Conserved Orthology in Mitochondrial Genomes of Distantly Related Nematodes

Nima PL[#], Riju A[$], Reena N[€] and Eapen SJ[¥]
Bioinformatics Centre, Indian Institute of Spices Research, Calicut- 12
Telephone. No. +914952731566
Email id: [#] nimanov1@gmail.com; [$] riju.bioinfo@gmail.com;
[€] reena_narayanan@yahoo.co.in; [¥] sjeapen@spices.res.in

ABSTRACT
Identification of orthologous segments plays a very important role in comparative genomics studies. In the present study, we have identified orthologous segments shared between *Radopholus similis* and 15 other nematodes. Complete genomes of 16 nematodes were used for the study. OSfinder was used to find the orthologous segments shared between *R. similis* and other 15 nematodes. Orthologous segments were visualized with the help of GTK powered Murasaki Visualizer (GMV) programme. Extremely AT rich genome of the burrowing nematode *R. similis*, which has the largest mitochondrial genome, was found to have orthologous segments from start position, 4 to end position 16791 with 15 nematodes. *Brugia malayi, Dirofilaria immitis, Onchocerca volvulus*, and *Xiphinema americanum* share similar orthologous segment with that of *R. similis*. The mitochondrial genome analysis revealed the presence of conserved gene locations in mitochondrion and the close evolutionary relationship of nematodes belonging to different clades and different parasitic habitats. This study has many practical implications like reconstruction of ancestral genome of nematode and calculation of evolutionary time.

Categories and Subject Descriptors
A.0 [General]: Conference proceedings

General Terms
Experimentation

Keywords
Comparative genomics, Nematode, Synteny, Orthologous segments, Mitochondrial genome.

1. INTRODUCTION
Comparative genomics is an exciting new field of biological research in which the genome sequences of different species are compared. It also provides a powerful tool for studying evolutionary changes among organisms, to identify genes that are conserved among species, as well as genes that give each organism its unique characteristics. The rapidly emerging field of comparative genomics has already yielded dramatic results. For example, by comparing the fruit fly genome with the human genome discovered that about 60 percent of genes are conserved between fly and human. More recently, a comparative genomic analysis of six species of yeast prompted scientists to significantly revise their initial catalog of yeast genes and to predict a new set of functional elements thought to play a role in regulating genome activity. Comparative genomics can be done using organelles like mitochondria, mitochondrial plasmids and chloroplast.

The accurate detection of orthologous segments plays a key role in comparative genomics as it is useful for the following: inferring rearrangement-based phylogenies [2, 21], reconstructing ancestral genomes [1, 17], computing whole-genome alignments [5, 6], identifying orthologous genes [11, 24] and detecting non-coding functional elements such as regulatory elements [7]. In the present study, mitochondrial genomes of nematodes were analyzed to detect orthologous segments. Nematodes or roundworms are an ancient and diverse group of organisms and the most abundant of all metazoans. Molecular phylogenetics defines three major nematodes classes which are further divided into five clades [16]. Generally mitochondrial genomes are small (~13-26kb), circular, compact and haploid. They contain 12-13 protein genes that encode enzymes required for oxidative phosphorylation, two ribosomal rRNA genes encoding the RNA components of the mt ribosome, and 22 transfer tRNA genes required for translation of the different mt proteins. Most nematode mt genomes contain 12 protein-coding genes, 22 tRNA genes and two rRNA genes [10].

The problem of identifying orthologous segments is referred to as orthology mapping [5]. The general strategy of orthology mapping include (1) detect the positions of short homologous regions called anchors (2) detect collinear anchors which are distributed in the same order and have the same orientation (3) connect closely located collinear anchors (4) output connected components as orthologous segments [8]. A number of algorithms for detecting orthologous segments have been proposed. ADHoRe [22] and SyMAP [20] are tools for detecting orthologous segments that are capable of automatically determining the distance threshold value. Both ADHoRe and SyMAP define the quality of the orthologous segments on the basis of the diagonal properties of the anchor positions. Although these programs can determine the distance threshold automatically, they require a quality threshold to be set manually. In addition to the above programs, other

orthology-mapping algorithms, including DAGChainer [9], AXTCHAIN [14-15], DiagHunter [4], FISH [3] and Cinteny [19] also require the manual setting of key threshold parameters [8].
In the present study, we used an orthology-mapping algorithm, named OSfinder (Orthologous Segment finder) [8], which uses a novel scoring scheme based on stochastic models.

2. MATERIALS AND METHODS

2.1 Mitochondrial Genome Retrieval

A total of 16 mitochondrial genomes of nematodes belonging to different clades including human, domestic, animal, plant and free living parasites were downloaded from NCBI namely, *Radopholus similis*, *Brugia malayi, Dirofilaria immitis, Ancylostoma caninum, Ascaris suum, Caenorhabditis briggsae, Caenorhabditis elegans, Cooperia onchophora, Haemonchus contortus, Necator americanus, Onchocerca volvulus, Strongyloides stercoralis, Steinernema carpocapsae, Toxocara canis, Trichinella spiralis* and *Xiphinema americanum*.

R. similis and *X. americanum* are plant parasites. *N. americanus*, *S. stercoralis, B. malayi, O.volvulus*, and *T. spiralis* are human parasites. *H. contortus, A. caninum, C. onchophora, A. suum,* and *D. immitis* are domestic animal parasites. *C. elegans* and *C. briggsae* are free living parasites.

R. similis is one of the most important root pathogen attacking banana, black pepper and many other tropical plants. It is widespread and can cause yield losses of up to 30-60% in many countries. Among the nematodes considered, *R. similis* has got the largest genome of about 16791 nucleotides. Mitochondrial genome of *R. similis* is extremely AT rich and has a unique genetic code [12-13].

R. similis was used as the base nematode for the study. To detect orthologous segments between *R. similis* and 15 nematodes, position of short homologous regions called anchors were generated using Murasaki [18].

2.2 Anchor Detection

The term anchor generally refers to well-conserved short regions between two or more genomes, and is biologically defined as a group of homologous genes or a set of homologous sequence matches. Genome sequences of *R. similis* with each of the 15 nematodes were given as input to murasaki and anchors were generated. A total of 15 anchor files were generated. After format transformation, the anchors generated by Murasaki were given as input to OSfinder [8].

2.3 Constructing Orthologous Segments and Visualization

OSfinder is a tool for accurate orthology mapping. OSfinder performs accurate orthology mapping by using Markov chain models and machine learning techniques. The originality of OSfinder lies in the ability to compare multiple genomes simultaneously, and to automatically optimize the parameters used in the Markov chain models, while the users of the other existing orthology mapping programs suffer from the manual setting of the parameter values. OSfinder takes the genomic locations of anchors as input, and reports the genomic locations of chains (set of genomic segments in which anchors are distributed densely in off-diagonal positions) and orthologous segments as output. Anchor files generated by Murasaki were format transformed and were given as input to OSfinder. It generated the orthologous segments between *R. similis* and 15 nematodes in text format. The output need to be changed into anchor file format in order to visualize it through GMV. GMV is a comparative genome browser for Murasaki. GMV visualizes anchors from Murasaki, annotation data from GenBank files, and expression / prediction score from GFF (General Feature Format) files.

3. RESULTS AND DISCUSSION

In this study, orthologous regions were found out among the nematode genomes. The mitochondrial genome of the burrowing nematode, *R. similis* was compared with 15 nematode genomes (**Table 1**).

Mitochondrial genome of *R. similis* is found to be orthologous with 15 nematodes from start position 4 to end position 16791. As the above table shows genomes of *B. malayi, D. immitis, O. volvulus*, and *X. americanum* shares similar orthologous segment with that of *R. similis*. Similarly orthologous segments of *C. elegans* and *N. americanus* were found to be similar. The extremely AT rich burrowing nematode *R. similis* has the largest genome of about 16791 nucleotides.

In the mitochondrial genome of *R. similis*, several genes were detected in the orthologous region shared by the 15 nematodes. Some of these genes have similar position with the compared nematodes (**Table 2**). Genes ND1, ND5, COX2, COX1 and ND4 present in *R. similis* have got similar position in other nine nematodes. ND1- Mitochondrially encoded NADH dehydorgenase 1 provides instructions for making a protein called NADH dehydrogenase 1. ND4 encoding NADH dehydrogenase 4 helps to make the protein NADH dehydrogenase 4 which is a part of complex 1. All these three genes help in the process of oxidative phosphorylation by mitochondria.

Any changes or mutation to these genes can cause several conditions like Leber hereditary optic neuropathy, mytocho ndrial encephalomyopathy, lactic acidosis and stroke like episodes.

COX1-Cytochrome c oxidase subunit 1 and COX2 - Cytochrome c oxidase subunit 2 are the terminal components of the mitochondrial respiratory chain, catelysing the electron transfer from reduced cytochrome c to oxygen and are involved in the biosynthesis of heme A. COX deficiency is a clinically heterogeneous disorder. The clinical features range from isolated myopathy to severe multisystem disease with onset from infancy to adulthood.

Table 1. Orthologus segments detected between *R. similis* and other 15 nematodes

Radopholus similis start position	*Radopholus similis* end position	Strand	Nematodes	Start position	End position	Strand
335	16777	+	*D. immitis*	16	13638	+
108	16785	+	*A. caninum*	82	13682	+
4	16597	+	*A. suum*	2	14239	+
157	16777	+	*B. malayi*	5	13638	+
364	16502	+	*C. briggsae*	87	14416	-
14	16386	+	*C. elegans*	68	13784	+
107	16791	+	*C. onchophora*	112	13500	+
626	16568	+	*H. contortus*	11	13961	+
14	16686	+	*N. americanus*	68	13467	+
845	16781	+	*O. volvulus*	3	13714	+
599	16685	+	*S. stercoralis*	1	13575	+
37	16774	+	*S. carpocapsae*	181	13866	+
422	16660	+	*T. canis*	40	14219	+
1512	16648	+	*T. spiralis*	8	16706	-
37	16224	+	*X. americanum*	62	12399	-

Table 2. Conserved genes with positional information.

Species	**ND1**	**ND5**	**COX2**	**COX1**	**ND4**
R. similis	1029-1889	2505-4040	4148-4825	6113-7669	12734-13951
A. Caninum		3948-5523			12388-13617
C. briggsae			3526-4449		
C. elegans	1763-2638				
C. onchophora	1808-2678				
H. contortus		3912-5493			12746-13975
T. Canis				6046-7623	
T. spiralis		1008-2683			
N. americanus	1805-2677				
S. stercoralis				5380-6921	

The orthologous segments in detected between mitochondrial genome of *R. similis* and *B. malayi, O. volvulus* and *D. immitis* are as follows:

R. similis and B. malayi
157-16777 and 5-13638
R. similis and D. immitis
335-16777 and 16-13638
R. similis and O. volvulus
845-16781 and 3-13714

This study revealed the conserved orthology and positional information of genes between distantly related mitochondrial genomes of nematodes considered for the study.

4. ACKNOWLEDGEMENTS

This work was supported by the grant from Department of Biotechnology (BTISnet), Government of India, New Delhi.

5. REFERENCES

[1] Bourque, G., Zdobnov, EM., Bork, P., Pevzner, P.A., and Tesler, G. Comparative architectures of mammalian and chicken genomes reveal highly variable rates of genomic rearrangements across different lineages. *Genome Res*., 15 (2005), 98–110.

[2] Bourque, G.. Pevzner, P.A., and Tesler, G. Reconstructing the genomic architecture of ancestral mammals: lessons from human, mouse, and rat genomes. *Genome Res*., 14 (2004), 507–516.

[3] Calabrese, P.P., Chakravarty, S. and Vision, J.T. Fast identification and statistical evaluation of segmental homologies in comparative maps. *Bioinformatics*, 19 (Suppl. 1) (2003), 74–80.

[4] Cannon, S.B., Kozik, A., Chan, B., Michelmore, R. and Nevin D Young DiagHunter and GenoPix2D: programs for genomic comparisons, large-scale homology discovery and visualization. *Genome Biol.,* 4 (2003), R68.

[5] Dewey, CN. Huggins, P.M., Woods, K., Sturmfels, B. and Pachter,L. Parametric alignment of Drosophila genomes. *PLoS Comput. Biol.,* (2006) 2, e73.

[6] Dewey, C.N. Aligning multiple whole genomes with Mercator and MAVID. *Methods Mol. Biol*., 395 (2007), 221–236.

[7] Frazer, K.A., Pachter, L., Poliakov, A., Rubin, ,E.M. and Dubchak, I. VIATA: computational tools for comparative genomics. *Nucleic Acids Res*., 32 (2004), W273–W279.

[8] Hachiya, T., Osana, Y., Popendorf, K., and Sakakibara, Y. Accurate identification of orthologus segments among multiple genomes. *Bioinformatics*. 25, 7(2009), 853-860.

[9] Hass, B.J., Delcher, A.L., Wortman, J.R. and Salzberg, S.L. DAGchainer: a tool for mining segmental genome duplications and synteny. *Bioinformatics*, 20 (2004), 3643–3646.

[10] Hu, M., and Gasser, R.B. Mitochondrial genomes of parasitic nematodes-progress and perspectives. *TRENDS Parasitol*. 22(2) (2006),78-84.

[11] Hubbard, T,J., Aken, B.L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S.C., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Ouverdin, B., Parker, A., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X.M., Flicek, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal, A., Birney, E. Ensembl 2007. *Nucleic Acids Res*., 35 (2007), 610–617.

[12] Jacob, J., Mitreva, M., Vanholme, B., Gheysen, G. Exploring the transcripome of the burrowing nematode *Radopholus similis*. *Mol Genetics Genomics.* (2008), 280:1-17.

[13] Jacob, JEM, Vanholme, B., Leeuwen, T.V., and Gheysen, G. A unique genetic code change in the parasitic nematode *Radopholus similis. BMC Research Notes* 2**:**192 (2009), doi:10.1186/1756-0500-2-192.

[14] Kent, W.J., Baertsch, R, Hinrichs, A, Miller, W., and Haussler, D. Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. *Proc. Natl Acad. Sci. USA*, 100 (2003), 11484–11489.

[15] Kent, W.J., and Zahler, A. Conservation, Regulation, Synteny and Introns in a Large-scale *C. briggsae-C.elegans* Genomic Alignment. *Genome Res*. (2009), 1115-1125.

[16] Mitreva, M., Blaxter, M.L., Bird, D.M., and McCarter, J.P. Comparative genomics of nematodes. TRENDS in Genetics. 21(10) (2005), 573-81.

[17] Murphy, W.J, Larkin, D.M., Wind, A.E., Bourque, G., Tesler, G., Auvil, L., Beever, J.E., Chowdhary, B.P., Galibert, F., Gatzke, L., Hitte, C., Meyers, S.N., Milan, D., Ostrander, E.A., Pape, G., Parker, H.G., Raudsepp, T., Rogatcheva, M.B., Schook, L.B., Skow, L.C., Welge, M., Womack, J.E., O'Brien, S.J., Pevzner, P.A., and Lewin, H.A. Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. *Science,* 309(5734) (2005), 613–617.

[18] Popendorf, K., Osana, Y., Hachiya, T., and Sakakibara, Y. Murasaki-homology detection across mutliple large-scale genomes. In Fifth Annual RECOMB Satellite Workshop on Comparative Genomics. San Diego, USA (2007).

[19] Sinha, A.U., and Meller, J. Cinteny: flexible analysis and visualization of synteny and genome rearrangements in multiple organisms. *BMC Bioinformatics*, 8, 82 (2007). doi:10.1186/1471-2105-8-82

[20] Soderlund, C., Nelson, W., Shoemaker, A. and Paterson A. SyMAP: A system for discovering and viewing syntenic regions of FPC maps. *Genome Res.,* 16, (2006)1159–1168.

[21] Tesler, G. GRIMM: genome rearrangements web server. *Bioinformatics,* 18 (2002), 492–493.

[22] Vandepoele, K, Saeys Y, Simillion C, Raes J and Van De Peer, Y. The automatic detection of homologous regions (ADHoRe) and its application to microcolinearity betweenArabidopsis and rice. *Genome Res*., 12 (2002), 1792–1801.

[23] Zheng, X.H, Lu, F., Wang, Z-Y., Zhong, F, Hoover, J., and Mural, R. Using shared genomic synteny and shared protein functions to enhance the identification of orthologous gene pairs. *Bioinformatics*, 21(2005), 703–710.