1) **Introduction**

Nematode.net was originally developed as a community resource for investigating gene sequences from nematodes at a time when the primary vehicle for this study was Sanger EST sequencing. The primary data hosted back then were EST contigs & the reads that comprised them (via FTP), along with some simple annotation based on homology.

Over the years Nematode.net has grown in the amount of data hosted, the types of data hosted as well as the breadth of analytical results available to users. But our goal to be a community resource has remained.

In the time since, Nematode.net has grown to host resources for over 50 nematode species (~56 I think), and has expanded its analytical repertoire from being simply a repository of ESTs to hosting the results from numerous analyses and providing a number of useful tools for comparative genomics studies and data mining. Only our goal to be a useful community resource remains unchanged. That said, we’re always interested in feedback. Please feel free to comment here, or email me (my email address is in the handout for the talk).

2) **Outline of Talk**

This presentation will provide an overview of the site, and present the various resources available within. This talk will cover:

**Data Navigation**: describe the 3 ‘entry-portals’ used to find site features and resources

**Site Overview**: describe the site in some detail, including how to use the various tools

**Hosted Datasets**: describe the resources hosted on the site

**Upcoming Projects**: a brief mention of upcoming data

The handout provided is indexed by slide number, and will provide URLs & instructions to assist you in following along on the site itself. That’s probably the best way to absorb this talk, to bounce around the site on your own, trying things out as I describe them. And please don’t be afraid to ask questions!

3) **Data** **Navigation: On-Ramps**

Data & resources within Nematode.net are accessed using 3 types of on-ramps. These portals provide an ‘entrance’ into the content of the site. Specifically these 3 kinds of entry portals are:

1. Organism-centric portals
2. Portals for data-mining
3. Data analysis & comparative genomics portals

Which on-ramp you take into the site depends on what information you are looking for, and what information you already have.

<READ FROM SLIDE>

4) **Data Navigation: Organism-Centric Portals**

The ORGANISM-CENTRIC portals are most useful if you want to explore all the available resources for a specific organism. An expandable menu of species ‘hubs’ will always be visible along the left side of each page. Clicking on a species will take you to a summary of information about that organism. Also listed will be available data and analysis for that organism. Links to data and/or analysis results are provided allowing you to jump from this page to whatever catches your interest.

5) **Data Navigation: Species Hubs**

The species hubs are the best way to navigate Nematode.net resources for a particular organism of interest. Available resources for the selected organism will have links to their respective pages from these hubs. It’s the best place to look for a top down view of what information is available for each species.

6) **Data Navigation: Data-Mining Portals**

DATA-MINING portals provide access into the large datasets hosted on Nematode.net. The primary DATA-MINING portals are:

**Transcript assembly repository** – Provides information about the worm cDNA assemblies we’ve made locally, as well as providing links for downloading the data

**NemaGene** – This is a database of genes & transcript clusters. It can be searched by stage (resulting in all isotigs/contigs containing any reads from the selected life cycle stage), isogroup/cluster name, gene/isotig/contig name or cDNA read name on a per species basis. Results are organized by isogroup/cluster, and provide available information for the resulting set of transcripts

Not yet integrated into NemaGene proper are the Gene Table resources we have for B.malayi. The pilot example of the **NemaGene Gene Table** shows the kind of information we plan to host in the NemaGene database later this year. The NemaGene Gene Table for B.malayi is a significant resource in itself, and offers a great deal of information on the B.malayi geneset

**NemaBLAST** – Offers the ability to align your sequence(s) of interest against our various transcript database collections.

**NemaBrowse** – NemaBrowse provides an annotated view of locally assembled worms using the GBrowse interface

**NemaSNP** – This is a stand-alone tool for annotations revolving around population genetic studies. While this will probably be wrapped into NemaBrowse with our impending site update, it currently hosts variant calls from populations of worms against a transcriptomic reference

**NemFam** – This is a collection of nematode related, homologous peptides that were not (originally) represented in Pfam. The NemFam interface allows the exploration of these conserved regions by clade, species, specific group name, contig or read name. Information on secondary structure, domain organization, signal peptide presence & transmembrane information across the NemFam groups is presented via GBrowse views.

**Codon usage information repository** – Our repository hosts codon usage tables for over 30 nematodes, available for download as excel files or in text-based tab-delimited format

**EST Cluster data repository** – This is a storehouse of information gathered from our Sanger EST assemblies. We host cluster & contig information, lists of component EST reads as well as protein translations of each cluster for over 30 nematodes that still represent the only available view of the transcriptome for some of these organisms

**Gene Expression Data** – We also host expression data, both microarray & RNAseq –based, in our Gene Expression Data page. Expression data is available for download along with experimental information and supplemental files as provided by the contributors

7) **Data Navigation: Data Analysis & Comparative Genomics Portals**

These portals offer the ability to make custom queries against some of the resources hosted on Nematode.net. These tools require some familiarity to use well, but offer the most advanced views of hosted data that we provide.

**NemaPath** – The NemaPath viewer allows users to explore & compare KO based annotations across species, hosts, clades or stages. Results are presented graphically, using pathway maps provided by KEGG. Enzyme nodes are painted according to the comparison being made, and further information per node is made available on mouse-over

**HelmCoP** – HelmCoP (Helminth Control and Prevention) provides a repository for storage, annotation and comparative genomics data on helminth proteomes. This resource facilitates drug and target discovery for parasitic helminthes which can be used to prioritize drugs, pesticides, and vaccines, as well as enabling researchers to compare hosted nematode genomes. Functional, structural and comparative genomic data from plant, animal and human helminths, as well as hosts and model organisms are hosted and query-able within the HelmCoP resource.

8) **Site Overview: Home Page**

Now I’m going to spend some time describing how to interact with Nematode.net. This section of the talk will present the interfaces & views you’ll see as you move around the site.

When you first arrive at <http://nematode.net>, you land on our home page. Primary navigation within the site is handled by using the ‘Analysis navigation menus’ along the top of this page or the ‘Species hub menus’ along the left-hand side. Those 2 banks of menus, along with our news banner & community resource menus, will always be available to you from wherever you are in Nematode.net.

The ‘Community resources’ menu includes the ‘COMMUNITY’ link, which will take you to our forums, wherein you can post questions about the site, or data contained within the site. The ‘EDUCATION’ menu currently leads to a page hosting the ‘Introduction to Nematodes’ teaching package, put together by McGawley et al. This package includes a quicktime video presentation, as well as over 100 slides meant to be used as a teaching aid for students new to the field of nematology. The ‘FAQ’ menu lists 3 FAQs: ‘Nematode.net FAQ’ (which is a general FAQ about the site and its hosted resources), ‘NemaGene FAQ’ (which is a FAQ about the NemaGene data collection) and the ‘HelmCoP FAQ’ (which holds information about HelmCoP and how to use the query interface for that resource).

In the center of the page, just below the mission statement is a link to the current citation for Nematode.net. Immediately beneath that is our PROJECT STATUS table, which lists the current progress on parasitic nematode genome sequencing projects at both the TGI and the Sanger Institute Pathogens Unit. The species names in that table are links that jump to summary pages further detailing each project.

Lastly, towards the bottom of the page (off the screenshot shown in the slide) is a listing of already completed nematode genomes, with links to their project pages around the net.

9) **Site Overview: Home Menu**

Under the HOME MENU, we have a number of ‘housekeeping’ items including summaries of our nematode sequencing efforts across various platforms, publication, grant, collaborator & staff lists and other useful but not necessarily exciting bits of information.

Of special note is the ‘Data Download’ link, which takes you to a page on which we host supplemental data for ourselves and our collaborators, as well as various files that for whatever reason we deem important enough to be made public, but that don’t quite fit in other sections of the website.

10) **Site Overview: Species Hub**

As I touched on a minute ago, one of the central pillars of Nematode.net are the species hubs. These pages are always a click away, and can both summarize whats available for your worm of interest, as well as get you started exploring a particular set of analytical data.

The upper half of the page typically shows a picture of the worm, and some basic information such as common name, clade (**NOTE:** *as defined by the Blaxter et al. 1998 paper*), primary project contacts and a summary of disease(s) the worm is involved in. Note that the pictures are not all open source, we got many of them from NemaPix volume 1 from Mactode Publications. If obtained from an external source we credit the image in the picture frame. If you’re unsure about whom to contact but want to use a specific image, you can just contact me and I’ll point you in the right direction.

The lower part of the page displays the available resources for the chosen organism. Links only appear if the worm you are viewing has data of the type listed. For some analytical results a summary may be displayed right on this page, but for most you’ll be provided a link to the proper section of Nematode.net to view those results.

Finally, we provide links to external resources for the organism as well. We try to provide appropriate links into WormBase, Sanger, NemBase4 as well as to the NCBI Bioproject page when available.

11) **Site Overview: NemaGene Cluster Search**

NemaGene is a collection of clustered transcripts & genes. The NemaGene Cluster Search page allows users to query this database and return lists of these transcripts & genes, organized by cluster or isogroup. Note that much of the access into NemaGene comes from other integrated applications within Nematode.net. For example contigs displayed within the NemaPath tool link out to basic contig information hosted within the NemaGene db.

The NemaGene Cluster Search page is the place to come when you have identified a contig, isotig or gene from one of the other Nematode.net resources (eg. a pan-phylum NemaBLAST result or a cluster of interest from one of our transcript assemblies you’ve downloaded and analyzed) and you want more detail on that sequence entity. This tool also is useful in identifying a stage specific set of isotigs/contigs for a given organism using the ‘Stage’ search selection

Present towards the bottom of the NemaGene Cluster Search page, beneath the interface form, are tables summarizing the numbers of clusters/isogroups, contigs/isotigs and cDNA reads used in the transcript assembly of each listed species divided up by sequencing platform. It also displays the number of Sanger ESTs present in dbEST as of the listed date to give visitors an idea of how complete hosted Sanger-based EST assemblies are compared to the currently available dataset.

12) **Site Overview: NemaGene Cluster Search Interface**

Queries are constructed using this simple form, in which you specify the species you want to query against, the ‘Search Type’ and the search string itself entered in the bottom field. Note that some queries that return long lists of results may take some time to run.

Above the search form are a number of examples of information that is appropriate for each search type, and can be used to help familiarize yourself with the form.

13) **Site Overview: Querying NemaGene Clusters**

After submitting your query, results will be returned to you in the form of a list of genes, isotigs or contigs (depending on the available data for the organism you’re querying). These will be organized by cluster (or isogroup for 454 cDNA assemblies).

Most results are specifically generated based on the list of genes/contigs/isotigs returned by your query. Unless you are specifically searching for clusters or isogroups by name, the tool will only report the exact slice of data (genes/contigs/isotigs) you requested, organized by cluster or isogroup. This means that while you will see the cluster or isogroup name in your return list, the displayed data (genes/contigs/isotigs) under that group is not necessarily the FULL list of all genes/contigs/isotigs that belong to that group. To see EVERY member of a cluster or isogroup, you need to specifically search by cluster/isogroup (set the ‘Search Type’ to cluster/isogroup).

From your return list, select either the ‘group link’, or one of the component gene/contig/isotig links to get to the information display for that entity. If you selected the ‘group link’, you’ll get output displaying all the data (genes/contigs/isotigs) reported in your return list for that group. (which means not necessarily ALL members of the group…only the members defined by your query… Remember that if you want to see everything under a group you need to re-query using the specific group name)

The information display for each gene, contig or isotig shows the nucleotide sequence & protein translation (if available), and shows all the libraries of reads used in the construction of contigs or isotigs. It also reports the number of reads supporting the current entity.

If your query is returning actual genes, additional information may be available such as IPR ids, KO annotations, GO classification and gene description.

14) **Site Overview: NemaGene Gene Table**

The most recent addition to Nematode.net is the inclusion of NemaGene Gene Tables. Currently populated with data from B.malayi, these tables are collections of various annotations on final genesets. Available annotation can vary, but in general will provide clustering information, KO, IPR, GO annotations, expression data and more.

The first page of the NemaGene Gene Table application simply lists all available final gene sets.

15) Site Overview: NemaGene Gene Table Data

Gene annotations are displayed in a sortable table format within a scrollable viewing area. All genes are listed vertically along the page (you need to scroll up-down to see them all), with assorted annotations listed in the columns along the top. Depending on how many annotations were provided with the geneset being displayed, this table may span several page widths. The table is sortable on each column by clicking on the column header.

Its important to note that the table may feel unresponsive when first loaded. If you allow the page to fully load before trying to navigate the information it will scroll & sort much faster.

16) **Site Overview: NemaGene Gene Table Legend**

If you click on the “**Table legend**” link at the top of the table, you’ll open a separate browser window summarizing the various columns present in the Gene Table you are viewing. As I mentioned before each project can potentially have its own unique set of annotations, but the legend displayed will be applicable to the dataset you are viewing.

In this case the table shows these annotations:

<READ FROM SLIDE>

17) **Site Overview: NemaGene Gene Table Display**

I just want to stress again that to see all the data for most annotated genesets you will need to scroll horizontally. For B.malayi, the data table spans more than 3 full page widths.

18) **Site Overview: NemaPath**

NemaPath is a tool for visualizing the presence, absence and composition of enzymatic pathways in given organsisms. It also supports comparative views between organisms, clades, hosts and stages (although the stage viewer is currently in the pilot stage). This allows users to visually see & explore enzymatic pathway differences between these entities based on actual transcriptomic data.

This is useful in many ways, from identifying potential drug targets, to helping understand the differences between species utilizing different survival strategies. Browsing clade or host –specific comparisons can offer insight into taxa defining differences, while the ability to directly compare any two species provides a visualization of enzymatic differences between the two.

I’ll provide a description of how clustered transcripts are mapped into the enzymatic pathways a little later. For now I will just describe the basic navigation through the tool.

19) **Site Overview: NemaPath View Selection**

The first step in using this tool is to select which view you want to use:

**Species-specific comparisons** – Allows for the viewing of the transcript supported pathway composition of a single species, and also allows the comparison of 2 species

**Clade-specific comparisons** – Shows the comparison of transcript supported pathway composition across the 4 clades I, III, IV & V

**Host-specific comparisons** – Shows the comparison of supported pathway composition across host type (plant, animal and free living)

**Stage-specific comparisons** – Allows for the comparison of pathway composition supported by transcripts from 2 specific life cycle stages

20) **Site Overview: NemaPath Species Comparisons**

To view species-specific comparisons, the first task is to select the baseline species to display. After clicking the ‘Species-specific comparison’ link you’re brought to this page, where you select the organism you want to start with. The selection table also includes the # of genes, contigs or isotigs used to construct the pathway views for each species in the list.

21) **Site Overview: NemaPath Species Cutoff Selection**

After choosing the first species you want to explore, you will arrive at a page showing the distribution of mapping scores your species’ collection of genes or transcripts made when aligned to the KEGG genes db. On this page you need to set the e-value cutoff you want to apply to the mappings. Setting a weaker e-value will increase the sensitivity of the mapped KOs, recruiting more mapped genes or transcripts into the map view. But at the expense of more false positive mappings. Setting stronger e-value cutoffs will reduce false positives at the expense of sensitivity.

You can use the distribution to assist you in choosing a cutoff that balances KO mappings with accuracy. In this example I’d probably go with ~1e-05 (i.e. enter ‘5’ in the text field and submit). The next page will be a confirmation of your entered e-value. You can hit submit again to continue, or reset the value and select again

22) **Site Overview: NemaPath Pathway Selection**

This page is where you select the initial KEGG pathway you want to view. Our current implementation supports pathways from the 4 major categories:

Metabolism

Genetic Information Processing

Environmental Information Processing

Cellular Processes

Choose from the menu which specific pathway you’d like to view first. Note that you will be able to jump pathways & select new species (and species to compare against) from the visualization page as well.

23) **Site Overview: NemaPath Pathway Visualization**

This page displays KEGG’s map of the selected pathway and paints the enzyme nodes in deepening shades of green depending on how many of your selected species’ genes/transcripts map to KOs associated with each node.

Mousing over a node will display a table with information about which genes/transcripts hit that node, what KO was hit that spawned the assignment, the strength of the hit and so forth. Clicking on the KO number will navigate to the KEGG Orthology entry for it. The subject link takes you to the KEGG genes entry in KEGG’s GenomeNet, and the query name itself is a link into NemaGene, where you will find the translated sequence and other information.

In the upper right corner of the page is the comparative interface. The upper dropbox allows the user to flip between pathways without backing out to the pathway selection page. And beneath that is another dropbox menu allowing the user to select a 2nd species for a comparative display.

24) **Site Overview: NemaPath Pathway Comparative View**

When 2 species are selected, the genes/transcripts of both are painted onto the map according to their KO assignments. The yellow/blue color scheme is assigned to the ‘current’ (first species chosen when you entered the map view) and the ‘second’ species (the species you chose in the drop menu for comparison).

Mouse-over of a node will now show the assigning hits from both organisms, with all the same link-outs available.

Choosing a new pathway will maintain the currently selected species. And you are free to choose a different comparative species at any time. Your current base species, the second, comparative species, the pathway & the e-value cutoff selected are displayed in the upper right corner of the page.

To reset your view back to a single organism you’ll need to back out of the page and reselect a primary species from the selection page.

25) **Site Overview: NemaPath Pathway Clade View**

Selecting the Clade-specific comparison view summarizes all gene/transcript hits per clades I, III, IV & V. For this view the hit cutoff is preset to 1e-05, and enzyme nodes in the pathway map are painted if the node is touched by at least 1 query from the given clade. Mouse-over displays the same information as before, now grouped by clade. All the same link-outs are available.

26) **Site Overview: NemaPath Pathway Host View**

Selecting the Host-specific comparison view summarizes the gene/transcript hits based on the class of host each organism uses. Organisms are categorized as having a plant host, animal host or being free-living. The view itself functions in the same fashion as the previous comparative views.

27) **Site Overview: NemaPath Pathway Stage View**

The NemaPath Stage-specific view summarizes the populated nodes by life cycle stage. This view functions the same as the others, but with transcript data aligned & painted onto the KEGG map by stage. Notice that the breadth & depth of coverage that the NemaGene transcripts achieves across the mapped KO is displayed in the mouse-over for this view.

Currently this resource is in its pilot stage, and we have only one dataset loaded, that being the L3 vs. Adult stages of Necator americanus. But we plan to expand this resource with our impending site update.

28) **Site Overview: HelmCoP**

HelmCoP (Helminth Control and Prevention) is a database of integrated functional, structural and comparative genomics data from plant, animal & human helminthes, as well as model organisms & several host organisms, all capped by a query interface that allows users to ask complex questions of this data. HelmCoP’s primary goal is to assist researchers in the process of identifying drug, pesticide & vaccine targets in helminthes.

HelmCoP has the versatility to enable users to search for drug targets for specific helminthes or for a group of helminth pathogens, and also to allow the user to search for broad-spectrum drug targets that span multiple species.

A summary of what data is available for which of the included organisms is shown towards the bottom of the HelmCoP splash page, as well as on each of the query interface pages. In general, the host organisms are present only to help characterize orthologous groups. Most Helminth species have information of every type we collected.

Querying the HelmCop db is done using one of two forms. One for users interested in building gene-based custom queries, and the other for users wanting to search the database using orthology based queries. The choice of which form to use really depends on what you want to do. If you just want to retrieve information about a specific gene of interest, the gene-search page is for you. If you are more interested in casting a net into HelmCoP and fishing for genes that meet your criteria, the ortholog-search page would probably be better for you.

29) **Site Overview: HelmCoP Gene Search**

This is the top of the form for building gene-based queries. This capture doesn’t include the filters & output settings that appear a little further down the page.

If you have a specific gene of interest, and are using the HelmCoP interface to mine additional information about that gene, you can enter the gene name in the provided box. Note that entering a specific gene name short circuits any other filter selection you may have applied on this page. If you request a gene, you are given that single gene as your output set. Output options still function normally though.

If you do not have a specific gene in mind, but are interested in genes from a specific set of organisms, select the species you want to include in your output in the ‘Species’ section.

Note that you are allowed to set additional filters on your query further down the page, and I’ll show those filters in a minute. This upper section just defines the cosmos of genes that the filters will be applied to.

Be aware that due to the size of the expected return data sets that can be defined in this form, that the HTML view of your output will be truncated at 20,000 rows. Your full output will still be available as a tab-delimited text file available via a button on the results page, but the HTML will never display more than 20k rows.

30) **Site Overview: HelmCoP Ortholog Search**

If you are building an ortholog based query, this form is the one you’ll be using. If you know the name of the orthologous group you are interested in, you can enter it directly in the ortholog name field. As before, setting an explicit ortholog name short-circuits any other filters you may have applied on the page, and returns your selected output fields for the exact ortholog you requested.

If you don’t have a specific ortholog in mind, then you’ll need to define the initial set of orthologs that your filters (set later) will apply to. But in this case, unlike the gene search, you are setting organisms to be included or excluded in the orthologs returned by your query.

Each species can be set to ‘include’, ‘exclude’ or ‘Do not filter on this’, with the default state being not to use the species in the filter. Your query will return all orthologs in HelmCoP that have at least 1 gene from any of the ‘included’ species. But, the presence of 1 gene from any of the ‘excluded’ species will remove that entire ortholog from your return set. Note that you must select at least 1 species to be ‘included’.

Another key difference between the results you get from an ortholog based search and a gene based search is that due to the behavior of the SQL ‘limit’ statement, which is used to truncate results to the first 20,000 rows for display in HTML, if your output dataset DOES get truncated, you are NOT guaranteed to consistently have every member of filter-passing orthologous groups reported in the HTML output. In order to see all members of all orthologous groups returned by your query, you’ll need to download the full results text file. Of course if your query returns fewer than 20k rows of output, all members of each orthologous group will be properly displayed within the HTML view.

31) **Site Overview: HelmCoP Filters & Output settings**

For both the gene & ortholog based searches, after defining the set of species to query, the next task is to apply desired filters to limit your output down to only those genes or orthologs that interest you. This panel shows all the filter settings available in HelmCoP, as well as the available output columns you can request.

The **Functional** filters allow you to set specific GO, KO and/or IPR ids that you require to be present in returned genes, or at least in one of the members of returned orthologous groups.

The **Essentiality** filter screens your return set by inferred RNAi phenotype. This filter will require your returned genes to have evidence supporting the essentiality you chose, or for orthologous groups, at least one member of the group must show the phenotype.

The **Structure** based filters allow you to limit your return set by requiring them to have shown sequence similarity to the PDB id you enter. This section also allows you to filter your results based on the presence or absence of a detected signal peptide.

The **Drugs** section allows you to filter on genes with homology to targets in Drugbank. It also allows you to filter based on whether or not your returned gene (or at least 1 of the genes per each returned ortholog) is considered a ‘Hopkins druggable target’.

The **Vaccine candidates** section allows the user to filter based on the presence of various structural based hints that may imply epitopes that are vaccine candidates. You can pick and choose specific traits, or just turn on the ‘Show all vaccine candidates’ switch to apply them all.

Finally, the last thing the user needs to set are the **Output options**. These allow the user to customize the output to include only information they are interested in. Some columns will always be returned, such as gene and ortholog name, as well as species of origin. But otherwise the user needs to select the information they want reported.

Be aware that given the size of some return sets, selecting many or all of the available outputs can cause your query to take a long time to build. In some cases this may even cause your website to time out before the information can be fed back to your browser. When you first submit your query we do run a query manager that tries to estimate if your query is complete-able within an amount of time that should not time out your browser, but the manager is not infallible. If the manager deems your query would not complete it will let you know, and suggest which of your options are the most time-intensive, allowing you to reconstruct your query with fewer outputs, or by defining a smaller starting set to look through (i.e. choosing fewer species or a more restrictive set of orthologs in the species section). Typically if you construct a query request that is too slow, you can run the tool in 2 parts, funneling half your outputs into each part to end up with everything you were wanting.

32) **Site Overview: HelmCoP Results**

This is what the HTML results of a HelmCoP ortholog search query looks like. The output table is gene-based, with ortholog information provided as well for ortholog-based searches, or if the user requests orthologous group annotation for their gene-based search.

Every requested output will be a column, so requesting many outputs can result in your data spanning multiple page widths.

Where appropriate, the HTML view provides link-outs to the various resources each output type is based upon. And for orthologous group output, it tries to aggregate cases where multiple members of an ortholog report the same information.

Towards the top of the page you will be shown the query filters & output requests you specified that resulted in the output table you see. The button for accessing a full, tab-delimited text version of your output table is also at the top up there.

Again I want to stress that the HTML table you see will be truncated if your query returned more than 20000 rows. For a gene-based query this is a simple truncation, but for ortholog based requests this can result in some reported orthologs not displaying every member in the viewable area of the HTML table.

To be safe its always best to grab the full text download and parse the results on your end.

33) **Site Overview: HelmCoP BLAST**

One additional utility we’ve made available to help with mining the HelmCoP db is our HelmCoP BLAST service. This page allows the user to enter nucleotide or protein sequence and run wu-blast (blastx or blastp) against the HelmCoP collection of proteins. You can select which species you want to search against, as well as set whether or not to apply a complexity filter or repeatmasking. Results (in default blast output format) are mailed directly to the user.

34) **Site Overview: NemaBLAST**

In the same vein as HelmCoP blast, Nematode.net also hosts a WU-BLAST service useable against all our other transcript & gene data collections. This service uses WU-BLAST 2.0 for alignments.

The subject databases are divided for convenience into 2 collections. You can target your BLAST again:

**EST reads grouped by library:** This database represents all the EST reads produced for species we’ve sequenced. The reads are organized by species & library, allowing the user to mix & match the reads that will serve as the subject.

**Transcript contigs, isotigs & genes:** This database is used to screen for the presence or absence of your query across all the species hosted on Nematode.net. This database comprises EST assembly contigs, gene CDS and/or transcript assembly isotigs organized by species.

35) **Site Overview: NemaBLAST vs Reads**

To align against EST reads grouped by library, enter your sequence in the text field at the top of the screen, then check the boxes next to the species you wish to include in the alignment. This first page only allows you to select entire species, you can filter on specific libraries on the next page. The buttons on the right side of the page offer convenient ways to select entire clades at once.

After entering your query sequence, and making your choice of species, the “Build BLAST Query Page” button will bring you to a 2nd form in which you can specify the exact libraries you want to include. By default, all available libraries for the selected organisms will be included, but you can freely click on or off whatever you like.

Here you can also select between a blastn alignment & a tblastn translated alignment, as well as choose to RepeatMask the query. You need to enter an email address to receive results, then click ‘BLAST Search’ to launch the job.

36) **Site Overview: NemaBLAST vs Transcripts & Genes**

To align against EST contigs, genes and/or transcript assembly isotigs, enter your sequence in the text field at the top of the screen, then select the specific datasets you want to include in the subject database. The image here shows a truncated view of the page, but all clades & available organisms are listed on the site itself. The actual geneset resources are noted in bold. Note that transcript assembly isotigs and sanger EST contigs can be discerned in your output by subject name. Transcript assembly isotigs are annotated in the output with a 4 letter prefix specific to each species. Likewise gene CDS can be identified by the prefix CDS\_ and then the 4 letter species code. These prefixes are displayed in the NemaGene FAQ, which is accessible from anywhere in the site via the left navbar menu.

On this page you also need to set the flavor of WU-BLAST to use (blastn or tblastn), and whether to apply a complexity filter or RepeatMask your query. You also need to provide an email address for the results. Hit the ‘BLAST Search’ button to launch your job.

37) **Site Overview: NemaBLAST Request Status**

After launching your blast search, the site will return a request status page letting you know your job has been submitted. The status page summarizes your alignment request, and in the event of a problem with your request it would give details on the problem.

38) **Site Overview: NemaSNP**

The NemaSNP resource hosts SNP loci annotations made by mapping distinct populations of cDNA reads back to specific worm transcriptomic assemblies. Comparative annotations are then displayed using GBrowse, providing information such as the synonymous or non-synonymous state of each loci for each population, or snp enrichment in specific pathways, or etc…

Use of the information provided in this resource will vary depending on the kinds of annotation for the reference and populations of query data. This tool is meant to highlight genes or transcript contigs/isotigs that differ between distinct populations of worms, for example a drug resistant vs. normal population.

Currently only the pilot project data has been loaded, but we have more data on the way.

39) **Site Overview: NemaSNP Region Selection**

After selecting the project to view, you are next brought to a page on which you will select a specific region on the reference to view. The exact form of this table can very with each project, but generally the reference will be navigable by some sensible collection of regions, and some metric will be provided to help discriminate regions.

In this case we’re comparing an inbred & field strain of T.circumcincta, and each contig represents a putative transcript. So the pieces are organized by total number of detected loci per reference, with the counts of inbred vs field originated snp loci reported on each line.

40) **Site Overview: NemaSNP Detail View**

After selecting a region to view, you’re brought to a GBrowse display of that region. The exact layout of this image depends on the annotations made in the project. In this example, our pilot data is annotated by 2 populations of T.circumcincta, an inbred and a field population. In addition to some ‘freebie’ content we get just from loading sequence, we’ve built a track showing the SNP loci and their synonymous, non-synonymous or non-coding position states, as well as labeling them according to their origin from the inbred or field populations. We’re also showing the read alignments for each population of worm in separate tracks.

41) **Site Overview: NemaBrowse**

While Nematode.net’s primary focus has been providing navigation of nematode transcriptomic data, we also provide genome annotations from draft assemblies that are part of the Parasitic Nematode Genome Project. Note that we are not intending to become a repository for model organisms, WormBase already has that covered. We just wanted a place to host select parasitic nematodes & their annotations that we think may be useful to nematologists asking certain kinds of questions. With our upcoming site update we plan to merge NemaSNP data into this repository which will offer interesting comparisons between different isolates. Getting our initial variant data into NemaBrowse will require the mapping of transcript contigs onto genome references so that variant loci can be forwarded, but this is really just an intermediate solution. Once good genomic references become available, snps will be recalled on full genomes. We will also be adding Rfam predictions & gene-calling evidence into the mix.

To navigate this resource, just select the genome you want to explore and click the GBrowse annotation link.

42) **Site Overview: NemaBrowse Gene View**

After selecting the genome to view, you are presented with a table that guides your selection to a predicted gene. The entire draft genomes are available within GBrowse, this table simply directs your entry point onto a specific gene. Once in GBrowse you can navigate normally.

The GBrowse view itself currently provides fairly standard gene annotations. But more information will be available soon, as will more annotated genomes.

43) **Site Overview: GO Associations**

GO classifications assigned by homologies identified using interproscan are available for most of our transcript resources. AmiGO provides a means to identify these resources available for a given biological process, cellular component or molecular function of interest.

First select the organism you want to explore.

44) **Site Overview: GO Associations AmiGO**

You’re then presented with an AmiGO overview showing the number of transcripts annotated under each GO category. Categories expand to display child categories along with their transcript assignment counts.

The pie-images next to each GO category are clickable, and expand into views showing counts in all child categories beneath the level at which you clicked.

45) **Site Overview: GO Association Term Search**

If you have a specific GO term of interest, and want to explore the transcripts assigned to that term for your selected species, enter the GO id into the “Search GO” field, leave the ‘Terms’ box checked, and click.

You’ll be presented with the view shown in the top panel here. Click on the ‘GO Term’ to open a new window listing all the transcripts assigned to this term for your chosen species.

Note that had you directly clicked on one of the GO terms on the first AmiGO page it would also have brought you to this same view.

46) **Site Overview: NemFam**

NemFam is a database of homologous, nematode-derived families of peptides that were not, when the resource was originally built, represented in Pfam. The original NemFam set included families built from over 214,000 peptides from 32 nematode species (27 of which are parasitic) clustered using MCL.

The NemFam pages provide an interface for exploring this data. Typical users either start with a known transcript assembly contig they are interested in, or use the ‘Advanced search’ link to identify a set of NemFam families of interest to them. This resource provides structural information for each family viewable through a GBrowse interface.

Since the construction of the original NemFam collection in 2009, several additional sets of NemFams have been built, but not yet been loaded into this repository (…such as a set of conserved Nematode families from *full length genes* built from B.malayi, M.incognita, M.hapla and T.spiralis, or the HelmCoP orthologous groups based on 23 species). These additional datasets will be versioned when they do get added into the repository. Old versions will remain available for researchers that need them.

47) **Site Overview: NemFam Advanced Search**

A commonly used portal into the NemFam repository is the advanced search form.

This form allows the user to select any combination of species included in the original NemFam collection and produce a table of all NemFam families that contain at least one member from every species selected.

Once you’ve selected the organisms to recruit, you’re presented with a table listing all the NemFam families meeting your requirements (i.e. families with at least one member from every selected species). This table also indicates features detected for each family, to give the user an idea of which families may be most interesting to them. Click on one of the family links to get to the NemFam View.

48) **Site Overview: NemFam View**

Annotated structural features are displayed via GBrowse for the family you selected. The references from each NemFam family are artificially constructed spacers, so normal GBrowse navigation through the data is not possible. Instead its easiest to fall back to the group selection page to move to a different NemFam.

The NemFam view shows the alignment of family members within each group, and offers structural information divided into 3 tracks:

**Secondary structure**:

Gray – region of disorder

Yellow – beta sheet

Green – alpha helix

**SignalP & transmembrane**:

Pink – transmembrane region

Purple – signal peptide

**Domains**:

Red – predicted domain

White box with cross – this is a gap in the alignment to the family

Additional information about the NemFam group itself is provided in the header. This field displays the NemFam group ID, # clades spanned by group, # species spanned by group & the # of group members

49) **Site Overview: Transcript Assembly Data**

This page hosts our transcript assemblies. Expanding any of the rows provides information on the numbers & platform types of reads used in each assembly. Also shown are the numbers of isotigs (putative transcripts), isogroups (groups of isotigs putatively representing all the expressed isoforms for a gene locus) and numbers of reads per stage if that information is available.

For each assembly download links are provided for:

**Isotig nucleotide fasta** – Isotigs refer to alternatively spliced isoforms of genes

**Isotig protein translations** – These are protein translations (typically made using prot4EST) of the isotigs produced by the assembler

**Isogroup membership file** – Isogroup refers to the grouping of isotigs that putatively represent multiple isoforms for a gene locus. This file lists the isogroups predicted by the assembler, and provides the list of member isotigs per isogroup.

**Read membership file** – The read membership file lists the read members for each isotig in the isotig file.

Note that our currently hosted transcript assemblies are built from 454 data using the newbler assembler, which generates isotig & isogroup information.

50) **Site Overview: Isotigs & Isogroups**

For transcript assemblies, assemblers need to deal with the fact that a single gene locus can produce alternatively spliced transcripts. In an assembly, these isoforms can share ‘contigs’, but also have deletions or additions of contigs that make them unique. The newbler assembler we used for our transcript assemblies builds isotigs (which are representations of putative isoforms) for every isoform it believes to exist for a gene by stitching together component contigs based on evidence it sees (such as specific breaks in the alignment of reads that imply a missing contig).

Isogroups are simply the collection of isotigs that putatively belong to the same gene locus. Collectively the Isogroup represents the set of observed transcripts for the gene.

51) **Site Overview: Codon Usage Tables**

Codon usage tables are available both in excel format and as tab-delimited text files on this page.

The upper table provides links to codon usage frequencies calculated locally, using transcript contigs & isotigs available in NemaGene as of Aug. 2011.

The lower table hosts codon usage patterns generated via methods described in the Mitreva et al. 2006 Genome Biology paper whose link heads the table. The files offered in the upper table are the more current and should be used preferentially. The information from the 2006 paper is kept mainly as an archive version, although a few species only have tables available from that older calculation.

52) **Site Overview: EST Cluster Data**

This page hosts Sanger EST based contig & cluster designations. Download links are available for:

**CLUSTER\_HISTORY files** – These files list each EST contig and its EST members in a text format

**CLUSTER\_SUMMARY files** – These provide the nucleotide fasta sequence of each EST contig

**Cluster\_Groups files** – These files provide EST contig clustering information. These files list the EST contig members of identified cluster groups for each species. Analogous with isotigs & isogroups, EST contigs represent single putative transcripts, whereas the ‘Cluster\_group’ represents the putative collection of transcripts for a single gene locus

**Translation files** – These are protein fasta files translated from the EST contigs using prot4EST

53) **Site Overview: Intestinal Transcriptome Data**

This section of Nematode.net defines a collection of nematode intestinal genes from the worms Ascaris suum, Haemonchus contortus and Caenorhabditis elegans. This data was gathered during the course of an analysis project by one of our lab alumni, and it proved useful enough that we decided to host it as a unique data collection. In addition to the ~10,000 genes identified across the 3 species, a core set of ~2k genes grouped into 241 families was also identified and is hosted here.

These datasets have been assigned GO categorizations, visible via the ‘AmiGO’ links, and have also been mapped to the KEGG genes db (v39, which was the current version at the time the mapping was done), with the KO mappings available in a downloadable spreadsheet.

Note that the spreadsheet & AmiGO browser only provide the names for these Intestinal transcripts (represented by Sanger EST contigs). To access their sequence data you’ll need to search for these names in NemaGene, or download the full EST contig dataset from the EST Cluster Data section and parse out the specific subset of contigs defining the Intestinal data.

54) **Site Overview: Gene Expression Data**

This section hosts the results of a number of Microarray & RNAseq-based gene expression experiments either performed in our lab, or in the labs of our collaborators. Where possible we try to host experimental results with the goal of meeting MIAME standards, although obviously some of the component files don’t exist for non-microarray based experiments. When data becomes available in a more formal repository we do try and update this resource with the appropriate link to that repository.

55) **Hosted Datasets: Summary**

For the next part of my talk I’m going to describe in more detail the datasets hosted on Nematode.net. While all of these resources have been touched upon during the site overview, I’m going to focus now on the data itself, and less on the web interfaces. I’ll define how the data was generated & annotated, and will give a few example publications that made use of the data.

This image is a summary of the main repositories hosted on our site, along with their respective portals for exploring & accessing the data.

56) **Hosted Datasets: NemaGene**

The NemaGene resource is collectively the sum of all our hosted gene & transcript data. Three types of data is currently recruited into NemaGene:

-Sanger EST contigs grouped into clusters

-454 cDNA isotigs grouped into isogroups

-Gene CDS from finished genomes

For some of these worms, this data represents the only view of the transcriptome that will be available for some time. At least until full genomes can be sequenced, which is usually limited to the most prominent model organisms, and those with high impact on human and/or livestock health

**Sanger EST contigs** were built by assembling Sanger ESTs using the phrap assembler, then clustered by primary sequence similarity, grouping contigs with at least 100 bases of alignment at 93% identity into clusters. The phrap assembler used at that time was primarily meant for genomic assemblies, but we ran a number of iterative assemblies with manual review in between to select parameters best suited for use with ESTs.

The EST contig component of NemaGene is made up of more than 170,000 contigs, grouped into 121,696 clusters spanning 33 species (although one of those is on ‘archive’ status due to newer RNAseq based isotigs being available)

**454 cDNA isotigs** were created by assembling 454 reads using the Newbler transcriptome assembler. This assembler uses an overlap layout consensus approach to generate splice graphs to recreate putative isoforms (referred to as isotigs) belonging to a gene locus (the locus being referred to as an isogroup).

Almost 12 million 454 reads totaling around 4 billion bases spanning 9 species & multiple stages went into these 454 transcript assemblies.

**Gene CDS from finished genomes** are also included within NemaGene. The CDS for the Brugia malayi geneset (21,332 genes … WS230) has been loaded into NemaGene proper, and CDS for 9 additional worms is available for download (although not yet searchable via the cluster search interface).

In addition to those 3 types, for upcoming projects we’ll be introducing **Illumina RNAseq** data into NemaGene. These RNAseq reads will be assembled and the *contigs* will be loaded in as putative transcripts alongside our other data.

**NOTE TO PRESENTER:** *Illumina RNAseq assemblers will not be done using the newbler transcriptome aligner, and thus presumably will NOT be organized into isotigs & isogroups. These assemblies will simply be contigs.*

As we’ve touched on before, a lot of information is available to users for all of these transcripts & genes. **Codon Usage** & **GO annotations** are available, as well as primary sequence data & in many cases protein translations via **FTP**. The **Intestinal subset** has been defined, as well as extensive information on the finished **Brugia malayi geneset in the Gene Table**, which offers expression data across 7 stages, structural information and orthoMCL based clustering, as well as KO, IPR & independently assigned GO annotations.

Pathway information based on KO mappings can be explored using blast-based annotations against the last, publically available release of KEGG (v58.0) via **NemaPath**, which as shown previously offers comparative tools. And **NemaBLAST** allows wu-blast based alignments to be made against all the data in NemaGene.

**57) Hosted Datasets: NemaGene Related Publications**

This is a selection of just a few publications that reference data hosted in NemaGene

--1--) Defines sanger EST clustering & describes the building of the M.incognita transcripts

***Analysis and functional classification of transcripts from the nematode Meloidogyne incognita.***

[*McCarter JP*](http://www.ncbi.nlm.nih.gov/pubmed?term=McCarter%20JP%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Mitreva MD*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20MD%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Martin J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Martin%20J%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Dante M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Dante%20M%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Wylie T*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wylie%20T%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Rao U*](http://www.ncbi.nlm.nih.gov/pubmed?term=Rao%20U%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Pape D*](http://www.ncbi.nlm.nih.gov/pubmed?term=Pape%20D%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Bowers Y*](http://www.ncbi.nlm.nih.gov/pubmed?term=Bowers%20Y%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Theising B*](http://www.ncbi.nlm.nih.gov/pubmed?term=Theising%20B%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Murphy CV*](http://www.ncbi.nlm.nih.gov/pubmed?term=Murphy%20CV%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Kloek AP*](http://www.ncbi.nlm.nih.gov/pubmed?term=Kloek%20AP%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Chiapelli BJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Chiapelli%20BJ%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Clifton SW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Clifton%20SW%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Bird DM*](http://www.ncbi.nlm.nih.gov/pubmed?term=Bird%20DM%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*,* [*Waterston RH*](http://www.ncbi.nlm.nih.gov/pubmed?term=Waterston%20RH%5BAuthor%5D&cauthor=true&cauthor_uid=12702207)*.*

*Genome Biol. 2003;4(4):R26. Epub 2003 Mar 31.*

--2--) This 2nd paper looks at the effects of radiation and culture on gene expression in infective larvae (described transcripts available in NemaGene)

*Brugia malayi: effects of radiation and culture on gene expression in infective larvae.*

[*Li BW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Li%20BW%5BAuthor%5D&cauthor=true&cauthor_uid=16824625)*,* [*Rush AC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Rush%20AC%5BAuthor%5D&cauthor=true&cauthor_uid=16824625)*,* [*Weil GJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Weil%20GJ%5BAuthor%5D&cauthor=true&cauthor_uid=16824625)*,* [*McCarter JP*](http://www.ncbi.nlm.nih.gov/pubmed?term=McCarter%20JP%5BAuthor%5D&cauthor=true&cauthor_uid=16824625)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=16824625)*.*

*Mol Biochem Parasitol. 2006 Oct;149(2):201-7. Epub 2006 Jun 22.*

--3--) The 3rd one here examines the life cycle stages of B.malayi (references data on the B.malayi Gene Table)

*A deep sequencing approach to comparatively analyze the transcriptome of lifecycle stages of the filarial worm, Brugia malayi.*

[*Choi YJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Choi%20YJ%5BAuthor%5D&cauthor=true&cauthor_uid=22180794)*,* [*Ghedin E*](http://www.ncbi.nlm.nih.gov/pubmed?term=Ghedin%20E%5BAuthor%5D&cauthor=true&cauthor_uid=22180794)*,* [*Berriman M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Berriman%20M%5BAuthor%5D&cauthor=true&cauthor_uid=22180794)*,* [*McQuillan J*](http://www.ncbi.nlm.nih.gov/pubmed?term=McQuillan%20J%5BAuthor%5D&cauthor=true&cauthor_uid=22180794)*,* [*Holroyd N*](http://www.ncbi.nlm.nih.gov/pubmed?term=Holroyd%20N%5BAuthor%5D&cauthor=true&cauthor_uid=22180794)*,* [*Mayhew GF*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mayhew%20GF%5BAuthor%5D&cauthor=true&cauthor_uid=22180794)*,* [*Christensen BM*](http://www.ncbi.nlm.nih.gov/pubmed?term=Christensen%20BM%5BAuthor%5D&cauthor=true&cauthor_uid=22180794)*,* [*Michalski ML*](http://www.ncbi.nlm.nih.gov/pubmed?term=Michalski%20ML%5BAuthor%5D&cauthor=true&cauthor_uid=22180794)*.*

*PLoS Negl Trop Dis. 2011 Dec;5(12):e1409. doi: 10.1371/journal.pntd.0001409. Epub 2011 Dec 13.*

--4--) This 4th paper makes use of NemaGene EST contigs from S.ratti & S.strongyloides as a db in a search for some mass spec data they had produced.

*Life cycle stage-resolved proteomic analysis of the excretome/secretome from Strongyloides ratti--identification of stage-specific proteases.*

[*Soblik H*](http://www.ncbi.nlm.nih.gov/pubmed?term=Soblik%20H%5BAuthor%5D&cauthor=true&cauthor_uid=21964353)*,* [*Younis AE*](http://www.ncbi.nlm.nih.gov/pubmed?term=Younis%20AE%5BAuthor%5D&cauthor=true&cauthor_uid=21964353)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21964353)*,* [*Renard BY*](http://www.ncbi.nlm.nih.gov/pubmed?term=Renard%20BY%5BAuthor%5D&cauthor=true&cauthor_uid=21964353)*,* [*Kirchner M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Kirchner%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21964353)*,* [*Geisinger F*](http://www.ncbi.nlm.nih.gov/pubmed?term=Geisinger%20F%5BAuthor%5D&cauthor=true&cauthor_uid=21964353)*,* [*Steen H*](http://www.ncbi.nlm.nih.gov/pubmed?term=Steen%20H%5BAuthor%5D&cauthor=true&cauthor_uid=21964353)*,* [*Brattig NW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Brattig%20NW%5BAuthor%5D&cauthor=true&cauthor_uid=21964353)*.*

*Mol Cell Proteomics. 2011 Dec;10(12):M111.010157. doi: 10.1074/mcp.M111.010157. Epub 2011 Sep 30.*

--5--) This paper also demonstrates the use of NemaGene EST contigs from S.ratti for investigating mass spec data.

*Stage-specific excretory-secretory small heat shock proteins from the parasitic nematode Strongyloides ratti--putative links to host's intestinal mucosal defense system.*

[*Younis AE*](http://www.ncbi.nlm.nih.gov/pubmed?term=Younis%20AE%5BAuthor%5D&cauthor=true&cauthor_uid=21762402)*,* [*Geisinger F*](http://www.ncbi.nlm.nih.gov/pubmed?term=Geisinger%20F%5BAuthor%5D&cauthor=true&cauthor_uid=21762402)*,* [*Ajonina-Ekoti I*](http://www.ncbi.nlm.nih.gov/pubmed?term=Ajonina-Ekoti%20I%5BAuthor%5D&cauthor=true&cauthor_uid=21762402)*,* [*Soblik H*](http://www.ncbi.nlm.nih.gov/pubmed?term=Soblik%20H%5BAuthor%5D&cauthor=true&cauthor_uid=21762402)*,* [*Steen H*](http://www.ncbi.nlm.nih.gov/pubmed?term=Steen%20H%5BAuthor%5D&cauthor=true&cauthor_uid=21762402)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21762402)*,* [*Erttmann KD*](http://www.ncbi.nlm.nih.gov/pubmed?term=Erttmann%20KD%5BAuthor%5D&cauthor=true&cauthor_uid=21762402)*,* [*Perbandt M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Perbandt%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21762402)*,* [*Liebau E*](http://www.ncbi.nlm.nih.gov/pubmed?term=Liebau%20E%5BAuthor%5D&cauthor=true&cauthor_uid=21762402)*,* [*Brattig NW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Brattig%20NW%5BAuthor%5D&cauthor=true&cauthor_uid=21762402)*.*

*FEBS J. 2011 Sep;278(18):3319-36. doi: 10.1111/j.1742-4658.2011.08248.x. Epub 2011 Aug 24.*

58) **Hosted Datasets: HelmCoP**

HelmCoP (stands for Helminth Control and Prevention) comprises 38,776 orthologous groups built using orthoMCL clustering of the proteomes of 18 organisms, including 4 parasitic nematodes, 6 free-living nematodes, 2 flatworms & 6 outgroups.

Access to HelmCoP is primarily through the **main interface** described above, but all the proteomes used in HelmCoP are also available as a blastable database via the **HelmCoP BLAST** service.

HelmCoP offers a comparative genomics tool on top of one of the most comprehensive sets of helminth proteomes available. One of the more powerful uses of this resource is the ability to look for orthologous groups with representatives from some organisms while excluding others.

**NOTE TO PRESENTER***: If running out of time, SKIP the* ***BLUE*** *text below and just direct the audience to view the* ***HelmCoP FAQ*** *to see details on how all the various annotations in HelmCoP were assigned.*

The data hosted within this resource includes:

**EC & KO numbers** are assignedbased on mapping to the KEGG genes db (v58) providing information about the annotated enzymes & orthologous groups of enzymes specific to each species.

**IPR ids** are assigned based on homology to the Interpro db.

**Gene Ontology id** assignments are generated using interproscan, and provide classification of the genes & groups with regards to their *cellular component*, *biological process* & *molecular function*.

**Protein-Protein interactions** are based on the presence of C.elegans, H.sapians, D.melanogaster, S.cerevisiae, A.thaliana and M.musculus genes within orthologous groups. These 5 organisms are annotated within the MINT & IntAct PPI databases and are used to infer PPI.

The presence of **InDels specific to nematodes in relation to mammalian & metazoan outgroups** is available in some cases when defined by previous studies.

**RNAi phenotype** is available for orthologs to C.elegans genes.

The **PDB gene structure names** are annotated when a homology to a protein in PDB is found.

The **DrugBank structure name** is annotated if homology to a DrugBank protein is found.

Additionally, for genes with homology to a DrubBank protein, **Chemoinformatics** information is reported. And if genes are considered **Hopkins druggable targets** if an IPR id assigned to them shows up in the list of ids found in Hopkin’s study.

Finally structural information is reported, such as the presence of detected **signal peptide or transmembrane regions**, the presence of **Coiled Coils**, **Secondary Structure** composition & **regions of Disorder**. These annotations were made using the programs Paircoil2, JUFO, PsiPred, PHDPROF, IUPRED & RONN.

59) **Hosted Datasets: HelmCoP Related Publications**

These are a couple example publications that discuss HelmCoP.

--1--) This is the HelmCoP paper, describing the resource in detail

*HelmCoP: an online resource for helminth functional genomics and drug and vaccine targets prioritization.*

[*Abubucker S*](http://www.ncbi.nlm.nih.gov/pubmed?term=Abubucker%20S%5BAuthor%5D&cauthor=true&cauthor_uid=21760913)*,* [*Martin J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Martin%20J%5BAuthor%5D&cauthor=true&cauthor_uid=21760913)*,* [*Taylor CM*](http://www.ncbi.nlm.nih.gov/pubmed?term=Taylor%20CM%5BAuthor%5D&cauthor=true&cauthor_uid=21760913)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21760913)*.*

*PLoS One. 2011;6(7):e21832. doi: 10.1371/journal.pone.0021832. Epub 2011 Jul 8.*

--2--) This paper talks about HelmCoP in the role of accelerating research & development of tools for schistosomiasis diagnosis & treatment

*New frontiers in schistosoma genomics and transcriptomics.*

[*Nahum LA*](http://www.ncbi.nlm.nih.gov/pubmed?term=Nahum%20LA%5BAuthor%5D&cauthor=true&cauthor_uid=23227308)*,* [*Mourão MM*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mour%C3%A3o%20MM%5BAuthor%5D&cauthor=true&cauthor_uid=23227308)*,* [*Oliveira G*](http://www.ncbi.nlm.nih.gov/pubmed?term=Oliveira%20G%5BAuthor%5D&cauthor=true&cauthor_uid=23227308)*.*

*J Parasitol Res. 2012;2012:849132. doi: 10.1155/2012/849132. Epub 2012 Nov 21.*

60) **Hosted Datasets: NemFam**

The original (2009) NemFam collection of nematode-related, conserved protein families was constructed by clustering a set of around 130,000 EST contigs, themselves built from more than 262,000 ESTs from 29 species, plus the complete genesets from 5 genome sequencing projects (3 Caenorhabditis species, B.malayi & A.caninum), with 2 of those genesets also having EST contigs available (for a total of 32 species included). With the genesets adding ~84,000 protein sequences, a total of 214,000 peptides were clustered to form NemFam.

Protein families were built using MCL, originally resulting in about 54,000 families. From that set we extracted only families with members from at least 3 different nematodes, which further filtered the set down to 5326 multi-species families. Then we filtered this set to keep only those families that did not share homology with any of the protein family models included in Pfam-A.

We further refined this set by generating an HMM for each remaining family, and in the case of having more than one member from the same species in said family, only one representative per species was kept based on which had the best alignment back to its parent model. Finally, we required each valid family to have at least 10% of its full alignment length contributed simultaneously be sequences from 3 or more species. This final refinement reduced the set further to 1593 groups built from 13,963 protein sequences.

61) **Hosted Datasets: NemFam Related Publications**

Here is a selection of papers that discuss NemFam.

--1--) This is the paper detailing the construction of NemFam

*Molecular determinants archetypical to the phylum Nematoda.*

[*Yin Y*](http://www.ncbi.nlm.nih.gov/pubmed?term=Yin%20Y%5BAuthor%5D&cauthor=true&cauthor_uid=19296854)*,* [*Martin J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Martin%20J%5BAuthor%5D&cauthor=true&cauthor_uid=19296854)*,* [*Abubucker S*](http://www.ncbi.nlm.nih.gov/pubmed?term=Abubucker%20S%5BAuthor%5D&cauthor=true&cauthor_uid=19296854)*,* [*Wang Z*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wang%20Z%5BAuthor%5D&cauthor=true&cauthor_uid=19296854)*,* [*Wyrwicz L*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wyrwicz%20L%5BAuthor%5D&cauthor=true&cauthor_uid=19296854)*,* [*Rychlewski L*](http://www.ncbi.nlm.nih.gov/pubmed?term=Rychlewski%20L%5BAuthor%5D&cauthor=true&cauthor_uid=19296854)*,* [*McCarter JP*](http://www.ncbi.nlm.nih.gov/pubmed?term=McCarter%20JP%5BAuthor%5D&cauthor=true&cauthor_uid=19296854)*,* [*Wilson RK*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wilson%20RK%5BAuthor%5D&cauthor=true&cauthor_uid=19296854)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=19296854)*.*

*BMC Genomics. 2009 Mar 18;10:114. doi: 10.1186/1471-2164-10-114.*

--2--) This paper references NemFam as one of the resources available that can help in the process of identifying molecular determinants that underlie species adaptation & evolution.

*The draft genome of the parasitic nematode Trichinella spiralis.*

[*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Jasmer DP*](http://www.ncbi.nlm.nih.gov/pubmed?term=Jasmer%20DP%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Zarlenga DS*](http://www.ncbi.nlm.nih.gov/pubmed?term=Zarlenga%20DS%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Wang Z*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wang%20Z%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Abubucker S*](http://www.ncbi.nlm.nih.gov/pubmed?term=Abubucker%20S%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Martin J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Martin%20J%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Taylor CM*](http://www.ncbi.nlm.nih.gov/pubmed?term=Taylor%20CM%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Yin Y*](http://www.ncbi.nlm.nih.gov/pubmed?term=Yin%20Y%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Fulton L*](http://www.ncbi.nlm.nih.gov/pubmed?term=Fulton%20L%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Minx P*](http://www.ncbi.nlm.nih.gov/pubmed?term=Minx%20P%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Yang SP*](http://www.ncbi.nlm.nih.gov/pubmed?term=Yang%20SP%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Warren WC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Warren%20WC%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Fulton RS*](http://www.ncbi.nlm.nih.gov/pubmed?term=Fulton%20RS%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Bhonagiri V*](http://www.ncbi.nlm.nih.gov/pubmed?term=Bhonagiri%20V%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Zhang X*](http://www.ncbi.nlm.nih.gov/pubmed?term=Zhang%20X%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Hallsworth-Pepin K*](http://www.ncbi.nlm.nih.gov/pubmed?term=Hallsworth-Pepin%20K%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Clifton SW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Clifton%20SW%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*McCarter JP*](http://www.ncbi.nlm.nih.gov/pubmed?term=McCarter%20JP%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Appleton J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Appleton%20J%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Mardis ER*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mardis%20ER%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Wilson RK*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wilson%20RK%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*.*

*Nat Genet. 2011 Mar;43(3):228-35. doi: 10.1038/ng.769. Epub 2011 Feb 20.*

62) **Hosted Datasets: Expression Data**

Nematode.net hosts both microarray & RNAseq –based expression data. The exact nature of each experiment can differ. So to fully explore these datasets your best bet is to refer to each experiment’s publication, which once available is linked into the repository.

The data you will typically be interested in is the **Processed data**, which are normalized expression values (either signal strengths for microarrays, or read counts for RNAseq based experiments). But you’ll also want to look over the other associated files for additional information. The **Raw data** holds the experimental values from before normalization, the **Sample annotations** typically describe the experimental conditions used on each sample in the analysis. **Experimental design** provides information on things such as biological & technical replicates performed in the experiment, which samples are applied to which arrays and so forth. **Array annotation** provides information about the probes used in microarrays (and is not applicable to RNAseq based experiments), and **Methods** describe the lab protocols & data processing techniques used.

63) **Hosted Datasets: Expression Data Related Publications**

These are a number of publications related to experimental results hosted on Nematode.net.

--1--) This is the paper describing our hosted microarray-based Ascaris suum experiment.

*Gene expression analysis distinguishes tissue-specific and gender-related functions among adult Ascaris suum tissues.*

[*Wang Z*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wang%20Z%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*,* [*Gao X*](http://www.ncbi.nlm.nih.gov/pubmed?term=Gao%20X%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*,* [*Martin J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Martin%20J%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*,* [*Yin Y*](http://www.ncbi.nlm.nih.gov/pubmed?term=Yin%20Y%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*,* [*Abubucker S*](http://www.ncbi.nlm.nih.gov/pubmed?term=Abubucker%20S%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*,* [*Rash AC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Rash%20AC%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*,* [*Li BW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Li%20BW%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*,* [*Nash B*](http://www.ncbi.nlm.nih.gov/pubmed?term=Nash%20B%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*,* [*Hallsworth-Pepin K*](http://www.ncbi.nlm.nih.gov/pubmed?term=Hallsworth-Pepin%20K%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*,* [*Jasmer DP*](http://www.ncbi.nlm.nih.gov/pubmed?term=Jasmer%20DP%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=23572074)*.*

*Mol Genet Genomics. 2013 Apr 10. [Epub ahead of print]*

--2--) This paper is the original one that defined the B.malayi version 2 array.

*Profiling of gender-regulated gene transcripts in the filarial nematode Brugia malayi by cDNA oligonucleotide array analysis.*

[*Li BW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Li%20BW%5BAuthor%5D&cauthor=true&cauthor_uid=15992941)*,* [*Rush AC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Rush%20AC%5BAuthor%5D&cauthor=true&cauthor_uid=15992941)*,* [*Crosby SD*](http://www.ncbi.nlm.nih.gov/pubmed?term=Crosby%20SD%5BAuthor%5D&cauthor=true&cauthor_uid=15992941)*,* [*Warren WC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Warren%20WC%5BAuthor%5D&cauthor=true&cauthor_uid=15992941)*,* [*Williams SA*](http://www.ncbi.nlm.nih.gov/pubmed?term=Williams%20SA%5BAuthor%5D&cauthor=true&cauthor_uid=15992941)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=15992941)*,* [*Weil GJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Weil%20GJ%5BAuthor%5D&cauthor=true&cauthor_uid=15992941)*.*

*Mol Biochem Parasitol. 2005 Sep;143(1):49-57.*

--3--) This is a more recent paper discussing gender-associated genes in filarial nematodes that references the B.malayi v2 array.

*Gender-associated genes in filarial nematodes are important for reproduction and potential intervention targets.*

[*Li BW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Li%20BW%5BAuthor%5D&cauthor=true&cauthor_uid=21283610)*,* [*Rush AC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Rush%20AC%5BAuthor%5D&cauthor=true&cauthor_uid=21283610)*,* [*Jiang DJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Jiang%20DJ%5BAuthor%5D&cauthor=true&cauthor_uid=21283610)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21283610)*,* [*Abubucker S*](http://www.ncbi.nlm.nih.gov/pubmed?term=Abubucker%20S%5BAuthor%5D&cauthor=true&cauthor_uid=21283610)*,* [*Weil GJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Weil%20GJ%5BAuthor%5D&cauthor=true&cauthor_uid=21283610)*.*

*PLoS Negl Trop Dis. 2011 Jan 25;5(1):e947. doi: 10.1371/journal.pntd.0000947.*

--4--) This is the paper describing the Brugia malayi lifecycle stage microarray.

*Transcriptomes and pathways associated with infectivity, survival and immunogenicity in Brugia malayi L3.*

[*Li BW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Li%20BW%5BAuthor%5D&cauthor=true&cauthor_uid=19527522)*,* [*Rush AC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Rush%20AC%5BAuthor%5D&cauthor=true&cauthor_uid=19527522)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=19527522)*,* [*Yin Y*](http://www.ncbi.nlm.nih.gov/pubmed?term=Yin%20Y%5BAuthor%5D&cauthor=true&cauthor_uid=19527522)*,* [*Spiro D*](http://www.ncbi.nlm.nih.gov/pubmed?term=Spiro%20D%5BAuthor%5D&cauthor=true&cauthor_uid=19527522)*,* [*Ghedin E*](http://www.ncbi.nlm.nih.gov/pubmed?term=Ghedin%20E%5BAuthor%5D&cauthor=true&cauthor_uid=19527522)*,* [*Weil GJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Weil%20GJ%5BAuthor%5D&cauthor=true&cauthor_uid=19527522)*.*

*BMC Genomics. 2009 Jun 15;10:267. doi: 10.1186/1471-2164-10-267.*

--5--) This is a more recent paper making use of the lifecycle stage microarray to explore stage & function dependent expression patterns in B.malayi

*Transcription profiling reveals stage- and function-dependent expression patterns in the filarial nematode Brugia malayi.*

[*Li BW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Li%20BW%5BAuthor%5D&cauthor=true&cauthor_uid=22583769)*,* [*Wang Z*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wang%20Z%5BAuthor%5D&cauthor=true&cauthor_uid=22583769)*,* [*Rush AC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Rush%20AC%5BAuthor%5D&cauthor=true&cauthor_uid=22583769)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=22583769)*,* [*Weil GJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Weil%20GJ%5BAuthor%5D&cauthor=true&cauthor_uid=22583769)*.*

*BMC Genomics. 2012 May 14;13:184. doi: 10.1186/1471-2164-13-184.*

--6--) This is the paper describing the Affymetrix Soybean Genome Array GeneChip.

*Sequence mining and transcript profiling to explore cyst nematode parasitism.*

[*Elling AA*](http://www.ncbi.nlm.nih.gov/pubmed?term=Elling%20AA%5BAuthor%5D&cauthor=true&cauthor_uid=19183474)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=19183474)*,* [*Gai X*](http://www.ncbi.nlm.nih.gov/pubmed?term=Gai%20X%5BAuthor%5D&cauthor=true&cauthor_uid=19183474)*,* [*Martin J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Martin%20J%5BAuthor%5D&cauthor=true&cauthor_uid=19183474)*,* [*Recknor J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Recknor%20J%5BAuthor%5D&cauthor=true&cauthor_uid=19183474)*,* [*Davis EL*](http://www.ncbi.nlm.nih.gov/pubmed?term=Davis%20EL%5BAuthor%5D&cauthor=true&cauthor_uid=19183474)*,* [*Hussey RS*](http://www.ncbi.nlm.nih.gov/pubmed?term=Hussey%20RS%5BAuthor%5D&cauthor=true&cauthor_uid=19183474)*,* [*Nettleton D*](http://www.ncbi.nlm.nih.gov/pubmed?term=Nettleton%20D%5BAuthor%5D&cauthor=true&cauthor_uid=19183474)*,* [*McCarter JP*](http://www.ncbi.nlm.nih.gov/pubmed?term=McCarter%20JP%5BAuthor%5D&cauthor=true&cauthor_uid=19183474)*,* [*Baum TJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Baum%20TJ%5BAuthor%5D&cauthor=true&cauthor_uid=19183474)*.*

*BMC Genomics. 2009 Jan 30;10:58. doi: 10.1186/1471-2164-10-58.*

64) **Hosted Datasets: Nematode Genomes**

Annotated **Nematode Genomes** are hosted in **NemaBrowse**. Currently we are hosting the Trichinella spiralis draft genome and a set of gene predictions on a draft assembly of Trichuris suis. As we finalize more parasitic worm genomes, we’ll annotate them and host them here as well. Necator americanus should be the next draft genome we upload into NemaBrowse. As I mentioned before, this repository will focus on non-model-organism parasitic nematodes, with the goal of hosting annotations that may help with pan-phylum parasitic-species derived questions (i.e. things related to paracitism)

The **NemaSNP** viewer for population genetic studies is currently hosting SNP annotations for 2 population isolates of Teladorsagia circumcincta against a transcriptomic assembly. This resource provides the loci positions for SNPs called from each population, as well as their population source (field or inbred in this case), and also the synonymous or non-synonymous state of each loci.

While **NemaSNP** is not technically a genomic resource, our intent is to roll this information into **NemaBrowse** as additional annotation on top of the usual tracks. Mapping the transcriptomic contigs onto the genomic reference will allow us to merge this information into **NemaBrowse** proper. Of course as full, annotated, genomic references become available, variant calls will be made directly against the annotated genomes. Mapping transcript-framed annotations is only an intermediate measure until genomic references become available.

65) **Hosted Datasets: Nematode Genomes Related Publications**

This is the T.spiralis draft genome paper.

--1--) This is the paper describing the Trichinella spiralis draft genome.

*The draft genome of the parasitic nematode Trichinella spiralis.*

[*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Jasmer DP*](http://www.ncbi.nlm.nih.gov/pubmed?term=Jasmer%20DP%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Zarlenga DS*](http://www.ncbi.nlm.nih.gov/pubmed?term=Zarlenga%20DS%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Wang Z*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wang%20Z%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Abubucker S*](http://www.ncbi.nlm.nih.gov/pubmed?term=Abubucker%20S%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Martin J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Martin%20J%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Taylor CM*](http://www.ncbi.nlm.nih.gov/pubmed?term=Taylor%20CM%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Yin Y*](http://www.ncbi.nlm.nih.gov/pubmed?term=Yin%20Y%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Fulton L*](http://www.ncbi.nlm.nih.gov/pubmed?term=Fulton%20L%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Minx P*](http://www.ncbi.nlm.nih.gov/pubmed?term=Minx%20P%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Yang SP*](http://www.ncbi.nlm.nih.gov/pubmed?term=Yang%20SP%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Warren WC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Warren%20WC%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Fulton RS*](http://www.ncbi.nlm.nih.gov/pubmed?term=Fulton%20RS%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Bhonagiri V*](http://www.ncbi.nlm.nih.gov/pubmed?term=Bhonagiri%20V%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Zhang X*](http://www.ncbi.nlm.nih.gov/pubmed?term=Zhang%20X%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Hallsworth-Pepin K*](http://www.ncbi.nlm.nih.gov/pubmed?term=Hallsworth-Pepin%20K%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Clifton SW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Clifton%20SW%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*McCarter JP*](http://www.ncbi.nlm.nih.gov/pubmed?term=McCarter%20JP%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Appleton J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Appleton%20J%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Mardis ER*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mardis%20ER%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*,* [*Wilson RK*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wilson%20RK%5BAuthor%5D&cauthor=true&cauthor_uid=21336279)*.*

*Nat Genet. 2011 Mar;43(3):228-35. doi: 10.1038/ng.769. Epub 2011 Feb 20.*

66) **Hosted Datasets: Publication Datasets**

The final resource I’ll discuss is our repository for datasets associated with nematode publications. Our **Publication Dataset FTP** hosts supplemental data & resources thought valuable enough to be made public for a number of projects.

This section is organized by species & publication, and hosts links to our FTP site allowing visitors to download these resources. Each resource is also briefly summarized in the text heading each link.

67) **Hosted Datasets: Several Publications With Hosted Data**

These are a couple publications for which we are hosting data on Nematode.net’s FTP.

--1--) This publication is on the effects of Doxycycline on gene expression in Wolbachia and Brugia malayi. Nematode.net hosts supplemental data for this paper.

*Effects of doxycycline on gene expression in Wolbachia and Brugia malayi adult female worms in vivo.*

[*Rao RU*](http://www.ncbi.nlm.nih.gov/pubmed?term=Rao%20RU%5BAuthor%5D&cauthor=true&cauthor_uid=22321609)*,* [*Huang Y*](http://www.ncbi.nlm.nih.gov/pubmed?term=Huang%20Y%5BAuthor%5D&cauthor=true&cauthor_uid=22321609)*,* [*Abubucker S*](http://www.ncbi.nlm.nih.gov/pubmed?term=Abubucker%20S%5BAuthor%5D&cauthor=true&cauthor_uid=22321609)*,* [*Heinz M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Heinz%20M%5BAuthor%5D&cauthor=true&cauthor_uid=22321609)*,* [*Crosby SD*](http://www.ncbi.nlm.nih.gov/pubmed?term=Crosby%20SD%5BAuthor%5D&cauthor=true&cauthor_uid=22321609)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=22321609)*,* [*Weil GJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Weil%20GJ%5BAuthor%5D&cauthor=true&cauthor_uid=22321609)*.*

*Source*

*J Biomed Sci. 2012 Feb 9;19:21. doi: 10.1186/1423-0127-19-21.*

--2--) Nematode.net hosts the assembly contigs for Wolbachia endosymbionts Onchocerca flexuosa & Acanthocheilonema viteae referenced in this paper looking at putative ancient Horizontal Gene Transfer.

*Endosymbiont DNA in endobacteria-free filarial nematodes indicates ancient horizontal genetic transfer.*

[*McNulty SN*](http://www.ncbi.nlm.nih.gov/pubmed?term=McNulty%20SN%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Foster JM*](http://www.ncbi.nlm.nih.gov/pubmed?term=Foster%20JM%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Dunning Hotopp JC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Dunning%20Hotopp%20JC%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Martin J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Martin%20J%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Fischer K*](http://www.ncbi.nlm.nih.gov/pubmed?term=Fischer%20K%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Wu B*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wu%20B%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Davis PJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Davis%20PJ%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Kumar S*](http://www.ncbi.nlm.nih.gov/pubmed?term=Kumar%20S%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Brattig NW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Brattig%20NW%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Slatko BE*](http://www.ncbi.nlm.nih.gov/pubmed?term=Slatko%20BE%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Weil GJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Weil%20GJ%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*,* [*Fischer PU*](http://www.ncbi.nlm.nih.gov/pubmed?term=Fischer%20PU%5BAuthor%5D&cauthor=true&cauthor_uid=20543958)*.*

*PLoS One. 2010 Jun 9;5(6):e11029. doi: 10.1371/journal.pone.0011029.*

--3--) Our FTP site hosts several resources referenced by this paper: i) The final assembly of the O.flexuosa transcriptome, ii) Protein translations from O.flexuosa contigs & singletons and iii) A list of proteins identified by mass spectrometry analysis of O.flexuosa worm lysate.

*Transcriptomic and proteomic analyses of a Wolbachia-free filarial parasite provide evidence of trans-kingdom horizontal gene transfer.*

[*McNulty SN*](http://www.ncbi.nlm.nih.gov/pubmed?term=McNulty%20SN%5BAuthor%5D&cauthor=true&cauthor_uid=23049857)*,* [*Abubucker S*](http://www.ncbi.nlm.nih.gov/pubmed?term=Abubucker%20S%5BAuthor%5D&cauthor=true&cauthor_uid=23049857)*,* [*Simon GM*](http://www.ncbi.nlm.nih.gov/pubmed?term=Simon%20GM%5BAuthor%5D&cauthor=true&cauthor_uid=23049857)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=23049857)*,* [*McNulty NP*](http://www.ncbi.nlm.nih.gov/pubmed?term=McNulty%20NP%5BAuthor%5D&cauthor=true&cauthor_uid=23049857)*,* [*Fischer K*](http://www.ncbi.nlm.nih.gov/pubmed?term=Fischer%20K%5BAuthor%5D&cauthor=true&cauthor_uid=23049857)*,* [*Curtis KC*](http://www.ncbi.nlm.nih.gov/pubmed?term=Curtis%20KC%5BAuthor%5D&cauthor=true&cauthor_uid=23049857)*,* [*Brattig NW*](http://www.ncbi.nlm.nih.gov/pubmed?term=Brattig%20NW%5BAuthor%5D&cauthor=true&cauthor_uid=23049857)*,* [*Weil GJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Weil%20GJ%5BAuthor%5D&cauthor=true&cauthor_uid=23049857)*,* [*Fischer PU*](http://www.ncbi.nlm.nih.gov/pubmed?term=Fischer%20PU%5BAuthor%5D&cauthor=true&cauthor_uid=23049857)*.*

*PLoS One. 2012;7(9):e45777. doi: 10.1371/journal.pone.0045777. Epub 2012 Sep 26.*

--4 & 5--) We host a number of resources for both these papers including Ascaris suum germline & somatic genome sequence assemblies, denovo cDNA assemblies, germline gene models, mrnas & proteins & more.

*Deep small RNA sequencing from the nematode Ascaris reveals conservation, functional diversification, and novel developmental profiles.*

[*Wang J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wang%20J%5BAuthor%5D&cauthor=true&cauthor_uid=21685128)*,* [*Czech B*](http://www.ncbi.nlm.nih.gov/pubmed?term=Czech%20B%5BAuthor%5D&cauthor=true&cauthor_uid=21685128)*,* [*Crunk A*](http://www.ncbi.nlm.nih.gov/pubmed?term=Crunk%20A%5BAuthor%5D&cauthor=true&cauthor_uid=21685128)*,* [*Wallace A*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wallace%20A%5BAuthor%5D&cauthor=true&cauthor_uid=21685128)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=21685128)*,* [*Hannon GJ*](http://www.ncbi.nlm.nih.gov/pubmed?term=Hannon%20GJ%5BAuthor%5D&cauthor=true&cauthor_uid=21685128)*,* [*Davis RE*](http://www.ncbi.nlm.nih.gov/pubmed?term=Davis%20RE%5BAuthor%5D&cauthor=true&cauthor_uid=21685128)*.*

*Genome Res. 2011 Sep;21(9):1462-77. doi: 10.1101/gr.121426.111. Epub 2011 Jun 17.*

*Silencing of germline-expressed genes by DNA elimination in somatic cells.*

[*Wang J*](http://www.ncbi.nlm.nih.gov/pubmed?term=Wang%20J%5BAuthor%5D&cauthor=true&cauthor_uid=23123092)*,* [*Mitreva M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Mitreva%20M%5BAuthor%5D&cauthor=true&cauthor_uid=23123092)*,* [*Berriman M*](http://www.ncbi.nlm.nih.gov/pubmed?term=Berriman%20M%5BAuthor%5D&cauthor=true&cauthor_uid=23123092)*,* [*Thorne A*](http://www.ncbi.nlm.nih.gov/pubmed?term=Thorne%20A%5BAuthor%5D&cauthor=true&cauthor_uid=23123092)*,* [*Magrini V*](http://www.ncbi.nlm.nih.gov/pubmed?term=Magrini%20V%5BAuthor%5D&cauthor=true&cauthor_uid=23123092)*,* [*Koutsovoulos G*](http://www.ncbi.nlm.nih.gov/pubmed?term=Koutsovoulos%20G%5BAuthor%5D&cauthor=true&cauthor_uid=23123092)*,* [*Kumar S*](http://www.ncbi.nlm.nih.gov/pubmed?term=Kumar%20S%5BAuthor%5D&cauthor=true&cauthor_uid=23123092)*,* [*Blaxter ML*](http://www.ncbi.nlm.nih.gov/pubmed?term=Blaxter%20ML%5BAuthor%5D&cauthor=true&cauthor_uid=23123092)*,* [*Davis RE*](http://www.ncbi.nlm.nih.gov/pubmed?term=Davis%20RE%5BAuthor%5D&cauthor=true&cauthor_uid=23123092)*.*

*Dev Cell. 2012 Nov 13;23(5):1072-80. doi: 10.1016/j.devcel.2012.09.020. Epub 2012 Nov 1.*

68) **Upcoming Nematode Projects**

The last thing I want to talk about is to mention the upcoming projects we are expecting that will be hosted on Nematode.net.

This table shows the nematodes that will be coming soon, or are in the very early stages of production (material collection and so forth). The Filarial species are highlighted in blue, since they are probably of particular interest to this audience. Data from these organisms will begin trickling into Nematode.net sometime later this year.

Of note is the fact that we’ll be getting multiple lab & field isolates from many of these worms, which will be used to study variation between different populations. The primary data generated for these projects will be Illumina reads, and the data will be assembled & annotated. Transcript contigs & related annotations will then be merged into NemaGene & made available to the public.

69) **Upcoming Nematode Projects: Collaborator & Collection Sites**

This image shows the location of our collaborators around the world, as well as sample collection sites.

These new projects, in addition to our planned site update, make this a particularly exciting time for Nematode.net.

70) **Nematode.net Staff & Alumni**

These are the primary people who have been involved in Nematode.net over the years. Todd was the original designer of Nematode.net back in 2000. Sahar is a colleague whose help has been invaluable over most of the life of the site, she very recently moved on and the site is quite challenging to manage without her.

Yong, Zhengyuan & Christy are content contributors who probably ended up knowing more about the inner workings of Nematode.net than they really wanted to, and Dr. Makedonka Mitreva is the current PI.

Site funding is provided by the NIAID.

Thank you for listening!