Nucleic Acids Research, 2003, Vol. 31, No. 10 ***2553–2560***
DOI: 10.1093/nar/gkg359

Limited microsynteny between the genomes of *Pristionchus pacificus* and *Caenorhabditis elegans*

Kwang-Zin Lee, Andreas Eizinger, Ramkumar Nandakumar[1], Stephan C. Schuster[1] and Ralf J. Sommer*

Abteilung für Evolutionsbiologie and [1]Genomzentrum, Max-Planck Institut für Entwicklungsbiologie, Spemannstrasse 37, 72076 Tübingen, Germany

Received February 12, 2003; Revised and Accepted March 20, 2003

ABSTRACT

Nematodes are an attractive group of organisms for studying the evolution of developmental processes. *Pristionchus pacificus* was established as a satellite organism for comparing vulva development and other processes to *Caenorhabditis elegans*. The generation of a genetic linkage map of *P.pacificus* has provided a first insight into the structure and organization of the genome of this species. *Pristionchus pacificus* and *C.elegans* are separated from one another by >100 000 000 years such that the structure of the genomes of these two nematodes might differ substantially. To evaluate the amount of synteny between the two genomes, we have obtained 126 kb of continuous genomic sequence of *P.pacificus*, flanking the developmental patterning gene *pal-1*. Of the 20 predicted open reading frames in this interval, 11 have *C.elegans* orthologs. Ten of these 11 orthologs are located on *C.elegans* chromosome III, indicating the existence of synteny. However, most of these genes are distributed over a 12 Mb interval of the *C.elegans* genome and only three pairs of genes show microsynteny. Thus, intrachromosomal rearrangements occur frequently in nematodes, limiting the likelihood of identifying orthologous genes of *P.pacificus* and *C.elegans* based on positional information within the two genomes.

INTRODUCTION

Pristionchus pacificus is a free-living nematode of the Diplogastridae family that has recently been developed as a satellite organism for functional studies in evolutionary developmental biology (1–3). By comparing *P.pacificus* to the model organism *Caenorhabditis elegans*, evolutionary alterations of developmental processes resulting in cellular and morphological changes can be identified (4). One developmental process that has been studied in great detail is the development of the vulva, the egg-laying structure of nematode females and hermaphrodites. In *C.elegans*, vulva formation is well characterized and at least four signaling systems are known to be required for the development of this organ (5). Comparative studies of vulva development between *P.pacificus* and *C.elegans* revealed the existence of many differences at the cellular level. Although homologous precursor cells are involved in generating vulval tissue in both species, the precise cell–cell interactions differ greatly (for review see 1). Genetic screens for vulva-defective mutants in *P.pacificus* identified many loci involved in the generation of vulval tissue (2). The comparison of mutations in *lin-39*, *mab-5* and *vab-7* between both species revealed substantial changes in the function of these homologous genes during vulva formation (6–9). These surprising results indicated that a candidate gene approach would not be sufficient to study *P.pacificus* mutations.

Genetic and molecular studies in *P.pacificus* have recently been complemented by a genomic approach. Wild isolates of *P.pacificus* from Pasadena (California, USA) and Port Angeles (Washington, USA) show a substantial number of DNA polymorphisms that can be used for the generation of a genetic linkage map of the organism (10,11). A BAC library was generated and completely end sequenced (12). Using single-stranded conformational polymorphism analyses (SSCP), a total of 133 polymorphisms were generated in BAC end and expressed sequence tags (EST) sequences. These markers were tested on a meiotic mapping panel, providing a first genetic linkage map of *P.pacificus* (12). Currently the genetic linkage map is complemented by the generation of a physical map (J.Srinivasan and R.J.Sommer, in preparation). Eventually, the genetic and physical maps will be integrated because both are derived from the analysis of BAC clones. In addition, an ongoing EST sequencing project has generated ~10 000 sequences (13).

Pristionchus pacificus and *C.elegans* have been separated from one another for ~100 000 000–200 000 000 years (Fig. 1) (6). Despite this relatively long evolutionary separation time, initial studies had provided some indication for the existence of synteny. For example, the homeotic genes *lin-39*, *mab-5* and *vab-7* are on chromosome III of *C.elegans*. In *P.pacificus*, the orthologous genes are also linked on one chromosome,

*To whom correspondence should be addressed. Tel: +49 7071 601 371; Fax: +49 7071 601 498; Email: ralf.sommer@tuebingen.mpg.de
Present address:
Andreas Eizinger, Department of Developmental Biology, Stanford University, Beckman Center, 279 Campus Drive B-341, Stanford, CA 94305, USA

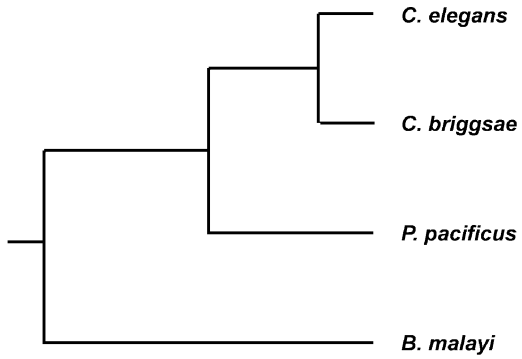

Figure 1. Phylogenetic relationship of the four nematodes compared at the genomic level. *Caenorhabditis elegans* and *C.briggsae* are sister species. Their estimated separation time is ~50 000 000 years. *Pristionchus pacificus* belongs to the Diplogastridae, another family of the Rhabditida and is thought to have separated from *Caenorhabditis* ~100 000 000–200 000 000 years ago. The parasite *B.malayi* shared the last common ancestor with *C.elegans* ~300 000 000 years ago.

which has been designated as chromosome III (12). A similar situation is given for chromosome IV, where *let-60*/Ras, the Twitchin genes *Cel-unc-22* and *Ppa-unc-1* and others are present on the same chromosome in both species (12). Also, ESTs of several *P.pacificus* vitellogenin have been shown to be located on the X chromosome, as in *C.elegans* (12).

When genomes of related organisms are compared, the level of conservation of gene order can be expressed as synteny. The term microsynteny indicates that the order of individual genes in considered genomes is exactly conserved. For example, the comparison of the genomes of *Caenorhabditis briggsae* (14) and *Brugia malayi* (15) with *C.elegans* revealed high rates of rearrangements and only limited microsynteny. Consistent with the different evolutionary separation times of these two species from *C.elegans* (Fig. 1), however, the *C.briggsae* genome has a higher degree of microsynteny to *C.elegans* than *B.malayi*, i.e. the order of genes is more similar between *C.elegans* and *C.briggsae* than between *C.elegans* and *B.malayi*. In this context, *P.pacificus* has an intermediate phylogenetic position, but is more distantly related to *C.elegans* than *C.briggsae* (Fig. 1). Here, we have extended our analysis of the *P.pacificus* genome to evaluate the degree of synteny between the genomes of *C.elegans* and *P.pacificus*. The existence of microsynteny would be a useful tool for the identification of developmental control genes, some of which are not highly conserved at the sequence level. If microsynteny is present between the genomes of both species, rapidly evolving genes could be cloned using positional information for these genes in *C.elegans*. We describe the complete sequence of a BAC clone 7E22 that provides a continuous sequence of 126 kb, the largest continuous genomic non-rhabditid nematode DNA sequence. It contains a region flanking the *caudal* homolog *pal-1* of *P.pacificus*, and contains 20 predicted open reading frames (ORFs), 11 of which have putative orthologs in *C.elegans*. Ten of these 11 orthologs are located on *C.elegans* chromosome III, indicating the existence of synteny. However, most of these genes are distributed over more than a 12 Mb interval of the *C.elegans* genome and only three pairs of genes show microsynteny.

MATERIALS AND METHODS

Ppa-pal-1 cloning

Ppa-pal-1 was cloned by PCR using the degenerate primers 5′-CGTCAGMGTACTGCNTAYAC-3′ in the N-terminal and 5′-CATGCKACKRTTYTGRAACCA-3′ in the C-terminal region of the homeodomain. To clone the 5′ and 3′ regions of *Ppa-pal-1*, we used RACE (rapid amplification of cDNA ends) experiments (16–20). The GenBank accession no. for the *Ppa-pal-1* sequence is AY216679.

BAC clone preparation and sequencing

PPBAC-7E22 was sequenced using the shotgun approach (21). BAC DNA was sheared, blunt end cloned into the pCR4-TOPO vector and transformed into *Escherichia coli* (Shotgun Subcloning Kit; Invitrogen). The plasmid DNA from the shotgun library was prepared using the TurboPrep Kit from Qiagen. Sequencing reactions were perfomed by using the BigDye Sequencing Kit (Version 2; Applied Biosystems) on an ABI prism 3700. 1000 plasmid clones were sequenced from both ends to yield 8× coverage of the BAC clone. The sequencing reads were analyzed and assembled using the Phred/Phrap/Consed package (Washington University) (22,23). Gaps were resolved by resequencing of clones under conditions that give increased read lengths and by primer walking. The annotated sequence is available in GenBank (accession nos 4716387–4716404).

Bioinformatics

The finished sequence of PPBAC-7E22 was compared to the GenBank non-redundant database (nucleic acid and protein) and dbEST and the *C.elegans* database WormBase, using BLAST (24,25). GeneFinder was used to determine predicted ORFs.

Verification of gene predictions

To confirm gene predictions from PPBAC7E22, primers were designed and PCR was performed on oligo(dT)-primed *P.pacificus* mixed stage cDNA with gene-specific primers. Secondary PCR experiments have been performed using nested primers. RT–PCR experiments for predicted ORFs that could not be confirmed have been repeated at least three times independently. The predicted exon–intron structure of the genes PPBAC7E22.9 and PPBAC7E22.10 has been confirmed by sequencing RT–PCR products after cloning.

RESULTS

Cloning of *Ppa-pal-1*

The nematode *caudal* ortholog *pal-1* encodes a transcription factor with a homeodomain and is an important regulator of several embryonic and post-embryonic developmental processes in *C.elegans* (18–20). *Cel-pal-1* is located on chromosome III of *C.elegans* and contains several interesting genes in its neighborhood, including the cell death adaptor protein *ced-4* (Fig. 2A). Various attempts to clone *Ppa-ced-4* by low stringency hybridization and PCR approaches have proven unsuccessful (data not shown). Therefore, we have used this region of the genome to investigate the degree of synteny between the genomes of *P.pacificus* and *C.elegans.* We cloned

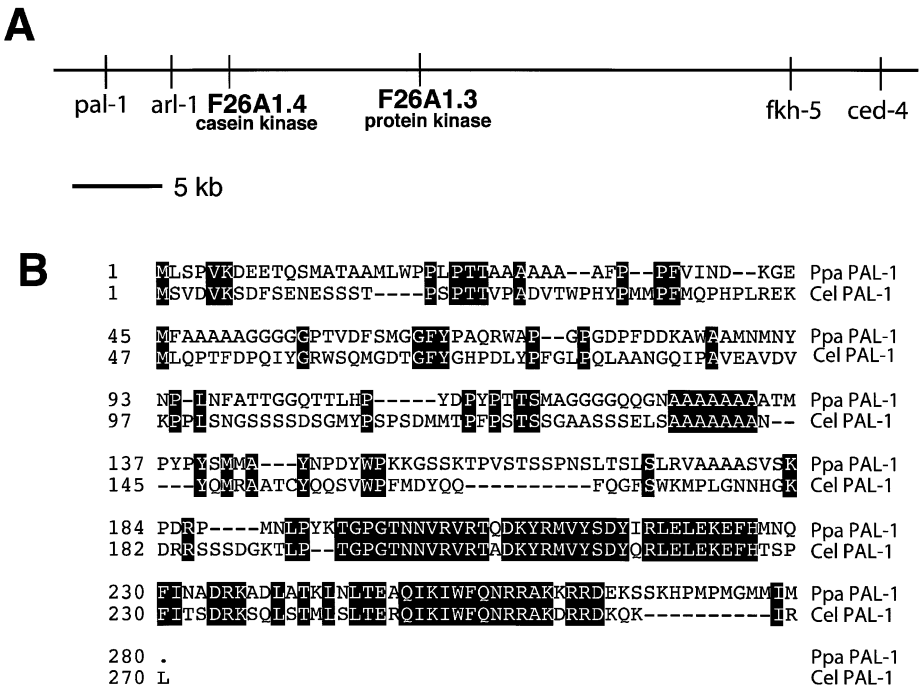

Figure 2. (**A**) Organization of the *pal-1* region of *C.elegans* chromosome III. (**B**) Sequence alignment of the *P.pacificus* and *C.elegans* PAL-1 proteins. Identical amino acids are shown in black.

pal-1 from *P.pacificus* using a PCR approach and obtained the complete ORF by 5′-RACE and 3′-RACE experiments (17). The *pal-1* proteins of *P.pacificus* and *C.elegans* are most similar in the homeodomain, but also contain small stretches of similarity in the N-terminal and the C-terminal parts of the protein (Fig. 2B). In total, the overall amino acid identity between the PAL-1 proteins of the two nematodes is 36%.

Next, we generated an SSCP marker in one of the introns of *Ppa-pal-1* by comparing the SSCP pattern of *P.pacificus* wild-type DNA from California with the DNA of the polymorphic strain from Washington. This SSCP marker was placed on the genetic linkage map and showed that *Ppa-pal-1* is located on chromosome III, as is *Cel-pal-1* (data not shown). Therefore, the *pal-1* region qualifies for the analysis of synteny, as there is no obvious translocation in that region of the genome.

Sequencing of the BAC clone 7E22

Hybridization experiments with a small fragment of *Ppa-pal-1* identified six BAC clones from the *P.pacificus Hin*dIII BAC library (12) which contain this locus. We chose PPBAC-7E22 for further analysis and sequenced this BAC using the shotgun method followed by gap filling using specific oligonucleotide primers. We obtained a total of 126 255 bp of continuous sequence (Fig. 3). The GC content is 43% and is similar to the value obtained previously from BAC end sequencing as a representation for the complete genome (12,14). In total, 20 ORFs have been predicted on PPBAC-7E22 with an average gene density of one gene every 6.3 kb.

Comparison of predicted ORFs

Next, we wondered if the predicted ORFs on PPBAC-7E22 are true genes and if they have homologs or even putative orthologs in *C.elegans*. Among the 20 predicted ORFs on PPBAC-7E22, three were known from existing *P.pacificus* EST clones (PPBAC-7E22.17, PPBAC-7E22.18 and PPBAC-7E22.19) and 12 were confirmed from RT–PCR experiments (Table 1). For the remaining five predicted ORFs, PPBAC-7E22.2, PPBAC-7E22.4, PPBAC-7E22.5, PPBAC-7E22.6 and PPBAC-7E22.8, we were unable to confirm these predicted ORFs by RT–PCR and there are no EST sequences available (Table 1). Interestingly, no clear *C.elegans* homolog could be assigned to these predicted ORFs, and the *C.elegans* sequence with the best hit for these ORFs have *e* values between 0.6 and 0.045 (Table 1). Therefore, there is no experimental support for these five gene predictions.

Fifteen of the predicted ORFs have a *C.elegans* homolog with high sequence similarity (*e* values < –10). Putative orthologs could be assigned to 11 of these genes, whereas for the four other predicted ORFs, several paralogous genes of *C.elegans* show equivalent sequence similarities (Table 1). In particular, PPBAC-7E22.1, PPBAC-7E22.7, PPBAC-7E22.9, PPBAC-7E22.10, PPBAC-7E22.12, PPBAC-7E22.13, PPBAC-7E22.14, PPBAC-7E22.15, PPBAC-7E22.17, PPBAC-7E22.19 and PPBAC-7E22.20 show sequence identities in the range of up to 80% to their *C.elegans* orthologs (Table 1). Given the amount of sequence similarity, these *P.pacificus* ORFs are most likely 1:1 orthologs of the corresponding *C.elegans* genes. In contrast, for the predicted ORFs PPBAC-7E22.3, PPBAC-7E22.11, PPBAC-7E22.16 and PPBAC-7E22.18, no orthology relationship could be assigned. For example, the predicted gene PPBAC-7E22.16, which encodes a putative *S*-adenosylmethionine synthetase, has equivalent sequence similarity to *C.elegans* genes C06E7.1a, C06E7.3 and T13A10.11a, respectively (Table 1). As the complete sequence of the *P.pacificus* genome is not known, it is possible that there are other *S*-adenosylmethionine

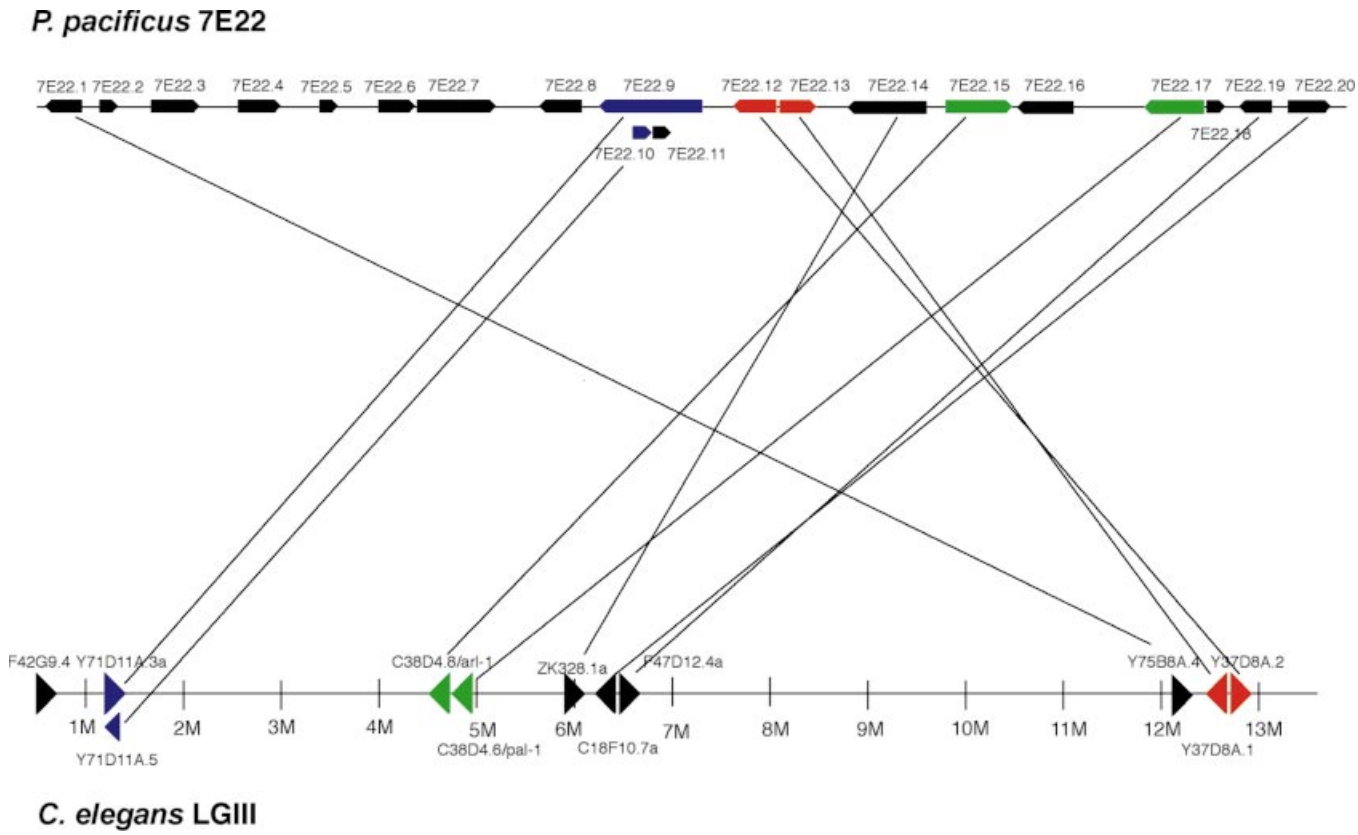

Figure 3. The PPBAC-7E22 contig compared to the *C.elegans* genome. Eleven of the 20 predicted genes of PPBAC-7E22 have a *C.elegans* ortholog. All but one of these orthologs are located on *C.elegans* chromosome III and span a region of >12 Mb. For detailed information on the predicted *P.pacificus* genes see Table 1.

synthetases with higher sequence similarities to these particular *C.elegans* genes. However, the comparison of the genomes of *C.elegans* and *C.briggsae* (14) indicated a high amount of gene duplication within nematodes. Therefore, it is more likely that gene duplications of *S*-adenosylmethionine synthetases have occurred in the *Pristionchus* and/or *Caenorhabditis* lineage such that a 1:1 orthology relationship of individual genes cannot be assigned.

In summary, the 20 predicted ORFs on PPBAC-7E22 can be divided into three categories: (i) genes with a putative *C.elegans* ortholog (11 of 20 predicted ORFs); (ii) genes with several *C.elegans* paralogs of similar sequence similarity (4 of 20 predicted ORFs); (iii) predicted ORFs that show no obvious sequence similarity to *C.elegans* genes and that have not been confirmed by EST and RT–PCR experiments (5 of 20 predicted ORFs).

Gene organization in relation to *C.elegans*

To evaluate the degree of synteny between the genomes of *P.pacificus* and *C.elegans*, we determined the genomic localization of the *C.elegans* orthologs of those genes identified on PPBAC-7E22. As indicated above, 11 ORFs in this BAC have a putative *C.elegans* ortholog (Table 1). Ten of these 11 orthologs are located on chromosome III in *C.elegans*. The only exception is the *C.elegans* gene F56C11.2, which is located on chromosome I and is the *C.elegans* gene with the highest similarity to PPBAC-7E22.7 (Fig. 3 and Table 1). These results support the existence of synteny between the genomes of the two nematode species. However, more detailed analysis indicates that 9 of the 10 *C.elegans* orthologs are not from the *pal-1* region (Fig. 3 and Table 1). For example, Y37D8A.2, the ortholog of PPBAC-7E22.12, is localized at position 12.81 Mb, whereas *Cel-pal-1* is at position 4.80 Mb (Table 1 and Fig. 3). Together, the *C.elegans* orthologs and homologs of *P.pacificus* genes of PPBAC-7E22 span a region of >12 Mb on chromosome III (Fig. 3).

Microsynteny is only observed for a few genes on PPBAC-7E22. PPBAC-7E22.15 encodes the *arl-1* (C38D4.8) ortholog of *C.elegans*. *arl-1* is the only gene to be found in the neighborhood of *pal-1* in both species. In *C.elegans*, *arl-1* and *pal-1* are ~10 kb apart, whereas in *P.pacificus* they are only 1.6 kb apart. Besides *pal-1* and *arl-1*, two additional pairs of genes exist that show linkage in both species. The *P.pacificus* genes PPBAC-7E22.9 and PPBAC-7E22.10 are closely linked, like their *C.elegans* orthologs Y71D11A.3a and Y71D11A.5 (see below for detailed description). Similarly, PPBAC-7E22.12 and PPBAC-7E22.13 are closely linked in *P.pacificus*, as are their *C.elegans* orthologs Y37D8A.2 and Y37D8A.1 (Fig. 3). Thus, the PPBAC-7E22 interval of the *P.pacificus* genome contains three pairs of genes that are also linked to one another in *C.elegans*. However, these three pairs of genes are distributed across 13 Mb of the 14 Mb *C.elegans* chromosome III (Fig. 3). Taken together, these data indicate

Table 1. Gene predictions on PPBAC-7E22

PPBAC7E22.x	Protein length (amino acids)	*C.elegans* homolog	Location	e-value	Confirmation	Gene prediction
1	502	Y75B8A.4	LGIII	5.9e-52	RT–PCR	ATPases associated with various cellular activities (AAA)
2	143	F42G9.4	LGIII	0.46	Not confirmed	Similarity to *E.coli* hypothetical protein YDBL precursor
3	932	F20B4.7	LGX	5.6e-88	RT–PCR	Reverse transcriptase
–	–	F09C3.6	LGI	3.8e-82	–	Reverse transcriptase
–	–	Y58G8A.2	LGV	9.7e-81	–	Reverse transcriptase
4	231	Y40B1A.2	LGI	0.42	Not confirmed	Similarity to Pfam domain: PF00397 (WW domain)
5	144	F56H11.3	LGIV	0.045	Not confirmed	Similarity to yeast hypothetical protein GNS1
6	458	K09A11.2	LGX	7.6e-32	Not confirmed	Similarity to Cytochrome P450
–	–	K09A11.4	LGX	1.4e-29	–	Similarity to Cytochrome P450
–	–	R04D3.1	LGX	2.2e-25	–	Similarity to Cytochrome P450
7	848	F56C11.2	LGI	9.5e-48	RT–PCR	*ptr-1*
8	378	–	LGX	0.68	Not confirmed	–
9	371	Y71D11A.3a	LGIII	8.0e-48	RT–PCR	2-Amino-3-carboxylmuconate-6-semialdehyde decarboxylase
10	519	Y71D11A.5	LGIII	2.2e-37	RT–PCR	Ion channel protein
11	447	F25H5.7	LGI	1.8e-19	RT–PCR	Protein-tyrosine phosphatase
–	–	F55F8.7	LGI	1.1e-17	–	Protein-tyrosine phosphatase
–	–	F47B3.1	LGI	2.7e-12	–	Protein-tyrosine phosphatase
12	440	Y37D8A.2	LGIII	5.5e-74	RT–PCR	–
13	409	Y37D8A.1	LGIII	9.5e-38	RT–PCR	Similarity to Arp2/3 complex subunit p21-Arc
14	1314	ZK328.1a	LGIII	1.1e-115	RT–PCR	EF hand/ubiquitin C-terminal hydrolases family 2
15	193	C38D4.8	LGIII	3.4e-26	RT–PCR	*arl-1*
16	527	C06E7.1a	LGIV	1.6e-56	RT–PCR	S-adenosylmethionine synthetase
–	–	C06E7.3	LGIV	1.6e-54	–	S-adenosylmethionine synthetase
–	–	T13A10.11a	LGIV	2.3e-53	–	S-adenosylmethionine synthetase
17	563	C38D4.6	LGIII	2.4e-23	EST	*pal-1*
18	147	E03H12.10	LGIV	2.2e-29	EST	Member of the sperm-specific family, class P gene class
–	–	C35D10.11	LGIII	9.7e-29	–	MSP (major sperm protein)
19	247	F47D12.4a	LGIII	1.5e-58	EST	hmg-1.2-like protein
20	357	C18F10.7a	LGIII	5.8e-53	RT–PCR	Ankyrin-repeat ubiquitin interaction motif

Putative ORFs are numbered as in Figure 3, and the *C.elegans* homolog and their location in the *C.elegans* genome are indicated. Gene predictions have been confirmed by EST clones and RT–PCR experiments.

that microsynteny is limited between the *P.pacificus* and *C.elegans* genomes, a notion that is in agreement to previous analysis in *B.malayi* (15). Thus, attempts at cloning *P.pacificus* genes by microsynteny to *C.elegans* is not a tool likely to be successful in most cases.

Genes within genes

Despite the absence of massive microsynteny, the organization and structure of individual genes can be well conserved between *P.pacificus* and *C.elegans*. Large-scale genome sequencing in different organisms has revealed many cases in which a gene is located in the intron of another gene that is transcribed in the antiparallel direction (26–29). The two genes PPBAC-7E22.9 and PPBAC-7E22.10 provide such examples (Fig. 4A). Interestingly, both genes have orthologs in *C.elegans*, Y71D11A.3a and Y71D11A.5, respectively, that are organized in similar ways. The two gene pairs show sequence identities of 65% (PPBAC-7E22.9/ Y71D11A.3a) and 70% (PPBAC-7E22.10/ Y71D11A.5) at the amino acid level (Fig. 4B and C). PPBAC-7E22.9 and Y71D11A.3 encode a 2-amino-3-carboxylmuconate 6-semialdehyde decarboxylase and contain a large intron of ~8.6 kb in *P.pacificus* and 17 kb in *C.elegans*. PPBAC-7E22.10 and Y71D11A.5 encode ion channels and are located in the two large introns of PPBAC-7E22.9 and Y71D11A.3a, respectively (Fig. 4A). In *C.elegans*, another predicted gene, Y71D11A.4, is also located in the same intron (Fig. 4A). This predicted gene encodes a transposase and has not yet been confirmed experimentally. No sequence similarity to Y71D11A.4 has been identified on PPBAC-7E22, further supporting the notion that Y71D11A.4 might not be a true gene. However, the predicted and confirmed gene PPBAC-7E22.11 is also located in an intron of PPBAC-7E22.9 and is transcribed in the same direction as PPBAC-7E22.10 (Fig. 4A). Interestingly, PPBAC-7E22.11 encodes a protein-tyrosine phosphatase with three similar genes in *C.elegans*, all of which are located on chromosome I (Table 1). Thus, this complex gene arrangement shows conserved as well as non-conserved elements such that it remains unknown what the ancestral gene arrangement looked like.

In addition to gene organization, the exon–intron structures of PPBAC-7E22.9 and Y71D11A.3a as well as PPBAC-7E22.10 and Y71D11A.5 are also very similar. More than half of the exon–intron boundaries are exactly conserved between *P.pacificus* and *C.elegans* in both gene pairs (Fig. 4). However, the *P.pacificus* genes contain more introns than the *C.elegans* genes. Specifically, PPBAC-7E22.9 contains 10 introns, whereas Y71D11A.3a contains only seven. Similarly, PPBAC-7E22.10 contains 14 introns, whereas there are only 11 introns in Y71D11A.5. The additional introns in *P.pacificus* are usually small with a size of ~100 bp. Both, the presence of additional introns and their smaller size seem to represent a general trend of the *P.pacificus* genome compared to *C.elegans*, and similar observations have been made previously during the analysis of several developmental control genes (4). Taken together, this example indicates that

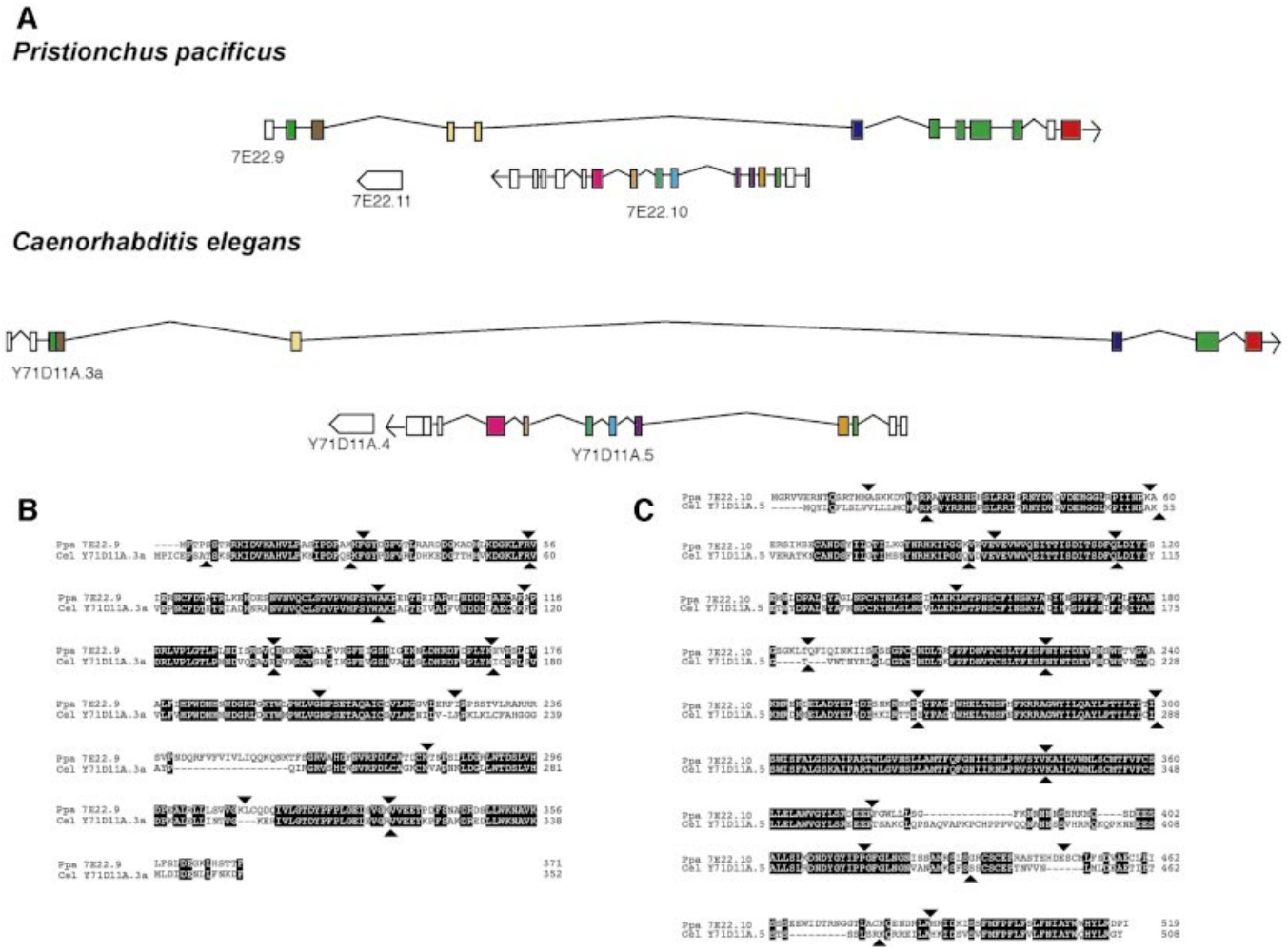

Figure 4. (**A**) Gene in gene organization of PPBAC-7E22.9 and PPBAC-7E22.10 and Y71D11A.3 and Y71D11A.5. PPBAC-7E22.9 and Y71D11A.3 encode a 2-amino-3-carboxylmuconate 6-semialdehyde decarboxylase and contain a large intron of 8.6 and 17 kb, respectively. PPBAC-7E22.10 and Y71D11A.5 encode ion channels and are located in the large intron of PPBAC-7E22.9 and Y71D11A.3, respectively. In *C.elegans*, Y71D11A.4 is a predicted reverse transcriptase that is also localized in the large intron of Y71D11A.3. No sequence similarity to Y71D11A.4 is observed in the large intron of PPBAC-7E22.9 or anywhere else on PPBAC-7E22. More than half of the exon–intron boundaries are conserved in both gene pairs. Identical colors of exons indicate homologous sequence. Exon–intron borders are often conserved, i.e. exon 3 of PPBAC-7E22.9 and exon 4 of Y71D11A.3a. In some cases one exon of Y71D11A.3a, such as exon 7, corresponds to several exons in PPBAC-7E22.9. It remains unknown which of the two patterns represents the ancestral situation. Both intron gain or intron loss could account for the observed patterns. Exons shown in white cannot be clearly aligned to one another by sequence and intron boundaries. (**B** and **C**) Sequence alignement of the *P.pacificus* PPBAC-7E22.9 and *C.elegans* Y71D11A.3 proteins (B) and *P.pacificus* PPBAC-7E22.10 and *C.elegans* Y71D11A.5 proteins (C). Identical amino acids are shown in black. Intron positions are indicated by triangles.

although microsynteny in general is limited between *P.pacificus* and *C.elegans*, complex gene arrangements and exon–intron structures can be conserved between both organisms.

DISCUSSION

Pristionchus pacificus is an important satellite organism in evolutionary developmental biology (3). After the generation of a genetic linkage map (12) and a physical map (J.Srinivasan and R.J.Sommer, in preparation), this study initiates a second step in the genomic analysis of this organism. The 126 kb fragment of genomic sequence of *P.pacificus* is the largest continuous genomic fragment of a non-rhabditid nematode known to date. Fifteen of the 20 predicted ORFs on PPBAC-7E22 have a *C.elegans* homolog, with 11 of these 15 ORFs having putative *C.elegans* orthologs. For the other five predicted ORFs, it remains unknown if they correspond to real genes, as we have not been able to confirm them by RT–PCR. Thus, this study does not provide evidence for unique *P.pacificus* genes. However, some of the *P.pacificus* genes might be the result of independent gene duplications that occurred after the phylogenetic separation of *C.elegans* and *P.pacificus*. It should be noted, however, that some of the predicted ORFs that were not confirmed experimentally are clustered together and are flanked by genes that are syntenic. One could speculate that this pattern is the result of a major insertion/deletion event.

If the 126 kb contig is a good representation of the *P.pacificus* genome, the gene complement is likely to be equivalent to that of *C.elegans*, an observation that has recently also been made for *B.malayi*, in which a 83 kb

fragment has been sequenced (15). However, smaller differences in genome size and gene complement cannot be excluded and might result from the different organization of smaller regions of the genomes. So far, the analysis of the *P.pacificus* genome predicted a genome size of ~100 Mb (12,30). However, genetic analysis revealed substantial differences of the genetic distances of the different chromosomes of *P.pacificus*. Unlike in *C.elegans*, where all six chromosomes are ~50 cM, SSCP marker analysis indicated that at least two *P.pacificus* chromosomes are substantially larger than 50 cM (12). At the same time, two other chromosomes are genetically smaller than 50 cM (12) (W.Sinz and R.J.Sommer, unpublished observation). How these different genetic properties of *C.elegans* and *P.pacificus* affect the genome size and the gene complement remains unknown.

The comparison of the genes on PPBAC-7E22 with the genomic localization of their putative *C.elegans* orthologs revealed that synteny is common, although microsynteny is limited. There are only three pairs of genes that show microsynteny, *pal-1/arl-1*, Y71D11A.3, PPBAC-7E22.9/Y71D11A.5, PPBAC-7E22.10 and Y37D8A.2, PPBAC-7E22.12/Y37D8A.1, PPBAC-7E22.13, respectively. Various genetic mapping experiments of vulva-defective mutants in *P.pacificus*, using the end-sequenced *Hin*dIII BAC library and EST clones support this observation: putative orthologs of various *C.elegans* genes are linked on one chromosome as they are in *C.elegans*. However, their exact order at a microscale level differs strongly from *C.elegans* (W. Sinz and R.J. Sommer, unpublished observation). These results are similar to observations in other invertebrates, but differ strongly from findings in vertebrates. For human, mouse and pufferfish, a substantial amount of microsynteny has been observed, indicating that intrachromosomal exchange is rare (31). In contrast, the data from *C.elegans*, *C.briggsae*, *P.pacificus* and *B.malayi* suggest that intrachromosomal exchanges occurred frequently within nematodes. This high level of rearrangement limits the likelihood of identifying orthologous genes of *P.pacificus* and *C.elegans* based on position information within the two genomes. For example, *Ppa-ced-4* is not found in the neighborhood of *Ppa-pal-1*, as is the case in *C.elegans*. Furthermore, this result strengthens the previous observation, mainly based on a comparison of *Caenorhabditis* and *Drosophila* species, that in general, nematodes show a faster rate of genome rearrangements than other phyla (32).

The recent generation of a genetic linkage map in *P.pacificus* provides the basis for extending the analysis of this species from developmental biology to other areas of biology, including genomics. Given the phylogenetic separation time of 100 000 000–200 000 000 years between *C.elegans* and *P.pacificus*, genomic comparisons of these two organisms may provide information on structural properties of nematode genomes that cannot be deduced from within-genus comparisons, such as between *C.elegans* and *C.briggsae*. Consistent with the different evolutionary separation times (Fig. 1), the organization and sequence of orthologous genes differs more strongly between *P.pacificus* and *C.elegans* than between *C.briggsae* and *C.elegans.* A larger sequencing effort in *P.pacificus* might therefore be informative for *C.elegans* biology itself in addition to providing information on more general aspects of nematode genomes. In *C.elegans*, the identification and analysis of regulatory control regions is often hampered by the fact that important regulatory elements can be located in the 5′ untranslated region (UTR), in introns and/or in the 3′ UTR (33). Sequence similarities in potential regulatory regions (promoters/introns) between two species such as *P.pacificus* and *C.elegans* that may have been separated for as long as 200 000 000 years are most likely based on functional restrictions. Therefore, sequence conservation in untranslated regions of orthologous genes between *P.pacificus* and *C.elegans* might provide a way to identify important regulatory elements (R. Hong and R.J. Sommer, unpublished observation). Such candidate elements could then be tested using GFP constructs in *C.elegans*. In a similar way, the sequence comparison between *C.elegans* and *C.briggsae* has been shown to be a useful tool for gene prediction and the identification of gene structure (14).

ACKNOWLEDGEMENTS

We thank the members of the Sommer laboratory for discussions and R. Hong, D. Rudel and J. Srinivasan for critically reading the manuscript. The work described in this manuscript has been carried out in compliance with the current law governing genetic experimentation in Germany.

REFERENCES

1. Sommer,R.J. (2001) As good as they get: cells in nematode vulva development and evolution. *Curr. Opin. Cell Biol.*, **13**, 715–720.
2. Eizinger,A., Jungblut,B. and Sommer,R.J. (1999) Evolutionary change in the functional specificity of genes. *Trends Genet.*, **15**, 197–202.
3. Simpson,P. (2002) Evolution of development in closely related species of flies and worms. *Nature Rev. Genet.*, **3**, 907–991.
4. Sommer,R.J. (2000) Evolution of nematode development. *Curr. Opin. Genet. Dev.*, **10**, 443–448.
5. Wang,M. and Sternberg,P.W. (2001) Pattern formation during *C. elegans* vulval induction. *Curr. Topics Dev. Biol.*, **51**, 189–220.
6. Eizinger,A. and Sommer,R.J. (1997) The homeotic gene *lin-39* and the evolution of nematode epidermal cell fates. *Science*, **278**, 452–455.
7. Sommer,R.J., Eizinger,A., Lee,K.Z., Jungblut,B., Bubeck,A. and Schlak,I. (1998) The *Pristionchus* HOX gene *Ppa-lin-39* inhibits programmed cell death to specify the vulva equivalence group and is not required during vulval induction. *Development*, **125**, 3865–3873.
8. Jungblut,B. and Sommer,R.J. (1998) The *Pristionchus pacificus mab-5* gene is involved in the regulation of ventral epidermal cell fates. *Curr. Biol.*, **8**, 775–778.
9. Jungblut,B. and Sommer,R.J. (2001) The nematode *even-skipped* homolog *vab-7* regulates gonad and vulva position in *Pristionchus pacificus. Development*, **128**, 253–261.
10. Srinivasan,J., Pires-daSilva,A., Gutierrez,A., Zheng,M., Jungblut,B., Witte,H., Schlak,I. and Sommer,R.J. (2001) Microevolutionary analysis of the nematode genus *Pristionchus* suggests a recent evolution of redundant developmental mechanisms during vulva formation. *Evol. Dev.*, **3**, 229–240.
11. Schlak,I., Eizinger,A. and Sommer,R.J. (1997) High rate of restriction fragment length polymorphisms between two populations of the nematode *Pristionchus pacificus* (Diplogastridae). *J. Zool. Syst. Evol.*, **35**, 137–142.
12. Srinivasan,J., Sinz,W., Lanz,C., Brand,A., Nandakumar,R., Raddatz,G., Witte,H., Keller,H., Kipping,I., Pires-DaSilva,A. *et al*. (2002) A bacterial artificial chromosome-based genetic linkage map of the nematode *Pristionchus pacificus*. *Genetics*, **162**, 129–134.
13. McCarter,J.P., Clifton,S.W., Bird,D.M. and Waterston,R.H. (2002) Nematode gene sequences, update for June 2002. *J. Nematol.*, **34**, 71–74.
14. Kent,W.J. and Zahler,A.M. (2000) Conservation, regulation, synteny, and introns in a large-scale *C. briggsae–C. elegans* genomic alignment. *Genome Res.*, **10**, 1115–1125.

15. Guiliano,D.B., Hall,N., Jones,S.J., Clark,L.N., Corton,C.H., Barrell,B.G. and Blaxter,M.L. (2002) Conservation of long-range synteny and microsynteny between the genomes of two distantly related nematodes. *Genome Biol.*, **3**, 57.1–57.14.
16. Ranasinghe,C. and Hobbs,A.A. (1998) Isolation and characterization of two cytochrome P450 cDNA clones for CYP6B6 and CYP6B7 from *Helicoverpa armigera* (Hubner): possible involvement of CYP6B7 in pyrethroid resistance. *Insect Biochem. Mol. Biol.*, **28**, 571–580.
17. Frohman,M.A. (1994) On beyond classic RACE (rapid amplification of cDNA ends). *PCR Methods Appl.*, **4**, S40–S58.
18. Hunter,C.P. and Kenyon,C. (1996) Spatial and temporal controls target *pal-1* blastomere-specification activity to a single blastomere lineage in *C. elegans* embryos. *Cell*, **87**, 217–226.
19. Waring,D.A. and Kenyon,C. (1990) Selective silencing of cell communication influences anteroposterior pattern formation in *C. elegans*. *Cell*, **60**, 123–131.
20. Waring,D.A. and Kenyon,C. (1991) Regulation of cellular responsiveness to inductive signals in the developing *C. elegans* nervous system. *Nature*, **350**, 712–715.
21. Fleischmann,R.D., Adams,M.D., White,O., Clayton,R.A., Kirkness,E.F., Kerlavage,A.R., Bult,C.J., Tomb,J.F., Dougherty,B.A., Merrick,J.M. *et al*. (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science*, **269**, 496–512.
22. Ewing,B. and Green,P. (1998) Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res.*, **8**, 186–194.
23. Brosch,R., Gordon,S.V., Billault,A., Garnier,T., Eiglmeier,K., Soravito,C., Barrell,B.G. and Cole,S.T. (1998) Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing and comparative genomics. *Infect. Immun.*, **66**, 2221–2229.
24. Harris,T.W., Lee,R., Schwarz,E., Bradnam,K., Lawson,D., Chen,W., Blasier,D., Kenny,E., Cunningham,F., Kishore,R. *et al.* (2003) WormBase: a cross-species database for comparative genomics. *Nucleic Acids Res.*, **31**, 133–137.
25. Altschul,S.F., Gish,W., Miller,W., Myers,E.W. and Lipman,D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.*, **215**, 403–410.
26. Waterston,R.H., Lander,E.S. and Sulston,J.E. (2002) On the sequencing of the human genome. *Proc. Natl Acad. Sci. USA*, **99**, 3712–3716.
27. Venter,J.C., Adams,M.D., Myers,E.W., Li,P.W., Mural,R.J., Sutton,G.G., Smith,H.O., Yandell,M., Evans,C.A., Holt,R.A. *et al*. (2001) The sequence of the human genome. *Science*, **291**, 1304–1351.
28. Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D., Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science*, **287**, 2185–2195.
29. The *C. elegans* Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, **282**, 2012–2018.
30. Sommer,R.J., Carta,L.K., Kim,S.-y. and Sternberg,P.W. (1996) Morphological, genetic and molecular description of *Pristionchus pacificus* sp. n. (Nematoda: Neodiplogastridae). *Fundam. Appl. Nematol.*, **19**, 511–521.
31. Aparicio,S., Chapman,J., Stupka,E., Putnam,N., Chia,J.M., Dehal,P., Christoffels,A., Rash,S., Hoon,S., Smit,A. *et al*. (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, **297**, 1301–1310.
32. Coghlan,A. and Wolfe,K.H. (2002) Fourfold faster rate of genome rearrangement in nematodes than in *Drosophila*. *Genome Res.*, **12**, 857–867.
33. Grandien,K. and Sommer,R.J. (2001) Functional comparison of the nematode Hox gene *lin-39* in *C. elegans* and *P. pacificus* reveals evolutionary conservation of protein function despite divergence of primary sequences. *Genes Dev.*, **15**, 2161–2172.