

Nematode.net update 2011: addition of data sets and tools featuring next-generation sequencing data

John Martin¹, Sahar Abubucker¹, Esley Heizer¹, Christina M. Taylor¹ and Makedonka Mitreva^{1,2,*}

¹The Genome Institute and ²Department of Genetics, Washington University School of Medicine, St Louis, MO 63108, USA

Received September 14, 2011; Revised October 27, 2011; Accepted November 16, 2011

ABSTRACT

Nematode.net (<http://nematode.net>) has been a publicly available resource for studying nematodes for over a decade. In the past 3 years, we reorganized Nematode.net to provide more user-friendly navigation through the site, a necessity due to the explosion of data from next-generation sequencing platforms. Organism-centric portals containing dynamically generated data are available for over 56 different nematode species. Next-generation data has been added to the various data-mining portals hosted, including NemaBLAST and NemaBrowse. The NemaPath metabolic pathway viewer builds associations using KOs, rather than ECs to provide more accurate and fine-grained descriptions of proteins. Two new features for data analysis and comparative genomics have been added to the site. NemaSNP enables the user to perform population genetics studies in various nematode populations using next-generation sequencing data. HelmCoP (Helminth Control and Prevention) as an independent component of Nematode.net provides an integrated resource for storage, annotation and comparative genomics of helminth genomes to aid in learning more about nematode genomes, as well as drug, pesticide, vaccine and drug target discovery. With this update, Nematode.net will continue to realize its original goal to disseminate diverse bioinformatic data sets and provide analysis tools to the broad scientific community in a useful and user-friendly manner.

INTRODUCTION

Next-generation sequencing (NGS) technology has revolutionized genome and transcriptome sequencing

and related applications, enabling greater exploration of the highly diverse phylum, Nematoda (1). As the most abundant animal on earth (2), nematodes live in diverse environments in both land and water, ranging from hot springs to polar habitats, and in myriad hosts, from insects to higher mammals. Nematode species can be free-living or parasitic, infecting plants, animals and humans. Human parasitic nematodes are a scourge on humanity, with the World Health Organization estimating 2.9 billion people infected and 10.6 million people potentially at risk for just three of the major parasitic nematode infections (Ascariasis, Trichuriasis and Hookworm infection) (3). Parasitic nematode infections can lead to disfigurement, impaired growth and development in children, and poor pregnancy outcomes, often infecting women and children and predisposing the population for other diseases (3). In addition, the parasitic nematodes of veterinary and agricultural importance elevate levels of poverty worldwide. Anthelmintic drugs for pets and life stock represented a \$3.2 billion industry (1,4) and plant parasitic nematodes cost an inflation-adjusted \$10 billion in US and \$125 globally (5). The ever-increasing cost for controlling these parasites is mainly a result of emerging resistance due to the heavy use of the limited number of anthelmintics used to control them (6–8). Given their huge impact on human health and economics it is critical to enable use of existing and upcoming ‘omics’ data to better understand these devastating pathogens at a molecular level and to aid in discovering novel control and prevention programs by the nematode research community.

Nematode.net (<http://nematode.net>) has provided the nematode community with access to annotation of genomes, expression and related genome-scale data, as well as data-mining tools and their comparative data products, for over a decade. Hosting ‘omics’ data that span more than 50 species, Nematode.net is an invaluable tool for the nematode research community serving 3060 unique visitors and 4420 returning visitors between 1 May 2011 and 31 August 2011. To continue serving the parasitology community, this update of the Nematode.net

*To whom correspondence should be addressed. Tel: +1 314 286 2005; Fax: +1 314 286 1810; Email: mmitreva@genome.wustl.edu

Table 1. Growth of data hosted by Nematode.net

	2004	2008	2011
cDNA sequencing totals	215 127	509 161	11 880 572
NemaGene transcripts ^a	11 185	109 268	233 125
NemaProt gene totals	0	12 881	225 519
Codon Usage table codon counts	730 585	2 708 625	17 463 274

^aNemaGene Transcript number is accounting for 8% fragmentation (non-overlapping parts of the same gene) for conventional capillary ESTs and 20% fragmentation for the Roche/454 cDNAs.

resource describes new species and data types, such as the addition of pyrosequencing cDNA data for nine species (including assembly and analysis), including hookworms (>1 billion infections) and veterinary pathogens such as lungworm and trichostrongylids. New expression data based on microarrays has been made available under the Microarray section of the site (9,10). New comparative analysis tools, NemaSNP and HelmCoP, have been added, and earlier tools, NemaBrowse, NemaBlast and NemaPath have been updated and improved. Table 1 highlights some of the massive expansion of data hosted by Nematode.net over the last few years.

RECENT IMPROVEMENTS

To aid in better flow through our site, a reorganization has been done. A species-specific navigation menu is available on the left-hand side of all pages. Due to the large number of species now available (8 free-living nematode species, 10 human-parasitic, 15 animal-parasitic, 17 plant-parasitic and 2 entomopathogenic species), the menu has been grouped based on host type, and clicking on the host expands to show species. Further, additional resources relevant to users of the site are available under the News, Community, Education and FAQ menus near the top of the left-hand navigation panel. The page header now contains dropdown menus for all major types of analysis available on Nematode.net. The website organization was improved, enabling the user to more easily access information. The progress of various helminth genome sequences being researched as part of the Parasitic Nematode Genome program (PNGP) at The Genome Institute (TGI) is also shown on the front page, allowing visitors to see the status of various genome-sequencing projects. We provide links to helminth genomes that have already been sequenced by other institutions, creating an information hub to all researchers interested in nematode biology. An overview of the information flow through the site is shown in Figure 1.

ORGANISM-CENTRIC PORTALS

The previous species pages have been expanded into organism-specific portals hosting all available data sets, information and analyses available for that species. The organisms represented have been reorganized and expanded, resulting in inclusion of 56 nematode species. The portals host basic information such as species name,

disease, sequencing center, collaborators and contact information as well as available sequence information, with links to the assembly data, NemaBLAST database, codon usage tables, GO associations, codon usage statistics, NemaPath results (Figure 2A) and where available, a link to the NCBI Bioproject page is provided (example is shown in Figure 2B). We also link out to corresponding organism pages at WormBase (11), Nembase4 (12) and the Sanger Institute (<http://www.sanger.ac.uk/resources/downloads/helminths/>) when applicable. These pages offer a user that is interested in a specific nematode a central portal from which they can obtain all experimental and analysis data available for that species. A MySQL table is used on the backend, allowing the species pages to be dynamically generated as the database gets populated with new information.

DATA-MINING PORTALS: UPDATES AND IMPROVEMENTS

NemaGene

NemaGene has been expanded by the addition of transcriptome assemblies from nine species (Figure 3) as part of our goal to include the large amount of data generated using next-generation sequencing platforms. A total of 11.8 million reads totaling around 4 billion bp, are included in these new assemblies. We have been able to take advantage of the newbler transcriptome assembler (pre-release version 2.5) (13) and provide alternatively spliced transcripts generated from these assemblies. While there are many assemblers for cDNA pyrosequencing data, only the newbler assembler provides information on alternative splicing (14). An example of this is shown in Figure 3B and C, illustrating the type of assembled loci users can expect. The assembler uses an overlap layout consensus approach to generate splice graphs and recreates putative isoforms (referred to as 'isotigs') belonging to a gene locus (referred to as 'isogroups'). New search options have been added to this analysis tool. The page now supports searches based on isotig, isogroup and stage (L1, L2, L3, L4, egg, microfilariae, etc.) in addition to its previous capabilities. The addition of the 'Stage' search criteria provides users a handle on transcripts expressed in specific life cycle stages. NemaGene pages resulting from these searches have also been streamlined to present information in a more concise manner. The NemaGene FAQ has been expanded to better describe this resource.

NemaBLAST

NemaBLAST has been updated to include all available Roche/454 cDNA isotigs. The incorporation of this new data increased the total number of species with available transcriptome data in the 'nemablast versus contigs' tool to 45 and added data to 4 pre-existing species (*Ascaris suum*, *Onchocerca flexuosa*, *Ancylostoma caninum*, *Ostertagia osteragi*). This page uses the WU-BLAST (15) service to perform similarity searches of protein or nucleotide queries against locally curated nematode databases. The entries in the database have

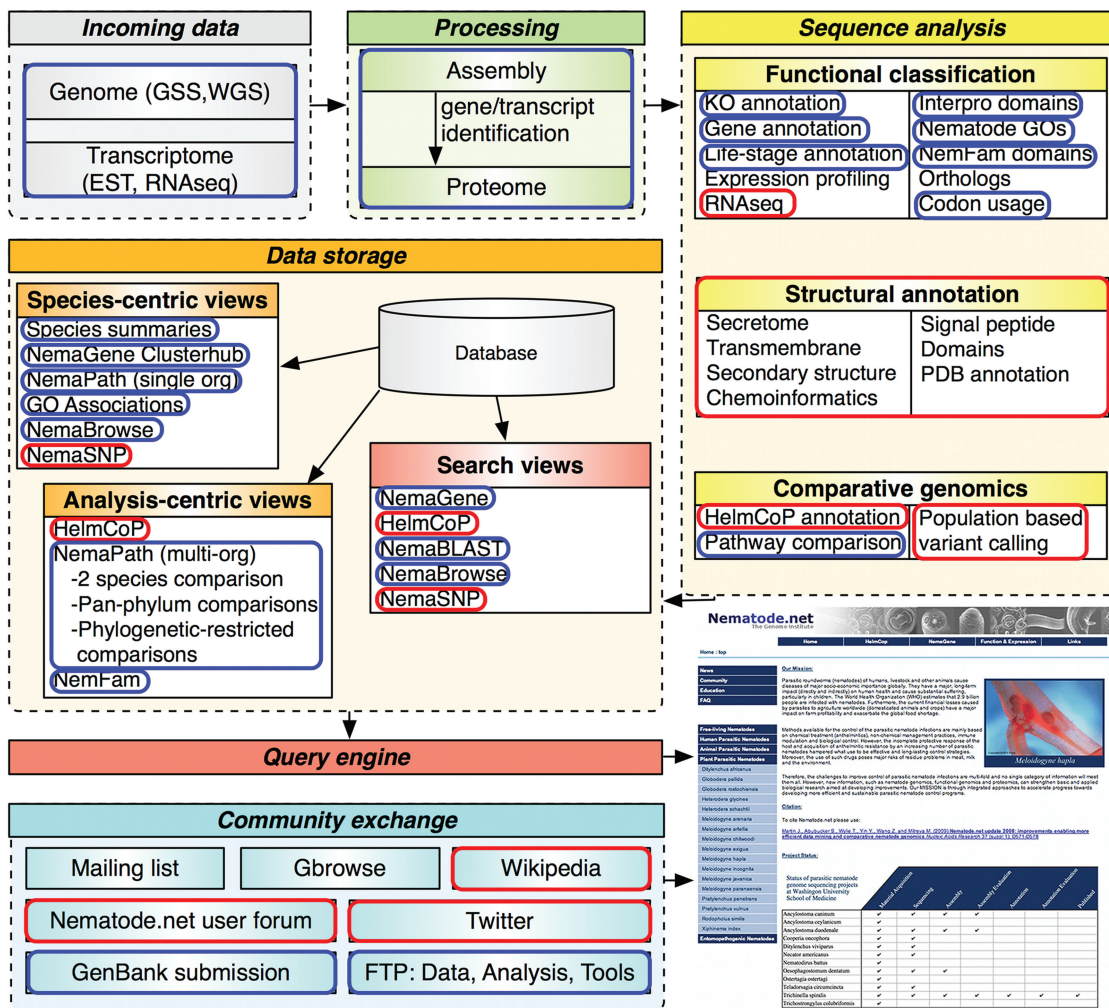


Figure 1. Information flow on Nematode.net. The flowchart outlines the passage of information from raw data to analysis and output onto Nematode.net. New features (boxed in red) as well as resources with data added since the 2008 revision (boxed in blue) have also been indicated.

also been annotated to reflect the platform on which they were generated (Sanger capillary or Roche/454).

NemaBrowse

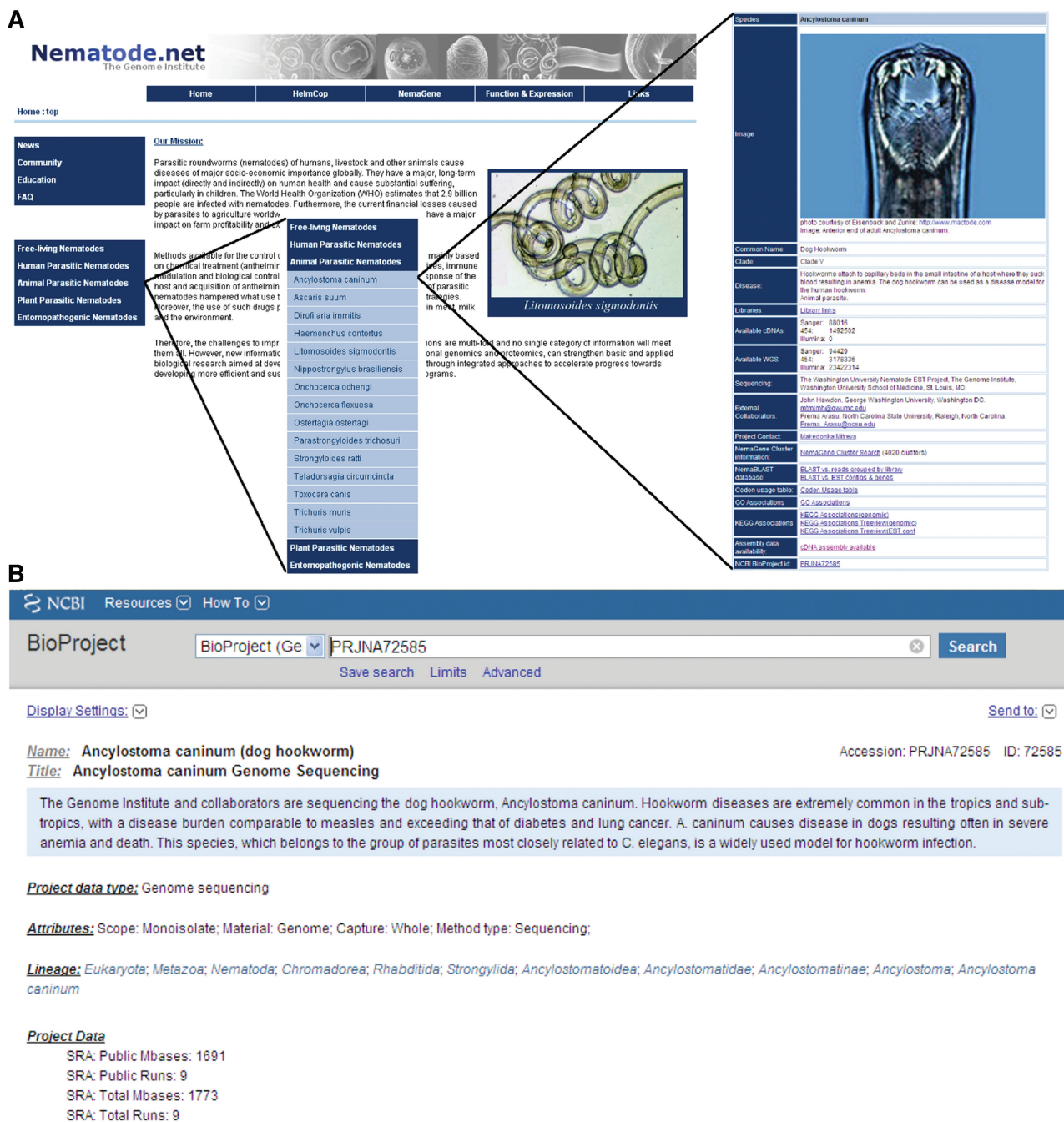
While Nematode.net’s focus has primarily been providing navigation of nematode transcriptome data (16), the resource has also been providing genome annotations from intermediate draft assemblies that are part of the PNGP. Using the GBrowse interface, NemaBrowse contains a collection of nematode genomes that have gene predictions. For this release of Nematode.net, we have added gene annotations for *Trichinella spiralis* (17). As the PNGP (18,19) progresses and new genome resources are made available, we plan to add them into the NemaBrowse, along with Rfam predictions, NemaSNP variants and gene-calling evidence as analysis becomes available. The end goal would be to include the final assembly and annotation in WormBase (11) as it is a resource that has already started hosting non-*Caenorhabditis* species, however the completion of a genome project is a lengthy process (even when

next-generation sequence data is used), therefore an immediate release of the interim products enables use of the data by the research community as soon as they are generated.

NemaPath

The NemaPath metabolic pathway viewer has been modified to build associations using KOs, rather than ECs. KO mapping provides a more focused and accurate representation of enzymes involved in various metabolic pathways and more genes have KO assignments compared to enzyme assignments (2426322 versus 1310435). The NemaPath pipeline utilizes WU-BLAST to align the nematode transcriptome data to the genes in the KEGG Genes database (20). The web viewing application, NemaPath (21), colors the KEGG metabolic pathway image maps based on user-defined cutoffs.

In addition to the metabolic pathway maps, three new high-level pathway categories— genetic information processing, environmental information processing and cellular processes—have been added. The maps have



been updated to KEGG version 58, a recent release. In addition, for all species with new Roche/454 cDNA data, KEGG associations have been generated and added to the portal. The previous version of the KEGG annotations has been archived.

NemFam

The NemFam collection of conserved, nematode protein families has been extended by the addition 274 nematode

specific, full length gene families. Proteins from four nematode species spanning the phylum Nematoda (*Brugia malayi*, *Meloidogyne incognita*, *M. hapla* and *T. spiralis*) and two outgroups (*Drosophila melanogaster* and *Saccharomyces cerevisiae*) were clustered using OrthoMCL to build orthologous families. From the groups identified, 2517 were conserved across the nematodes and 274 were found only in nematodes, out of which 85 were found to be single-copy genes (17).

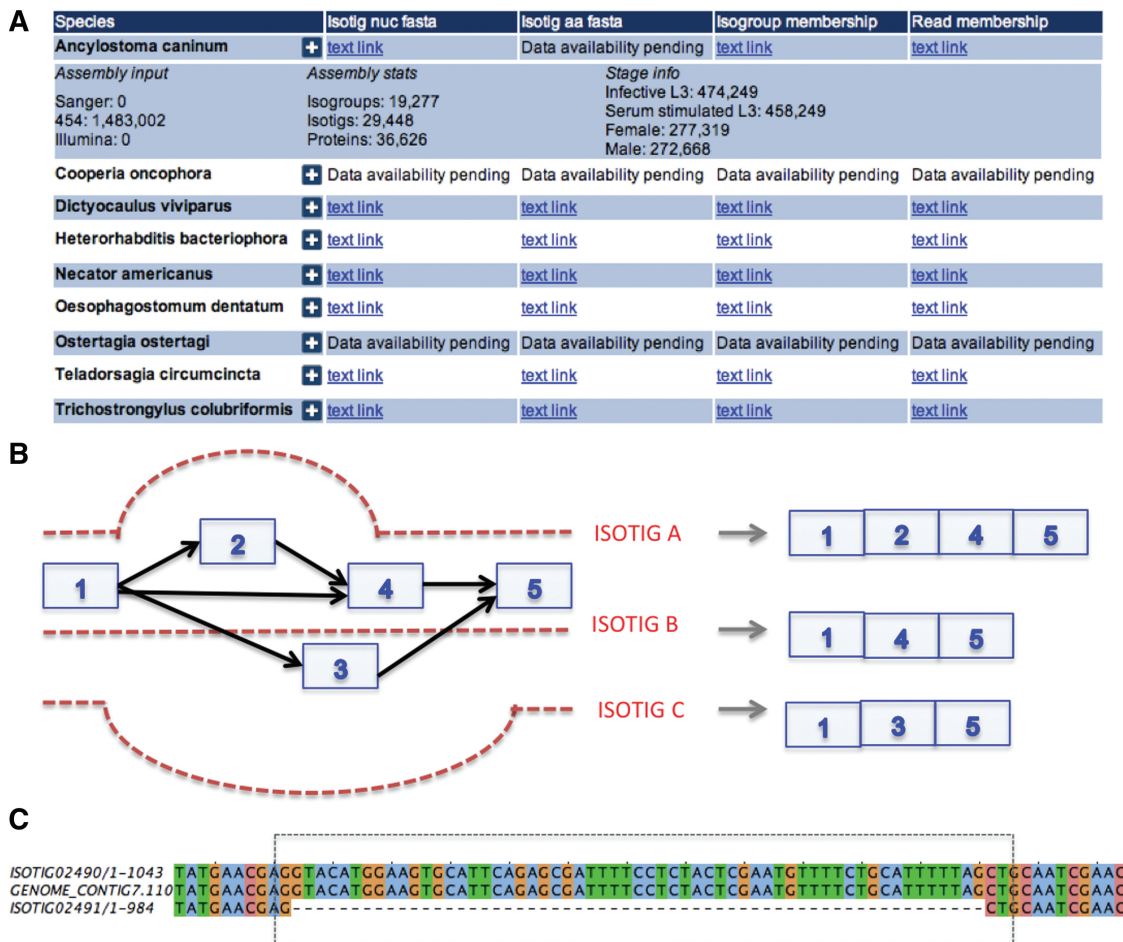


Figure 3. Transcript assemblies using next-generation sequencing data. (A) Transcript assembly downloads are accessible through this page. The text links can be followed to download the isotig nucleic acid fasta, isotig amino acid fasta, isogroup membership and read membership. The species can be expanded to obtain various assembly statistics and for information regarding the raw data used to assemble the sequences. (B) Putative isoforms (isotigs) built from contigs (which are labeled 1–5). This image shows examples of alternatively spliced transcripts that a user can generate using the cDNA isotig/isogroup information provided. (C) Alignment of a putative gene (isogroup) from *A. caninum* against a portion of the *A. caninum* draft genome. Acan_isogroup00809 has two alternatively spliced isotigs, Acan_isotig04444 and Acan_isotig04445 from three contigs. The contig which is present only in one isotig and the genome but not in the other isotig is outlined.

DATA ANALYSIS AND COMPARATIVE GENOMICS PORTAL

HelmCoP

Helminth Control and Prevention (HelmCoP) (22) is a significant addition to the comparative genomics tools offered on Nematode.net. This site provides a central resource for storage, annotation and comparative genomics of helminth proteomes. The site facilitates drug and target discovery for parasitic helminths which can be used to prioritize drug, pesticide and vaccine, as well as enabling researchers to compare the rapidly growing number of nematode genomes available. The site incorporates functional, structural and comparative genomic data from plant, animal and human helminths, as well as hosts and model organisms. A MySQL relational database is used to store pertinent information defining HelmCoP, and a perl based query engine constructs SQL queries from user-defined searches. Perl CGI scripts mediate the display.

HelmCoP can be searched using genes as the atomic unit, or by orthologous groups. Data from a number of parasitic and free-living nematodes can be screened, along with several flatworms and outgroup/host organisms, in a combinatorial manner. The query can be limited to species-specific proteins by excluding a species from the search, which is very useful for excluding host proteins to find species-specific information. The site hosts functional annotations, such as Gene Ontology (23), KEGG Orthology and InterPro (24) domains. The KO numbers can be mapped to pathways and drugs within the KEGG database. The site incorporates RNAi information regarding *C. elegans* to determine gene essentiality in other species through orthology. The site also provides structural annotation, as well as a searchable field, including signal peptide information which could determine if the protein is secreted / excreted, indicating that the protein is potentially involved in a host–parasite interaction. Structural annotation also includes associated transcript information, which can provide details about the tissues

where the protein is expressed. Data can be searched for proteins that have homology to the PDB in the various genomes. The database can be searched based on DrugBank IDs (25) and proteins from species with homology to the DrugBank target will be returned. Cheminformatic information can also be output to enable researchers to do compound prioritization. The database can also be parsed for vaccine candidates in different species or those that are not also present in the host. The data can be annotated with coiled coil, secondary structure and regions of disorder information. Results are displayed via HTML, but are also available as tab-delimited text for download. A HelmCoP FAQ has been provided to offer users more detail regarding the construction, composition and typical usage of this new resource.

NemaSNP

With the wide use of NGS data, one could easily perform population genetics studies utilizing multiple cohorts of worms. The genomes of nematodes display different levels of geographic diversity and identification of genetic markers suitable for fingerprinting of nematode isolates is required in order to differentiate, for example, persistent or reintroduced infections or determine polymorphic genes/gene families (for drug or vaccine development heavily polymorphic targets may affect efficacy).

Furthermore, heavy use of anthelmintics to control parasitic nematodes has led to nematode resistance against all drug classes (6), but research into the mechanism responsible for this resistance is subsequently in a more advanced stage is some classes, such as benzimidazoles, than for other classes (26). To aid in research focused on genetic variations among nematode isolates/populations on a genome-wide level, we have developed a new portal NemaSNP. The features that the portal displays are demonstrated through the comparison of two populations of the cattle stomach parasite *Teladorsagia circumcincta*. To estimate the level of reduction of polymorphism in the selected inbred susceptible strain (i.e. 'inbred population'), we compared its adult transcriptome to the adult transcriptome of a population that has also been passaged for years but with no attempts to reduce polymorphism (i.e. 'non-inbred population'). Single nucleotide polymorphisms (SNPs) were enumerated using samtools mpileup & bcftools (27), and are displayed in Gbrowse (<http://gmod.org>). Differences between the two populations can be easily visualized as shown in Figure 4.

DATA INTEGRATION

The two major sister sites serving the nematode community are WormBase and Nembase4. WormBase is primarily focused on *C. elegans* genomics and slowly

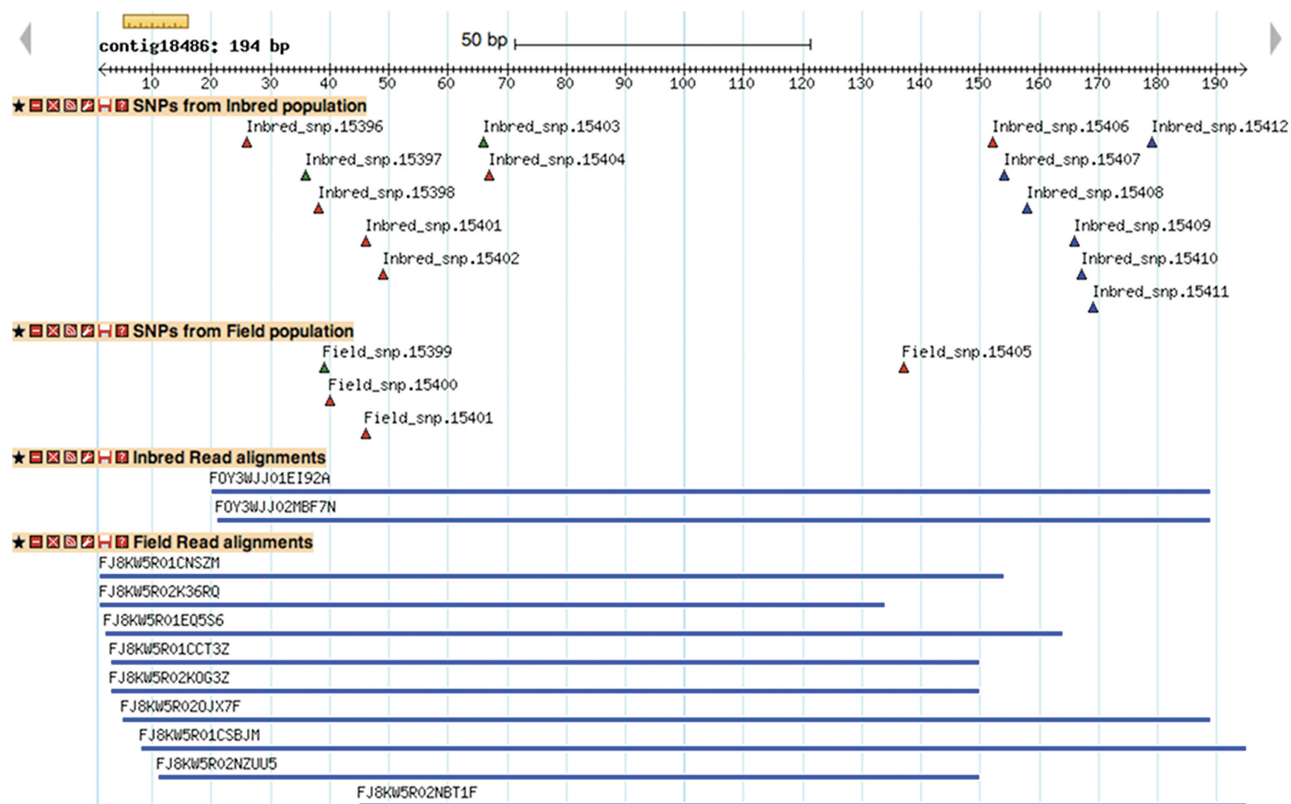


Figure 4. NemaSNP viewer. This screen shot shows a typical SNP annotation using GBrowse. This example shows a screen shot of SNPs in *T. circumcincta* (contig02609) in an inbred versus field population represented with triangles along the contig (red triangles indicate non-synonymous, green indicate synonymous SNPs, yellow is a site with multiple alleles that can result in a silent or non-synonymous change, and in blue is a site that lies outside of validated CDS). The inbred and field alignments are shown below the SNP annotation.

incorporates other non-*Caenorhabditis* genomes and Nembase4 serves transcriptome data from multiple species and hosts various analyses on these data sets. While there is some overlap in data these analyses are independent of work hosted at Nematode.net. Nematode.net resources are integrated with these external resources on several levels. WormBase genomes annotated with 'Nematode ESTs (non-*Caenorhabditis*)' link directly into the NemaGene database, and users can seamlessly move from the WormBase genome browser into the NemaGene resource anytime mappings between WormBase genes and NemaGene transcripts are found. Organisms listed in the Nematode.net organism portals provide link-outs to genome browser views in WormBase, species pages at Nembase4 and genome resource pages at the Sanger Institute website where available. Nematode genome projects in progress at both the Genome and Sanger Institutes are tracked in the Project Status table on the front page of Nematode.net, and links are provided to the appropriate project host page from that table. Additionally, HelmCoP search results provide links into PDB and DrugBank when relationships exist. HelmCoP also provides links to GO descriptions and KEGG KO and EC definitions for all results having such annotations.

DATA MANAGEMENT AND NEMATODE COMMUNITY

Data archive and download

We have built a new archive page for previous versions of data. The new data archives include the earlier release of NemaGene Clusters and KEGG annotations from the previous mappings to KEGG. The user can refer to the earlier release of data sets that may have been used for a particular analysis. Much data has been added to the data download repository, including transcript assemblies in the FTP repository, which includes: nucleotide and protein fasta for isotigs, isogroups membership files and read membership files.

FAQ and nematode community discussion

We have created a Nematode.net user forum on google groups to aid in discussions regarding helminths and the Nematode.net site. This mechanism will also allow us to gain additional feedback regarding user preferences and request by users that will help us implement additional useful features to the site. A FAQ has been generated to provide useful information to navigate the site and help users take advantage of available tools and resources. We have also incorporated the google analytics tool that tracks the most frequently used pages on Nematode.net. This has led us to focus more on these frequently accessed pages to provide a better user experience. We also track keywords that led users to the website, helping us improve the content. Based on these analytics, we have also tested the site on the four most popular browsers favored by nearly 99% of our users (data not shown).

We have also embraced social media outlets to keep the research community informed about the latest advances

in neglected tropical disease from The Genome Institute at Washington University. We have established a presence on twitter (<http://twitter.com/#!/nematodenet>) and also on Facebook, where we post nematode-related news on the Genome Institute's page. In addition, we have also generated a Wikipedia page to increase the visibility of Nematode.net among the general population (<http://en.wikipedia.org/wiki/Nematode.net>).

CONCLUSIONS AND FUTURE PLANS

The major goal of our work is to provide the nematode research community with a site that integrates omics data, providing a springboard for subsequent investigation of these neglected tropical diseases. In the last quarter, 2556 users from 105 countries have visited our site. With the deluge of data coming from next-generation sequencers, the importance of a site that organizes the data into a manageable format will be paramount. The focus of this Nematode.net release was, among others, the incorporation of pyrosequencing transcriptome data, however the next release of Nematode.net will have to accommodate this huge increase in RNAseq data generated on the Illumina platform. In the current release, we have improved upon the infrastructure necessary to support the increase in new data and necessary analysis brought about by next-generation sequencing and also improved the site interface to ease navigation. HelmCoP has enabled users to make meaningful hypothesis using the genomic data that can be translated directly into the laboratory. The integration of HelmCoP into the Nematode.net site enables users to go from a protein sequence or gene name all the way to testable protein and drug predictions that can be tested in the laboratory without any programming experience. We have also added data to several data-mining portals, such as NemaBLAST, NemaBrowse, NemaGene and NemaPath. NemaPath has also been overhauled to utilize the more descriptive KO numbers, rather than EC.

In this new release, we have incorporated several unique features, like incorporation of transcriptome data from next-generation sequencing increasing the total number of transcripts from 110 thousands to 230 thousands (represented by half a million to 11.8 million reads, respectively), and tools for comparative genomics, and drug target and vaccine prioritization like NemaSNP and HelmCoP.

In the future, we will expand in the following areas: (i) The Genome Institute current has a dozen parasitic nematode whole genome sequencing projects in progress and over 30 clinical strains of anthelmintics resistant populations (<http://www.genome.gov/10002154>). All the genome projects include RNAseq data from Illumina platforms, which offer a less-expensive way to get expression data from multiple stages and tissues and it is becoming an important and useful tool for studying nematode genomics. The obtained expression profiles yield a great deal of valuable information that can be used to understand important proteins involved in nematode development and parasitism. An example illustrating the utility

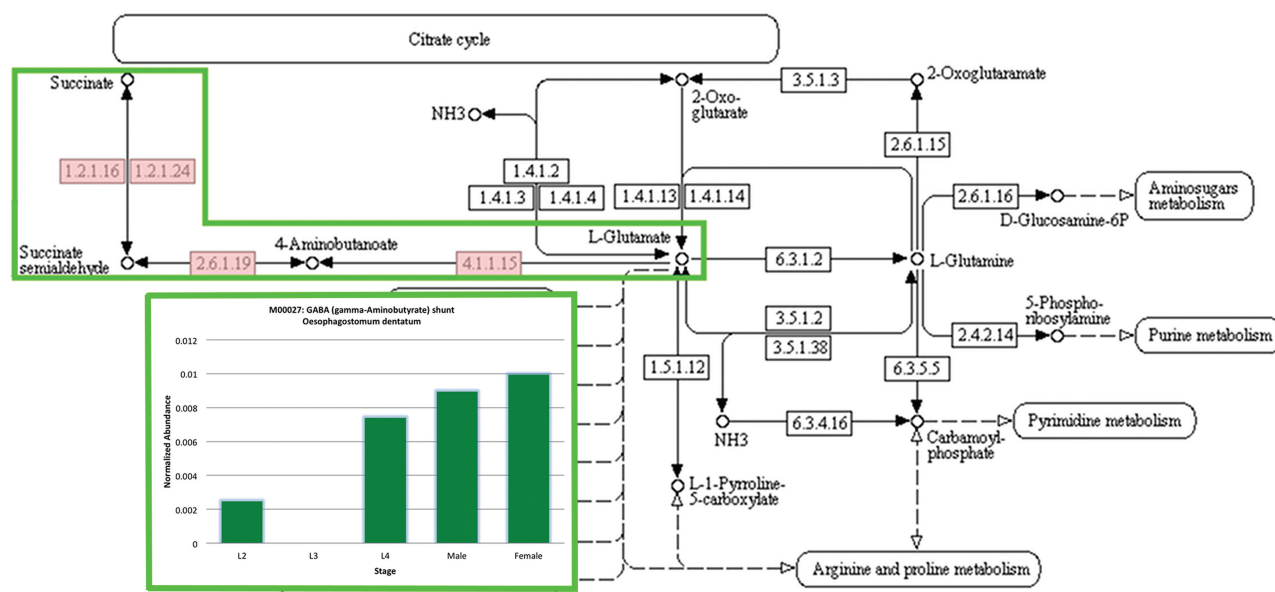


Figure 5. Stage-based metabolic profiling of *O. dentatum*. KEGG module M00027 (GABA (gamma-Aminobutyrate) shunt) from the pathway Alanine, Aspartate and glutamate Metabolism is shown with a plot displaying stage specific expression. This module was found to be significantly abundant in males, females and the L4 stage in *O. dentatum*.

of such data when implemented is shown in Figure 5, which represents a planned extension of our NemaPath tool. Expression data from L2, L3, L4, and adult male and female stages within *Oesophagostomum dentatum*'s life cycle have been sequenced at unprecedented breadth and depth at the TGI. The data was used to reconstruct the metabolic pathways and a binomial test ($P < 0.05$) was used to identify condition-specific modules and/or pathways that may play a role in parasitism. In Figure 5, the highlighted module, GABA (gamma-Aminobutyrate) shunt (M00027) is shown to be significantly abundant in males, females and in the L4 stage in *O. dentatum*. Using this pathway *O. dentatum* likely obtains glutamate from the host to produce succinate, which feeds directly into the Krebs Cycle. Further, leucine degradation (M00036, not shown) is almost significant in males and is significantly abundant in females. The worm probably takes leucine from the host, then breaks it down to acetoacetate. Acetoacetate is turned into acetoacetyl CoA, then into acetyl CoA and fed directly into the Krebs Cycle. Both modules have nearly full coverage in their respective stages. As more stage-specific RNA data becomes available, Nematode.net will host similar data for relevant species. (ii) We plan to merge data from NemaBrowse data with the NemaSNP data to offer a comprehensive collection of multiple data sources for robust analysis and visualization. (iii) Systems biology analysis will be incorporated into Nematode.net. As more protein-interaction network data comes available, we will map protein-protein interaction data from *C. elegans*, as well as orthologs, to various nematode genomes. MINT (28) and IntAct (29) are constantly expanding sites that provide experimentally based protein-protein interaction data. This will provide the community with invaluable systems biology data which

will yield insights regarding host-pathogen interactions and nematode biology. (iv) We plan to expand tools for drug discovery and comparative genomics by adding genomics tools to HelmCoP and providing chokepoint analysis on certain genomes. These new tools will make Nematode.net an invaluable tool for data acquisition and analysis within the nematode research community. (v) Finally, we plan to introduce a gene-centric search portal to the site that centralizes information about all NemaProt members and allows users interested in a specific protein to easily find all available analysis.

ACKNOWLEDGEMENTS

We would like to thank the numerous collaborators in the nematology community (nematode.net/collaborators.html) for providing worm material and the dedicated members of the production group at the TGI for the cDNA library construction and sequencing. We would like to thank Gaia Remerowski for technical advice and John Spieth, who heads the Wormbase efforts at the TGI.

FUNDING

Funding for open access charge: National Institutes of Health (NIAID AI081803 to M.M.).

Conflict of interest statement. None declared.

REFERENCES

- De Ley, P. (2006) A quick tour of nematode diversity and the backbone of nematode phylogeny. In: WormBook. (ed.), The *C. elegans* Research Community, WormBook, doi/10.1895/wormbook.1.41.1.

2. Coghlan, A. (2005) Nematode genome evolution. In: WormBook. (ed.), The C. elegans Research Community, WormBook, doi/10.1895/wormbook.1.15.1.
3. Hotez, P.J., Molyneux, D.H., Fenwick, A., Kumaresan, J., Sachs, S.E., Sachs, J.D. and Savioli, L. (2007) Control of neglected tropical diseases. *N. Engl. J. Med.*, **357**, 1018–1027.
4. Coles, G.C. (2001) The future of veterinary parasitology. *Vet. Parasitol.*, **98**, 31–39.
5. Chitwood, D.J. (2003) Research on plant-parasitic nematode biology conducted by the United States Department of Agriculture-Agricultural Research Service. *Pest Manag. Sci.*, **59**, 748–753.
6. Gilleard, J.S. (2006) Understanding anthelmintic resistance: the need for genomics and genetics. *Int. J. Parasitol.*, **36**, 1227–1239.
7. Awadzi, K., Edwards, G., Opoku, N.O., Ardrey, A.E., Favager, S., Addy, E.T., Attah, S.K., Yamuah, L.K. and Quartey, B.T. (2004) The safety, tolerability and pharmacokinetics of levamisole alone, levamisole plus ivermectin, and levamisole plus albendazole, and their efficacy against *Onchocerca volvulus*. *Ann. Trop. Med. Parasitol.*, **98**, 595–614.
8. Osei-Atweneboana, M.Y., Eng, J.K., Boakye, D.A., Gyapong, J.O. and Prichard, R.K. (2007) Prevalence and intensity of *Onchocerca volvulus* infection and efficacy of ivermectin in endemic communities in Ghana: a two-phase epidemiological study. *Lancet*, **369**, 2021–2029.
9. Li, B.W., Rush, A.C., Mitreva, M., Yin, Y., Spiro, D., Ghedin, E. and Weil, G.J. (2009) Transcriptomes and pathways associated with infectivity, survival and immunogenicity in *Brugia malayi* L3. *BMC Genomics*, **10**, 267.
10. Li, B.W., Rush, A.C., Jiang, D.J., Mitreva, M., Abubucker, S. and Weil, G.J. (2011) Gender-associated genes in filarial nematodes are important for reproduction and potential intervention targets. *PLoS Negl. Trop. Dis.*, **5**, e947.
11. Harris, T.W., Antoshechkin, I., Bieri, T., Blasiar, D., Chan, J., Chen, W.J., De La Cruz, N., Davis, P., Duesbury, M., Fang, R. *et al.* (2010) WormBase: a comprehensive resource for nematode research. *Nucleic Acids Res.*, **38**, D463–D467.
12. Elsworth, B., Wasmuth, J. and Blaxter, M. (2011) NEMBASE4: the nematode transcriptome resource. *Int. J. Parasitol.*, **41**, 881–894.
13. Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z. *et al.* (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, **437**, 376–380.
14. Kumar, S. and Blaxter, M.L. (2010) Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics*, **11**, 571.
15. Gish, W. *WU BLAST*. <http://blast.wustl.edu> (4 May 2006, date last accessed).
16. Martin, J., Abubucker, S., Wylie, T., Yin, Y., Wang, Z. and Mitreva, M. (2009) Nematode.net update 2008: improvements enabling more efficient data mining and comparative nematode genomics. *Nucleic Acids Res.*, **37**, D571–D578.
17. Mitreva, M., Jasmer, D.P., Zarlenga, D.S., Wang, Z., Abubucker, S., Martin, J., Taylor, C.M., Yin, Y., Fulton, L., Minx, P. *et al.* (2011) The draft genome of the parasitic nematode *Trichinella spiralis*. *Nat. Genet.*, **43**, 228–235.
18. Mitreva, M., Zarlenga, D.S., McCarter, J.P. and Jasmer, D.P. (2007) Parasitic nematodes - from genomes to control. *Vet. Parasitol.*, **148**, 31–42.
19. Brindley, P.J., Mitreva, M., Ghedin, E. and Lustigman, S. (2009) Helminth genomics: the implications for human health. *PLoS Negl. Trop. Dis.*, **3**, e538.
20. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. and Hirakawa, M. (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res.*, **38**, D355–D360.
21. Wylie, T., Martin, J., Abubucker, S., Yin, Y., Messina, D., Wang, Z., McCarter, J.P. and Mitreva, M. (2008) NemaPath: online exploration of KEGG-based metabolic pathways for nematodes. *BMC Genomics*, **9**, 525.
22. Abubucker, S., Martin, J., Taylor, C.M. and Mitreva, M. (2011) HelmCoP: an online resource for helminth functional genomics and drug and vaccine targets prioritization. *PLoS One*, **6**, e21832.
23. (2010) The Gene Ontology in 2010: extensions and refinements. *Nucleic Acids Res.*, **38**, D331–D335.
24. Hunter, S., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bork, P., Das, U., Daugherty, L., Duquenne, L. *et al.* (2009) InterPro: the integrative protein signature database. *Nucleic Acids Res.*, **37**, D211–D215.
25. Knox, C., Law, V., Jewison, T., Liu, P., Ly, S., Frolkis, A., Pon, A., Banco, K., Mak, C., Neveu, V. *et al.* (2011) DrugBank 3.0: a comprehensive resource for 'omics' research on drugs. *Nucleic Acids Res.*, **39**, D1035–D1041.
26. Von Samson-Himmelstjerna, G., Blackhall, W.J., McCarthy, J.S. and Skuce, P.J. (2007) Single nucleotide polymorphism (SNP) markers for benzimidazole resistance in veterinary nematodes. *Parasitology*, **134**, 1077–1086.
27. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, **25**, 2078–2079.
28. Ceol, A., Chatr Aryamontri, A., Licata, L., Peluso, D., Briganti, L., Perfetto, L., Castagnoli, L. and Cesareni, G. (2010) MINT, the molecular interaction database: 2009 update. *Nucleic Acids Res.*, **38**, D532–D539.
29. Aranda, B., Achuthan, P., Alam-Faruque, Y., Armean, I., Bridge, A., Derow, C., Feuermann, M., Ghanbarian, A.T., Kerrien, S., Khadake, J. *et al.* (2010) The IntAct molecular interaction database in 2010. *Nucleic Acids Res.*, **38**, D525–D531.