Available online at www.sciencedirect.com
SCIENCE DIRECT®
ELSEVIER
Molecular & Biochemical Parasitology 145 (2006) 239–244
MOLECULAR & BIOCHEMICAL PARASITOLOGY

Comparative analysis of the expressed genome of the infective juvenile entomopathogenic nematode, *Heterorhabditis bacteriophora* ☆

Sukhinder K. Sandhu, Ganpati B. Jagdale, Saskia A. Hogenhout, Parwinder S. Grewal *

Department of Entomology, The Ohio State University, OARDC, Wooster, OH 44691, USA

Abstract

We report the first cDNA-sequencing project of the entomopathogenic nematode, *Heterorhabditis bacteriophora*. A total of 1246 expressed sequence tags (ESTs) were generated by random sequencing of clones from a cDNA library of the infective juvenile stage. The ESTs were annotated resulting in 1072 useful ESTs that were categorized into functional categories according to Kyoto Encyclopedia of Genes and Genomes. Approximately 459 of 1072 ESTs (43%) had significant similarities to annotated sequences in GenBank. Of these, 417 had significant similarities to the free-living nematode *Caenorhanditis elegans* proteins. Most ESTs (18%) belonged to the genetic information processing category followed by metabolism (15% ESTs) and environmental information processing (15%) pathways. Several interesting ESTs were found that may have roles in the infectivity and survival of infective juveniles. These included proteases, dauer pathway genes (*akt-1, pdk-1 & daf-7*) and aging and stress resistance genes such as superoxide dismutase (*sod-4*), heat shock genes (*hsp-4 & hsp-6*), and *eat* genes, and signaling proteins like G-protein coupled receptors, regulators of G-protein signaling (*rgs*), and serine/threonine kinases. Other interesting ESTs include systemic RNAi defective protein (*sid-1*), ribonuclease III family members (*rnh-2 & rnc*) and transposase gene (Tc3A). About 67% of the ESTs did not find matches in any of the searched databases suggesting potentially novel genes in this enomopathogenic nematode.
© 2006 Published by Elsevier B.V.

Keywords: *Heterorhabditis bacteriophora*; GPS11 strain; Entomopathogenic nematodes; Dauer juveniles; ESTs

Entomopathogenic nematodes in the families Heterorhabditidae (Strongyloidea) and Steinernematidae (Strongyloidoidea sensu [1]) are unique models for the study of parasitism, pathogenicity, and symbiosis. These nematodes form mutualistic symbiosis with insect pathogenic bacteria in the family Enterobacteriaceae: heterorhabditids are associated with *Photorhabdus* and steinernematids with *Xenorhabdus*, respectively [2]. The infective juveniles (IJs) or dauer (enduring) juveniles persist in soil in search of a suitable insect host [3]. Following entry through the cuticle or natural body openings, the IJs release the symbiotic bacteria into the insect hemocoel, which rapidly kill the host, usually within 24–48 h [4]. Nematodes feed on symbiotic bacteria, complete 1–3 generations in the host cadaver, and as food resources are depleted new IJs are produced which disperse in search of new hosts. In the laboratory, each partner can be cultured separately, but when combined they present a high degree of specificity.

Entomopathogenic nematodes (EPNs) are also extremely important biological control agents of insect pests [5]. However, to realize their full potential improvements are needed in the IJ longevity, bacterial retention, tolerance to environmental extremes (particularly heat, ultra violet radiation, and desiccation), resistance to encapsulation in the hemocoel encountered in some key insects, and trait stability. Recent developments have opened new possibilities for improving the biological control potential of EPNs. First, the genome of *Photorhabdus luminescens* subsp *luminescens*, the bacterial symbiont of *Heterorhabditis bacteriophora*, has been fully sequenced. The 5.7 Mb bacterial genome contains 4839 predicted genes and a plethora of putative virulence factors, including Type I–IV and TPS secretion systems, adhesion proteins, toxins, hemolysins, proteases, lipases, and antibiotics [6]. Second, the availability of the genome sequences of two closely related free-living nema-

Abbreviations: EST, expressed sequence tags; EPN, entomopathogenic nematodes; KEGG, Kyoto encyclopedia of genes and genomes; BLAST, basic local alignment search tool; RNAi, RNA interference; cDNA, complementary DNA; dsRNA, double stranded RNA; IJs, infective juveniles

☆ *Note:* Sequences described in this paper have been deposited in Genbank under the accessions DN152655–DN152999 and DN153000–DN153726.

* Corresponding author. Tel.: +1 330 263 3963; fax: +1 330 263 3686.
E-mail address: grewal.4@osu.edu (P.S. Grewal).

0166-6851/$ – see front matter © 2006 Published by Elsevier B.V.
doi:10.1016/j.molbiopara.2006.01.002

todes, *Caenorhabditis elegans* and *C. briggsae*, together with a full complement of genetic and molecular tools [7,8] offers new means for improving the biological control potential of *H. bacteriophora*.

In order to establish genomics as a tool to fully exploit the unique biology and biological control potential of the entomopathogenic nematodes, we describe a pilot scale cDNA-sequencing project that provides the first foray into the gene diversity and discovery in *H. bacteriophora*. We selected *H. bacteriophora* for this study because it is closely related to *C. elegans* [9], the availability of the genome sequence of its bacterial symbiont [6], and the demonstration of the expression of *C. elegans* gene sequences in *H. bacteriophora* [10]. The main goal of this study was to determine sequence similarities of the expressed genomes of *H. bacteriophora* and *C. elegans*. We focused on the expressed genome of the infective stage of *H. bacteriophora* because this stage is similar to *C. elegans* dauer stage in ontogeny, and serves as the basis for parasitism and mutualism in *H. bacteriophora*. Despite its comparative genomics value, functional comparisons between stress responses, aging, innate immunity, olfaction, sex determination and dauer formation will be of great interest in relationship to mutualism, parasitism, and biological control.

The *H. bacteriophora* strain GPS11 was cultured in the third instar wax moth, *Galleria mellonella* larvae, as described by Kaya and Stock [11]. The nematodes were harvested from the white traps [12] in water, concentrated on filter paper and immediately immersed in liquid nitrogen before storage at −80 °C till further processing. For total RNA isolation, nematodes were taken from the –80 °C freezer and grounded in liquid nitrogen to a fine powder using a pestle and mortar. Total RNA was extracted with the Totally RNA kit (Ambion Inc. Austin, TX) following the instructions provided by the manufacturer. RNA quality and quantity was determined by electrophoresis [13]. Briefly, 10 μg of total RNA in a final volume of 15 μl in a PCR tube was mixed with 15 μl of Ambion Glyoxal Sample Loading Dye and heated for 35 min at 50 °C in the PCR machine. The tubes were chilled on ice. The RNA was separated by electrophoresis at 70 V for 2 h in 1.2% agarose gel in 1× BPTE buffer. About 19.2 μg of good quality total RNA was sent to Evrogen for cDNA library construction and sequencing.

The cDNA library was constructed from total RNA using SMART approach [14] and the cDNA library was normalized using the Duplex Specific Nucleotide (DSN) normalization method [15] to correct for over abundance of highly expressed transcripts. In the DSN method, the hybridization reaction contained 3 μl of purified double stranded cDNA and 1 μl 4× hybridization buffer (200 mM HEPES-HCL, pH 8.0; 2 M NaCl) overlaid with one drop of mineral oil and incubated at 95 °C for 5 min and 68 °C for 4 h. The DSN treatment included addition of 3.5 μl milliQ water, 1 μl of 5× DNAse buffer (500 mM Tris–HCl, pH 8.0; 50 mM $MgCl_2$; 10 mM DTT) and 0.5 μl DSN enzyme to the 68 °C hybridization reaction and subsequent incubation at 65 °C for 25 min. The DSN enzyme was heat-inactivated at 97 °C for 5 min. The cDNA sample was diluted with 30 μl milliQ water and used for PCR amplification. The PCR reaction (50 μl) contained: 1 μl diluted cDNA, 1× Advantage 2 reaction buffer (BD Biosciences Clontech, Palo Alto, CA), 200 μM dNTPs, 0.3 μM SMART PCR primer, 1× Advantage 2 Polymerize mix (BD Biosciences Clontech, Palo Alto, CA) and was carried in a MJ Research PTC-200 DNA Thermal Cycler. The PCR conditions were 19 PCR cycles, 95 °C for 7 s, 65 °C for 20 s and 72 °C for 3 min. Amplified normalized cDNA was purified using QIAquick PCR Purification Kit (QIAGEN, CA), ligated into modified pAL 16 vector (Evrogen) and used for *Escherichia coli* transformation with the BioRad Micropulser (Bio-Rad). To confirm successful transformation of *E. coli* cells, 11 randomly picked clones were tested by PCR with Sp6 and T7 primers and all were found to contain the insert. The library was arrayed in 10 plates (approximately 20,000 white colonies per plate) and about 200,000 independent clones were obtained. Colonies were eluted by LB/Amp medium. Colonies were picked and incubated in 50 μl of distilled water at 94 °C for 2 min. PCR amplifications were performed in a 20 μl volume on 5 μl of colony extract using M13-forward and reverse-primers and AmpliTaq Gold DNA Polymerase (Perkin-Elmer) following PCR profile, 94C°C; 10 min; then 10 cycles of 94 °C for 30 s; 65 °C for 30 s with a decrease of 1 °C per cycle and 72 °C for 2.5 min; followed by 30 similar cycles with an annealing temperature kept at 55 °C. PCR products were separated by agarose gel electrophoresis. Amplification products were used as templates for sequencing reactions by using the PRISM^{TM} Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems) and the SL1 primer. In general the sequence reads start approximately 25 nucleotides from the SL1 sequence. For selected clones, the 5′ end of the cDNA was sequenced by using the M13-forward or –reverse primer. The sequencing reactions were analyzed on an Applied Biosystems 373A DNA sequencer.

A total of 1246 ESTs were sequenced from cDNA library. Quality scores of sequence data were obtained with the McPhred/McPhrap software (Phil Green, University of Washington, Seattle, WA) and further sequence processing (vector and end trimming) was conducted with Sequencher software (Gene Codes Corp., Ann Arbor, MI). Trimming parameters were a.) trim from the 5′ end, trimming no more than 25%, trim until the first 15 bases containing less than 3 bases with confidences below 20 and b.) trim from the 3’end, trimming no more than 25%, trim until the last 15 bases contain less than 3 bases with confidence below 20. Sequences from plasmid and linker/adapters were trimmed as well and the ESTs shorter than 50 bp were removed from the dataset. This resulted in 1072 ESTs with an average insert size of 502 ± 162 bp. The 1072 sequences were then searched against existing protein databases using BLASTX installed on a local linux workstation. The ESTs were compared against the following databases: wormpep131 (*C. elegans* protein database at The Wellcome Trust Sanger Institute, Cambridge, UK), the fully curated Universal Protein Resource (uniprot_SWISSPROT) [16] and the non-redundant (nr) protein database at the National Center for Biotechnology Information using the BLASTX algorithm [17] (release 2004 with BLOSUM62 substitution matrix and the standard genetic code). Sequence similarities with e-value less than or equal to 10^{-5} were considered significant. To categorize the transcripts

Table 1
Functional categorisation of 417 ESTs of *H. bacteriophora* with significant similarity to *C. elegans* proteins

Regulatory pathway	Number of ESTs
1. Metabolism	60
1.1. Carbohydrate metabolism	20
1.2. Energy metabolism	1
1.3. Lipid metabolism	8
1.4. Nucleotide metabolism	4
1.5. Amino acid metabolism	15
1.6. Metabolism of other amino acids	4
1.7. Glycan biosynthesis and metabolism	4
1.8. Biosynthesis of polyketides and non ribosomal peptides	1
1.9. Metabolism of Cofactors and Vitamins	3
1.10. Biosynthesis of secondary metabolites	0
1.11. Biodegradation of xenobiotics	0
2. Genetic information processing	76
2.1. Transcription	19
2.2. Translation	18
2.3. Sorting and degradation	29
2.4 Replication and repair	10
3. Environmental information processing	59
3.1 Membrane transport	3
3.2. Signal transduction	32
3.3. Ligand-receptor interaction	24
3.4. Immune system	0
4. Cellular processes	14
4.1 Cell motility	3
4.2 Cell growth and death	6
4.3 Cell communication	2
4.4 Development	3
5. Unassigned	74
6. Hypothetical proteins	134

by function, the ESTs were assigned to metabolic pathways following the Kyoto Encyclopedia of Genes and Genomes [18] (KEGG) database and the ESTs with the same hit were grouped.

The BLAST searches showed that 613 ESTs (57%) did not match any protein in the searched databases and were hence novel. A total of 459 ESTs (43%) matched to one or more known sequences in all the searched databases. Of these, 417 had significant similarities to *C. elegans* proteins and the remaining 42 were different. The 417 ESTs that showed similarities to *C. elegans* proteins were categorized into functional groups following KEGG pathway database (see supplementary Table 1). Most ESTs (18%) belonged to the genetic information processing category with most of them having role in the sorting and degradation pathways (Fig. 1). The proteases found in this category may have important role in insect pathogenesis by the EPNs as shown by previous work on effect of protease inhibitors on reduced penetration of *Steinernema glaseri* into insect hosts [19]. The second most represented category was metabolism (15% ESTs) with carbohydrate, amino acid and lipid metabolism in the order of representation. The lipid metabolism pathway in the dauer stage EPNs is the most active pathway as the fat reserves are the main reservoirs for energy [20]. The major genes found in the overall metabolism category of the infective stage include co UDP-glucuronosyltransferase (R11A8.3), hexokinase (H25P06.1), fructose-1,6-biphosphatase (*fbp-1*) (K07A3.1), glycerol-3-phosphate dehydrogenase (K11H3.1a), and cysteine protease (*cpr-1*) (F36D3.9). Most of these transcripts are common to those found in SAGE [21] and microarray [22] studies of *C. elegans* dauer stage. A large number of ESTs (14%) belonged to the category of environmental information processing. About 17% of the ESTs remained unassigned to any functional category and 31% ESTs were similar to *C. elegans* hypothetical proteins (Table 1, Fig. 1). The least number of transcripts were found in the cellular processes category suggestive of developmental arrest at this stage. Indeed, dauer formation is an adaptation to tide over the unfavorable conditions of low food by staying in an arrested state of development and reproduction [23]. The GC-content of 1072 *H. bacteriophora* ESTs obtained in this study was 38%.

Our data show the existence of several genes in *H. bacteriophora* library similar to *C. elegans* aging and dauer pathway genes. A member of Transforming growth factor-β (TGF-β) pathway, *daf-7*, and two homologs of IGF-1 (insulin-like growth factor) signaling pathway, *akt-1* and *pdk-1* [24], found in *H. bacteriophora* ESTs seem to support the fact that there is a common pathway driving entry into dauer stage between *C. elegans* and EPNs. The programming of the above-mentioned genes is regulated by IGF-1 signaling pathway and TGF-β signaling pathway. In *C. elegans*, a signal transduction cascade from the DAF-2 insulin/IGF-1 receptor homolog to the DAF-16 forkhead transcription factor regulates metabolism, development, and longevity through *akt-1* and *pdk-1* receptors [24]. Further studies on dauer pathway genes in *H. bacteriophora* will help unravel the infective juvenile formation mechanisms in EPNs and may also enable improvements in infective juvenile recovery, production and longevity.

Other genes which are not specific to the dauer stage but known to be more highly represented in the dauer stage of *C. elegans* included heat shock genes such as *hsp-70* and protein phosphatase I (PP1C) γ protein. We found ESTs similar to stress-resistance proteins like superoxide dismutase (*sod-4*), heat shock proteins (*hsp-4, hsp-6*), and orthologs of yeast sirutin (*sir-4*) and *C. elegans eat* genes in the *H. bacteriophora* cDNA library. It has been shown that superoxide dismutase plays a significant role in longevity of *C. elegans* adults by depleting the superoxide radicals responsible for serious oxidative damage, which results in aging [25]. Recent experiments with certain

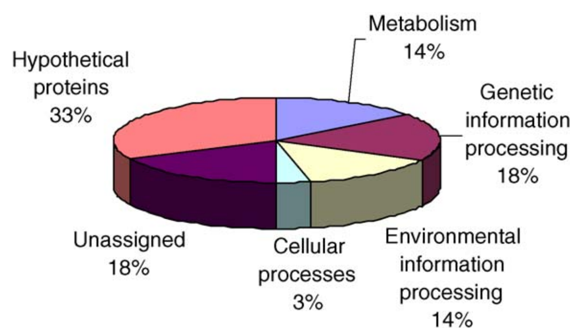

Fig. 1. Percent representation of *H. bacteriophora* ESTs in KEGG regulatory pathways.

Table 2
H. bacteriophora ESTs with no homolog in wormpep130

EST	Genbank protein hit	Genbank protein ID	e-value
Bacterial proteins			
HbGPS11.10K02	Hypothetical protein plu3613 [*Photorhabdus luminescens*]	ref|NP_930825.1|	92, 7e-18
HbGPS11.6I15	Fatty acid metabolism regulator protein [*Photorhabdus luminescens*]	ref|NP_929798.1|	69, 7e-22
HbGPS11.6D19	TcdA2 Insecticidal toxin protein[*Photorhabdus luminescens*]	gb|AAO17201.1|	120, 6e-26
HbGPS11.14E12	Putative outer membrane protein, probably involved in nutrient binding [*Bacteroides thetaiotaomicron*]	ref|NP_810788.1|	100, 3e-20
HbGPS11.14K23	Type I restriction-modification enzyme subunit M; HsdM [*Klebsiella pneumoniae*]	gb|AAQ84546.1|	107, 3e-34
HbGPS11.10F07	COG1639: Predicted signal transduction protein [*Microbulbifer degradans*]	ref|ZP_00317041.1|	89, 5e-19
HbGPS11.10E17	Histidine kinase response regulator hybrid protein [*Leptospira interrogans*]	ref|YP_002994.1|	48, 2e-09
HbGPS11.14O23	Orf122 [*Chlorobium tepidum*]	gb|AAG12204.1|	52, 4e-07
Eukaryotic proteins			
HbGPS11.14E20	Hypothetical protein FG00211.1 [*Gibberella zeae*]	gb|EAA68933.1|	82, 7e-15
HbGPS11.14G01	Putative MAPK [*Tetrahymena thermophila*]	gb|AAR04352.1|	77, 3e-13
HbGPS11.14I17	Hypothetical protein MG08934.4 [*Magnaporthe grisea*]	gb|EAA46591.1|	114, 8e-25
HbGPS11.14G11	Putative leishmanolysin-like protein [*Oryza sativa*]	gb|AAT77877.1|	62, 1e-08
HbGPS11.6E21	At5g48020 [*Arabidopsis thaliana*]	gb|AAP37672.1|	55, 2e-06
HbGPS11.14F10	PP2C protein phosphatase [*Arabidopsis thaliana*]	ref|NP_680572.1|	53, 2e-06
HbGPS11.14K05	Hypothetical protein UM01909.1 [*Ustilago maydis*]	ref|XP_399524.1|	234, 4e-62
HbGPS11.14K07	C15orf26 protein [*Homo sapiens*]	gb|AAH62471.1|	55, 3e-06
HbGPS11.10G18	Unnamed protein product [*Homo sapiens*]	dbj|BAC86735.1|	64, 1e-09
HbGPS11.14M05	PREDICTED: hypothetical protein [*Canis familiaris*]	ref|XP_541720.1|	59, 9e-08
HbGPS11.14M12	Choriogenin Hminor [*Oryzias latipes*]	dbj|BAA76901.1|	60, 5e-08
HbGPS11.14N14	Mitochondrial protein of unknown function, overexpression suppresses an rpo41 mutation affecting mitochondrial RNA polymerase; encoded within the 25S rRNA gene on the opposite strand [*Saccharomyces cerevisiae*]	ref|NP_690845.1|	57, 1e-07
HbGPS11.14O10	Similar to KIAA1749 protein [*Rattus norvegicus*]	ref|XP_236385.2|	56, 8e-07
HbGPS11.6A08	Similar to *Plasmodium falciparum* Hypothetical protein	gb|AAS45365.1|	55, 1e-06
HbGPS11.6I05	Hypothetical protein [*Plasmodium falciparum*]	ref|NP_704513.1|	72, 7e-12
HbGPS11.14D20	ENSANGP00000005606 [*Anopheles gambiae*]	ref|XP_310378.1|	67, 6e-10
HbGPS11.14O01	ENSANGP00000017298 [*Anopheles gambiae*]	ref|XP_310523.1	54, 6e-06
HbGPS11.10J19	ENSANGP00000019133 [*Anopheles gambiae*]	ref|XP_308373.1|	55, 6e-10
HbGPS11.6D11	CG31802-PA [*Drosophila melanogaster*]	ref|NP_724103.1|	41, 3e-07
HbGPS11.14N22	Flj13081-prov protein [*Xenopus laevis*]	gb|AAH42336.1|	83, 5e-15
HbGPS11.6I21	Unknown (protein for MGC:89513) [*Xenopus tropicalis*]	gb|AAH84996.1|	75, 1e-12
HbGPS11.6K14	MHR23A [*Mus musculus*]	emb|CAA63145.1|	49, 1e-16
HbGPS11.10G04	Unnamed protein product [*Mus musculus*]	dbj|BAC33939|	52, 6e-06
HbGPS11.6I03	Synaptobrevin like 1 [*Mus musculus*]	ref|NP_035645.1|	66, 7e-10
HbGPS11.6O05	Hypothetical protein AN4711.2 [*Aspergillus nidulans*]	gb|EAA60753.1|	53, 3e-10
HbGPS11.6P23	Unnamed protein product [*Tetraodon nigroviridis*]	emb|CAG01698.1|	75, 1e-12
HbGPS11.10B22	Unnamed protein product [*Tetraodon nigroviridis*]	emb|CAG01396|	52, 3e-06
HbGPS11.6H04	Unnamed protein product [*Tetraodon nigroviridis*]	emb|CAF89842.1|	88, 8e-17
HbGPS11.14O12	Unnamed protein product [*Tetraodon nigroviridis*]	emb|CAG01948.1|	72, 1e-11
HbGPS11.14N02	PakC [*Dictyostelium discoideum*]	gb|AAF82310.3|	67, 2e-10
HbGPS11.10A22	Similar to *Dictyostelium discoideum* (Slime mold)	gb|AAS38755.1|	82, 8e-19
HbGPS11.10B20	ENSANGP00000006608 [*Anopheles gambiae*]	ref|XP_317008.1|	70, 1e-11
HbGPS11.10D04	Pupal cuticle protein [*Tenebrio molitor*]	|P80685|	47, 8e-06
HbGPS11.10E24	GLP_64_7313_4665 [*Giardia lamblia* ATCC 50803]	gb|EAA37964.1|	105, 2e-21

hsp-70 proteins have shown that *hsp70F/mot-2/hsp-6* and *hsp70A* when present in extra copies lengthen the life span of *C. elegans* [26]. Previous experiments on transformation of *H. bacteriophora* with *hsp70* gene from *C. elegans* have shown increase in thermotolerance [10]. According to one of the theories of aging [25], dietary restriction (DR) increases life span in a wide variety of animals ranging from yeast to vertebrates. Mutations in *eat* genes result in partial starvation of *C. elegans* by disrupting function of the feeding organ, the pharynx and lengthening of life span by up to 20–50% (*eat-2*, *eat-6*, and other *eat* mutants) [27]. The discovery of these important stress genes in *H. bacteriophora* is useful for genetic improvement of EPNs to increase their survival under broad environmental stress conditions.

Several *H. bacteriophora* genes are found which are similar to genes known to have role in RNA interference (RNAi). Our sequence data have shown the existence of ribonuclease III family members, *rnh-2* and *rnc* genes, which may prove useful candidates in the ongoing RNAi endeavors for functional genomic studies of EPNs. Another RNAi protein found in our dataset is the systemic RNA Interference Defective protein (SID-1), a transmembrane protein expressed in cells sensitive to RNAi in *C. elegans* [28]. The dsRNA-mediated gene interference in *C. elegans* systemically inhibits gene expression throughout the organism and thus supplying dsRNA in the diet is sufficient to silence the corresponding target gene in all tissues and even in the progeny. *sid-1* is one of the few genes (*rde-1*, *rde-4*, *rrf-1*, *sid-1*) required for systemic RNAi as it enables transportation of dsRNA between cells [29].

One of the interesting discoveries of this project is the active transposase genes in *H. bacteriophora*, which include Tc3A transposase and two other ESTs having transposase domains. One earlier report of transposable elements in *H. bacteriophora* has described existence of mariner-like transposable elements that belong to the Tc1/Mariner superfamily [30], which are transpositionally inactive due to accumulation of mutations. The discovery of the active Tc3A, transposable element in this nematode can serve as a useful tool in reverse genetics in EPNs and will help start the era of transposon mediated mutagenesis in EPNs to explore functions of important genes in these interesting parasites.

There are large number of ESTs having homology to transcripts encoding for signaling proteins such as serine/threonine kinases, protein phosphatases, G-protein coupled receptors, and regulators of G-proteins signaling (*rgs*). We predict these genes to have a role in the important life processes of this stage mainly chemotaxis to locate insect hosts, exit from dauer stage (recovery), and molting. Animal- and plant-parasitic nematodes use a variety of host-related and environmental cues for host finding and resumption of development. Important cues may include host-related volatiles, carbon dioxide, a change in pH or temperature and exposure to the hosts' digestive enzymes or to plant root diffusates [31]. Host finding in *H. bacteriophora* is stimulated by host-related volatiles and by carbon dioxide. In addition, heterorhabditid infective juvenile recovery in-vitro is stimulated by carbon dioxide and a food signal secreted by the nematode's symbiotic bacterium. In *C. elegans*, the entry into dauer stage is initiated by pheromones secreted during shortage of food whereas recovery of dauer occurs by a food signal (carbohydrate like material) produced by *E. coli* [23]. The discovery of more signaling transcripts from *H. bacteriophora* infective juvenile stage can add to the existing knowledge base of dauer-related genes in *C. elegans* and further help explore the importance of this life stage in increased longevity and stress resistance in nematodes.

Interestingly, 42 of the *H. bacteriophora* ESTs did not match to any of the *C. elegans* proteins (Table 2) but matched those of other organisms including bacteria and eukaryotes like mammals, insects, fungi, and plants. About 8 of the 42 ESTs were similar to bacterial genes and hence could be derived from the symbiotic bacteria of *H. bacteriophora*, *P. luminescens*. These 8 ESTs included insecticidal proteins from *P. luminescens* and some hypothetical proteins. The remaining 34 ESTs were similar to insects, plants and animal proteins. These may be novel nematode symbiosis or virulence genes or contaminants from insect cadaver. The origin of these genes can be easily determined by PCR amplification of the sequences from host (*G. mellonella*) or the bacterial symbiont DNA. The absence of amplification of these ESTs from both the organisms will confirm these to be novel nematode genes. These ESTs are of particular interest as they might have a role in EPN life cycle that is more diverse than that of *C. elegans*. Indeed, unlike *C. elegans*, the EPN life cycle includes mutualistic association with the bacterium *P. luminescens* (symbiosis) and parasitism of insects.

There were several ESTs that occurred multiple times in the *H. bacteriophora* cDNA library. These ESTs included G-protein coupled receptors and other signaling proteins (25), uncoordinated locomotion proteins (*uncs*) (11), proteases (12), ribosomal proteins (9), ubiquitin conjugating enzymes (6) and kinesin like proteins (4). These likely correspond to highly expressed gene families in current cDNA library of *H. bacteriophora* infective stage. The role played by these genes in the infective stage of *H. bacteriophora* is not predicted given that they may vary from housekeeping to stage specific functions. Additional evidence to support either hypothesis is not available at this time. Other interesting proteins found are those similar to proteins having role in secretion process like *sec-61*, cell-surface proteins like tenascin [32] and other membrane associated proteins, which are anticipated to have a role in nematode-host interaction during the infection process. A complete genome sequencing of *H. bacteriophora* will result in the discovery of more genes and deep insight into the complex nematode-insect-bacterial interactions in this tripartite association. The estimated genome size of *H. bacteriophora* is about one third the size of *C. elegans* genome (39 MB versus 97 MB) and 51% of this genome is composed of repetitive DNA content [33]. Hence, the diverse physiological and functional attributes of this species are intriguing contrary to the size of its genome.

Acknowledgements

We thank the Ohio Agricultural Research and Development Center for funding, Sharon Reed for help in RNA isolation, and Dr. Xiaodong Bai for help with data analysis and sequence submission. We also thank Dr. Ian Holford for his help in bioinformatics scripts, Dr. David Francis for permission to use their workstation and Dr. Sophien Kamoun for the use of the Sequencher software. We thank Dr. Todd Ciche and Dr. Paul Sternberg for reviewing the manuscript and Dr. Erich Schwarz (wormpep) for his help in data access.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.molbiopara.2006.01.002.

References

[1] De Ley P, Blaxter M. Systematic position and phylogeny. In: Lee D, editor. The biology of nematodes. Taylor & Francis; 2002. p. 1–30.

[2] Boemare N. Interactions between the partners of the entomopathogenic bacterium nematode complexes, *Steinernema-Xenorhabdus* and *Heterorhabditis-Photorhabdus*. Nematology 2002;4:601–3.

[3] Grewal PS, Lewis EE, Gaugler R. Response of infective stage parasites (Nematoda: Steinernematidae) to volatile cues from infected hosts. J Chem Ecol 1997;23:503–15.

[4] Ciche TA, Ensign JC. For the insect pathogen *Photorhabdus luminescens*, which end of a nematode is out? Appl Environ Microbiol 2003;69:1890–7.

[5] Grewal PS, Ehlers RU, Shapiro-Ilan D, editors. Nematodes as biocontrol agents. Wallingford, UK: CABI Publishing; 2005.

[6] Duchaud E, Rusniok C, Frangeul L, et al. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. Nat Biotech 2004;21:1307–13.

[7] The *C. elegans* Sequencing Consortium. Genome Sequence of the nematode *C. elegans*: A platform for investigating biology. Science 1998; 282: 2012–2018.

[8] Stein LD, Bao Z, Blasiar D, et al. The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol 2003;1:E45.

[9] Blaxter ML, De Ley P, Garey JR, et al. A molecular evolutionary framework for the phylum Nematoda. Nature 1998;392:71–5.

[10] Hashmi S, Hashmi G, Glazer I, Gaugler R. Thermal response of Heterorhabditis bacteriophora transformed with the *Caenorhabditis elegans* HSP70 encoding gene. J Exp Zool 1998;281:164–70.

[11] Kaya HK, Stock SP. Techniques in insect nematology. In: Lacey LA, editor. Manual of techniques in insect pathology. London: Academic Press; 1997. p. 281–324.

[12] White GF. A method for obtaining infective nematode larvae from cultures. Science 1927;66:302–3.

[13] Bai X, Fazzolari T, Hogenhout SA. Identification and characterization of traE genes of *Spiroplasma kunkelii*. Gene 2004;336:81–91.

[14] Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques 2001;30:892–7.

[15] Zhulidov PA, Bogdanova EA, Altshuler IM, et al. Simple full length cDNA normalization using kamchatka crab duplex-specific nucleases. Nucl Acids Res 2004;32(3):e37.

[16] Apweiler R, Bairoch A, Wu CH, et al. UniProt: the Universal Protein knowledgebase. Nucl Acids Res 2004;32:D115–9.

[17] Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol 1990;215:403–10.

[18] Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucl Acids Res 2000;28:27–30.

[19] AbuHatab M, Selvan S, Gaugler R. Role of proteases in penetration of insect gut by the entomopathogenic nematode *Steinernema glaseri* (Nematoda: Steinernematidae). J Invertebr Pathol 1995;66:125–30.

[20] Selvan S, Gaugler R, Lewis EE. Biochemical energy reserves of entomopathogenic nematodes. J Parasitol 1993;79:167–72.

[21] Jones SJ, Riddle DL, Pouzyrev AT, et al. Changes in gene expression associated with developmental arrest and longevity in *Caenorhabditis elegans*. Genome Res 2001;11:1346–52.

[22] Wang J, Kim SK. Global analysis of dauer gene expression in *Caenorhabditis elegans*. Development 2003;130:1621–34.

[23] Riddle DL. The dauer larva. In: Wood WB, editor. The nematode: caenorhabditis elegans. Cold Spring Harbor Laboratory Press; 1988. p. 393–412.

[24] Paradis S, Ruvkun G. *Caenorhabditis elegans* Akt/PKB transduces insulin receptors-like signals from AGE-1 PI3 kinase to the DAF-16 transcription factor. Genes Dev 1998;12:2488–98.

[25] Kenyon C. Environmental factors and gene activities that influence life span. In: Riddle DL, Blumenthal T, Meyer BJ, Priess JR, editors. *C. elegans* II. New York: Cold Spring Harbor Press; 1997. p. 791–814.

[26] Yokoyama K, Fukumoto K, Murakami T, et al. Extended longevity of *Caenorhabditis elegans* by knocking in extra copies of *hsp70F*, a homolog of *mot-2* (mortalin/mthsp70/grp75). FEBS Lett 2002;516:53–7.

[27] Houthoofd K, Braeckman BP, Johnson TE, Vanfleteren JR. Life extension via dietary restriction is independent of the Ins/IGF-1 signalling pathway in *Caenorhabditis elegans*. Exp Gerontol 2003;38:947–54.

[28] Winston WM, Molodowitch C, Hunter CP. Systemic RNAi in *C. elegans* requires the putative transmembrane protein SID-1. Science 2002;295:2456–9.

[29] Chen CC, McCullough J, Mello C. Identification of RNAi deficient genes (rde) genes in *C. elegans*. International Worm Meeting, 2003.

[30] Grenier E, Abadon M, Brunet F, Capy P, Abad P. A mariner-like transposable element in the insect parasite nematode *Heterorhabditis bacteriophora*. J Mol Evol 1999;48:328–36.

[31] Griffin CT, Boemare NE, Lewis EE. Biology and behavior. In: Grewal PS, Ehlers R-U, Shapiro-Ilan DI, editors. Nematodes as biocontrol agents. Wallingford, UK: CABI Publishing; 2005. p. 47–64.

[32] Donald EN, Timothy JH, Ji-Hyon K, Jeffrey RG, Kari S. The complete cDNA sequence of human hexabrachion (Tenascin) a multidomain protein containing unique epidermal growth factor repeats. J Biol Chem 1991;266:2818–23.

[33] Grenier E, Catzeflis FM, Abad P. Genome sizes of the entomopathogenic nematodes *Steinernema carpocapsae* and *Heterorhabditis bacteriophora* (Nematoda:Rhabditida). Parasitology 1997;114:497–501.