orm Base The Biology and Genome of C. elegans.

. Gene Summaries

The Gene Sum-	 ✓ Gene Summary for goa-1 - Mozilla ▲ Eile Edit View Go Bookmarks Tools Window Help 			
mary Page pro-	So Search			
vides concise in-				
formation about	Find: Any Gene Summary Locus Summary Sequence Summary Protein Summary EST Alignments Genome Browser Genetic Map Nearby Genes Bibliography Tree Display XML Schema Acedb Image			
each C. elegans				
gene organized	Gene Summary for goa-1 Specify a gene using a gene name (unc-26), a predicted gene id (R13A5.9), or a protein ID (CE02711): goa-1			
into distinct sec-	[identification] [location] [function] [gene ontology] [reactome knowledgebase] [alleles] [similarities] [reagents] [bibliography]			
tions: identifica-	Identification IDs: CGC name Sequence name Other name(s) WB Gene ID Version goa-1 - (G protein,O, Alpha subunit) C26C6.2 11101 WB Gene00001648 1			

4. Synteny Viewer

- One way to start the Synteny Viewer is to click on the "Syntenic Alignment" link on a Gene Page:
 - Caenorhabditis briggsae: CBG02462 [syntenic alignment] (Stein LD et al. PubMed; best reciprocal ortholog(s): blastp match-seg-off)

Another way is to click right on a gene's *briggsae* alignment in the Genome Browser:

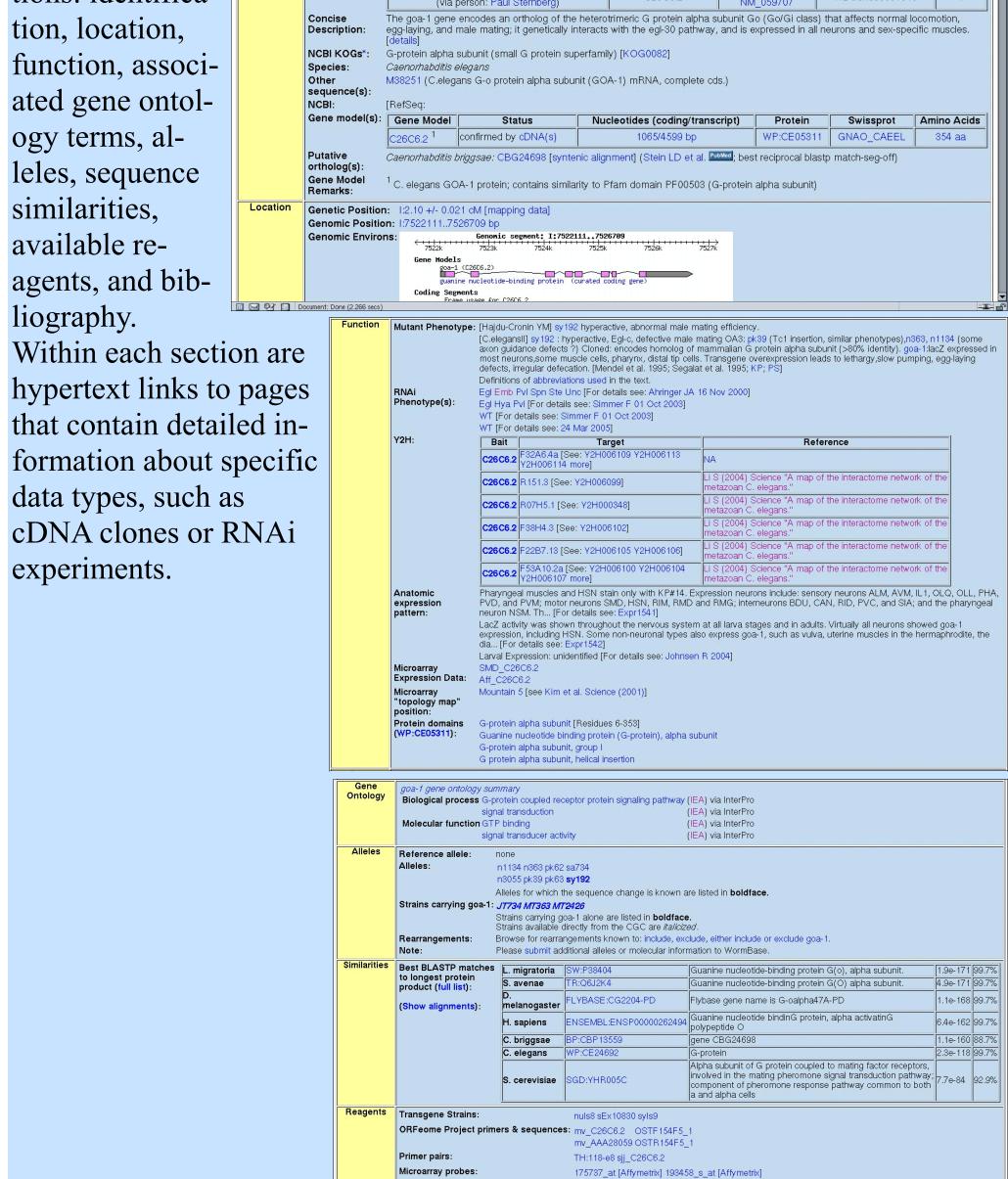
8. Gene Ontology Summaries

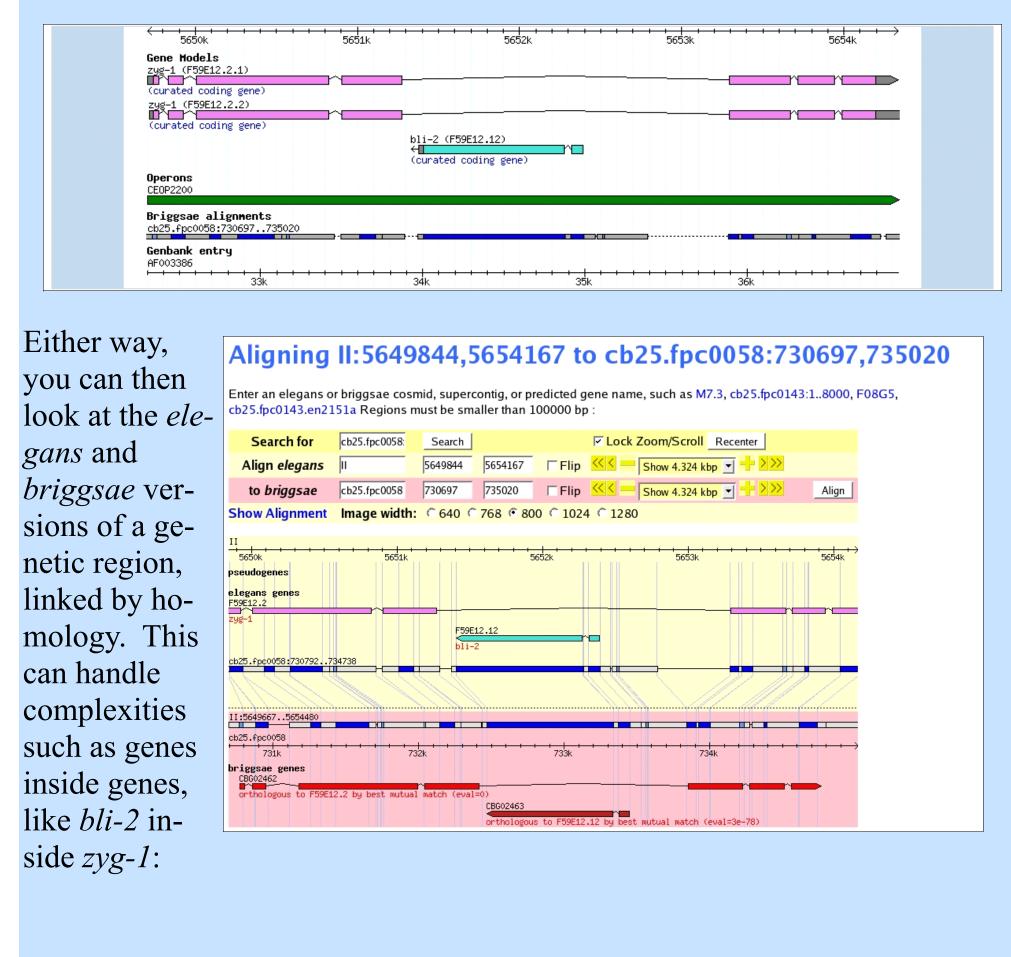
Gene Ontology terms are categorized according to Biological Process, Cellular Component, or Molecular Function. Evidence, such as IEA (inferred from electronic annotation) or IMP (inferred from mutant phenotype), indicates the type of data used to assign the GO terms. More in-depth searches can also be performed on this page.

Gone Ontology Summary for gone: lot-23

11.Textpresso

The literature search engine in WormBase contains over 7,000 full-text journal articles pertaining to worm biology (>95% from the last 6 years) and is updated weekly. Abstracts and full text are searchable either by keywords or by categories selected from drop down menus. Categories are biologically meaningful words and phrases that have been pre-organized into groups for searching, and can be combined with keywords to increase the accuracy of a search. The search engine provides a number of useful links for each citation returned from a search including: (i) a link to the on-line journal access if you have a subscription to that journal, (ii) a link to the PubMed report page for that citation and (iii) a downloadable file that can be read into the EndNote citation manager. Most importantly, the user can follow a link to the actual underlying sentence in full text that matched the search terms. The system is powered by Textpresso (http://www.textpresso.org), an information retrieval and extraction system developed at WormBase.

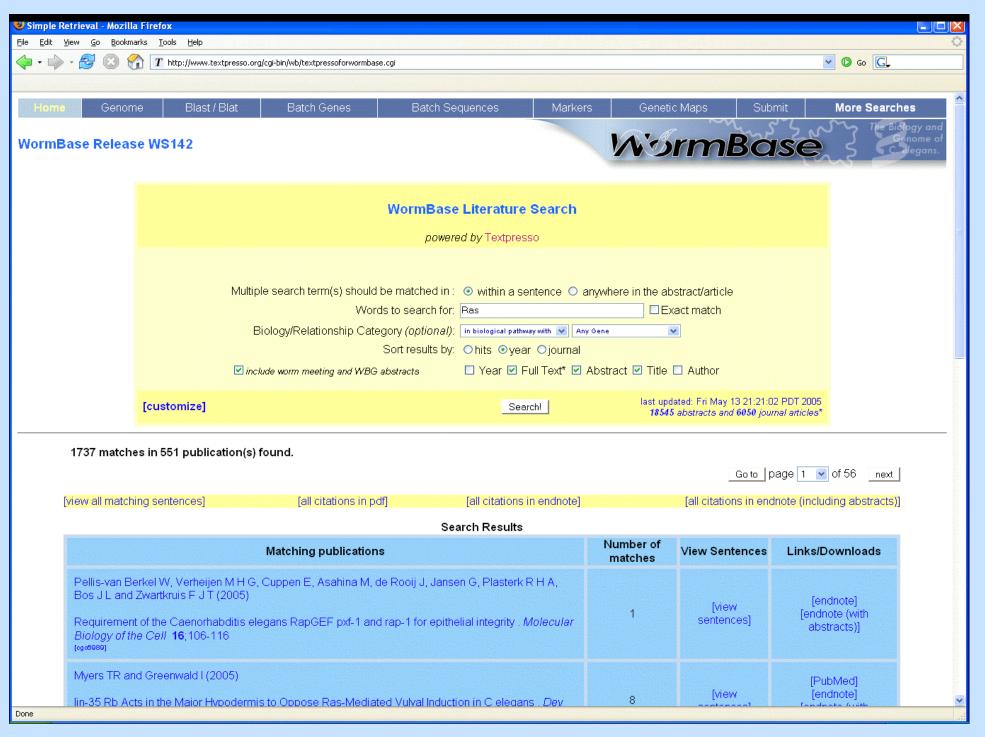




Biological process			
Term	Definition	Evidence	
larval development (sensu Nematoda) (GO:0002119)	Development from the newly hatched first-stage larva to the end of the last larval stage (e. g. fourth- stage in C. elegans) before molting to adulthood.	Inferred from mutant phenotype (IM via Paper Herman RK 1978	
male genital morphogenesis (GO:0030539)		Inferred from mutant phenotype (IMP) via Paper Chamberlin HM and Sternberg PW 1994	
ovulation (sensu Mammalia) (GO:0001542)	Processes leading to rupture the of the follicle releasing the central located oocyte.	Inferred from mutant phenotype (IMP) via Paper Clandinin TR et al. 1998	
positive regulation of vulval development (sensu Nematoda) (GO:0040026)	Any process that activates or increases the rate of development of the vulva in nematodes.	Inferred from mutant phenotype (IMP) via Paper Ferguson EL and Horvitz HR 1985	
protein amino acid phosphorylation (GO:0006468)	The process of introducing a phosphoric group on to a protein.	Inferred from electronic annotation (IEA) via InterPro	
Cellular component			
Term	Definition	Evidence	
basolateral plasma membrane (GO:0016323)	Part of the plasma membrane that includes the basal end and sides of the cell. Often used in reference to animal polarized epithelial membranes, where the basal membrane is the part attached to the extracellular matrix, or in plant cells, where the basal membrane is defined with respect to the zygotic axis.	Inferred from direct assay (IDA) via Paper Kaech SM et al. 1998	
membrane (GO:0016020)	Double layer of lipid molecules that encloses all cells, and, in eukaryotes, many organelles; may be a single or double lipid bilayer, also includes associated proteins.	Inferred from electronic annotation (IEA) via InterPro	
Molecular function			
Term	Definition	Evidence	
ATP binding (GO:0005524)	Interacting selectively with ATP, adenosine 5'- triphosphate, a universally important coenzyme and enzyme regulator.	Inferred from electronic annotation (IEA) via InterPro	
epidermal growth factor receptor activity (GO:0005006)		Inferred from electronic annotation (IEA) via InterPro	
protein kinase activity (GO:0004672)	Catalysis of the transfer of a phosphate group, usually from ATP, to a protein substrate.	Inferred from electronic annotation (IEA) via InterPro	
protein-tyrosine kinase activity (GO:0004713)	Catalysis of the reaction: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.	Inferred from electronic annotation (IEA) via InterPro	
transmembrane receptor protein tyrosine kinase activity (GO:0004714)		Inferred from electronic annotation (IEA) via InterPro	
Queŋ	Ana CDS Cell	S	
	scription (endocytosis)	ression patterns es nology groups Submit	
WBGene00002299	Moti Phe Psu Pub		

9.Literature Curation

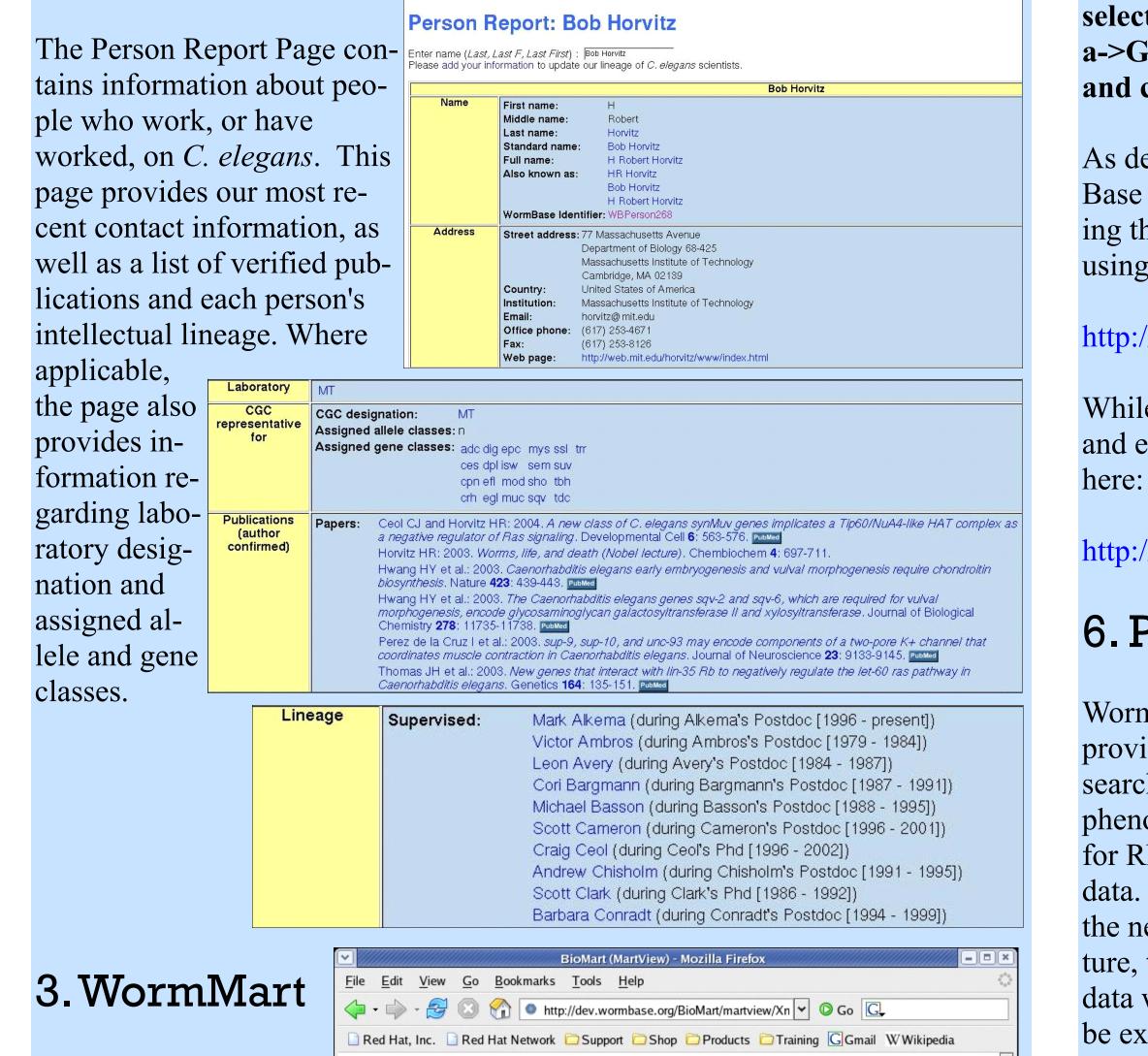
Literature curaion is an important part of the data collection in WormBase. There are ~11,400 *C. elegans* papers in Medline and there is a steady increase in the number of new papers published each year. To facilitate data extraction from the literature we have implemented a two-stage approach. During the



12. Future Plans



2. Person Reports



types that were carried out against genes located on chromosome IV by the Julie Ahringer group:

AQL query language provides users with flexible data retrieval based on cus-

tom criteria. Queries can aggregate data from multiple database classes, mak-

ing AQL a powerful data mining tool. For instance, the following query will

return a table containing RNAi ID, associated phenotype, gene ID and corre-

sponding sequence name for all RNAi experiments with non-wildtype pheno-

select a, a->Phenotype, b, b->Sequence name from a in class RNAi, b in a->Gene, c in a->Reference where b->Interpolated_map_position="IV" and c->Author like "*Ahringer*" and not a->Phenotype="WT"

As demonstrated by this example, using AQL requires knowledge of Worm-Base data models. The data model for each object can be accessed by following the Schema link on most WormBase pages. Models can also be browsed using the Class Browser:

http://www.wormbase.org/db/searches/class_query?class=model

While AQL takes some time to learn, mastering it provides users with a new and efficient way of data extraction. A good tutorial on AQL can be found

http://www.acedb.org/Software/whelp/AQL/

6. Phenotype

lassi

nutan

eles.`

5. Using AQL

WormBase **RNAi Phenotype Search** provides Select one or more of the following maternal, embryonic or post-embryonic phenotypes. Press Search when done searchable Embryonic lethal: © yes © no @ don't care Partly or completely sterile: 🕤 yes 🕤 no 🔅 don't care phenotypes General Postembryonic Phenotype Growth defect Morphology defect Movement defect Reproductive defect for RNAi data. (In (Bli) blistered (Egl) egg laying defect (Mlt) molt defect (Hya) HYperActive (Bmd) body (Fem) feminization of XX (Let) larval lethal (Muv) MUltiVulva (Stp) sterile the near fu-(Rol) roller norphology defect and XO animals progeny (Fog) feminization of ture, these (Clr) clear (Lon) long body 🔽 (Prl) paralyzed 🔽 (Rup) exploded germline (Unc) uncoordinate (Daf) DAue (Lva) larva (Gro) slow growth (Vul) vulvaless lata will (Sck) sick (Prz) PaRaLyzed (Pvl) protruding (Him) high incidence of (Lvl) larval lethal ☐ (Sma) small (Dpy) dumpy be exmale progeny

nom me merature we	- nave implemente	eu a two-stage ap	proach. During the
first stage, the pa-			
pers are scanned by		e Info Falls in	
a curator who identi-	(only b	iological data is she	own below)
fies the data of inter-	Gene data:	Gene product data:	Sequence data:
est. He flags the pa-	egene-sequence connection	e gene product function	 gene structure update/corrections DNA features
pers with the moni-	e mapping data	 expression data transgene 	 BNA leadures sequenced mutations
tored data types.	* mutant phenotype RNAi (non-large scale) 	 structure info 	○ SNPs
This information is	 RNAi (large-scale) 	\bigcirc in vitro protein analysis	
passed on to the cu-	 mosaic analysis 	 antibodies covalent modification 	Cell data:
rators who actually	\bigcirc site of action analysis	* gene product interactions	nomenclature, function, ablation
extract the informa-	🟓 allele info	microarrays	n 🤟 and a sharen a sharen birthe Askar a birthe sharen birthe santasin 🦇 a sharen birthe bi
	* gene interactions	★○ overexpression data	
tion during the sec-		igodoldoldoldoldoldoldoldoldoldoldoldoldol	* used for functional
ond stage of the cu-	DB entry completeness:		annotations
ration. First-pass	●>90% complete ● ~5-25%	complete 🔘 Not being extrac	ted yet
curation is evolving			
as new data types are		tion pipeline. Cu	rrently, some data
types are flagged but			
not extracted due to	Data Type	Distribution (3,590 papers)
limited staff; these		```	• • •
data are stored for	1200 hine		
future extraction.			
First-pass curation has covered all <i>C</i> . <i>elegans</i> papers since January 2002 and all information-rich pa- pers since January 2001. So far, ~3,700 papers have gone through first-pass	800 - 814 600 - 606 589 400 - 505 49 200 - 0		128 121 84 57 56 54 39 23 19 12 86 57 56 54 39 23 19 12 96 57 56 54 39 23 19 12 97 56 54 39 23 19 12
curation. Curatable			
curation. Curatable			

WormBase hopes to accomplish the following tasks in the near future:

- Incorporation of the genomes of several *Caenorhabditis* species (*C. re*manei, C. japonica, C. sp. CB5161), that are currently in the sequencing pipeline, as they become available.
- Building tools essential for viewing and analysing newly incorporated nematode genomes, such as synteny viewers, motif browsers, orthologous groups, etc.
- Curation of phenotypes and attaching phenotype information to alleles, transgenes and RNAi experiments
- Completing the anatomy ontology and building a phenotype ontology • Building queries for gene expression using the anatomy ontology • Building links between WormBase and WormBook
- Continuing gene function curation through updating concise descriptions and assigning GO (Gene Ontology) annotations for all the named genes in C. elegans.in *C. elegans*
- Reorganizing complex Web pages to simplify access to data and to improve speed.

13. WormBase Help via email and Wiki

WormBase wants your comments, complaints, questions, or praises. Please send e-mail to the general list help@wormbase.org, or anyone on this list of people <http://wormbase.org/about/people.html> if you desire more privacy. We strive for a response time of less than 24 hours during working days.

Decumentation & User Foodback

WormMart provides	new START
easy and interactive ac-	
cess to data en masse.	DATASET 1
In particular, it enables	IDENTIFICATION
users to retrieve se-	Species
quences and genome	CGC name
annotations using a va-	
riety of filters to con-	WB Gene ID
strain retrieved results.	
Some of the available	Prediction status
filters include gene	
names, genomic loca-	UTR status
tion and the presence	C Orthologous gene
or absence of specific	
annotations. Data can	
	Chromosome_nam
be returned in a variety	Chromosome coo From (base pair)
of formats, including	To (base pair)
HTML, text, FASTA,	Chromosome_stra
and Microsoft Excel.	FUNCTION
	Annotated with

START & FILTER	+ OUTPUT	export	worm		p
		(dack (next ►	(refresh) (Help Desk)		C
r1			Summary		n
ICATION			▶ start		16
ecies	Caenorhabditis el	egans 🗾	Dataset: Genes(WS142) 44279 Entries Total		
notated with GC name	Only Excluded		▶ filter		
ecified identifiers of type			Species: Caenorhabditis elegans		
/B Gene ID			chromosome name: II		
			From (base pair): 20000		7
		Browse	 To (base pair): 40000 RNAi: Only 		
ediction status	Confirmed	•	9 Entries pass Filters		
anscript type	Coding		output		Γ
'R status	UTR5+UTR3		🙂 Not yet initialised		a
thologous gene	Caenorhabditis br	riggsae 🔽			C
ON					
romosome_name	II •				
romosome coordinates					
om (base pair) (base pair)	20000 40000				
romosome_strand	-1 💌				
ON					
notated with	Only				
NAi	C Excluded			*	
	111			>	

140	Combine conditions with @ AND @ OR
d to	Prominent datasets
1	Limit results to the following datasets:
cal	🕝 Ahringer, 2003 🖓 Hyman, 2000 🖓 Maeda, 2001 🖓 Piano, 2002 🖓 Simmer, 2003 🖓 Vidal, 2004 🖓 other
4 -1	Chromosomal limits
it al-	Limit results to the following chromosomal range in genetic map values (ie -1.0 or 4.3):
	Reset Search
	List genes with wild type RNAi assays List genes with positive RNAi assays List genes with ANY RNAi assay

7. Phenotype Ontology

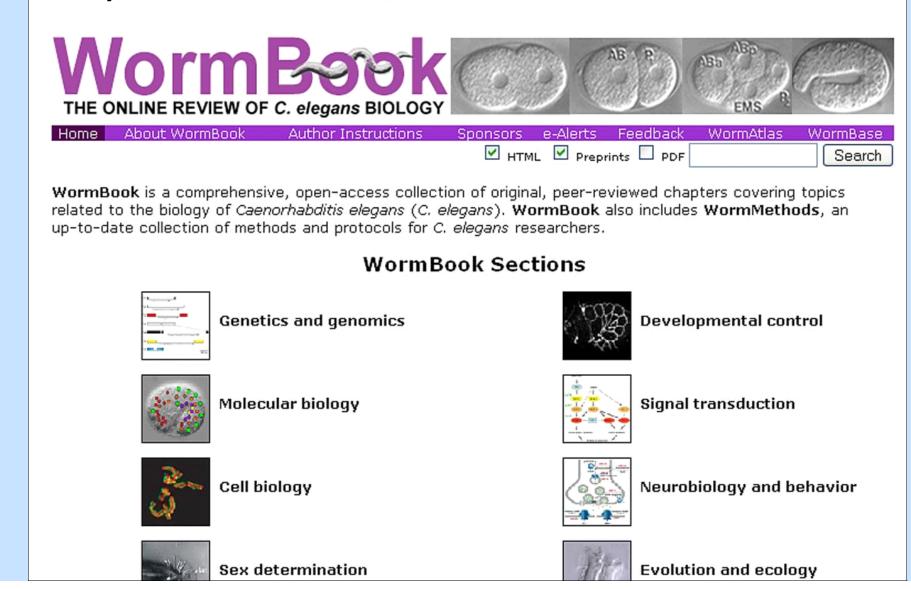
Developing a controlled vocabulary for the description of the phenotypes will allow high level and more detailed descriptions to be contained in a hierarchical browsable structure.

egg_laying_abnormal	ID		WBPhenotype:0001074	
egg_laying_defective	Namespace	C_elegans_pheno	otype_ontology	
E + gg_laying_serotonin_insensitive	OBOClass name			
vulval_muscie_unresponsive_to_serotonin	vulval_muscle_unres	ponsive_to_serotonin		
□←@ eggs_retained	(Dec.		and the deficitions	
← @ bag_ot_worms	Definition	* Comment Co	mplete definitions	
	Text		Dbxrefs Edit	
fewer_egg_laying_events_during_active	The vulval muscle does not respond typically to serotonin, based on imaging		WB:cab	
□←∅ inactive_phase_iong	of calcium transient		pmid:14588249	
active_phase_switch_defective	serotonin.			
← egg_laying_irregular				
E+@ egg_laying_phases_abnormal	A State of the second s		Charles and the second second	

information falls into 29 different fields as shown below. The percentage of the data extraction is shown for different data types.

10.Wormbook

Comprehensive Review: WormBook



,	antation & User Feedback
WormBase Navigation Main Page User Guide FAQs Meetings Recent changes Help	article discussion view source history Main Page This is the official wiki site for WormBase &, the database of the model organism Caenorhabditis elegans @ and related nematodes. In addit on using WormBase, this Wiki contains articles on C. elegans biology and experimental protocols. WormBase is developed by an an internat researchers and computer scientists. Before contributing, please see the WormBaseWiki usage guidelines. Contents Enow Community Resources This section of the WormBase Wiki is open to editing by all members of the community. Feel free to make contributions here.
search	Using WormBase
Go Search coolbox What links here Related changes Upload file Special pages Printable version	 User Guide FAQs Glossary of terms Downloads - Download commonly requested files and datasets Data mining - Learn about the various methods of mining WormBase systematically For Developers Linking To WormBase CVS Access To WormBase Suggested pre-release data checks
	About WormBase
	1. Presentations Publications