Article

Genomic Analyses Identify Novel Molecular Signatures Specific for the *Caenorhabditis* and other Nematode Taxa Providing Novel Means for Genetic and Biochemical Studies

Bijendra Khadka 1,†, Tonuka Chatterjee 1,‡, Bhagwati P. Gupta 2 and Radhey S. Gupta 1,*

1 Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, Ontario L9H 6K5, Canada; khadkab@mcmaster.ca (B.K.); tinaburi02@hotmail.com (T.C.)
2 Department of Biology, McMaster University, Hamilton, Ontario L8N 3Z5, Canada; guptab@mcmaster.ca
* Correspondence: gupta@mcmaster.ca
† These authors contributed equally to this work.

Received: 26 July 2019; Accepted: 17 September 2019; Published: 24 September 2019

**Abstract:** The phylum Nematoda encompasses numerous free-living as well as parasitic members, including the widely used animal model *Caenorhabditis elegans*, with significant impact on human health, agriculture, and environment. In view of the importance of nematodes, it is of much interest to identify novel molecular characteristics that are distinctive features of this phylum, or specific taxonomic groups/clades within it, thereby providing innovative means for diagnostics as well as genetic and biochemical studies. Using genome sequences for 52 available nematodes, a robust phylogenetic tree was constructed based on concatenated sequences of 17 conserved proteins. The branching of species in this tree provides important insights into the evolutionary relationships among the studied nematode species. In parallel, detailed comparative analyses on protein sequences from nematodes (*Caenorhabditis*) species reported here have identified 52 novel molecular signatures (or synapomorphies) consisting of conserved signature indels (CSIs) in different proteins, which are uniquely shared by the homologs from either all genome-sequenced *Caenorhabditis* species or a number of higher taxonomic clades of nematodes encompassing this genus. Of these molecular signatures, 39 CSIs in proteins involved in diverse functions are uniquely present in all *Caenorhabditis* species providing reliable means for distinguishing this group of nematodes in molecular terms. The remainder of the CSIs are specific for a number of higher clades of nematodes and offer important insights into the evolutionary relationships among these species. The structural locations of some of the nematodes-specific CSIs were also mapped in the structural models of the corresponding proteins. All of the studied CSIs are localized within the surface-exposed loops of the proteins suggesting that they may potentially be involved in mediating novel protein–protein or protein–ligand interactions, which are specific for these groups of nematodes. The identified CSIs, due to their exclusivity for the indicated groups, provide reliable means for the identification of species within these nematodes groups in molecular terms. Further, due to the predicted roles of these CSIs in cellular functions, they provide important tools for genetic and biochemical studies in *Caenorhabditis* and other nematodes.

**Keywords:** genome sequences; molecular markers (synapomorphies); phylogenetic trees; conserved signature indels; *Caenorhabditis elegans*; Chromadorea; structural analysis of *Caenorhabditis/nematodes-specific* indels; evolutionary relationships among nematodes
1. Introduction

Animals of the phylum Nematoda represent a large and diverse group of eukaryotes present in various marine, freshwater, and terrestrial ecosystems [1–3]. Of the >1 million nematode species that are indicated to exist, approximately 25000 species are currently recognized [3–5]. Most nematodes are transparent and small organisms. However, some can grow to lengths of several meters [3,6]. There are two major types of nematodes in terms of their trophic ecology, i.e., those which are free-living and others which are parasitic [1,5–7]. Free-living nematodes are found in all types of environments and feed on bacteria, algae, or fungi [5,6,8]. Parasitic nematodes occupy and obtain nutrients from various host organisms including animals, plants, and insects [1,6,8–11].

Nematodes play significant roles in diverse environments and several parasitic species cause extensive damage to agricultural crops, harm livestock and are also a threat to humans [1,2,5]. For example, root-knot nematodes from the genus *Meloidogyne* damage soybean, potato, and sugar beet crops, resulting in large losses to the agricultural industry [2]. *Haemonchus contortus* and *Ascaris suum* are animal parasitic species known to infect sheep and pigs, respectively [6]. Moreover, some nematode species belonging to the genera *Trichinella* and *Trichuris* infect humans and can cause severe gastrointestinal problems, which, in some cases, can result in death [12,13]. However, some nematode species, specifically *Caenorhabditis elegans* (*C. elegans*), have found wide-spread usage as important model organisms for studies related to cellular development, aging and other genetic, biochemical, and cell biological studies [14–20]. Among the many advantages of *C. elegans* as a model system, it is a transparent microscopic organism with nervous system and all its neurons have been mapped. Further, it has a short life cycle and it is easy and inexpensive to maintain in a lab, and can be readily manipulated genetically [2,3,21,22]. More importantly, *C. elegans* contains a number of genes that are similar and homologous to the human disease genes, making it an ideal organism to study human diseases in an animal model [2,3,15,17,22–26]. The global concerns of animal and plant-parasitic nematodes, as well as the medical and agricultural applications of other free-living and parasitic nematodes, underscore the need to understand the evolutionary relationships as well as novel characteristics of different groups of nematodes.

Earlier studies on the classification of nematodes were based on morphological characteristics [1,5–7]. However, as most of the studied morphological traits were homoplasious (i.e., shared presence was not due to common ancestry), the resulting classification was misleading [1,6,27]. In recent years, phylogenetic studies employing predominantly 18S and 28S ribosomal RNA (rRNA) genes and mitochondrial DNA have been used to examine the evolutionary relationships among nematodes [1,2,4–6,21,28–32]. However, these studies are often unable to discriminate between species of higher-level nematode taxa [33,34]. Further, as the branching of species in phylogenetic trees is affected by large numbers of variables [35–39], additional more reliable means for distinguishing different main groups of nematodes are needed. Currently, very few reliable molecular characteristics are known that are specific for the genus *Caenorhabditis* or other nematodes, which could be used to confidently discriminate important groups of nematodes in molecular terms.

Genome sequences are currently available for 52 nematode species, providing good coverage of several important groups within the phylum Nematoda [40]. These sequences serve as a valuable resource for a more reliable understanding of the evolutionary relationships amongst the species [41] and for identifying novel molecular characteristics that are uniquely shared within specific groups/clades of nematodes. Additionally, the sequence data offers powerful means for genetic, biochemical studies, and other types of studies including identification of novel drug targets [9,11,30,42–44]. One important class of molecular markers whose discovery has been facilitated by genome sequence analyses is comprised of conserved signature indels (insertions/deletions) (CSIs) in gene/protein sequences that are uniquely shared by an evolutionarily related group of species [37,38,45,46]. The CSIs that are useful for evolutionary studies are generally of specific lengths, present at specific positions in particular genes/proteins, and they are flanked on both sides by conserved regions to ensure that they constitute reliable characteristics [37,47–50]. The CSIs in gene/protein sequences generally result from rare genetic
changes and they have provided important means for demarcation of different groups of organisms in molecular terms [37,38,45,48,51]. Further, based upon their presence or absence in different species, important inferences regarding the evolutionary relationships among a given group of species can be derived [37,38,45].

In the present study, we have used the genome sequences of 52 nematode species to construct a phylogenetic tree for the nematodes based on concatenated sequences of 17 conserved proteins. This tree provides important insights into the evolutionary relationships amongst the nematodes, and a number of major groups/taxa within the phylum Nematoda are reliably resolved. More importantly, our comparative genomic analysis of the protein sequences of Caenorhabditis species has uncovered 52 molecular signatures comprising of CSIs in diverse proteins that are uniquely shared by either all sequenced Caenorhabditis species or by several higher taxa of nematodes encompassing this genus. Of these molecular markers, 39 CSIs in proteins involved in diverse functions are distinctive characteristics of homologs from all six genome-sequenced Caenorhabditis species. The described molecular markers, due to their exclusivity for the specific groups of nematodes, provide useful means for the development of novel diagnostics as well as for genetic and biochemical studies on this important group of organisms.

2. Materials and Methods

2.1. Construction of Phylogenetic Trees

To construct a phylogenetic tree for 52 genome sequenced nematode species, sequences of 17 conserved proteins involved in a variety of cellular functions, which were present in a single copy in these genomes were identified (Table S1). Sequences for four outgroup species viz. Cryptosporidium muris, Plasmodium falciparum, Babesia sp. Xinjiang and Eimeria necatrix, were used for the rooting of the tree. The phylogenetic tree construction was carried out using an internally developed pipeline described in our earlier work [52]. Briefly, the CD-HIT program was used [53] to identify protein families sharing a minimum of 50% in sequence identity and sequence length and which were found in at least 80% of the input genomes. The Clustal Omega [54] algorithm was used to generate multiple sequence alignment (MSA) of these protein families. The aligned protein families were trimmed with TrimAl [55] to remove poorly aligned regions [56] before concatenation to the other proteins. This concatenated sequence alignment consisting of 10764 aligned amino acids positions was used for phylogenetic analysis. An approximate maximum likelihood (ML) tree based on this sequence alignment was initially constructed in FastTree 2 [57] using the Whelan and Goldman model of protein sequence evolution [58]. The resulting tree was then used as input for RAxML [59], where the Le and Gascuel model of protein sequence evolution [60] in RAxML 8 to optimize individual branch lengths and to identify the optimal maximum-likelihood topology. Optimization of the robustness of the tree was completed by conducting SH tests [61] in RAxML 8 [59]. The sequence alignment created by the above program was also used to construct an ML tree based on 100 bootstrap replicates in MEGA6 [62] using Whelan and Goldman +Freq. model [58] and JTT matrix-based model [62].

2.2. Identification of Conserved Signature Indels (CSIs)

To identify potential CSIs specific for different groups within the phylum Nematoda, BLASTp searches were performed on >11800 proteins from Caenorhabditis elegans genome (from accession number NP_001033396.1 to NP_001343573.1) covering approximately 40% of the annotated proteins. Based on these blast searches, proteins for which high scoring homologs (E-values less than 1e−20) were present in multiple nematode species, as well as several non-nematode organisms, were identified and the sequences of these proteins from 15–25 species were retrieved. It was not essential that sequences from any Apicomplexa species (used to root the phylogenetic tree) be present among the sequences for non-nematode species. Multiple sequence alignments for the selected protein sequences were created using CLUSTAL_X 2.1 [63] and these alignments were examined manually to identify insertions or
deletions (indels), which were flanked by at least four to five conserved amino acid residues on both sides within the neighboring 40–50 residues [37,47,64]. Indels which were not flanked by conserved regions were not further considered as they do not provide reliable molecular characteristics. Query sequences encompassing the indel and its flanking 40–50 amino acids were collected for all potential CSIs. Afterward, the query sequences underwent another BLASTp search carried out against the NCBI nr database. The resulting top 250–500 hits for all queries were examined to identify CSIs that are uniquely found in the nematode species as well as to evaluate the group specificities of these CSIs. Signature files for all useful CSIs, which were specifically found in the indicated nematode groups, were created using SIG_CREATE and SIGSTYLE programs described in our earlier work [47] that are available on the GLEANS (Gleans.net) server. The CSIs reported here, unless otherwise indicated, are specific for all members of the indicated groups whose homologs were detected by BLASTp searches.

2.3. Homology Modelling and Analysis of Protein Structures

Homology models of some proteins which contain the CSIs were created for the C. elegans homolog to map the locations of the CSIs in the proteins’ structures. The homology models of the C. elegans Rab44 protein, poly ADP-ribose glycohydrolase protein and tRNA guanine N methyltransferase proteins were created using the solved structures of the following template proteins PDB ID: 2p5s (human), PDB ID: 6hmm (human) [65] and PDB ID: 4jwg (from (Schizosaccharomyces pombe) [66], respectively. Homology modeling was performed using MODELLER v9.15 [67] and the top 500 models were ranked on the basis of their discrete optimized protein energy (DOPE) scores [68]. The stereo-chemical properties of the final models were assessed using three independent servers: ERRAT, PROSA, and Verify3D [69–72]. These applications utilize a dataset of refined structures to evaluate the statistical significance of the models’ conformation, location, environment of each amino acid sequence and overall structural stability. Selected models were then refined using ModRefiner [73]. These resultant models were then used to explore the structural changes associated with the insertion. The superimposition of the validated models with the template structures was carried out using PyMOL (http://www.pymol.org) to examine the structure and location of identified CSIs in the modeled protein structures.

3. Results

3.1. Phylogenetic Analysis of Nematodes Based on Concatenated Sequences of Conserved Proteins

Evolutionary relationships of the nematodes species in the past have been mainly studied based on gene sequences for 18S or 28S rRNA and mitochondrial proteins [2,4,6,34]. Genome sequences are now available for 52 nematodes species covering a number of major groups/taxa within this phylum. These sequences can be used to examine the phylogenetic relationships among nematode species based on concatenated sequences for multiple conserved proteins. Hence, a maximum-likelihood (ML) phylogenetic tree was constructed for the 52 genome-sequenced nematodes species based on concatenated sequences of 17 conserved proteins. The proteins employed in these analyses, listed in Table S1, are present in a single copy in the available nematodes genomes. The resulting bootstrapped tree, which was rooted using homologous sequences from representative Apicomplexa species, is shown in Figure 1. In this tree, members from the two main classes within the phylum Nematoda, i.e., Chromadorea and Enoplea, were clearly separated from each other.
Figure 1. Maximum-likelihood tree for 52 genome-sequenced nematode species. The tree was constructed based on the concatenated alignment of 17 orthologous proteins present in a single copy in these genomes as described in the Methods. Bootstrap scores for each node are indicated at the branch points. The bar indicates 0.2 changes per position. The major nematode groups at different phylogenetic levels are labeled. The tree was rooted using the outgroup species shown.
Within the class *Chromadorea*, the species from the two suborders *Rhabditina* and *Spirulina* were also separated from each other. Additionally, species from a number of nematode genera for which sequences were available from multiple species viz. *Caenorhabditis*, *Ancyclostoma*, *Trichinella*, *Trichuris*, and *Brugia*, also formed monophyletic clades supporting the close relationships of species within these genera. However, in the constructed tree, species from the superfamilies *Rhabditioidea*, *Strongyloidea*, *Trichostrongyloidea*, *Filarioidea*, and *Ascarioidea* exhibited polyphyletic branching within each other. Thus, these families cannot be reliably demarcated on the basis of constructed phylogenetic tree and the interrelationships as well as grouping of species within these families remains unclear at present. The branching patterns, as well as the interrelationships among different nematode species observed in our tree, are similar to that reported recently by Smythe et al. [41] based on phylogenomic analysis using a conservative orthology inference strategy. We have also constructed ML trees based on our protein sequences using MEGA6 program employing two different amino acid substitution models, and the results obtained (Figure S1) are very similar to that seen in Figure 1. However, despite the noted limitations of the tree shown in Figure 1, it provides a good phylogenetic framework for understanding and analyzing the results obtained from comparative genomic analysis, which are discussed below.

3.2. Identification of Conserved Signature Indels Specific for Different Nematode Groups

While the phylogenetic tree shown in Figure 1 allows some inferences to be drawn regarding the evolutionary relationships amongst the nematode species, it is important to confirm these inferences using other independent approaches that are also capable of providing further insights into the evolutionary relationships among nematode species. As noted in the introduction, CSIs in protein sequences that are uniquely shared by a given group of organisms provide an important class of molecular markers that have been proven very useful for evolutionary/taxonomic studies [45,46,48–51]. Due to the rare and discrete nature of the genetic changes that give rise to CSIs, the presence or absence of CSIs in different lineages (or proteins) is generally not affected by the factors that can confound or limit the reliability of inferences from phylogenetic trees [41,45,47,48,50]. Hence, the CSIs provide powerful means for demarcating different groups of organisms in molecular terms and for understanding evolutionary relationships. Therefore, a major focus of the present study was to perform comprehensive genomic analysis of protein sequences from *Caenorhabditis* species to identify CSIs that are specific for this genus as well as other higher taxonomic groups/taxa of nematodes encompassing these organisms. The results of our analysis, reported below, have led to the identification of 52 novel molecular signatures in the form of CSIs that are uniquely shared by either all *Caenorhabditis* species or different nematode groups belonging to this phylum. A brief description of the characteristics of the identified CSIs is provided below.

Of the identified CSIs, 39 CSIs within proteins involved in diverse cellular functions are specifically found in the protein homologs of *Caenorhabditis* species, which form a strongly supported monophyletic clade in our phylogenetic tree. Two examples of such CSIs, one consisting of a 1 amino acid (aa) insertion in Rab44 protein (*C. elegans* gene number 4R79.2) and another comprising a 5 aa insertion in a poly ADP-ribose glycohydrolase protein (PARG-1) are shown in Figure 2A,B, respectively. As seen from Figure 2, both these CSIs are present in conserved regions of the proteins and they are commonly shared by the homologs of all six *Caenorhabditis* species with available genome sequences, but not found in the homologs from other nematodes or non-nematode organisms. Of the two proteins harboring these CSIs, Rab44 is a GTPase of Rab family (Ras superfamily). Although 4R79.2 is yet to be genetically characterized (www.wormbase.org), members of the Rab family act as molecular switches in vesicle trafficking and are known to interact with several other molecules at different trafficking stage [74–76]. The protein PARG-1 is a member of poly ADP-ribose glycohydrolase (PARG) family. PARG is a primary enzyme responsible for hydrolyzing the poly(ADP-ribose) polymer synthesized by poly-(ADP-ribose) polymerases and is involved in a variety of nuclear processes such as DNA damage response, development, programmed cell death and aging [65,77].
Figure 2. Partial sequence alignments of the proteins (A) Rab44 and (B) poly ADP-ribose glycohydrolase showing two CSIs (boxed) that are specific for the genus Caenorhabditis. Dashes (-) in these as well as all other alignments denote identity with the amino acid shown in the top sequence. Sequence information for only limited numbers of species is presented in this figure. More detailed alignments for these CSIs are shown in Figure S2. Sequence information for 37 additional CSIs, which are also specific for the genus Caenorhabditis is provided in Figures S3–S39 and a summary of these CSIs is provided in Table 1.
Table 1. Characteristics of the CSIs specific for the Genus Caenorhabditis.

| Protein Name | C. elegans Gene Name | Accession No. | Figure No. | Indel Size | Indel Position |
|--------------|----------------------|---------------|------------|------------|----------------|
| Rab44        | 4R79.2               | AF233163      | Figure 2A, Figure S2A | 1 aa ins  | 233–263 |
| Poly ADP-ribose Glycohydrolase | par-1               | NP_001253532 | Figure 2B, Figure S2B | 5 aa ins  | 411–454 |
| Poly (ADP-ribose) polymerase 2 | parp-2               | NP_001022057 | Figure S3 | 2 aa del  | 389–420 |
| DnaJ-domain containing chaperone protein | djaq-16             | OZ80352       | Figure S4 | 1 aa del  | 186–207 |
| Cyclin-dependent kinase 12 | cdk-12               | NP_001254914 | Figure S5 | 1 aa del  | 456–487 |
| CRAL-TRIO domain-containing Sec14 protein | T2365.2             | NP_001040875 | Figure S6 | 2 aa ins  | 448–474 |
| Mammalian ZAK kinase homolog | zak-1               | NP_001254942 | Figure S7 | 1 aa del  | 80–109 |
| Probable 3.5'–cyclic phosphodiesterase | pde-2               | NP_001022706 | Figure S8 | 2 aa ins  | 448–495 |
| Nuclear Hormone Receptor | nhr-68              | NP_001256335 | Figure S9 | 1 aa del  | 1–35 |
| SMA2-like | sma-1               | NP_001256383 | Figure S10 | 2 aa ins  | 1353–1393 |
| Glutathione transferase omega-1 | parg-1             | NP_001022706 | Figure S11 | 1 aa del  | 65–103 |
| Probable 26S proteosme regulatory subunit | rpn-6.2             | NP_001022706 | Figure S12 | 1 aa del  | 92–130 |
| Serine/Threonine protein phosphatase 2A Regulatory Subunit | parp-2             | NP_001022706 | Figure S13 | 1 aa del  | 46–90 |
| Failed axon connections-like protein | fsc-12.9           | NP_001293265 | Figure S14 | 1 aa del  | 176–211 |
| NADH dehydrogenase [ubiquinone] 1 alpha subcomplex assembly factor 2 | parp-2             | NP_001022706 | Figure S15 | 1 aa del  | 62–97 |
| Disorganized muscle protein 1 | Cbn-dim-1           | EGT45899      | Figure S16 | 1 aa del  | 135–170 |
| ETS (E2F transformation-specific) class transcription factor | ets-9              | NP_001022482 | Figure S17 | 1 aa del  | 54–78 |
| Glycine-rich domain-containing protein | F32B5.7            | EGT8541       | Figure S18 | 1 aa del  | 430–466 |
| Heat shock protein 70 | F11F1.1             | NP_001255199 | Figure S19 | 2 aa del  | 364–399 |
| Heat shock protein 70 | F11F1.1             | NP_001255199 | Figure S20 | 1 aa del  | 437–481 |
| Abnormal cell migration protein 13 | mig-13             | NP_001024661 | Figure S21 | 1 aa del  | 123–151 |
| Regulatory-associated protein of mTOR-like protein | daf-15             | NP_001255199 | Figure S22 | 1 aa del  | 143–175 |
| Abnormal cell migration protein 13 | mig-13             | NP_001024661 | Figure S23 | 1 aa del  | 141–170 |
| Abnormal cell migration protein 13 | mig-13             | NP_001024661 | Figure S24 | 1 aa del  | 220–251 |
| Flexin | plc-1               | NP_001024661 | Figure S25 | 1 aa del  | 460–497 |
| Piwi-like protein | erp-1              | NP_001024661 | Figure S26 | 1 aa del  | 1020–1070 |
| Stomatin | sto-1               | NP_001024661 | Figure S27 | 1 aa del  | 70–99 |
| Ral guanine nucleotide dissociation stimulator | rgs-1            | NP_001024661 | Figure S28 | 1 aa del  | 257–290 |
| Transglutaminase/ protease homolog | ltd-1              | NP_001024661 | Figure S29 | 1 aa del  | 261–290 |
| Vacuum protein sorting-associated protein 41 homolog | vps-41             | NP_001033544 | Figure S30 | 1 aa del  | 209–242 |
| Serine/arginine-rich splicing factor | rsp-1             | NP_001024661 | Figure S31 | 1 aa del  | 13–36 |
| Serine/Threonine-protein phosphatase PP1 | Cni-W03D8.2       | PIC40784      | Figure S32 | 1 aa del  | 159–191 |
| NPEPrilysin metallopeptidase * | npe-20            | NP_001024661 | Figure S33 | 1 aa del  | 761–804 |
| DNA PRlase homolog | ptr-2             | NP_001024661 | Figure S34 | 1 aa del  | 224–262 |
| Probable maleylacetate isomerase | Y103E8A.21         | NP_001253272 | Figure S35 | 3 aa del  | 56–91 |
| Glutathione S-transferase * | C25H3.7           | NP_001254102 | Figure S36 | 1 aa del  | 39–61 |
| CTD nuclear envelope phosphatase 1 homolog | cne-1             | NP_001254102 | Figure S37 | 1 aa del  | 32–92 |
| Kelch-domain protein | F53E4.1           | NP_001024661 | Figure S38 | 6 aa del  | 206–248 |
| Intermediate filament protein | ifc-2             | NP_001024661 | Figure S39 | 2 aa del  | 946–983 |

* Two isoforms of this protein are present in Caenorhabditis species.

In addition to the two CSIs shown in Figure 2, our study has identified 37 other CSIs in different proteins, which are also specifically found in the homologs from Caenorhabditis species. Sequence information for these other CSIs is provided in Figures S3–S39 and some of their characteristics are summarized in Table 1. For some proteins containing these CSIs (viz. an intermediate filament protein, Figure S39), two homologs are present in Caenorhabditis species and the described CSI was found in only one of the two homologs. In such cases, it is likely that the two sets of homologs originated from a gene duplication event in a common ancestor of Caenorhabditis and the genetic change leading to the observed CSI occurred at this stage in the ancestor of one of the homologs. Due to the exclusive presence of different CSIs listed in Table 1 in the protein homologs for Caenorhabditis species, the described CSIs provide reliable molecular markers for distinguishing the members of this genus. The genetic changes responsible for these CSIs are postulated to have occurred in a common ancestor of the genus Caenorhabditis during its divergence from other nematodes.

Our work has also identified 4 CSIs, which, in addition to the Caenorhabditis species, are also commonly shared by the species Diploscapter pachys. The species from both these genera are part of the family Rhabditoidae [3]. One such CSI is a 2 aa insertion in a protein annotated as abnormal cell
migration protein 13 (MIG-13), which is specifically found in the homologs from the family *Rhabditoida* and it is not present in the homologous proteins from other nematodes or other species (see Figure 3).

Although the exact function of the abnormal cell migration protein MIG-13 has not been elucidated, cell migration and morphogenesis are key events in tissue development and organogenesis [18,20]. MIG-13 is an evolutionarily conserved transmembrane protein that has been shown to play an important role in cell migration in the Q neuroblast lineage [78]. MIG-13 acts cell-autonomously to regulate the asymmetric distribution of the actin cytoskeleton in the leading edge of QR descendants [79,80]. Thus, the presence of a CSI in this protein, which is specific for the family *Rhabditoida* is of much interest.

Sequence information for the other three CSIs, which are also specific for the family *Rhabditoida* is provided in Figures S41–S43 and information for them is summarized in Table 2. These CSIs provide reliable evidence supporting a grouping of *Diploscapter pachys* with the *Caenorhabditis* species and they can be used to distinguish members of the family *Rhabditoida* from other nematodes in molecular terms.

**Figure 3.** Partial sequence alignment of a conserved region from a protein annotated as abnormal cell migration protein 13 (MIG-13) containing a 2 aa insertion (boxed) which is specific for the family *Rhabditoida*. This insertion is not present in the homologous proteins from other nematodes as well as other eukaryotic species. Sequence information for three additional CSIs, which are also specific for the family *Rhabditoida* is provided in Figures S41–S43 and a summary of these CSIs is provided in Table 2. Other details are the same as in the legend to Figure 2.
Table 2. Characteristics of the CSIs specific for the nematode suborder Rhabditoidea and class Chromadorea.

| Protein Name                                      | C. elegans Gene Name | Accession No. | Figure (Fig. Sup) No. | Indel Size | Indel Position | Specificity          |
|---------------------------------------------------|----------------------|---------------|-----------------------|------------|----------------|---------------------|
| Cleavage Factor I$_m$ homolog                     | cfim-2               | NP_001255355  | Figure S40            | 2 aa ins   | 87–130         | Rhabditoidea        |
| Methyl-CpG-binding protein                        | mbd-2                | NP_001021012  | Figure S41            | 2 aa ins   | 158–200        |                     |
| Abnormal cell migration protein 13                | mig-13               | NP_001024660  | Figure 3              | 2 aa ins   | 71–105         |                     |
| PAX3- and PAX7 binding protein 1                   | F43G9.12             | NP_001250840  | Figure S42            | 1 aa del   | 126–164        | Chromadorea         |
| tRNA (guanine-N(1))-methyltransferase             | F46F11.10            | NP_491647     | Figure S42            | 2 aa ins   | 632–669        |                     |
| Palmitoyltransferase *                            | spe-10               | KHJ83757      | Figure S44            | 1 aa del   | 234–270        |                     |
| Palmitoyltransferase                              | spe-10               | KHJ83757      | Figure S45            | 2 aa del   | 255–282        |                     |
| ETS (E26 transformation-specific) class transcription factor | ets-5               | KJH47557      | Figure S47            | 1 aa ins   | 122–155        |                     |
| Heterogeneous nuclear ribonucleoprotein A1 *      | H28G03.1             | KJH46562      | Figure S48            | 1 aa ins   | 93–122         |                     |
| Heterogeneous nuclear ribonucleoprotein A1 *      | H28G03.1             | XP_013302959  | Figure S49            | 5 aa del   | 139–171        |                     |
| Regulator of G-protein signaling 7 *              | Cbn-rgs-7            | EGT30339      | Figure S50            | 1 aa ins   | 221–252        |                     |
| Na(+)/H(+) Exchange Regulatory Factor *           | nrfl-1               | NP_001294068  | Figure 5              | 1 aa ins   | 210–245        | Nematoda            |

* Two isoforms of this protein are present in *Rhabditida* species. * These CSIs are not found in *Strongyloides ratti*, which branches deeply in comparison to the other *Chromadorea* species.
The genus *Caenorhabditis* is embedded within the class *Chromadorea*, which constitutes one of the two main classes within the phylum Nematoda [3,81]. Our analysis has identified eight CSIs in different proteins that are uniquely shared by the homologs from different *Chromadorea* species but absent in nematodes belonging to the class *Enoplea* as well as other organisms. Of these eight CSIs, six CSIs are commonly shared in most cases by all genome-sequenced *Chromadorea* species, whereas in two of them, the described CSIs lack in the species *Strongyloides ratti* (belonging to the suborder *Tylenchina*) [30], which in our phylogenetic tree branches are between the classes *Chromadorea* and *Enoplea*. One example of a CSI that is specific for the class *Chromadorea* is presented in Figure 4.

**Figure 4.** Excerpts from the sequence alignment of a conserved region of the protein tRNA (guanine-N(1))-methyltransferase protein containing a 4 aa CSI (boxed) which is specifically found in the homologs from the class *Chromadorea*. Sequence information for seven additional CSIs, which are also specific for the class *Chromadorea* is provided in Figures S44–S50 and a summary of these CSIs is provided in Table 2.
In the CSI shown in Figure 4, which is specific for the class Chromadorea, a four aa insertion is present in a conserved region of tRNA (guanine-N(1))-methyltransferase, which is encoded by F46F11.10 gene in C. elegans. The human homolog of this protein plays an essential role in the methylation of specific guanine residues in tRNA molecules [82]. This CSI is uniquely shared by different Chromadorea species, but it is absent in other nematodes as well as different other organisms. Sequence information for the other seven CSIs, which are also specific for the class Chromadorea is presented in Figures S44–S50 and information for them is summarized in Table 2.

Lastly, our analysis has also identified one CSI in a Na(+)/H(+) exchange regulatory factor protein NRFL-1 that appears to be specific for the phylum Nematoda. The nrfl-1 gene is expressed in many cells and tissues including excretory cell, intestine, pharynx, and tail [83]. NRFL-1 binds to an amino acid transporter AAT-6 to help retain localization of AAT-6 on the intestinal luminal membrane in older worms. Partial sequence alignment of NRFL-1 from nematodes species as well as representative outgroups species are shown in Figure 5. Most nematodes species contain two homologs of nrfl-1. Of these two homologs, one contains a single aa insertion within a conserved region that is specifically found in all nematodes species (Figure 5). This insert is absent in the other protein homolog as well as in the homologous protein from different outgroup species. The absence of this insertion in the outgroup species indicates that this indel is an insert and the genetic change leading to it was introduced in a common ancestor of the phylum Nematoda. More detailed information regarding the species distribution of this CSI is provided in Figure S51.

3.3. Localizations of the CSIs in Protein Structures

Earlier work on CSIs in proteins shows that most of the studied CSIs in proteins are located on the surface exposed loops of proteins [84–86]. The surface-exposed loops in proteins are known to play important roles in mediating novel protein–protein or protein–ligands interaction [84,87,88]. In view of these earlier studies, we have also examined the locations of some of the nematodes-specific CSIs identified in the structures of the nematodes proteins. The mapping of the CSIs in protein structures was carried out for three different proteins. These proteins included Rab-44 (4R79.2) and poly ADP-ribose glycohydrolase (PARG-1), which contain one and five aa insertions, respectively, that are specific for the Caenorhabditis species (Figure 2), and a four aa insertion in the protein tRNA (guanine-N(1))-methyltransferase (F46F11.10) (Figure 4) that is specific for the class Chromadorea. The structural information for these proteins from Caenorhabditis or any other nematode species is presently lacking. However, the structures of their homologs from humans or other eukaryotic organisms exhibiting high sequence similarity to the C. elegans homologs are available [65,66] (see Materials and Methods). Using the available structures of these proteins as templates and by means of the homology modeling technique, the structures of the corresponding C. elegans proteins were constructed and validated as detailed in the Methods section. To visualize the locations of the identified CSIs in the structures of these proteins, structural overlaps of the modeled proteins containing the CSIs and the solved structures of the proteins lacking the CSIs were carried out. The results of these studies for the proteins Rab-44 (4R79.2), poly ADP-ribose glycohydrolase (PARG-1) and tRNA (guanine-N(1))-methyltransferase (F46F11.10) are presented in Figure 6A–C, respectively. The locations of the CSIs in the protein structures are shown in red color in this figure. As seen from the presented structural overlaps, the CSIs in all three studied proteins are localized within the surface-exposed loops of these proteins, which is in accordance with the results of earlier studies [84–87].
Figure 5. Partial sequence alignment from a conserved region of a Na(+)/H(+) exchange regulatory factor protein (NRFL-1) harboring a 1 aa insertion (boxed) which is specific for the phylum Nematoda. Most nematode species contain two homologs of this protein and this CSI is specifically present in one of these two homologs. More detailed information regarding the species distribution of this CSI is provided in Figure S51.
Figure 6. Homology models of the *C. elegans* proteins (A) Rab-44, (B) poly ADP-ribose glycohydrolase, and (C) tRNA (guanine-N(1))-methyltransferase showing the locations of the CSIs in the structures of these proteins. The CSIs are shown in red color in these figures. As seen from the presented structural overlap, the CSIs in all three studied proteins are localized within the surface-exposed loops of these proteins. More details regarding modeling of these structures are provided in the Methods section.

4. Discussion

Nematodes species are clinically, economically, and scientifically important organisms. In addition to their significance for human health and agricultural industry due to their animal and plant pathogenicity, they provide very useful model organisms for scientific research relevant to human [1–3,5,14,22,89]. Thus, it is of much importance to understand their evolutionary relationships and identify reliable molecular means capable of clearly distinguishing different important groups among nematodes. In this study, we have used available genome sequences of 52 diverse nematode species to examine their evolutionary relationships and have performed a comparative analysis on their protein sequences to identify novel molecular markers that are distinctive characteristics of the *Caenorhabditis* species as well as other groups of nematodes.

Phylogenetic trees based on concatenated sequences for multiple proteins are known to provide a more accurate depiction of the evolutionary relationships among a given group of species than trees based on a single gene/protein sequence [30,39,47,90,91]. Hence, a phylogenetic tree for the genome-sequenced nematodes species was constructed in this work based on concatenated sequences.
of 17 conserved proteins. The tree shows a clear separation of the two main classes, i.e., Chromadorea and Enoplea, within the nematodes [5]. Recently, Smythe et al. [41] have also reported phylogenomic analysis of 108 nematodes using a conservative orthology inference strategy. Their analyses also indicated that the class Enoplea formed a sister taxon to the rest of the Nematoda [41]. In the phylogenetic trees constructed in this work as well as by Smythe et al. [41], species from a number of nematode genera viz. Caenorhabditis, Ancylostoma, Trichinella, Trichuris, and Brugia, formed distinct clades supporting their expected close and specific groupings. However, in both these trees, the species from the superfamilies Strongyloidea, Trichostrongyloidea, and Metastrongyloidea were found to cluster closely together and exhibited polyphyletic branching within each other. Thus, the clades corresponding to these superfamilies are reliably discerned presently and their interrelationships are also not resolved.

However, the main focus of the present work was on species from the genus Caenorhabditis, which formed a strongly supported monophyletic clade in the tree. Our comparative genomic analysis was aimed at identifying molecular markers that are commonly and uniquely shared by the members of this genus or other larger clades of nematodes which included Caenorhabditis. These studies have identified for the first time 52 novel molecular markers (or synapomorphies) consisting of conserved signature indels (CSIs) in proteins involved in various biological processes, which are uniquely shared by either all available Caenorhabditis species or other higher taxa of nematodes encompassing this genus, provide novel and important tools for studying these organisms. It should be mentioned that Mitreva and coworkers [43,92,93] have previously carried out extensive work examining the presence of indels in nematode proteins. Although their work has identified large numbers of indels in nematode proteins, unlike the CSIs that are the focus of this work, the indels identified by these authors are not specific for a phylogenetically coherent group (i.e., species related by common ancestry), and in most cases, they were also not present in conserved regions. Extensive earlier work shows that only the indels of fixed lengths, which are flanked on both sides by conserved regions and are uniquely found in a monophyletic group of organisms, provide reliable molecular characteristics that are useful for evolutionary studies and for the demarcation of different groups of organisms in molecular terms [37,46–50,94]. The other indels in protein sequences not meeting these criteria, although they provide valuable tools for genetic and biochemical studies [43,92], their utility for evolutionary studies is limited.

A summary diagram showing the nematode groups’ specificities of different identified CSIs is presented in Figure 7. Of the 52 CSIs identified in this work, 39 CSIs in different proteins are uniquely shared by all members of the genus Caenorhabditis. Four CSIs are specific for the family Rhabditoidae, which, in addition to the genus Caenorhabditis, also includes the genome-sequenced species Diploscapter pachys whereas eight CSIs in unrelated proteins are distinguishing characteristics of the different species from the class Chromadorea. In addition, we have identified one CSI in an Na(+) /H(+) exchange regulatory factor, NRFL-1, that appears to be a common and unique characteristic of different species from the phylum Nematoda. Some molecular features specific for the phylum Nematoda have also been reported by Yin et al. [95]. However, our analysis did not identify any CSI that was specific for the Strongyloidea or Trichostrongyloidea superfamily, which also did not form well-resolved clades in our phylogenetic tree. Thus, the species distribution of the identified CSIs independently supports the different observed groupings of Chromadorea species in the phylogenetic tree. The specificities of the identified CSIs for different members of the indicated clades indicates that the genetic changes responsible for these CSIs initially occurred in the common ancestors of these groups and these genetic changes were then retained/inherited by various descendent species [47].
Figure 7. A conceptual diagram summarizing the species specificities of different nematodes-specific CSIs identified in this work and the evolutionary relationships inferred from them and the constructed phylogenetic tree. The numbers of CSIs that are specific for different clades or species-groupings are noted on the respective nodes.

The identified CSIs, due to their exclusive presence in the indicated groups of nematodes, provide novel and useful means for the identification of both known as well as novel species from these groups in molecular terms and for genetic, biochemical and evolutionary relationships. Extensive earlier work on CSIs for other groups of organisms strongly indicates that these molecular characteristics exhibit a high degree of constancy and predictive ability to be found in other members of the indicated groups [37,45,47,96]. It is expected that of the 39 CSIs identified in the present work which are specific for Caenorhabditis species, a large number of them should also be found in other non-genome sequenced or novel Caenorhabditis species. All of the described CSIs are present within conserved regions of the
genes/proteins. Thus, based on the conserved regions encompassing these CSIs, the presence/absence of these CSIs in other nematodes/Caenorhabditis species could be readily examined by means of different commonly used experimental techniques viz. PCR-based, q-PCR-based, as well as by in silico BLAST searches examining the presence of these CSIs in genomic sequence data. The CSIs-based approaches have been used previously for developing novel and highly specific diagnostic tests for a number of important bacterial pathogens [97,98].

The CSIs identified in this work are present in diverse proteins (see Tables 1 and 2) that are involved in important/essential functions in C. elegans that are likely to be conserved in other nematodes as well. Although the cellular functions of these CSIs are currently not known, earlier work on CSIs in other organisms has shown that these conserved molecular characteristics play important and often essential functions in the organisms where they are found [84,99]. Most of the studied CSIs in protein sequences are located in the surface loops of proteins, which are known to play important roles in mediating novel protein-protein or protein-ligand interactions that are essential or important for the CSI-containing organisms [84,87,99]. In the present work, using homology modeling technique and structural overlaps of the CSIs-containing and CSIs-lacking proteins, we have mapped the locations of the CSIs in three proteins viz. Rab-44 (4R79.2), poly ADP-ribose glycohydrolase (PARG-1) and tRNA (guanine-N(1))-methyltransferase (F46F11.10) that contain CSIs specific for the Caenorhabditis or Chormadorea species (Figures 2 and 4). In all three cases, the CSIs in these proteins in the modeled structures from C. elegans were localized in the surface-exposed regions of the proteins (Figure 6).

As noted in the introduction, C. elegans is an important model organism for studying developmental process, for aging research, and for examining the cellular functions of different genes/proteins in eukaryotic organisms [2,3,22,26]. Further, as many genes in C. elegans (Caenorhabditis species) are homologous to human proteins, it has also used as a model for studying the role of homologous genes/proteins involved in human diseases [14,15,17,23,24,100]. One important advantage of C. elegans is that, in addition to its ease of growth, transparency, and well-studied developmental pathways, it can also be readily manipulated genetically. Thus, it should be possible to investigate in this system the functional significance of the CSIs that are specific for nematode groups [18,19,23,89,100–103]. Earlier work on CSIs has shown that these genetic characteristics are functionally important and often play essential roles in the organisms for which they are specific [84,94,99]. Additionally, the conserved indels in protein sequences also provide potential drug targets [92,104]. In view of these considerations, further studies on understanding the functional significance of the CSIs which are specific for the Caenorhabditis/nematodes species should be of much interest and these could lead to the discovery of novel functional aspects of these important organisms.

**Supplementary Materials:** The following are available online at http://www.mdpi.com/2073-4425/10/10/739/s1, Table S1: Name and accession numbers of proteins used for phylogenetic analysis, Figure S1: Maximum-likelihood trees for genome-sequenced nematode species based on concatenated sequences of 17 conserved proteins. The trees were constructed in MEGA6 using (A) Whelan and Goldman + Freq. and (B) JTT matrix-based models, Figure S2: Partial sequence alignments of (A) Rab-44 protein containing a 1 aa CSI and (B) poly ADP-ribose glycohydrolase protein containing a 5 aa insertion, which re specific for the genus Caenorhabditis, Figure S3: Partial sequence alignments of poly (ADP-ribose) polymerase 2 protein with a 2 aa deletion which is specific for the Caenorhabditis genus, Figure S4: Partial sequence alignments of DnaJ-domain containing chaperone protein consisting of a 1 aa deletion which is specific for the Caenorhabditis genus, Figure S5: Partial sequence alignments of Cyclin dependent kinase 12 protein consisting of a 1 aa deletion that is specific for the Caenorhabditis genus, Figure S6: Partial sequence alignments of CRAL-TRIO domaincontaining Sec14 protein consisting of a 2 aa CSI which is specific for the Caenorhabditis genus, Figure S7: Partial sequence alignments of mammalian ZAK kinase homolog protein with a 1 aa CSI which is specific for the Caenorhabditis genus, Figure S8: Partial sequence alignments of probable 3’5’-cyclic phosphodiesterase pde-2 protein with a 2 aa CSI which is specific for the Caenorhabditis genus, Figure S9: Partial sequence alignments of nuclear hormone receptor protein with 1 aa deletion which is specific for the Caenorhabditis genus, Figure S10: Partial sequence alignments of SMAIL-like (spore membrane assembly protein 2-like) protein with 2 aa CSI which is specific for the Caenorhabditis genus, Figure S11: Partial sequence alignments of glutathione transferase omega-1 protein with a 1 aa CSI which is specific for the Caenorhabditis genus, Figure S12: Partial sequence alignments of 26S proteasome regulatory protein subunit with a 1 aa CSI which is specific for the Caenorhabditis genus, Figure S13: Partial sequence alignments of serine/threonine protein phosphatase 2A regulatory protein subunit with a 1 aa CSI which is specific for the Caenorhabditis genus, Figure S14: Partial sequence alignments of failed axon connections like protein with a 1 aa CSI which is specific for...
the *Caenorhabditis* genus, Figure S15: Partial sequence alignments of NADH dehydrogenase 1 alpha sub-complex assembly factor 2 protein with a 13 aa CSI which is specific for the *Caenorhabditis* genus, Figure S16: Partial sequence alignments of disorganized muscle protein with a 1 aa deletion that is specific for the *Caenorhabditis* genus, Figure S17: Partial sequence alignments of ETS (E26 transformation-specific) class transcription factor protein with a 1 aa CSI which is specific for the *Caenorhabditis* genus, Figure S18: Partial sequence alignments of glycine-rich domain containing protein with a 1 aa CSI which is specific for the *Caenorhabditis* genus, Figure S19: Partial sequence alignments of a heat shock protein 70 protein consisting of a 2 aa deletion that is specific for the *Caenorhabditis* genus, Figure S20: Partial sequence alignments of abnormal cell migration protein 13 protein with a 1 aa deletion which is specific for the *Caenorhabditis* genus, Figure S21: Partial sequence alignments of abnormal cell migration protein 13 protein consisting of a 1 aa deletion which is specific for the *Caenorhabditis* genus, Figure S22: Partial sequence alignments of regulatory associated protein of mTOR-like protein consisting of a 1 aa CSI that is specific for the *Caenorhabditis* genus, Figure S23: Partial sequence alignments of abnormal cell migration protein 13 protein consisting of a 3 aa deletion which is specific for the *Caenorhabditis* genus, Figure S24: Partial sequence alignments of abnormal cell migration protein 13 protein consisting of a 1 aa deletion which is specific for the *Caenorhabditis* genus, Figure S25: Partial sequence alignments of Plexin protein with a 1 aa CSI that is specific for the *Caenorhabditis* genus, Figure S26: Partial sequence alignments of a Piwi-like protein protein consisting of a 1 aa CSI that is specific for the *Caenorhabditis* genus, Figure S27: Partial sequence alignments of stomatin protein with a CSI consisting of a 1 aa deletion that is specific for the *Caenorhabditis* genus, Figure S28: Partial sequence alignments of Ral guanine nucleotide dissociation stimulator protein consisting of a 1 aa deletion that is specific for the *Caenorhabditis* genus, Figure S29: Partial sequence alignments of transglutaminase/protease homolog protein consisting of a 1 aa deletion that is specific for the *Caenorhabditis* genus, Figure S30: Partial sequence alignments of vacuolar protein sorting associated protein 41 homolog protein consisting of a 1 aa CSI which is specific for the *Caenorhabditis* genus, Figure S31: Partial sequence alignments of serine/arginine-rich splicing factor protein consisting of a 1 aa deletion that is specific for the *Caenorhabditis* genus, Figure S32: Partial sequence alignments of serine/threonineprotease homolog protein protein consisting of a 1 aa CSI that is specific for the *Caenorhabditis* genus, Figure S33: Partial sequence alignments of NEPRilysin metallopeptidase family protein consisting of a 1 aa deletion which is specific for the *Caenorhabditis* genus, Figure S34: Partial sequence alignments of DNA PRImase homolog protein consisting of a 1 aa CSI which is specific for the *Caenorhabditis* genus, Figure S35: Partial sequence alignments of a probable maleylacetoacetate isomerase protein with a 3 aa deletion that is specific for the *Caenorhabditis* genus, Figure S36: Partial sequence alignments of glutathione S-transferase protein consisting of a 1 aa CSI which is specific for the *Caenorhabditis* genus, Figure S37: Partial sequence alignments of CTD nuclear envelope phosphatase 1 homolog protein with a 1 aa CSI which is specific for the *Caenorhabditis* genus, Figure S38: Partial sequence alignments of Kelch-domain protein with a 6 aa CSI that is specific for the *Caenorhabditis* genus, Figure S39: Partial sequence alignments of intermediate filament protein consisting of a 2 aa deletion that is specific for the *Caenorhabditis* genus, Figure S40: Partial sequence alignments of cleavage factor IM (CFIm) homolog protein consisting of a 2 aa CSI which is specific for the *Rhabditoida* suborder, Figure S41: Partial sequence alignments of Methyl-CpG-binding protein consisting of a 2 aa CSI which is specific for the *Rhabditoida* suborder, Figure S42: Partial sequence alignments of abnormal cell migration protein 13 consisting of a 1 aa CSI which is specific for the *Rhabditoida* suborder, Figure S43: Partial sequence alignments of PAX3- and PAX7 binding protein 1 protein with a 1 aa deletion that is specific for the *Rhabditoida* suborder, Figure S44: Partial sequence alignments of palmitoyltransferase protein consisting of a 1 aa deletion that is specific for the *Chromadorea* class, Figure S45: Partial sequence alignments of palmitoyltransferase protein with a 2 aa deletion that is specific for the *Chromadorea* class, Figure S46: Partial sequence alignments of a battenin protein with a 3 aa CSI that is specific for the *Chromadorea* class, Figure S47: Partial sequence alignments of ETS (E26 transformation-specific) class transcription factor protein consisting of a 1 aa CSI that is specific for the *Chromadorea* class, Figure S48: Partial sequence alignments of heterogeneous nuclear ribonucleoprotein A1 protein with a 1 aa CSI that is specific for the *Chromadorea* class, Figure S49: Partial sequence alignments of heterogeneous nuclear ribonucleoprotein A1 protein consisting of a 5 aa deletion which is specific for the *Chromadorea* class, Figure S50: Partial sequence alignments of regulator of Gprotein signaling 7 protein consisting of a 1 aa CSI which is specific for the *Chromadorea* class, Figure S51: Partial sequence alignments of Na(+)/H(+) exchange regulatory factor protein with a 1 aa CSI which is specific for the entire Nematoda phylum.

**Author Contributions:** B.K. and T.C. carried out identification of CSIs, phyllogenetic analysis and homology modelling and localization of the CSIs in protein structures, and preparation of a draft manuscript. R.S.G., planning and supervision of the work, creation of sequence alignments of proteins for identification of the CSIs, obtained funding for the project and writing and finalizing of the manuscript. B.P.G., contributed to the editing of the manuscript and provided information regarding the functional significance of the proteins containing the CSIs.

**Funding:** This work was supported by Research Grant number 249924 from the Natural Science and Engineering Research Council of Canada awarded to Radhey S. Gupta.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

1. Dorris, M.; de Ley, P.; Blaxter, M.L. Molecular analysis of nematode diversity and the evolution of parasitism. *Parasitol. Today* 1999, 15, 188–193. [CrossRef]
2. Holterman, M.; van der Wurff, A.; van den Elsen, S.; van Megen, H.; Bongers, T.; Holovachov, O.; Helder, J. Phylum-wide analysis of SSU rDNA reveals deep phylogenetic relationships among nematodes and accelerated evolution toward crown Clades. *Mol. Biol. Evol.* **2006**, *23*, 1792–1800. [CrossRef] [PubMed]

3. Blaxter, M. *Nematodes*: The worm and its relatives. *PloS Biol.* **2011**, *9*, e1001050. [CrossRef]

4. Liu, G.H.; Shao, R.; Li, J.Y.; Zhou, D.H.; Li, H.; Zhu, X.Q. The complete mitochondrial genomes of three parasitic nematodes of birds: A unique gene order and insights into nematode phylogeny. *BMC Genom.* **2013**, *14*, 414. [CrossRef] [PubMed]

5. Meldal, B.H.; Debenham, N.J.; De Ley, P.; de Goede, R.G.; Georgieva, S.S. Feeding habits in soil *Nematode* families and genera-an outline for soil ecologists. *J. Nematol.* **1993**, *25*, 315–331.

6. Van den Elsen, S.; Holovachov, O.; Karssen, G.; van Megen, H.; Helder, J.; Bongers, T.; Mooyman, P. A phylogenetic tree of nematodes based on about 1200 full-length small subunit ribosomal DNA sequences. *Nematology* **2009**, *11*, 927–950. [CrossRef]

7. Adamson, M.L. Phylogenetic analysis of the higher classification of the *Nematoda*. *Can. J. Zool.* **1987**, *65*, 1478–1482. [CrossRef]

8. Yeates, G.W.; Bongers, T.; de Goede, R.G.; Freckman, D.W.; Georgieva, S.S. Feeding habits in soil *Nematode* families and genera-an outline for soil ecologists. *J. Nematol.* **1993**, *25*, 315–331.

9. Kikuchi, T.; Eves-van den Akker, S.; Jones, J.T. Genome Evolution of Plant-Parasitic Nematodes. *Annu. Rev. Phytopathol.* **2017**, *55*, 333–354. [CrossRef]

10. Blaxter, M.; Koutsovoulos, G. The evolution of parasitism in *Nematoda*. *Parasitology* **2015**, *1*, S26–S39. [CrossRef]

11. Blaxter, M.L. *Nematoda*: Genes, genomes and the evolution of parasitism. *Adv. Parasitol.* **2003**, *54*, 101–195. [PubMed]

12. Pozio, E.; Darwin, M.K. Systematics and epidemiology of *Trichinella*. *Adv. Parasitol.* **2006**, *63*, 367–439. [PubMed]

13. Rombout, Y.B.; Bosch, S.; Van Der Giessen, J.W. Detection and identification of eight *Trichinella* genotypes by reverse line blot hybridization. *J. Clin. Microbiol.* **2001**, *39*, 642–646. [CrossRef] [PubMed]

14. Litke, R.; Boulanger, E.; Fradin, C. *Caenorhabditis elegans* as a model organism for aging: Relevance, limitations and future. *Med. Sci.* **2018**, *34*, 571–579.

15. Ganner, A.; Neumann-Haefelin, E. Genetic kidney diseases: *Caenorhabditis elegans* as model system. *Cell Tissue Res.* **2017**, *369*, 105–118. [CrossRef] [PubMed]

16. Nigon, V.M.; Felix, M.A. History of research on *C. elegans* and other free-living *Nematodes* as model organisms. *WormBook* **2017**, 1–84. [PubMed]

17. Kyriakakis, E.; Markaki, M.; Tavernarakis, N. *Caenorhabditis elegans* as a model for cancer research. *Mol. Cell Oncol.* **2015**, *2*, e975027. [CrossRef] [PubMed]

18. Sommer, R.J.; Bumbarger, D.J. Nematode model systems in evolution and development. *Wiley. Interdiscip. Rev. Dev. Biol.* **2012**, *1*, 389–400. [CrossRef]

19. Richman, C.; Rashid, S.; Prashar, S.; Mishra, R.; Selvaganapathy, P.R.; Gupta, B.P. *C. elegans* MANF Homolog Is Necessary for the Protection of Dopaminergic Neurons and ER Unfolded Protein Response. *Front. Neurosci.* **2018**, *12*, 544. [CrossRef]

20. Ranawade, A.V.; Cumbo, P.; Gupta, B.P. *Caenorhabditis elegans* histone deacetylase hda-1 is required for morphogenesis of the vulva and LIN-12/Notch-mediated specification of uterine cell fates. *G3 Genes Genomes Genet.* **2013**, *3*, 1363–1374.

21. Kiontke, K.; Gavin, N.P.; Raynes, Y.; Roehrig, C.; Piano, F.; Fitch, D.H. *Caenorhabditis* phylogeny predicts convergence of hermaphroditism and extensive intron loss. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 9003–9008. [CrossRef] [PubMed]

22. Nass, R.; Merchant, K.M.; Ryan, T. *Caenorhabditis elegans* in Parkinson’s disease drug discovery: Addressing an unmet medical need. *Mol. Interv.* **2008**, *8*, 284–293. [CrossRef] [PubMed]

23. Elkabti, A.B.; Issi, L.; Rao, R.P. *Caenorhabditis elegans* as a Model Host to Monitor the *Candida* Infection Processes. *J. Fungi* **2018**, *4*, 123. [CrossRef] [PubMed]

24. Martinez, B.A.; Caldwell, K.A.; Caldwell, G.A. *C. elegans* as a model system to accelerate discovery for Parkinson disease. *Curr. Opin. Genet. Dev.* **2017**, *44*, 102–109. [CrossRef] [PubMed]
25. Lublin, A.L.; Link, C.D. Alzheimer’s disease drug discovery: In vivo screening using *Caenorhabditis elegans* as a model for beta-amyloid peptide-induced toxicity. *Drug Discov. Today Technol.* 2013, 10, e115–e119. [CrossRef] [PubMed]

26. Culetto, E.; Sattelle, D.B. A role for *Caenorhabditis elegans* in understanding the function and interactions of human disease genes. *Hum. Mol. Genet.* 2000, 9, 869–877. [CrossRef] [PubMed]

27. Viney, M.; Diaz, A. Phenotypic plasticity in nematodes: Evolutionary and ecological significance. *Worm* 2012, 1, 98–106. [CrossRef]

28. Callejon, R.; Nadler, S.; de Rojas, M.; Zurita, A.; Petrasova, J.; Cutillas, C. Molecular characterization and phylogeny of whipworm nematodes inferred from DNA sequences of cox1 mtDNA and 18S rDNA. *Parasitol. Res.* 2013, 112, 3933–3949. [CrossRef]

29. Aleshin, V.V.; Milyutina, I.A.; Kedrova, O.S.; Vladychenskaya, N.S.; Petrov, N.B. Phylogeny of *Cephalorhyncha* derived from 18S rDNA. *J. Mol. Evol* 1998, 47, 597–605. [CrossRef]

30. Hunt, V.L.; Tsai, I.J.; Coghlan, A.; Reid, A.J.; Holroyd, N.; Foth, B.J.; Tracey, A.; Cotton, J.A.; Stanley, E.J.; Beasley, H.; et al. The genomic basis of parasitism in the *Strongyloides* clade of nematodes. *Nat. Genet.* 2016, 48, 299–307. [CrossRef]

31. Nadler, S.A.; Hudshep, D.S. Phylogeny of the Ascardoidea (Nematoda: Ascaridida) based on three genes and morphology: Hypotheses of structural and sequence evolution. *J. Parasitol.* 2000, 86, 380–393. [CrossRef]

32. Bik, H.M.; Lambshead, P.J.; Thomas, W.K.; Lunt, D.H. Moving towards a complete molecular framework of the Nematoda: A focus on the Enoplida and early-branching clades. *BMC. Evol. Biol.* 2010, 10, 353. [CrossRef]

33. Kiontke, K.C.; Felix, M.A.; Allion, M.; Rockman, M.V.; Braendle, C.; Penigault, J.B.; Fitch, D.H. A phylogeny and molecular barcodes for *Caenorhabditis*, with numerous new species from rotting fruits. *BMC Evol. Biol.* 2011, 11, 339. [CrossRef]

34. Palomares-Rius, J.E.; Cantalapiedra-Navarrete, C.; Rchidona-Yuste, A.; Subbotin, S.A.; Castillo, P. The utility of mtDNA and rDNA for barcoding and phylogeny of plant-parasitic Nematodes from *Longidoridae* (Nematoda, Enoplea). *Sci. Rep.* 2017, 7, 10905. [CrossRef]

35. Smythe, A.B.; Sanderson, M.J.; Nadler, S.A. Nematode small subunit phylogeny correlates with alignment parameters. *Syst Biol.* 2006, 55, 972–992. [CrossRef]

36. Felsenstein, J. *Inferring Phylogenies*; Sinauer Associates, Inc.: Sunderland, MA, USA, 2004.

37. Gupta, R.S. Impact of genomics on the understanding of microbial evolution and classification: The importance of Darwin’s views on classification. *FEMS Microbiol. Rev.* 2016, 40, 520–553. [CrossRef]

38. Baldauf, S.L. Phylogeny for the faint of heart: A tutorial. *Trends Genet.* 2003, 19, 345–351. [CrossRef]

39. Rokas, A.; Williams, B.L.; King, N.; Carroll, S.B. Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* 2003, 425, 798–804. [CrossRef]

40. NCBI. NCBI Completed Microbial Genomes. Available online: [http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html](http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/micr.html) (accessed on 15 April 2019).

41. Smythe, A.B.; Holovachov, O.; Kocot, K.M. Improved phylogenomic sampling of free-living Nematodes enhances resolution of higher-level nematode phylogeny. *BMC Evol. Biol.* 2019, 19, 121. [CrossRef]

42. Ma, L.; Zhao, Y.; Chen, Y.; Cheng, B.; Peng, A.; Huang, K. *Caenorhabditis elegans* as a model system for target identification and drug screening against neurodegenerative diseases. *Eur. J. Pharmacol.* 2018, 819, 169–180. [CrossRef]

43. Wang, Z.; Martin, J.; Abubucker, S.; Yin, Y.; Gasser, R.B.; Mitreva, M. Systematic analysis of insertions and deletions specific to Nematode proteins and their proposed functional and evolutionary relevance. *BMC Evol. Biol.* 2009, 9, 23. [CrossRef]

44. Coghlan, A. *Nematode* genome evolution. *WormBook* 2005, 2005, 1–15. [CrossRef]

45. Gupta, R.S. Molecular signatures that are distinctive characteristics of the vertebrates and chordates and supporting a grouping of vertebrates with the tunicates. *Mol. Phylogenet. Evol.* 2016, 94, 383–391. [CrossRef]

46. Springer, M.S.; Stanhope, M.J.; Madsen, O.; de Jong, W.W. Molecules consolidate the placental mammal tree. *Trends Ecol. Evol.* 2004, 19, 430–438. [CrossRef]

47. Gupta, R.S. Identification of Conserved Indels that are Useful for Classification and Evolutionary Studies. *Methods Microbiol.* 2014, 41, 153–182.

48. Rokas, A.; Holland, P.W. Rare genomic changes as a tool for phylogenetics. *Trends Ecol. Evol.* 2000, 15, 454–459. [CrossRef]
49. Baldauf, S.L.; Palmer, J.D. Animals and fungi are each other's closest relatives: Congruent evidence from multiple proteins. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 11558–11562. [CrossRef]

50. Gupta, R.S. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among *Archaea*, *Eubacteria*, and *Eukaryotes*. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1435–1491.

51. Sharma, R.; Gupta, R.S. Novel molecular synapomorphies demarcate different main groups/subgroups of *Plasmodium* and *Piroplasma* species clarifying their evolutionary relationships. *Genes* **2019**, *10*, 490, in press. [CrossRef]

52. Adeolu, M.; Alnajar, S.; Naushad, S.; Gupta, S. Genome-based phylogeny and taxonomy of the 'Enterobacteriales': Proposal for Enterobacterales ord. nov. divided into the families *Enterobacteriaceae*, *Erwinia* fam. nov., *Pectobacteriaceae* fam. nov., *Yersiniaceae* fam. nov., *Hafniaceae* fam. nov., *Morganellaceae* fam. nov., and *Budviciaceae* fam. nov. *Int. J. Syst. Evol. Microbiol.* **2016**, *66*, 5575–5599.

53. Fu, L.; Niu, B.; Zhu, Z.; Wu, S.; Li, W. CD-HIT: Accelerated for clustering the next-generation sequencing data. *Bioinformatics* **2012**, *28*, 3150–3152. [CrossRef]

54. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T.J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Soding, J.; et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011**, *7*, 539. [CrossRef]

55. Capella-Gutierrez, S.; Silla-Martinez, J.M.; Gabaldon, T. TrimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **2009**, *25*, 1996–2003. [CrossRef]

56. Price, M.N.; Dehal, P.S.; Arkin, A.P. FastTree 2—Approximately maximum-likelihood trees for large alignments. *PLoS ONE* **2010**, *5*, e9490. [CrossRef]

57. Whelan, S.; Goldman, N. A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol. Biol. Evol.* **2001**, *18*, 691–699. [CrossRef]

58. Talavera, G.; Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst. Biol.* **2007**, *56*, 564–577. [CrossRef]

59. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef]

60. Le, S.Q.; Gascuel, O. An improved general amino acid replacement matrix. *Mol. Biol. Evol.* **2008**, *25*, 1307–1320. [CrossRef]

61. Guindon, S.; Dufayard, J.F.; Lefort, V.; Anisimova, M.; Hordijk, W.; Gascuel, O. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **2010**, *59*, 307–321. [CrossRef]

62. Tamura, K.; Stecher, G.; Peterson, D.; Filipski, A.; Kumar, S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* **2013**, *30*, 2725–2729. [CrossRef]

63. Larkin, M.A.; Blackshields, G.; Brown, N.P.; Chenna, R.; McGettigan, P.A.; McWilliam, H.; Valentin, F.; Wallace, I.M.; Baxtow, S.; McGowan, J.; Wilm, A.; Lopez, R.; et al. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947–2948. [CrossRef]

64. Naushad, H.S.; Lee, B.; Gupta, R.S. Conserved signature indels and signature proteins as novel tools for understanding microbial phylogeny and systematics: Identification of molecular signatures that are specific for the phytopathogenic genera *Dickeya*, *Pectobacterium* and *Brenneria*. *Int. J. Syst. Evol. Microbiol.* **2014**, *64*, 366–383. [CrossRef]

65. Waszkowycz, B.; Smith, K.M.; McGonagle, A.E.; Jordan, A.M.; Acton, B.; Fairweather, E.E.; Griffiths, L.A.; Hamilton, N.M.; Hamilton, N.S.; Hitchin, J.R.; et al. Cell-Active Small Molecule Inhibitors of the DNA-Damage Repair Enzyme Poly(ADP-ribose) Glycohydrolase (PARG): Discovery and Optimization of Orally Bioavailable Quinazolinedione Sulfonylides. *J. Med. Chem.* **2014**, *57*, 2947–2948. [CrossRef]

66. Waszkowycz, B.; Smith, K.M.; McGonagle, A.E.; Jordan, A.M.; Acton, B.; Fairweather, E.E.; Griffiths, L.A.; Hamilton, N.M.; Hamilton, N.S.; Hitchin, J.R.; et al. Cell-Active Small Molecule Inhibitors of the DNA-Damage Repair Enzyme Poly(ADP-ribose) Glycohydrolase (PARG): Discovery and Optimization of Orally Bioavailable Quinazolinedione Sulfonylides. *J. Med. Chem.* **2018**, *61*, 10767–10792. [CrossRef]

67. Shao, Z.; Yan, W.; Peng, J.; Zuo, X.; Zou, Y.; Li, F.; Gong, D.; Ma, R.; Wu, J.; Shi, Y.; et al. Crystal structure of tRNA m1G9 methyltransferase Trm10: Insight into the catalytic mechanism and recognition of tRNA substrate. *Nucleic Acids Res.* **2014**, *42*, 509–525. [CrossRef]

68. Eswar, N.; Webb, B.; Marti-Renom, M.A.; Madhusudhan, M.S.; Eramian, D.; Shen, M.Y.; Pieper, U.; Sali, A. Comparative protein structure modelling using Modeller. *Curr. Protoc. Bioinform.* **2007**, *5*, 5–6. [CrossRef]

69. Shen, M.Y.; Sali, A. Statistical potential for assessment and prediction of protein structures. *Protein Sci.* **2006**, *15*, 2507–2524. [CrossRef]
70. Colovos, C.; Yeates, T.O. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Sci.* 1993, 2, 1511–1519. [CrossRef]

71. Lovell, S.C.; Davis, I.W.; Arendall, W.B.; de Bakker, P.I.; Word, J.M.; Prisant, M.G.; Richardson, J.S.; Richardson, D.C. Structure validation by Calpha geometry: Phi, psi and Cbeta deviation. *Proteins* 2003, 50, 437–450. [CrossRef]

72. Luthy, R.; Bowie, J.U.; Eisenberg, D. Assessment of protein models with three-dimensional profiles. *Nature* 1992, 356, 83–85. [CrossRef]

73. Lee, G.R.; Heo, L.; Seok, C. Effective protein model structure refinement by loop modeling and overall relaxation. *Proteins* 2016, 84, 293–301. [CrossRef]

74. Segev, N. Ypt/rab gtpases: Regulators of protein trafficking. *Sci. STKE* 2001, re11. [CrossRef]

75. Li, G.; Marlin, M.C. Rab family of GTPases. *Methods Mol. Biol.* 2015, 1298, 1–15.

76. Lipatova, Z.; Hain, A.U.; Nazarko, V.Y.; Segev, N. Ypt/Adm-1 and Rab GTPases: Principles learned from yeast. *Crit Rev. Biochem Mol. Biol.* 2015, 50, 203–211. [CrossRef]

77. St Laurent, J.F.; Gagnon, S.N.; Dequen, F.; Hardy, I.; Desnoyers, S. Altered DNA damage response in *Caenorhabditis elegans* with impaired poly (ADP-ribose) glycohydrolases genes expression. *DNA Repair* 2007, 6, 329–343. [CrossRef]

78. Wang, X.; Zhou, F.; Lv, S.; Yi, P.; Zhu, Z.; Yang, Y.; Feng, G.; Li, W.; Ou, G. Transmembrane protein MIG-13 links the Wnt signaling and Hox genes to the cell polarity in neuronal migration. *Proc. Natl. Acad. Sci. USA* 2013, 110, 11175–11180. [CrossRef]

79. Masuda, H.; Nakamura, K.; Takata, N.; Itoh, B.; Hirose, T.; Moribe, H.; Mekada, E.; Okada, M. MIG-13 controls anteroposterior cell migration by interacting with UNC-71/ADM-1 and SRC-1 in *Caenorhabditis elegans*. *FEBS Lett.* 2012, 586, 740–746. [CrossRef]

80. Sym, M.; Robinson, N.; Kenyon, C. MIG-13 positions migrating cells along the anteroposterior body axis of *C. elegans*. *Cell* 1999, 98, 25–36. [CrossRef]

81. Kang, S.; Sultana, T.; Eom, K.S.; Park, Y.C.; Soothornpong, N.; Nadler, S.A.; Park, J.K. The mitochondrial genome sequence of Enterobius vermicularis (Nematoda: Oxyurida)—An idiosyncratic gene order and phylogenetic information for Chromadorean Nematodes. *Gene* 2009, 429, 87–97.

82. Brule, H.; Elliott, M.; Redlak, M.; Zehner, Z.E.; Holmes, W.M. Isolation and characterization of the human tRNA-(N1G37) methyltransferase (TRM5) and comparison to the *Escherichia coli* TrmD protein. *Biochemistry* 2004, 43, 9243–9255. [CrossRef]

83. Hagiwara, K.; Nagamori, S.; Umemura, Y.M.; Ohgaki, R.; Tanaka, H.; Murata, D.; Nakagomi, S.; Nomura, K.H.; Kage-Nakadai, E.; Mitani, S.; et al. NRFL-1, the *C. elegans* NHERF orthologue, interacts with amino acid transporter 6 (AAT-6) for age-dependent maintenance of AAT-6 on the membrane. *PLoS ONE* 2012, 7, e43050. [CrossRef]

84. Khadka, B.; Gupta, R.S. Identification of a conserved 8 aa insert in the PIP5K protein in the *Saccharomycetaceae* family of fungi and the molecular dynamics simulations and structural analysis to investigate its potential functional role. *Proteins* 2017, 85, 1454–1467. [CrossRef]

85. Alnajar, S.; Khadka, B.; Gupta, R.S. Ribonucleotide reductases from *Bifidobacteria* contain multiple conserved indels distinguishing them from all other organisms: In silico analysis of the possible role of a 43 aa *Bifidobacteria*-specific insert in the Class III RNR homolog. *Front. Microbiol.* 2017, 8, 1409. [CrossRef]

86. Gupta, R.S.; Nanda, A.; Khadka, B. Novel molecular, structural and evolutionary characteristics of the *Phosphoketolases* from *Bifidobacteria* and *Coriobacteriales*. *PLoS ONE* 2017, 12, e0172716. [CrossRef]

87. Akiva, E.; Itzhaki, Z.; Margalit, H. Built-in loops allow versatility in domain-domain interactions: Lessons from self-interacting domains. *Proc. Natl. Acad. Sci. USA* 2008, 105, 13292–13297. [CrossRef]

88. Hashimoto, K.; Panchenko, A.R. Mechanisms of protein oligomerization, the critical role of insertions and deletions in maintaining different oligomeric states. *Proc. Natl. Acad. Sci. USA* 2010, 107, 20352–20357. [CrossRef]

89. Tejeda-Benitez, L.; Olivero-Verbel, J. *Caenorhabditis elegans*, a Biological Model for Research in Toxicology. *Rev. Environ. Contam Toxicol.* 2016, 237, 1–35.

90. Martin, W.; Rujan, T.; Richly, E.; Hansen, A.; Cornelisen, S.; Lins, T.; Leister, D.; Stoebe, B.; Hasegawa, M.; Penny, D. Evolutionary analysis of *Arabidopsis*, cyanobacterial, and chloroplast genomes reveals plastid phylogeny and thousands of cyanobacterial genes in the nucleus. *Proc. Natl. Acad. Sci. USA* 2002, 99, 12246–12251. [CrossRef]
91. Ciccarelli, F.D.; Doerks, T.; von Mering, C.; Creevey, C.J.; Snel, B.; Bork, P. Toward automatic reconstruction of a highly resolved tree of life. *Science* 2006, 311, 1283–1287. [CrossRef]

92. Wang, Q.; Heizer, E.; Rosa, B.A.; Wildman, S.A.; Janetka, J.W.; Mitreva, M. Characterization of parasite-specific indels and their proposed relevance for selective anthelmintic drug targeting. *Infect. Genet. Evol.* 2016, 39, 201–211. [CrossRef]

93. Mitreva, M.; Smant, G.; Helder, J. Role of horizontal gene transfer in the evolution of plant parasitism among *Nematodes*. *Methods Mol. Biol.* 2009, 532, 517–535.

94. Khadka, B.; Gupta, R.S. Novel Molecular Signatures in the PIP4K/PIP5K Family of Proteins Specific for Different Isozymes and Subfamilies Provide Important Insights into the Evolutionary Divergence of this Protein Family. *Genes* 2019, 10, 312. [CrossRef]

95. Yin, Y.; Martin, J.; Abubucker, S.; Wang, Z.; Wyrwicz, L.; McCarter, J.P.; Wilson, R.K.; Mitreva, M. Molecular determinants archetypical to the phylum *Nematoda*. *BMC Genomics* 2009, 10, 114. [CrossRef]

96. Gupta, R.S.; Lo, B.; Son, J. Phylogenomics and Comparative Genomic Studies Robustly Support Division of the Genus *Mycobacterium* into an Emended Genus *Mycobacterium* and Four Novel Genera. *Front. Microbiol* 2018, 9, 67. [CrossRef]

97. Ahmod, N.Z.; Gupta, R.S.; Shah, H.N. Identification of a *Bacillus anthracis* specific indel in the yeaC gene and development of a rapid pyrosequencing assay for distinguishing *B. anthracis* from the *B. cereus* group. *J. Microbiol. Methods* 2011, 87, 278–285. [CrossRef]

98. Wong, S.Y.; Paschos, A.; Gupta, R.S.; Schellhorn, H.E. Insertion/deletion-based approach for the detection of *Escherichia coli* O157: H7 in freshwater environments. *Environ. Sci. Technol.* 2014, 48, 11462–11470. [CrossRef]

99. Singh, B.; Gupta, R.S. Conserved inserts in the Hsp60 (GroEL) and Hsp70 (DnaK) proteins are essential for cellular growth. *Mol. Genet. Genom.* 2009, 281, 361–373. [CrossRef]

100. Lans, H.; Vermeulen, W. Tissue specific response to DNA damage: *C. elegans* as role model. *DNA Repair* 2015, 32, 141–148. [CrossRef]

101. Wang, Y.A.; Kammenga, J.E.; Harvey, S.C. Genetic variation in neurodegenerative diseases and its accessibility in the model organism *Caenorhabditis elegans*. *Hum. Genom.* 2017, 11, 12. [CrossRef]

102. Sato, K.; Norris, A.; Sato, M.; Grant, B.D. *C. elegans* as a model for membrane traffic. *WormBook* 2014, 1–47. [CrossRef]

103. Ranawade, A.; Mallick, A.; Gupta, B.P. PRY-1/Axin signaling regulates lipid metabolism in *Caenorhabditis elegans*. *PLoS ONE* 2018, 13, e0206540. [CrossRef]

104. Nandan, D.; Lopez, M.; Ban, F.; Huang, M.; Li, Y.; Reiner, N.E.; Cherkasov, A. Indel-based targeting of essential proteins in human pathogens that have close host orthologue(s): Discovery of selective inhibitors for *Leishmania donovani* elongation factor-1α. *Proteins* 2007, 67, 53–64. [CrossRef]