issue):D121–D124.
125. Li L, Stoeckert CJ, Roos DS: **OrthoMCL: identification of ortholog groups for eukaryotic genomes.** *Genome Research* 2003, **13**(9):2178–2189.
126. van Dongen S: **A Cluster algorithm for graphs.** *Report - Information Syst* 2000, **10**:1–40.
127. van Dongen S: **Graph Clustering by Flow Simulation.** *PhD thesis*, University of Utrecht 2000. <http://www.wormbook.org>.
128. Seydoux G, Fire A: **Whole-mount in situ hybridization for the detection of RNA in *Caenorhabditis elegans* embryos.** *Methods in cell biology* 1995, **48**:323–337.

doi:10.1186/1471-2164-14-923

**Cite this article as:** Schiffer *et al.*: The genome of *Romanomermis culicivorax*: revealing fundamental changes in the core developmental genetic toolkit in Nematoda. *BMC Genomics* 2013 **14**:923.

**Submit your next manuscript to BioMed Central  
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

