Web resources for *C. elegans* studies*

Raymond Lee§, Division of Biology, California Institute of Technology, Pasadena, CA 91125, USA

Table of Contents

1. Introduction ............................................................................................................................ 1
2. Portal and knowledge environment .............................................................................................. 2
  2.1. WormBase ................................................................................................................... 2
  2.2. *Caenorhabditis elegans* WWW Server ........................................................................ 3
3. Literature search ...................................................................................................................... 4
  3.1. Textpresso ................................................................................................................... 4
  3.2. NCBI PubMed .............................................................................................................. 5
  3.3. *Caenorhabditis elegans* WWW Server Worm Literature Index ........................................... 5
4. Gene function ......................................................................................................................... 6
  4.1. WormBase Gene Summary ............................................................................................. 6
  4.2. NCBI AceView ............................................................................................................. 7
5. Sequence analysis-retrieval ........................................................................................................ 8
  5.1. NCBI GenBank ............................................................................................................ 8
  5.2. WormBase Genome Browser ........................................................................................... 9
6. Sequence analysis-search for homologous sequences ................................................................. 10
  6.1. NCBI Blast ................................................................................................................ 10
  6.2. WormBase Blat Server ................................................................................................. 10
7. Genetics .................................................................................................................................... 10
  7.1. *Caenorhabditis* Genetics Center (CGC) ........................................................................ 10
  7.2. Genetic mapping analysis ............................................................................................. 11
  7.3. Small interval genetic mapping ...................................................................................... 12
8. Cell and anatomy ................................................................................................................... 12
  8.1. WormAtlas ................................................................................................................ 12
  8.2. Gene Expression Pattern ............................................................................................... 13
  8.3. BCGSC Expression Pattern ........................................................................................... 16
9. Concluding remark ................................................................................................................. 16

1. Introduction

This chapter describes a list of core Web resources that I think are most useful to someone who is either new to studying *C. elegans* or who has not been using the Web much as a research tool. It does not contain a
comprehensive list of Web sites and services since links to other useful Web resources can usually be readily found at the sites discussed here.

Internet is the information super highway and the World Wide Web is the easiest way to access most of the information on the internet. Although one can always find some information about anything on the Web, it is not necessarily easy to find the most relevant quickly. The Web is vast. Since its debut for internal use at CERN in 1991, the Web has grown into a global network of 60 million sites as of March 2005 (<http://news.netcraft.com/archives/web_server_survey.html>). Furthermore, the format of Web resources and information pages is divergent, prevents highly efficient queries across all sites. Thus, Web search engines are billion-dollar industries and people are being paid for their Web surfing skills, researching the Web for answers of specific questions (<http://answers.google.com/answers/>).

Today, it is perhaps still possible to study *C. elegans* without using the Web, but likely with much reduced efficiency. One can use the Web to place an order for a deletion mutant, to read a summary of all that is known about one's favorite gene, to perform a sequence analysis or download a sequence for local analysis, or to look up a paper and read it online anywhere there is internet access. It is true that we cannot yet have worms delivered through fiber optics cables, but anything that can be reduced to a digital form is either available on the Web today or will be tomorrow. Via fast internet and hyperlinks, all digitized information around the world about the worm can be brought to one's computer terminal in an instance.

But one needs to know where to look. To most, it probably is an easier experience using the Web for the first time than to try to find something in a brick and mortar library-that's why the Web is popular. However, to efficiently take advantage of available Web resources, it is best to know exactly where the resources are and how to use them when one gets there. I try to write this chapter as an introductory guide for those who are just starting to tap into the Web for their worm research. It is by no means a complete guide but it should make it easier for people to get started.

### 2. Portal and knowledge environment

There are two types of Web services most suitable as one's starting points: portal and knowledge environment. A portal organizes Web links relevant to the domain of interest. From a portal, a user can quickly find sites and services of potential interest and via hyperlinks be brought there. In contrast, a knowledge environment is a Web site that directly offers many different types of information and services. A knowledge environment often emphasizes the interconnection among data types, which in a way is a web within the Web. In practice, most Web sites have portal pages.

#### 2.1. WormBase

WormBase [http://www.wormbase.org](http://www.wormbase.org) is a major repository for *C. elegans* information, including genomic, genetic, anatomy, people, and literature. WormBase is a knowledge environment. Access to information is via a set of Web pages, each of which specifically designed for a different type of biological knowledge. Further, different information types, when appropriate, are connected horizontally via hyperlinks. One can easily move from Web page of one type to another. For example, one can start by visiting a genome sequence, click a link to read about a gene that resides in this sequence on a gene page, click a link again to review an expression pattern description on an expression pattern page, click yet again to read about a cell on an anatomy page, and so on.

Some WormBase pages are complex. WormBase offers an online user guide. There is also a book chapter [http://www.mrw.interscience.wiley.com/cp/cpbi/articles/bi0108/frame.html](http://www.mrw.interscience.wiley.com/cp/cpbi/articles/bi0108/frame.html); although the access is restricted to subscribers of the journal. In this section I describe a few general purpose pages. In following sections, I discuss more specialized pages. Users are encouraged to visit the site and practice the most effective Web surfing technique-click and browse.

**Basic Query** [http://www.wormbase.org](http://www.wormbase.org)

On the front page of WormBase, type a query term (e.g. glucos) in the query box, select an information type from the pull down menu to the left and hit the “Search” button.
The information type "Anything" gives the broadest search option and should be selected unless one wishes to restrict the search to a specific type. The search term “glucos” is deliberately left in the ‘root’ form of related words so that the search includes all terms that start with “glucos”, such as glucose, glucosamin, glucosyl, etc.).

Site Map <http://www.wormbase.org/db/misc/site_map>

This page is like a portal which links to pages of major searches, data viewers and resources within WormBase, providing users a general idea of what's available.


Both the data and software of WormBase can be downloaded and worked with locally. This feature is not for the casual users but particularly useful for those who routinely perform large-scale data analysis. A README file <ftp://ftp.wormbase.org/pub/wormbase/README> indicates which data are available for download.

WormMart <http://dev.wormbase.org/BioMart/martview>

Presently still under development, WormMart is an advanced query tool which allows users customize complex queries. It is also designed to run fast, expediting results returned from complex and large-scale queries. WormMart focuses on sequence-related data types.

2.2. Caenorhabditis elegans WWW Server <http://elegans.swmed.edu/>

This site is a well-organized portal of many different types of information useful to C. elegans researchers. It is organized in two layers. The front page lists major topics or interests. Each topic has a hyperlink to a page that either offers a service (e.g. Literature Search) or is a list of links to other Web sites that offer services. Each page and link of this portal is generally self-explanatory; thus no user guide is necessary. Users can easily drill down the links to find out what is available at this site in a few minutes.
3. Literature search

3.1. Textpresso

Textpresso (<http://www.textpresso.org>) allows text searches on primarily *C. elegans* literature, including published papers, personal communications and meeting reports. Two major features distinguish Textpresso from other literature search tools: that it searches full-text contents of publications, and in addition to text strings, that it can search for groups of terms (categories).

The simplest way to search using Textpresso is to start with the default settings and type into the query box a text string. For example, if one wants to learn about regulation of kinases, one can search for “regulat kinase”.

*Caenorhabditis elegans* is a small (about 1 mm long) soil nematode found in temperate regions. In the 1960’s Sydney Brenner began using it to study the genetics of development and neurobiology. Since then the community of *C. elegans* researchers has expanded to over a thousand. The Riddle Lab has written an introduction to *C. elegans* for non-specialists. Biologists unfamiliar with *C. elegans* may find Mark Blaxter’s *The Genetics of Caenorhabditis elegans, An Introduction* useful.

Users in Europe may find access to the Greek mirror of the *C. elegans* WWW Server more convenient.
Textpresso treats the words independently. The default setting automatically appends a wild card to the end of each word thus expanding the search to include any word that begins with “regulat” or “kinase”. Further, the default setting is to search for sentences that simultaneously have both groups of words. Textpresso also offers category search and many other more advanced features. Users can read the user guide to learn how to use advanced features.

3.2. NCBI PubMed

**NCBI PubMed** allows queries on articles in a large collection of biomedical journals. Coverage of PubMed is broad and usually up to date but some literature relevant to *C. elegans* studies is not covered. Also, Pubmed searches are limited to citations and abstracts.

To search, one can start by typing a term of interest in the query box and click the “Go” button. Read the tutorial to learn how to perform more sophisticated queries.

3.3. *Caenorhabditis elegans* WWW Server Worm Literature Index

*Caenorhabditis elegans* WWW Server Worm Literature Index <http://elegans.swmed.edu/wli/> offers text searches on citations and abstracts of selected publications, Worm Breeder's Gazette articles, and Worm meeting abstracts. Coverage of publications is limited to those selected by CGC (*C. elegans* Genetic Center). Although the coverage of search space here is only a subset of that of Textpresso, the user interface is self-explanatory, easy to use, and one can search for phrases.
4. Gene function

Many different aspects of biological knowledge go into the description of gene functions, including mutant phenotype, site of action, interaction with other genes, and sequence similarity to other genes. A Web resource that has descriptions of gene function usually offers a type of pages that integrate these aspects so that a user can have a holistic view of gene function at a glance.

4.1. WormBase Gene Summary
<http://www.wormbase.org/db/gene/gene>

Each gene in WormBase has a summary page which collates together several different aspects of a gene, including identification, genetic and genomic location, function, reagents and bibliography.
4.2. NCBI AceView

To perform a simple search, type into the query box a key word (such as R13H8, a genomic cosmid name), click the “Go” button. Many navigation links are there on the result page. To get at the most information, users should click and browse around.

5. Sequence analysis-retrieval

5.1. NCBI GenBank

Notice the extensive links to other types of information here. Browse and click to explore.

5.2. **WormBase Genome Browser**

WormBase Genome Browser [http://www.wormbase.org/db/seq/gbrowse/wormbase] is a physical map browser. Using Genome Browser, one can search and display sequences and sequence-related features; one can also zoom in or out and move along on a chromosome.
With Genome Browser, one can search for sequences by name or sequence (restricted to oligo size pieces) match, display and explore selected tracks of sequence and sequence features, and export sequences, features and images. Interested users can either explore around the page or read the user guides (<http://www.its.caltech.edu/~wormbase/userguide/Menu/Sequence/index.html>;<http://www.mrw.interscience.wiley.com/cp/cpbi/articles/bi0108/frame.html>).

6. Sequence analysis-search for homologous sequences

6.1. NCBI Blast

NCBI Blast <http://www.ncbi.nlm.nih.gov/BLAST/> offers a very extensive set of blast services. Here, different types of blast searches can be performed against all available sequences in GenBank database. One can even download programs to install and run locally.

6.2. WormBase Blat Server

WormBase Blat Server <http://www.wormbase.org/db/searches/blat> is limited to basic Blast or Blat searches against C. elegans and C. Briggsae sequences. However, WormBase already stores information of homologous sequences from other species in its database. Such sequences may be displayed in Genome Browser under feature tracks.

7. Genetics

7.1. Caenorhabditis Genetics Center (CGC)

Caenorhabditis Genetics Center (CGC) <http://biosci.umn.edu/CGC/CGChomepage.htm> is a resource center for C. elegans genetics. It is responsible for gene nomenclature, strain collection and distribution, and genetic map construction. CGC homepage is a portal that has links to these and some other related services useful to C. elegans geneticists.
7.2. Genetic mapping analysis

WormBase integrates genetic map information with that of physical map. Two search tools are particular useful for genetic mapping analysis.

**Genetic Interval Search**

- This page will list genes inside a genetically-defined interval.
- Specify a genetic interval using the names of two genes (for example evi-7 to sup-23).
- Alternatively, define the endpoints using the notation chromosome, position, as in II-4-3 to II-4-0.
- Entering a single value into the start position will center the search on a window 0.3 cm wide.
- Interval size is currently limited to 2 cm. Longer intervals may take some time to collate.

From position/landmark: II-4-3 To: II-4-0

<table>
<thead>
<tr>
<th>Name</th>
<th>Map</th>
<th>Pos</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>C0304.3</td>
<td>II</td>
<td>-4.2541</td>
<td>named gene</td>
<td>The C0304.3 gene encodes a homolog of the human gene CYP7B1, which when mutated leads to giant cell hepatitis (OMIM 231100). RNA: WT [WBRNA00009688]</td>
</tr>
<tr>
<td>C0304.1</td>
<td>II</td>
<td>-4.2322</td>
<td>named gene</td>
<td>hhl-1 is orthologous to the human gene MYOGENIC FACTOR 6 (HERCULIN) (MYF6, OMIM 159991), which when mutated leads to disease. (C elegans) cc450am, recessive lethal, arrest at varying stages as lumpy dumpy with some muscle differentiation, disorganized myofilaments. Encodes member of MyoD family. Transcripts present transiently in MS descendants, stably in D daughters and C and MS lineages leading to body wall muscle. Antibody staining similar (nuclear staining, persists through development of body wall muscle cells). [Krause et al. 1994, 1990, Chen et al. 1992, KM, PD]. RNA: WT [WBRNA00009688], Unc [WBRNA00024379], Unc [WBRNA00027142]</td>
</tr>
<tr>
<td>C0304.4</td>
<td>II</td>
<td>-4.2233</td>
<td>predicted</td>
<td>No molecular description available.</td>
</tr>
</tbody>
</table>

WormBook.org
Genetic Interval Search <http://wormbase.org/db/searches/interval> can return a list of genes that have the potential to map within a specified genetic interval.

It should be noted that the genetic and physical map correlation is often solely based on statistical inferences thus should not be taken literally as factual.

7.3. Small interval genetic mapping

SNP, Visible Marker, And Strain Search <http://wormbase.org/db/searches/strains> is particularly useful for finding markers for genetic mapping experiments in a small interval.

**SNP, Visible Marker, And Strain Search**

This page is designed to assist in the selection of SNPS and visible genetic markers for mapping experiments. Provided with a genetic interval, it will display all visible markers and SNPs within that interval, as well as strains carrying those SNPs. The severity of markers can be restricted to avoid conflicts with the mutation of interest. To simplify in the acquisition of strains, strains can be restricted to those directly available from the CGC.

- Specify the interval using the names of two genetically defined loci (unc-70), or predicted genes (JC8.10a).
- Alternatively, you may provide genetic map positions using chromosome position notation (III-0.685).
- Please note: intervals are restricted to 2 cM.
- Entering a single value into the start position will center the search on a window 0.3 cM wide.
- Learn more about how this script works.

<table>
<thead>
<tr>
<th>Search Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>From position/landmark: bli-6 to unc-24</td>
<td></td>
</tr>
<tr>
<td>Loci</td>
<td>All</td>
</tr>
<tr>
<td>SNPs</td>
<td>All</td>
</tr>
<tr>
<td>SNP</td>
<td>All</td>
</tr>
<tr>
<td>Ease Of Scoring</td>
<td>3</td>
</tr>
<tr>
<td>Mating Efficiency</td>
<td>3</td>
</tr>
<tr>
<td>Exclude lethals?</td>
<td>Yes</td>
</tr>
<tr>
<td>Strains</td>
<td>CGC Strains Only</td>
</tr>
</tbody>
</table>

8. Cell and anatomy

8.1. WormAtlas

WormAtlas <http://www.wormatlas.org/> provides anatomical information of *C. elegans*. The front page lists several useful entry points.
One can use the simple text search tool to search the site for information that relates to anatomical terms (e.g. PVT, name of a neuron). Another good way to use this site is to read sections of the “handbook”.

**ATLAS OF C. elegans ANATOMY** - an illustrated handbook

**HERMAPHRODITE ANATOMY**
Edited by: Z. F. Altun & D. H. Hall

- **Chapter 1**: Overall description of body shape
- **Chapter 2**: Adult organs and tissues
  1. Epithelial System
     - Hypodermis
     - Specialized epith. cells Part I-Septum Cells
     - Specialized epith. cells Part II-Neuronal Support Cells
  2. Nervous System
     - Specialized epith. cells Part III-Other interfacial cells

**MALE ANATOMY**
Edited by: R. Lints & D. H. Hall

- **Chapter 1**: Introduction to the male
  1. Body shape, behavior and propagation
     1. Epithelial System
        - Hypodermis
        - Part I-Septum Cells
        - Part II-Neuronal Support Cells
        - Part III-Other interfacial cells
  2. Nervous System
     - General Description
     - Male sensory organs
     - Individual Neurons
     - Commissures
     - Connectives

**ADDENDUM**

- **Chapter 1**: *C. elegans* behaviors and processes
- **Chapter 2**: Phylogeny by David H. A. Fitch
- **APPENDIX**: References
- **A crawling *C. elegans* - from Goldstein lab**: A bird's eye view of *C. elegans* - quicktime movie.
  by Adam Hartley and Carolyn Marks

**8.2. Gene Expression Pattern**

Summary of published gene expression data can be found at WormBase, or *C. elegans* AceView. Two Web sites offer primary, sometimes unpublished expression data.
The Nematode Expression Pattern Database (NEXTDB) <http://nematode.lab.nig.ac.jp/db2/> provides access to *C. elegans* EST sequences obtained by Yuji Kohara's laboratory and some other experimental results derived from them, such as expression patterns determined by *in situ* hybridization, which can be searched via a text query tool.

As an example, let's query for cosmid K04H4. On the result page, one can either click on a group, such as CELK02617, to go to the summary page for this EST cluster group (roughly equivalent to a gene) directly, or one can select the cosmid K04H4 link to jump to a map view to begin browsing along the chromosome, gene-by-gene, for *in situ* data.
To go to the map view, select “display neighboring cosmids”. Note here that it is best to start with one cosmid on either side because more will be too crowded for the physical map display. Click on the “submit” button brings up an “Area map” page. On the left hand side of this page is a physical map which is actually an interactive graphical Java application.
On this page, clicking on any thumbnail will bring up links to full in situ images. To move ‘downward’ on the physical map, select the pink bar representing the lower cosmid; then click on the “shift” button on top of this page to effect the move. To move by more than one cosmid at a time, one can change the number (defaulted to 1) next to the “shift” button. To adjust the viewable area of the physical map display, one needs to resize the map. “Change map size” control strip is at the very bottom of the web page (not in view here). A warning here is that this Java application may not be stable at all times. Be gentle.

8.3. BCGSC Expression Pattern

BCGSC Expression Patterns <http://elegans.bcgsc.ca/perl/eprofile/browse> lists GFP expression data which can be browsed directly or searched by gene name, tissue pattern or life stage.

**Expression Patterns**
for C. elegans promoter::GFP fusions

![Web resources for C. elegans studies](image)

For each expression pattern, there are text annotations, images or even animations of series of images.

9. Concluding remark

The Web is still growing rapidly, both in terms of technology and content. We can expect that C. elegans Web resources will also grow and improve. As more people become accustomed to using the Web in their research, existing resources will improve and more will be made available. It will be a challenge for this chapter to keep up with future changes so that it remains to be useful.

All WormBook content, except where otherwise noted, is licensed under a [Creative Commons Attribution License](http://creativecommons.org/licenses/by/2.5/).