Chapter 5

Evolutionary Expansion of Nematode-Specific Glycine-Rich Secreted Peptides

Muying Ying, Mingyue Ren, Chenglin Liu and Ping Zhao

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/intechopen.68621

Abstract

A genome-wide survey across 10 species from algae *Guillardia theta* to mammals revealed that *Caenorhabditis elegans* and *Caenorhabditis briggsae* acquired a large number of glycine-rich secreted peptides (GRSPs, 110 GRSPs in *C. elegans* and 93 in *C. briggsae*) during evolution in this study. Chromosomal mapping indicated that most GRSPs were clustered on their genomes [103 (93.64%) in *C. elegans* and 82 (88.17%) in *C. briggsae*]. Totally, there are 18 GRSPs cluster units in *C. elegans* and 13 in *C. briggsae*. Except for four *C. elegans* where GRSP clusters lacking matching clusters in *C. briggsae*, all other GRSP clusters had its corresponding orthologous clusters between the two nematodes. Using eight transcriptomic datasets of Affmyetrix microarray, genome-wide association studies identified many co-expressed GRSPs clusters after *C. elegans* infections. Highly homologous coding sequences and conserved exon-intron organizations indicated that GRSP tight clusters might have originated from local DNA duplications. The conserved synteny blocks of GRSP clusters between their genomes, the co-expressed GRSPs clusters after *C. elegans* infections, and a strong purifying selection of protein-coding sequences suggested evolutionary constraint acting on *C. elegans* to ensure that *C. elegans* could rapidly launch and fulfill systematic responses against infections by co-expression, co-regulation, and co-functionality of GRSP clusters.

Keywords: glycine-rich secreted peptide, synteny block, co-expressed gene cluster, nematode infection

1. Introduction

According to the primary structure, glycine-rich proteins can be classified into two classes: (1) consisting of large glycine-rich proteins (GRPs >200 AA) with a length of over 200 amino

acids that typically function as cell wall structural components and (2) composed of small glycine-rich secreted peptides (GRSPs, <200 AA) that have a typical signal peptide followed by a mature peptide with a high glycine content. GRSPs represent a class of unique effectors of multicellular organisms, possessing relatively simple structures but exhibiting complex biological functions. According to previous research, almost all animals, plants, and microorganisms are enriched with GRPs, such as glycine-rich cold-induced proteins from zebrafish [1], glycine-rich keratin and keratin-associated proteins from 22 mammal genomes [2] and RNA-binding proteins with C-terminal glycine-rich domain from *Arabidopsis thaliana* [3]. Plant GRPs have shown diverse functions, including cell wall structure, plant defense, oleosin GRPs in pollen hydration and competition, extracellular ligands of kinase proteins, and RNA-binding GRPs in osmotic stress and cold stress [4]. Growing evidence suggests that these proteins play key roles in the adaptation of organisms to biotic and abiotic stresses including those resulting from pathogenesis, alterations in the osmotic, saline, and oxidative environment, and changes in temperature [3].

To our knowledge, total GRSPs encoded by genomes of different species are significantly distinct. GRSPs are enriched in some species, whereas in other species, no GRSPs have been identified. *Caenorhabditis elegans* and *Caenorhabditis briggsae* are highly enriched for GRSPs in this study. With relatively simple structures but complex biological functions, the importance of GRSPs in nematodes is highlighted by the observations that many members in the GRSP family were indicated to play important roles in *C. elegans* innate immunity. For example, *nlp-29* and *cnc-2* in the GRSP family were upregulated after *Serratia marcescens* infection of *C. elegans* [5]. *Nlp-29* and *nlp-31* in GRSP family were differentially expressed in response to fungal and bacterial infection [6]. Six members in GRSP family from *nlp-27* to *nlp-31* and *grsp-2* were upregulated after *Drechmeria coniospora* infection of *C. elegans* in vivo [7]. Expression of the family member *grsp-21* was upregulated twofold in response to *Microbacterium nematophilum* [8]. Evolutionary diversification of these GRSPs may enhance anti-fungal innate immunity of *C. elegans* [7]. Although these GRSPs are important for *C. elegans* innate immunity, we could not find its corresponding orthologs in human genome. As soil organisms and bacterial feeders, nematodes were constantly challenged by all the different species of soil bacteria, fungi, and other microbes, which have been driving the evolution of nematodes. We were impressed by published works about members of the GRSP family in immune responses of *C. elegans* and interested in knowing whether there were more GRSPs in nematodes and how GRSPs responded to *C. elegans* infections. We believed that free-living soil nematodes very likely to have developed unique components to adapt to the unique environment.

The importance of GRSP family in nematodes is further stressed by the fact that expression of certain GRSPs of *C. elegans* was upregulated by Gram−, Gram+, and fungi of natural infection. Supported by the above facts, we believed in the existence of additional GRSPs and hypothesized that analyzing the genomic sequence would identify novel GRSPs and provide a new global view of GRSP evolution in nematodes. To have a general knowledge of the two nematodes, in the present work, we particularly focused on (1) genome-wide identification and classification of GRSPs which would provide a global view of GRSPs evolution in the two nematodes, (2) mapping these GRSPs on their genomes which would provide a global view of GRSPs distributions on their chromosomes, (3) phylogenetic analyses based on signal

peptides of the two nematode GRSPs, and (4) integrated analysis of public transcriptome datasets about *C. elegans* infections would gain insights into the role of *C. elegans* GRSPs in innate immune.

2. Materials and methods

2.1. Identification of GRSPs in the two nematode genomes

Comprehensive comparison of GRSPs was conducted across 10 species of genomes: *Homo sapiens*, *Danio rerio*, *Drosophila melanogaster*, *C. elegans*, *C. briggsae*, *A. thaliana*, *Monosiga brevicollis*, *Saccharomyces cerevisiae*, *Dictyostelium discoideum,* and *G. theta*. Genome-wide protein sequences of the 10 species were downloaded from the UCSC database (https://genome.ucsc.edu/), and it used to construct two local protein sequence databases. Local-Blastp and PSI-Blast programs from NCBI were carried out to identify *C. elegans* GRSPs with the previously identified GRSPs: *nlp-29*, *nlp-31*, *nlp-33*, *cnc-2*, *cnc-4*, and *cnc-6* as initial queries. GRSPs of *C. briggsae* were identified by using all *C. elegans* GRSPs as initiation queries.

2.2. *C. elegans* GRSPs expression at transcriptional level

Gene expression omnibus (GEO) data sets in NCBI (http://www.ncbi.nlm.nih.gov/) and the reads of RNA sequencing project (PRJNA33023) in DRASearch (https://trace.ddbj.nig.ac.jp/DRASearch/) were used to confirm the transcriptional expression of *C. elegans* GRSPs and avoid false positive arising from genome annotation. This RNA sequencing project is a component of the *C. elegans* modENCODE project including 308 SRA experiments and 196 Biosamples. The total number of genes on each chromosome of *C. elegans* was obtained from UCSC (WS220/ce10) for the estimate of GRSPs density on each chromosome.

2.3. Mapping GRSPs to the genomes of the two nematodes

Characteristic parameters of GRSPs were obtained from WormBase (https://www.wormbase.org/). Configuration files were generated, and mapping of GRSPs to the genomes was performed by Circos [9]. Spacing was based on chromosomal units and the results were further manually modified for easier identification. Orthologous pairs were determined by the twoway reciprocal "best hits" and combining sequence similarity- and synteny-based approaches. Orthologous GRSPs pairs were mapped to their genomes and connected across their chromosomal maps by straight line to identify conserved orthologous synteny blocks of the two nematode genomes.

2.4. Transcriptomic analysis of *C. elegans* GRSPs following infection

Eight transcriptomic data sets related to *C. elegans* infections quantified by Affymetrix microarray (GSE20053, E-MEXP-696, GSE27867, GSE54212, GSE53732, GSE41058, GSE37266, and GSE2740) were downloaded from NCBI GEO database. Differentially expressed GRSPs were extracted to analyze using the GEO2R tool in the GEO database. The range of co-expression

clusters of *C. elegans* GRSPs was defined to be less than 500 kb. Due to the limited data sets of *C. briggsae* genome, we failed to confirm transcriptional expression of *C. briggsae* GRSPs to estimate GRSPs density on its chromosomes and to analyze the co-expressed *C. briggsae* GRSPs after infections.

2.5. Phylogenetic and evolutionary analysis

With the signal peptide sequences of the two nematode GRSPs, a phylogenetic tree was built to detect how the nematode GRSPs families had evolved by gene duplication by using the program Molecular Evolutionary Genetics Analysis package version 6 (MEGA 6) [10]. The bootstrap consensus tree inferred from 500 replicates was taken to represent the evolutionary history to assess the reliability of the phylogenetic tree using the neighbor-joining (NJ) method under p distance [11]. All sites bearing alignment gaps and missing information were retained initially, excluding them as necessary using the pairwise deletion option.

2.6. Analysis of the nucleotide sequences

Using MEGA 6, we estimated transition (Ti)/transversion (Tv) ratios (R) among nucleotides, the number of synonymous (dS) and nonsynonymous (dN) substitutions per site, and the codon-based Z-test for purifying selection. The program was operated under the model of the modified Nei-Gojobori (assumed Ti/Tv bias = 2,2) methods to calculate the difference of dN-dS, and the values were estimated by standard errors (SE) by the bootstrap methods (800 replicates; seed = 17,114) (for details, please refer to supplementary materials and methods in [12]).

3. Results

3.1. Genome-wide analysis of GRSPs across 10 species

The number of GRSPs in each genome of the 10 species was 4 for human, 6 for zebrafish, 53 for fruit fly, 110 for *C. elegans*, 93 for *C. briggsae*, 52 for *A. thaliana*, 0 for *M. brevicollis*, 0 for *S. cerevisiae*, 5 for *D. discoideum*, and 0 for *G.theta.* The two nematodes (110 for *C. elegans* and 93 for *C. briggsae*) are extremely enriched with GRSPs in this study. Analysis of *C. elegans* GRSPs in these species revealed that the number of twoway reciprocal "best hit" orthologs was respectively 0, 2, 8, 90, 3, 0, 0, 2, and 0 (**Table 1**) [12]. Few matching orthologs of *C. elegans* GRSPs in the other species may indicate that GRSPs were less vertically inherited. Besides the two nematodes, *D. melanogaster* and *A. thaliana* are also enriched for GRSPs when compared to the other species analyzed here, which may indicate that an evolutionary expansion of GRSPs happened in nematodes, arthropods, and plants over evolutionary adaption and speciation.

3.2. Identification and classification of the two nematode GRSPs

Based on sequence similarity and the conservation of intron position and phase, 203 GRSPs of the two nematodes were classified into 17 subfamilies (for details, please refer to Figure S1

Species name	Genome size (Mb)	Ref seq protein	Reference Bioproject	GRSPs	Orthologs of *C. elegans*
H. sapiens	3200	55968	PRJNA168	4	0
D. rerio	1371	47861	PRJNA13922	6	2
D. melanogaster	144	30275	PRJNA164	53	8
C. elegans	100	26047	PRJNA158	110	110
C. briggsae	104	17682	PRJNA20855	93	92
A. thaliana	120	35378	PRJNA116	52	3
M. brevicollis	42	9203	PRJNA28133	0	0
S. cerevisiae	12	5907	PRJNA128	0	0
D. discoideum	34	13315	PRJNA13925	5	2
G. theta	0.67	632	PRJNA210	0	0

Table 1. An estimated number of GRSPs in different species and the number of corresponding orthologs in *C. elegans*.

and S2 in [12]). GRSPs mature peptides are enriched for glycine with content ranging from 17 to 74% (For details, please refer to Table S3 in [12]). 62 GRSPs (30.54%) with glycine content from 30 to 40% are the most abundant (**Figure 1**). Among 110 *C. elegans* GRSPs, 36, 11, 14, and 2 have already been designated as "fungus-induced protein related" (FIPR) or "fungus-induced protein" (FIP), "*Caenorhabditis bacteriocin*" (CNC), "neuropeptide-like protein" (NLP), and "DAF-16/FOXO Controlled, germline Tumor affecting" (DCT) in public database. Based on

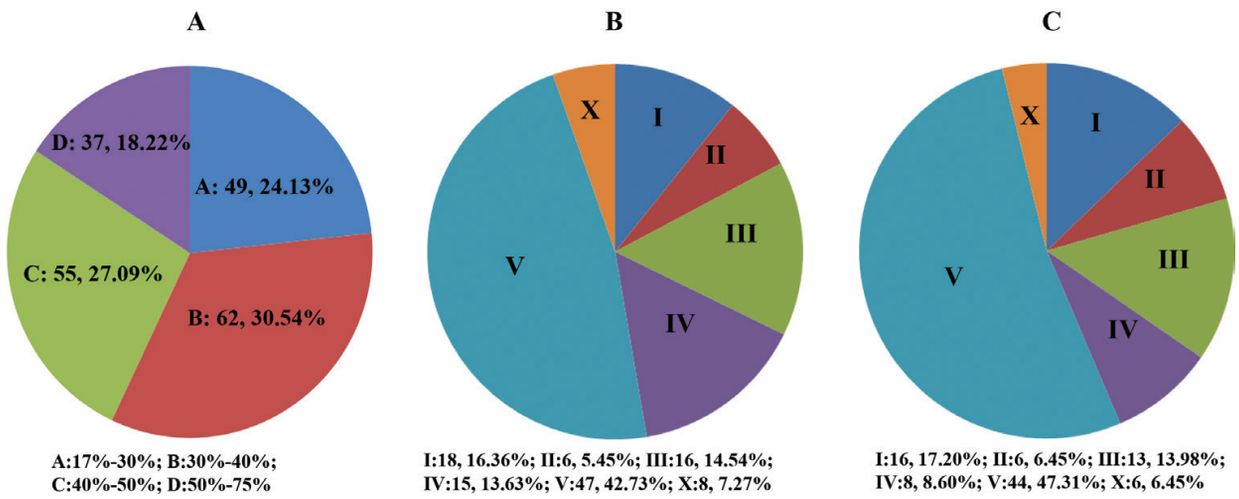

Figure 1. Statistic description of *C. elegans* and *C. briggsae* GRSPs. (A) The number of mature GRSPs peptides with different glycine contents: the number of mature GRSPs peptides with glycine content ranging from 17 to 30% is 49 (24.13%), from 30 to 40% is 62 (30.54%), from 40 to 50% is 55 (27.09%), and from 50 to 75% is 37 (18.23%). (B) The number and percentage of *C. elegans* GRSPs distributed on chromosomes: 18 (16.67%) GRSPs were found on chromosome I, 6 (5.45%) on chromosome II, 16 (14.54%) on chromosome III, 15 (13.63%) on chromosome IV, 47 (42.73%) on chromosome V, and 8 (7.27%) on chromosome X. (C) The number and percentage of *C. briggsae* GRSPs distributed on chromosomes: 16 (17.20%) GRSPs are found on chromosome I, 6 (6.45%) on chromosome II, 13 (13.98%) on chromosome III, 8 (8.60%) on chromosome IV, 44 (47.31%) on chromosome V, and 6 (6.45%) on chromosome X. Comparing S1B to S1C showed that the distribution ratio of GRSPs on its corresponding chromosomes of the two nematodes is similar.

the following shared characteristics: (1) a typical signal peptide located at the N-terminus, (2) a precursor peptide with less than 200 AA, (3) a predicted mature peptide with high glycine contents, and (4) by comparison with the three members (NP_001024238, NP_501117, and NP_504970) already named as GRSPs (*grsp-1*, *grsp-3*, and *grsp-4* in public database), we designated the other 47 unnamed peptides as GRSPs by these criteria. GRSPs identified in *C. briggsae* were referred to as "Cbr," representing the first three letters of the species name *C. briggsae*, plus the name of the corresponding orthologs in *C. elegans* following the previous study [7]. Except for *Cbr-grsp-32*, all the other *C. briggsae* GRSPs have its corresponding orthologs in *C. elegans*. The number of FIPR or FIP, CNC, NLP, and GRSPs family members in *C. briggsae* is, respectively, 31, 9, 12, and 41 (for details, please refer to Table S1 in [12]).

3.3. The evidence of transcriptional expression of *C. elegans* GRSPs

Highly homologous GRSPs are usually clustered together on the two nematode genomes. This is exemplified by GRSPs from *fipr-3* to *fipr-9* clustered on *C. elegans* chromosome V. Their percent identity of protein-coding sequence ranges from 86.1 to 100% (for details, please refer to Figure S4 in [12]). It is notorious that many short genes enriched for repeat sequences are frequently incorrect in genome annotation. To avoid false positive resulting from genome annotation, we further verified the transcriptional expression of all *C. elegans* GRSPs using the available public database. Evidence of transcriptional expression in GEO database showed that 65 *C. elegans* GRSPs were transcriptional expressions (for details, please refer to Table S1 in [12]). For the other 45 GRSPs without transcriptional evidence in GEO database, RNA reads from *C. elegans* transcriptome project were used to confirm their transcriptions, which showed that all GRSPs except for *fipr-12* had 100% matching reads in this project (for details, please refer to Figure S5 in [12]).

3.4. The clustered distribution of GRSPs on the two nematode genomes

GRSPs distribution on their genomes was marked by following qualities (**Figure 2** and **Table 2**): first, most of the GRSPs were clustered on their genomes. The criteria for the definition of GRSPs clusters are (1) the scale between closely adjacent GRSPs should be less than 1 Mb, (2) the number of GRSPs members are equal to or above 3, and (3) the scale of GRSPs clusters is less than 3 Mb. The number of GRSPs clustered on their genomes was 103 for *C. elegans* and 82 for *C. briggsae*. The number of GRSPs clusters is 18 for *C. elegans* and 13 for *C. briggsae*. Second, almost half of the GRSPs in the two nematodes were mapped on their chromosome V (47 in *C. elegans* and 44 in *C. briggsae*). The biggest cluster (from *fip-2* to *nlp-24*) on *C. elegans* chromosome V possesses 15 GRSPs. Of the total 3603 genes on *C. elegans* chromosome V, 47 GRSPs account for 1.30%.

Third, GRSPs clusters were maintained in relative conserved synteny blocks on the chromosomes of the two nematodes (**Figure 2** and **Table 2**). With the exception of four GRSPs clusters without the matching synteny clusters on *C. briggsae* genome, all the other GRSPs clusters possess the matching synteny clusters between the two nematodes. Generally, the lack of the four matching GRSPs synteny clusters in *C. briggsae* could be attributed to the following reasons: (1) no orthologs of *C. elegans* GRSPs were available in *C. briggsae,* (2) the orthologs of *C. elegans* GRSPs in *C. briggsae* were integrated into another unequal GRSPs cluster of *C. briggsae,* and (3) the map position of orthologs of *C. elegans* GRSPs on *C. briggsae* genome was changed. Some of the orthologous

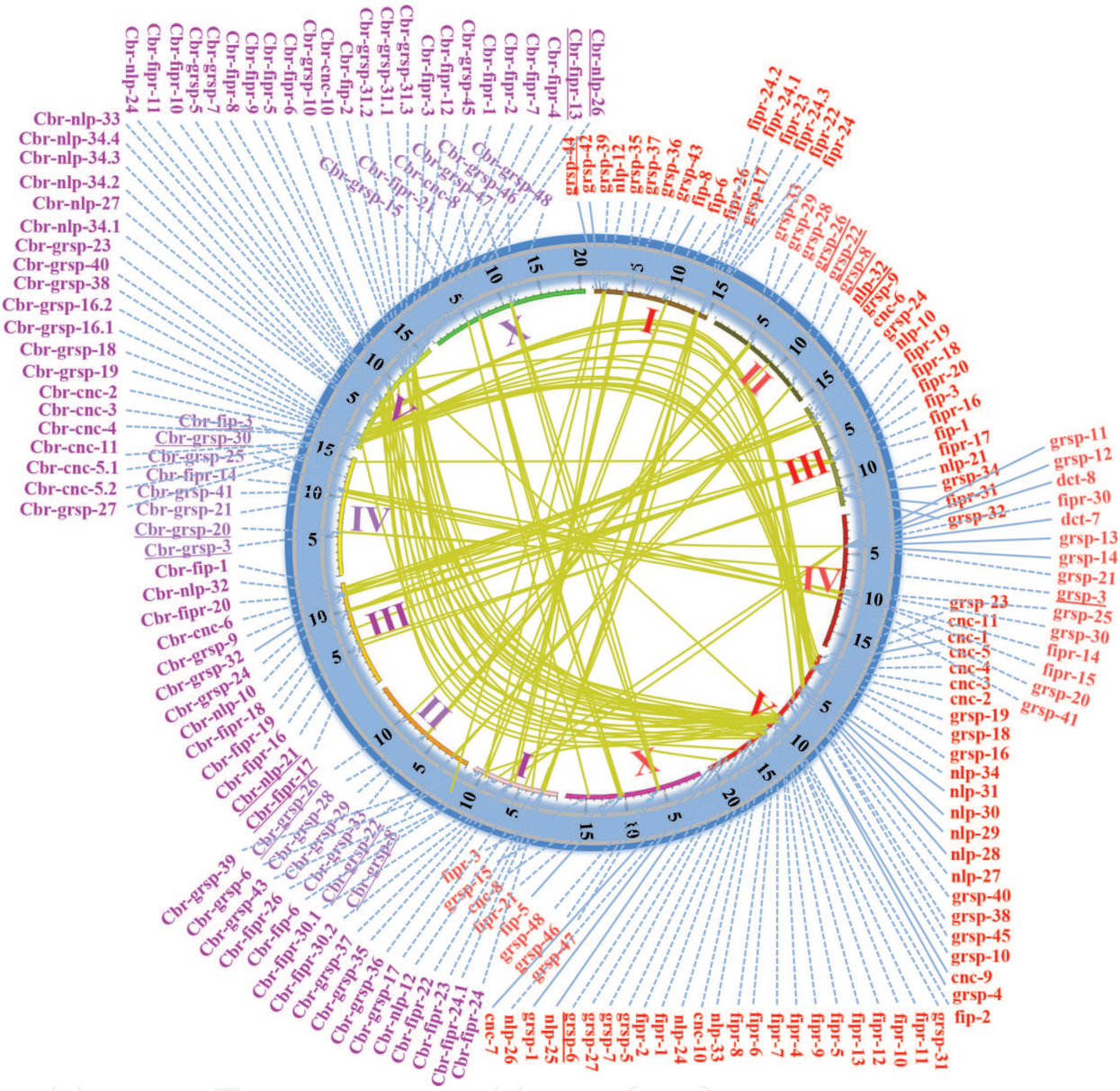

Figure 2. Mapping of GRSPs to genomes of the two nematodes is shown. *C.elegans* and *C. briggsae* GRSPs are indicated by red and purple letters, respectively, which are linked with their chromosomal location by a blue line. Letters from I-X represent chromosome serial numbers of *C. elegans* (red) and *C. briggsae* (purple). GRSPs orthologs between *C.elegans* and *C. briggsae* are linked by yellow beelines. GRSPs lacking orthologs between the two nematodes are linked by a blue solid line with their chromosomal location for easier identification. 7 *C. elegans* GRSPs (*grsp-44* on ChrI, *grsp-26*, *grsp-22,* and *grsp-8* on ChrII, *nlp-32* on ChrIII, *grsp-3* on ChrIV, and *grsp-6* on ChrV) and 11 *C. briggsae* GRSPs (*Cbr-grsp-26*, *Cbr-grsp-22* and *Cbr-grsp-8* on ChrII, *Cbr-fipr-17* and *Cbr-nlp-21* on ChrIII, *Cbr-grsp-3*, *Cbr-grsp-20*, *Cbr-grsp-30*, and *Cbr-fip-3* on ChrIV, *Cbr-fipr-13* and *Cbr-nlp-26* on ChrV) alone scattered on their respective genomes are indicated by an underline.

synteny clusters were observed one-to-two match on their genomes. For example, GRSPs cluster from *Cbr-grsp-27* to *Cbr-grsp-23* on *C. briggsae* chromosome V was matched to two orthologous synteny clusters (from *grsp-23* to *grsp-16* and from *grsp-40* to *grsp-4*) on *C. elegans* chromosome V.

In addition, the order of the orthologous synteny blocks of GRSPs clusters on chromosome V was more conserved than that on other chromosomes of the two nematodes. Orthologous pairs of GRSPs

Chromosome	Size(Mbp)		GRSPs/Gene	GRSPs cluster/ Scale	
	C. elegans	*C. briggsae*		*C. elegans*	*C. briggsae*
I	15.07	15.46	18/3807	From *grsp-42* to *nlp-12*/3GRSPs /904.87kb	From *Cbr-fipr-24* to *Cbr-nlp-12*/5GRSPs /563.98kb
				From *grsp-35* to *grsp-36*/3GRSPs /5.71kb	From *Cbr-grsp-17* to *Cbr-fipr-30.1*/6GRSPs /2.04Mb
				From *grsp43* to *grsp-17*/5GRSPs /2.53Mb	From *Cbr-fip-6* to *Cbr- grsp-39*/5GRSPs / 2.52Mb
				From *fipr-24.2* to *fipr-24*/6GRSPs / 6.6kb	
II	15.28	16.63	6/2648	From *grsp-33* to *grsp-28*/3GRSPs /1.7kb	From *Cbr-grsp-33* to *Cbr-grsp-28*/3GRSPs / 2.7kb
III	13.78	14.58	16/3665	From *grsp-9* to *nlp-10*/4GRSPs /2.17Mb	From *Cbr-fipr-16* to *Cbr-fipr-18*/3GRSPs /2.09Mb
				From *fipr-19* to *fip-1*/6GRSPs /1.18Mb	From *Cbr-nlp-10* to *Cbr-fip-1*/8GRSPs /1.19Mb
				From *fipr-17* to *grsp-32*/5GRSPs /802.9kb	
IV	17.49	17.49	15/3393	From *grsp-11* to *grsp-21*/8GRSPs /401.7kb	From *Cbr- grsp-21* to *Cbr-grsp-25*/4GRSPs /1.39Mb
				From *grsp-25* to *grsp-41*/6GRSPs /2.45Mb	
V	20.92	19.5	47/3603	From *grsp-23* to *grsp-16*/10GRSPs/781.2kb	From *Cbr-grsp-27* to *Cbr-grsp-23*/14GRSPs/ 1.52Mb
				From *nlp-34* to *nlp-27*/6GRSPs/10.3kb	From *Cbr-nlp-34.1* to *Cbr-nlp-34.4*/5GRSPs / 28.97kb
				From *grsp-40* to *grsp-4*/6GRSPs/2.71Mb	
				From *fip-2* to *nlp-24*/15GRSPs /2.54Mb	From *Cbr-nlp-33* to *Cbr-fipr-6*/10GRSPs /2.81Mb
				From *fipr-1* to *grsp-27*/5GRSPs/1.22Mb	From *Cbr-grsp-10* to *Cbr-fipr-4*/13GRSPs /1.78Mb
				From *nlp-25* to *cnc-7*/4GRSPs/786.58kb	
X	17.72	21.54	8/4818	From *grsp-47* to *grsp-48*/3GRSPs/3.16kb	From *Cbr-grsp-15* to *Cbr-cnc-8*/3GRSPs/2.22Mb
				From *fip-5* to *grsp-15*/5GRSPs/1.34Mb	From *Cbr-grsp-47* to *Cbr-grsp-48*/3GRSPs/3.08kb
Total	100.26	105.2	110/21932	18 cluster	13 cluster

Notes: Chr: Chromosome; Size: Chromosome size. GRSPs/Genes: GRSPs number/ total number of genes on each chromosome of *C. elegans*. GRSPs cluster/Scale, GRSPs in each cluster/ scale of GRSP clusters on chromosomes. Matching synteny clusters between the two nematodes were linked by dot lines. The criteria for the definition of GRSP clusters are: 1) distance between closely adjacent GRSPs is less than 1Mb; 2) GRSPs members are greater than or equal to 3; 3) scale of GRSP clusters is less than 3Mb.

Table 2. Summary of GRSPs clusters on the chromosomes of the two nematodes.

between the two nematodes were linked by straight lines on their genome mapping, which showed that the beelines of the orthologous GRSPs clusters on chromosomes V were more likely to be crossovers than those on other chromosomes (**Figure 2**). The crossover means that the order of orthologous synteny blocks of GRSPs clusters was maintained on the genomes of the two nematodes.

3.5. The transcriptional co-expression of *C. elegans* GRSPs clusters after infection

Genome-wide transcriptional analysis showed that many *C. elegans* genes that responded to infection were located in small genomic clusters [8]. All members of the GRSPs cluster from *nlp-27* to *nlp-34* were induced by *D. coniospora* infection of *C. elegans* [7]. Using the transcriptome data sets of *C. elegans* infection based on microarray quantification [7, 8, 13–16], we analyzed the transcriptional expression change of *C. elegans* GRSPs after *C. elegans* infection. The results showed that a total of 108 *C. elegans* GRSPs showed differential expressions at transcriptional levels after *C. elegans* infection in previous studies, which are indicated by blue letters in **Figure 3**. Co-expressed clusters of *C. elegans* GRSPs are shadowed by grey (**Table 3**)

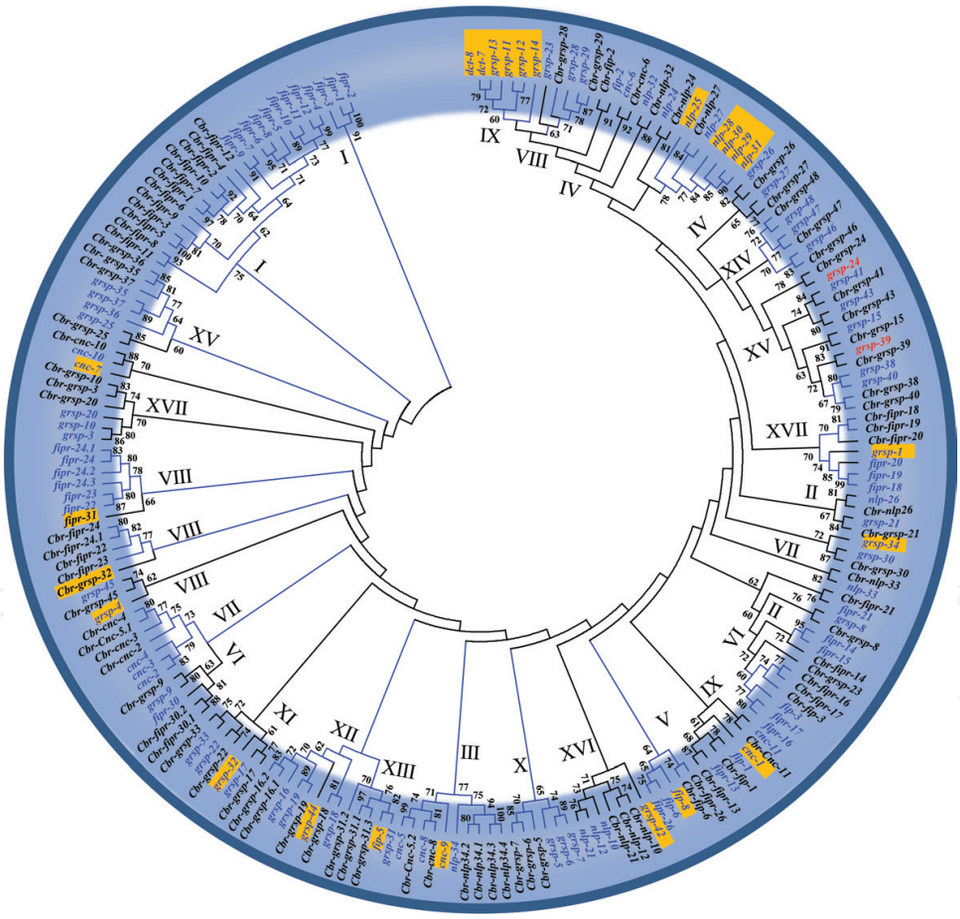

Figure 3. Phylogenetic analysis based on the typical signal peptides of GRSPs in *C. elegans* and *C. briggsae* is shown. The number from I-XVII represents different subfamilies. 24 GRSPs (23 *C. elegans* GRSPs and 1 *C. briggsae* GRSPs) lacking orthologs between the two nematodes are shadowed by orange color for easy identification. 108 of the 110 *C. elegans* GRSPs that had transcriptional expression after infection in previous studies are indicated by blue letters. Two *C. elegans* GRSPs (*grsp-24* and *grsp-39*) without detectable expression data in previous studies analyzed here are indicated by red letters.

Gene Name	Pathogenic	Reference
fip-8 fip-6; cnc-6 fip-3 fip-1; dct-7 grsp-21 (4.64kb) *grsp-3 grsp-25*; *cnc-4 cnc-3 cnc-2* (2.53kb) *nlp-34 nlp-31 nlp-30 nlp-29* (5.54kb) *grsp-38 grsp-10 fip-2 fipr-11 nlp-33 nlp-24 cnc-7; grsp-46 fip-5 fipr-21* (1.79kb) *grsp-15*	*D. coniospora*	Pujol et al., 2008
grsp-4 grsp-42 nlp-12 grsp-35 grsp-37 grsp-36 (5.13kb) *grsp-43 fipr-26 fipr-23 fipr-22 fipr-24* (4.35kb)*; grsp-8 grsp-26; nlp-32 nlp-10 fip-3 fipr-16 fip-1* (50.45kb) *nlp-21 fipr-30 dct-7 grsp-13 grsp-21*(20.91kb) *grsp-3 grsp-41 grsp-25 grsp-23 grsp-45 grsp-40 grsp-38* (1.05kb) *cnc-1 cnc-5 cnc-4 cnc-3 cnc-2* (5.75kb) *grsp-19 grsp-18 grsp-16* (8.41kb) *nlp-31 nlp-30 nlp-29 nlp-28 nlp-27* (9.24kb) *grsp-10 fip-2 grsp-31*(57.64kb) *nlp-33 fipr-2 grsp-7 (38.69kb) grsp-27 nlp-26 cnc-7 grsp-46 grsp-47 fip-5 fipr-21 cnc-8* (3.88kb) *grsp-15*	*M. nematophilum*	O'Rourke et al., 2006
grsp-44 grsp-42 nlp-12 grsp-35 grsp-37 grsp-36 (5.13kb) *grsp-43; grsp-41; nlp-30 grsp-40 grsp-38* (1.05kb) *grsp-45 nlp-33 grsp-5 grsp-7* (1.48kb)	*P. aeruginosa*	Sun et al., 2011
fipr-22 fipr-23 (2.19kb) *cnc-7 cnc-4*	*C.albicans*	Pukkila-Worley et al., 2012
grsp-44 grsp-42 nlp-12 grsp-35 grsp-37 grsp-36 (5.13kb) *grsp-43 fip-8 fip-6 fipr-26* (1.12kb) *grsp-17 fipr-23 fipr-22* (2.19kb); *grsp-33 grsp-29* (7.06kb) *grsp-26 grsp-22 grsp-8* (4.91kb); *nlp-32 grsp-9 cnc-6* (5.69kb) *nlp-10 fipr-18 fip-3 fipr-16 fip-1*(50.45kb) *fipr-17 nlp-21 grsp-32*(80.31kb); *grsp-11 grsp-12 dct-8* (2.18kb) *dct-7 grsp-13 grsp-14 grsp-21* (4.64kb) *grsp-30 fipr-14 fipr-15 grsp-20* (12.39kb) *grsp-41; grsp-23 cnc-11 cnc-1 cnc-5 cnc-4 cnc-3 cnc-2* (7.17kb) *grsp-19 grsp-18 grsp-16* (8,41kb) *nlp-34 nlp-31 nlp-30 nlp-29 nlp-28 nlp-27* (10.52kb) *grsp-40 grsp-38* (1,05kb) *grsp-45 grsp-10 cnc-9 grsp-31 fipr-11 fipr-10 fipr-12* (37,43kb) *fipr-13 fipr-5 fipr-4 fipr-7 fipr-6* (10.23kb) *nlp-33 cnc-10* (35.88kb) *nlp-24 fipr-1* grsp-5 grsp-7 (42.73kb) *grsp-27 grsp-6 nlp-25 nlp-26* (6,02kb) *cnc-7; grsp-47 grsp-46 grsp-48* (3.16kb) *fip-5 fipr-21 cnc-8* (3.88kb) *grsp-15*	*S.enterica*	Head & Aballay, 2014
grsp-42 nlp-12 grsp-35 grsp-37 grsp-36 (5.12kb) *grsp-43 fipr-26 fipr-22 fipr-24* (2.16kb); *grsp-33 grsp-29 grsp-28* (2.06kb) *grsp-26 grsp-8*; *nlp-32 cnc-6 nlp-10 fip-1 fip-3 fipr-16* (48.96kb) *fipr-17*; *fipr-30 dct-7 grsp-13 grsp-21*(213.78kb) *grsp-3 fipr-14* grsp-41; *grsp-23 cnc-1 cnc-11 cnc-5 cnc-4 cnc-3 cnc-2* (5.75kb) *grsp-19 grsp-16* (8.41kb) *nlp-31 nlp-30 nlp-29 nlp-28 nlp-27* (9.24kb) *fip-2 grsp-40 grsp-38* (1.05kb) *grsp-45 grsp-4 grsp-31 fipr-4 fipr-5 fipr-6 fipr-7 fipr-8 fipr-9* (18.61kb) *nlp-33 nlp-24* (469kb) *grsp-5 grsp-7* (1.48kb) *grsp-27 nlp-25 nlp-26* (6.02kb) *cnc-7*; *grsp-47 grsp-46 grsp-48* (3.16kb) *fip-5 fipr-21 cnc-8* (3.88kb)	*S. aureus*	Bond et al., 2014

Notes: Black letters: GRSPs were up-regulated. Blue letters: GRSPs were down-regulated. Co-expressed GRSP clusters were shadowed by grey. GRSPs on different chromosome were separated by a semicolon.

Table 3. Differential expression of GRSPs and co-expression of GRSPs clusters after *C. elegans* infection.

(for details, please refer to Table S4 in [12]). Certainly, it is possible that two *C. elegans* GRSPs (*grsp-24* and *grsp-39*) without detectable expression in previous studies analyzed here may be detectable in other studies, which we were unable to mine due to the limited length of this study [7].

3.6. The evolution of GRSPs multigene families by gene duplications

GRSPs subfamilies were classified based on the precursor sequences similarity and gene structure conservation. Phylogenetic analysis was performed using the signal peptide sequences. It is possible that the similarity between the two group sequences is not perfectly consistent among these GRSPs, which resulted in the observations that certain members within the same subfamilies were located in a different clade in the phylogenetic tree (**Figure 3**). Orthologous GRSPs of the two nematodes detected in the above could be well defined by phylogenetic analysis. Certain members of subfamilies (such as the members of subfamily I) were clustered together on their chromosomes and also the same clade on the phylogenetic tree (**Figure 3**). Five GRSPs from *nlp-27* to *nlp-31* were clustered on *C. elegans* genome. Phylogenetic analysis showed *nlp-27* clade was different from the clade formed by *nlp-28–nlp-31*, which was similar to previous results [7].

Subfamily	dN-dS	SE	Probability	R(Ti/Tv)
I	−5.323	0.073	0.000	1.81
II	−2.228	0.038	0.028	1.32
III	−3.626	0.087	0.011	1.21
IV	−3.321	0.035	0.000	5.54
V	−4.510	0.042	0.011	1.52
VI	−5.326	0.036	0.000	1.26
VII	−3.692	0.028	0.000	3.32
VIII	−2.649	0.053	0.022	1.78
IX	−3.451	0.038	0.000	1.67
X	−2.942	0.046	0.000	2.15
XI	−3.153	0.061	0.031	1.93
XII	−4.324	0.049	0.000	4.34
XIII	−3.256	0.027	0.000	1.52
XIV	−2.968	0.039	0.021	2.86

Notes: dN, non-synonymous substitutions; dS, synonymous substitutions; SE, standard error; Ti, transition; Tv, transversion; R, overall transition/transversion bias. The overall average difference of (dN-dS) was less than zero, and standard error value was less than 0.05.

Table 4. Estimates of overall average variance and pattern of nucleotide substitution.

3.7. Purifying selection of the two nematode GRSPs

Under the model of codon-based Z-test, the estimate of purifying selection was conducted directly to analyze sequence pairs and overall average. Its values are identically equal to zero and therefore rejected the null hypothesis of strict neutrality (dS = dN) and accepted the alternative hypothesis. The difference in average overall of dN-dS was less than zero. The standard error values were less than 0.05. Synonymous substitutions were clearly prevailing on protein-coding sequences of the nematode GRSPs, which indicated the occurrences of purifying selection. With an average ratio of R (Ti/Tv) > 1, the patterns of nucleotide substitution also showed a predominance of transitions over transversions (**Table 4**).

4. Discussion

Soil organisms (*A. thaliana*) and/or bacterial feeders (the two nematodes: *D. discoideum* and fruit fly, who feed on rotting fruit with a large number of bacteria) are relatively enriched for GRSPs in the current study. The environment and survival stress of soil living and/or bacterial feeding may be one of the main evolutionary driving forces for the expansion of lineage-specific GRSPs in the two nematodes. This was exemplified by the expansion of nematode-specific chemosensory genes (for *C. elegans* it is about 2000 and for human it is about 1000, about 2 times), which allowed it to mount a rapid response to environmental stimuli [17]. Comparing to the amplification of nematode-specific chemosensory genes, one may be more impressed by the amplification of nematode-specific GRSPs (for *C. elegans*, it is about 110 and for human, it is 4, about 28 times).

The conservation of precursor organizations, the unaltered position and phase of intron, together with the homologous sequence of DNA, suggested that the GRSPs clusters in the two nematodes might come from physically local DNA reproductions. The duplication of local genes came into being by gene clusters of paralogous genes whose products have similar functions. Paralogous genes with similar functions and expression patterns are frequent in *C. elegans* [18]. The co-expression of gene clusters encoding different proteins with similar functions in specific regions should provide effective combinatorial methods to coordinate complex biological systems [19]. The scales of most co-expression GRSPs clusters on their chromosomes are less than 10 kb and the smallest one is 1.05 kb (co-expression of *grsp-40* and *grsp-38*) (**Table 3**). Different GRSPs within the same cluster differentially responded to the same infection. For example, GRSPs from *cnc-1* to *cnc-5* (7.17 kb) and *cnc-11* in the same cluster showed co-expression with the upregulation of *cnc-11*, *cnc-1,* and *cnc-2* and the down-regulation from *cnc-3* to *cnc-5* after *C. elegans* infection [14]. GRSPs cluster from *grsp-35* to *grsp-36* (5.13 kb) were upregulated by *M. nematophilum* and *P. aeruginosa* infection of *C. elegans* [8, 16] and downregulated by *S. enterica* and *S. aureus* infection [13, 14]. A noticeable overlap of *C. elegans* GRSPs induced by different infections may indicate that the different sets of induced *C. elegans* GRSPs may still share some functionalities. Considering a large amount of operon regulation in *C. elegans*, we analyzed all *C. elegans* genes contained within operon by an internal Perl Scripts search to detect whether the small clusters of adjacent GRSPs could be co-regulated by operon regulation. While no *C. elegans* GRSPs were identified in operon regions (data not shown), the short genetic and physical distance on chromosomes and highly homologous

sequences suggest that neighboring GRSPs arising from duplicated GRSPs may share the same regulatory sequences. The same regulatory sequences on their promoters can be directly and coordinately activated by transcription factors binding to the shared regulatory elements.

With similar variance of (dn-dS), the two nematode GRSPs might have experienced similar selective stress during evolution, which is in concordance with the neutral mutation-random drift theory of molecular evolution. Relative conserved synteny blocks of the GRSPs orthologous clusters suggested that these GRSPs were subjected to functional restraint. With the increasing species complexity, the genome size and the members of a gene family usually undergo an evolutionary expansion in abundance for similar essential basic cellular mechanisms shared by eukaryotes [20]. The basic physiological process for *C. elegans* is similar to those observed in higher organisms. Few matching orthologs of *C. elegans* GRSPs in the other species may indirectly reflect nematode-specific biological functions of *C. elegans* GRSPs that are essential for nematode-specific environments such as soil living and bacterial feeding. The evolutionary diversification of these GRSPs might enhance the ability of *C. elegans* innate immunity to adapt to environmental stress [7].

This study built a full set of GRSPs from the algae *G. theta* to the mammal human by genome-wide comparison across 10 species. The two nematodes were enriched for GRSPs, which demonstrated a good example of DNA local reproductions and maintained a relative conserved synteny block on their genomes after speciation and separation. The phylogenetic conservation of synteny GRSPs clusters on their genomes, the co-expressed GRSPs clusters, and strong purifying selection may indicate evolutionary constraints acting on *C. elegans* to guarantee that *C. elegans* could mount a rapid systematical response to infection by co-expression of GRSPs clusters on the genomes. The mechanism of co-expression, co-regulation, and co-functionality behind these GRSPs clusters is still unknown. Our knowledge about it is expected to improve by the increasing comparative genomics of correlated expression patterns across different nematodes (such as *C. brenneri* and *C. remanei*), which holds promise to provide insights into the adaptive advantage of co-expressed GRSPs in nematodes.

Acknowledgements

This work was supported by grants from the National Nature Science Foundation of China (31160233), the Science and Technology Foundation of Jiangxi Province (20142BAB204013).

Author details

Muying Ying*, Mingyue Ren, Chenglin Liu and Ping Zhao

*Address all correspondence to: yingmuying@ncu.edu.cn

Department of Molecular Biology and Biochemistry, Basic Medical College of Nanchang University, Nanchang, PR China

References

[1] Tang SJ, Sun KH, Sun GH, Lin G, Lin WW, Chuang MJ. Cold-induced ependymin expression in zebrafish and carp brain: implications for cold acclimation. FEBS Letters. 1999;**459**:95-99. DOI: 10.1016/S0014-5793(99)01229-6

[2] Khan I, Maldonado E, Vasconcelos V, O&Brien SJ, Johnson WE, Antunes A. Mammalian keratin associated proteins (KRTAPs) subgenomes: Disentangling hair diversity and adaptation to terrestrial and aquatic environments. BMC Genomics. 2014;**15**:779. DOI: 10.1186/1471-2164-15-779

[3] Ciuzan O, Hancock J, Pamfil D, Wilson I, Ladomery M. The evolutionarily conserved multifunctional glycine-rich RNA binding proteins play key roles in development and stress adaptation. Physiologia Plantarum. 2015;**153**:1-11. doi: 10.1111/ppl.12286

[4] Mangeon A, Jungueira RM, Sachetto-Martins G. Functional diversity of the plant glycine-rich proteins superfamily. Plant Signaling & Behaviour. 2010;**5**:99-104

[5] Mallo GV, Kurz C, Couillault C, Pujol N, Granjeaud S, Kohara Y, Ewbank JJ. Inducible antibacterial defense system in *C. elegans*. Current Biology. 2002;**12**:1209-1214. DOI: 10.1016/S0960-9822(02)00928-4

[6] Couillault C, Pujol N, Reboul J, Sabatier L, Guichou JF, Kohara Y, Ewbank JJ. TLR-independent control of innate immunity in *Caenorhabditis elegans* by the TIR domain adaptor protein TIR-1, an ortholog of human SARM. Nature Immunology. 2004;**5**:488-494. DOI: 10.1038/ni1060

[7] Pujol N, Zugasti O, Wong D, Couillault C, Kurz CL, Schulenburg H, Ewbank JJ. Antifungal innate immunity in *C. elegans* is enhanced by evolutionary diversification of antimicrobial peptides. PLoS Pathogens. 2008;**4**:e1000105. DOI: 10.1371/journal.ppat.1000105

[8] O&Rourke D, Baban D, Demidova M, Mott R, Hodgkin J. Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*. Genome Research. 2006;**16**:1005-1016. DOI: 10.1101/gr.50823006

[9] Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: An information aesthetic for comparative genomics. Genome Research. 2009;**19**:1639-1645. doi: 10.1101/gr.092759.109

[10] Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution. 2013;**30**:2725-2729. DOI: 10.1093/molbev/mst197

[11] Pearson WR, Robins G, Zhang T. Generalized neighbor-joining: More reliable phylogenetic tree reconstruction. Molecular Biology and Evolution. 1999;**16**:806-816. DOI: 10.1007/978-1-62703-646-7_5

[12] Ying M, Qiao Y, Yu L. Evolutionary expansion of nematode-specific glycine-rich secreted peptides. Gene. 2016;**587**:76-82. DOI: 10.1016/j.gene.2016.04.049

[13] Bond MR, Ghosh S, Wang P, Hanover JA. Conserved nutrient sensor O-GlcNAc transferase is integral to *C. elegans* pathogen-specific immunity. PLoS One. 2014;**9**:e113231. DOI: 10.1371/journal.pone.0113231

[14] Head B, Aballay A. Recovery from an acute infection in *C. elegans* requires the GATA transcription factor ELT-2. PLoS Genetics. 2014;**10**:e1004609. DOI: 10.1371/journal.pgen.1004609

[15] Pukkila-Worley R, Rhonda FR, Kirienko NV, Larkins-Ford J, Conery AL, Ausubel FM. Stimulation of host immune defenses by a small molecule protects *C. elegans* from bacterial infection. PLoS Genetics. 2012;**8**:e1002733. DOI: 10.1371/journal.pgen.1002733

[16] Sun J, Singh V, Kajino-Sakamoto R, Aballay A. Neuronal GPCR controls innate immunity by regulating noncanonical unfolded protein response genes. Science. 2011;**332**:729-732. DOI: 10.1126/science.1203411

[17] Bargmann CI. Chemosensation in *C. elegans*. WormBook. 2006;**25**:1-29. DOI: 10.1895/wormbook.1.123.1

[18] *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science. 1998;**282**:2012-2018. DOI: 10.1126/science.282.5396.2012

[19] Sugimoto A. High-throughput RNAi in *Caenorhabditis elegans*: genome-wide screens and functional genomics. Differentiation. 2004;**72**:81-91. DOI: 10.1111/j.1432-0436.2004.07202004.x

[20] Ying M, Huang X, Zhao H, Wu Y, Wan F, Huang C, Jie K. Comprehensively surveying structure and function of RING domains from *Drosophila melanogaster*. PLoS One. 2011;**6**:e23863. DOI: 10.1371/journal.pone.0023863

IntechOpen