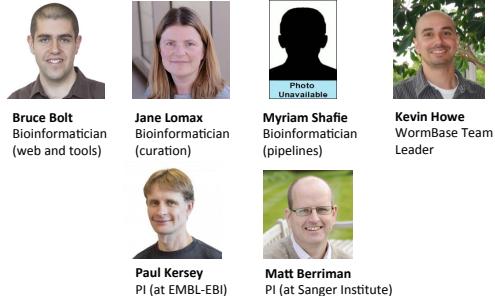


WormBase ParaSite Workshop

Edinburgh
9th March 2016

WormBase ParaSite Team



parasite.wormbase.org

- Features both nematodes (roundworms) and platyhelminthes (flatworms) genomes
- No additional curation for most genomes
- Focus on rapid availability of new data
- Automated pipelines run over all genomes

The Website

- Genome Browser
- Transcriptomic Data Display
- Gene, transcript and protein information pages
- Comparative Genomics
- Sequence Similarity Search (BLAST)
- Variant Effect Predictor (VEP) *
- Advanced Search Tool (BioMart)
- Access to BioMart data using R
- Programmatic Access (REST API)

* = Not covered today – speak to us for more information

The Data

- All genomes are shown “as supplied” by the submitter (except WormBase “core” genomes)
- Varying levels of coverage and quality
- Details of assembly and annotation displayed on information page
- “Core” parasitic genomes: *Brugia malayi*, *Onchocerca volvulus*, *Pristionchus pacificus* and *Strongyloides ratti*
- Receive more care and attention
- Community driven manual curation

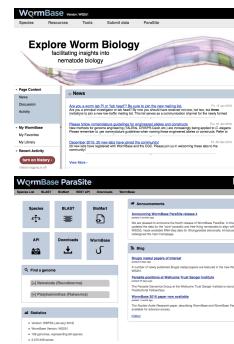
This screenshot shows the WormBase ParaSite website for the *Trichuris muris* genome (PRJEB120). The top navigation bar includes links for Species List, BLAST, BioMart, REST API, Downloads, and WormBase. The main content area displays the genome assembly information for *TMUR2.0*. It includes sections for Gene annotation, Comparative genomics, and Downloads. The gene annotation section shows a genomic track with various tracks for protein coding genes, non-coding RNA, and other genomic features. The comparative genomics section provides links to homologous genes across multiple species. The downloads section offers options for genomic sequence (FASTA) and other formats.

Your Data

- Publicly available transcriptomic data annotated and displayed on browser
- Website supports ad-hoc visualisation of your own data (e.g. RNA-Seq alignments, variations)
- We welcome submissions of your own data to display on genome browser – allow readers of your papers to easily visualise your data
- Please contact us (link at bottom of website) to discuss requirements

WormBase and WormBase ParaSite

- wormbase.org is the home for highly curated data from *C. elegans* and other related nematodes
- Genes from “core” parasites also displayed here
- More genomic data for parasites available from parasite.wormbase.org



This afternoon's agenda...

- 13:00 – 13:15
Introduction to WormBase ParaSite
- 13:15 – 13:45
Using the website (Part 1)
- 13:45 – 14:15
Using the website (Part 2)
- 14:15 – 15:00
Sequence Search with BLAST
- 15:00 – 15:30
Coffee Break
- 15:30 – 16:30
Data Mining with BioMart
- 16:30 – 16:45
Bulk downloads and programmatic access

After this workshop...

- Please contact us with any questions (contact form link at bottom of every page)
- Solutions to exercises on YouTube: parasite.wormbase.org/workshop



Workshop Feedback

- Your feedback helps tailor future workshops
- We would be very grateful if you could complete this before leaving

Post-workshop Feedback

We would be grateful if you could spare a few moments giving us some feedback about today's workshop. This will help us to improve future workshops.

1. Did you see WormBase ParaSite before the workshop?
2. Will you use WormBase ParaSite more often after this workshop?
3. Would you recommend this workshop to your colleagues?
4. How useful are each section of the workshop?

Browsing the website	Very Useful	Useful	Not Useful	Useless (not relevant, if possible)
BLAST				
Curator's Seminar				
Bulk Downloads				
5. How happy were you with each of the following?

Topics	Very Happy	Happy	Not Happy	Unsure (not relevant, if possible)
Outcomes Covered				
Level				
Interactivity				
Quality of presentations				
Exercises				
Balance of presentations and exercises				
6. Do you have any other comments or feedback?

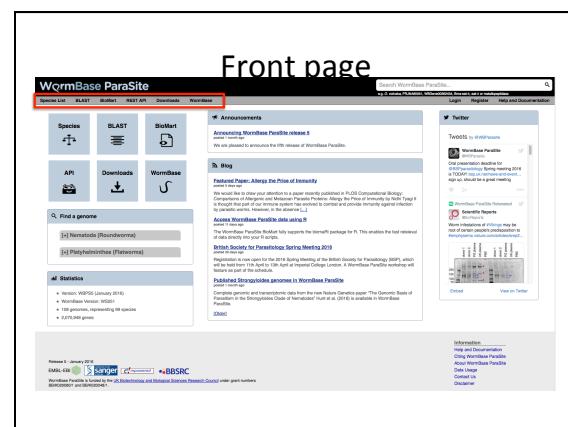
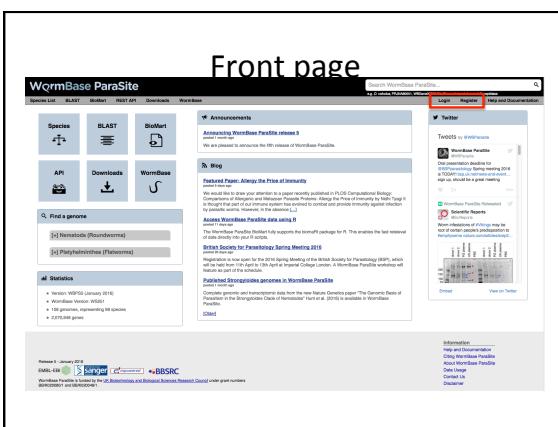
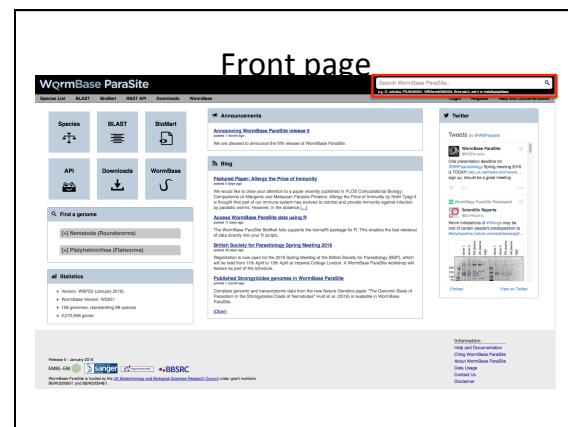
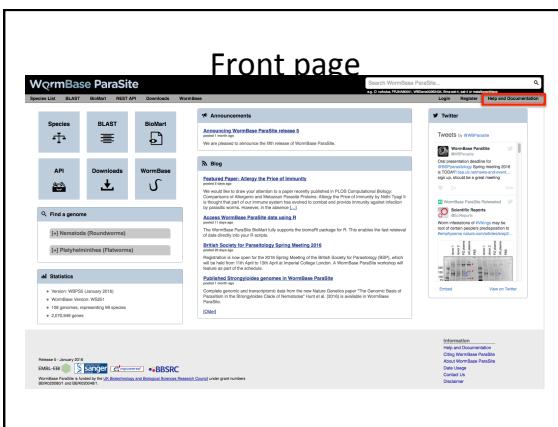
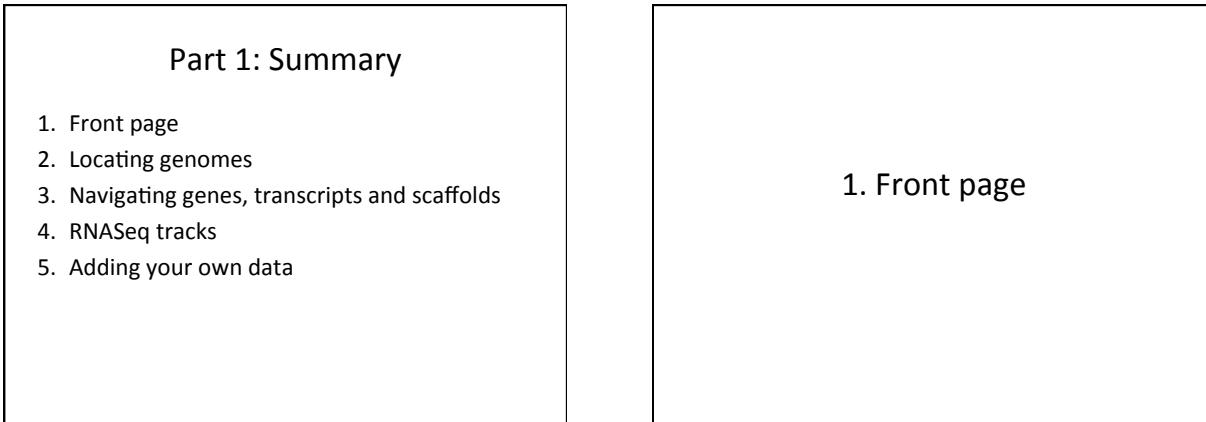
Thank you!

Part 1: Using the website

Part 1: Summary

1. Front page
2. Locating genomes
3. Navigating genes, transcripts and scaffolds
4. RNASeq tracks
5. Adding your own data

1. Front page



Front page

WormBase ParaSite

Species List BLAST BioMart PEST API Downloads Wormbase

Search Wormbase ParaSite...

Login Register Help and Documentation

Announcements

Announcing Wormbase ParaSite release 3

We are pleased to announce the 10th release of Wormbase ParaSite.

Blog

Published Paper: Allergy: the Price of Immunity

We would like to draw your attention to a paper recently published in PLOS Computational Biology. Correspondence of Allergen and Mite Parasite Proteins. This paper describes how the Wormbase ParaSite team has used the Wormbase ParaSite Perl module to support the analysis of the data presented in the paper.

Allergen

The Wormbase ParaSite Perl module supports the perl module package for P. This enables the fast retrieval of data directly from the source.

British Society for Parasitology Spring Meeting 2016

Registration is now open for the 2016 Spring Meeting of the British Society for Parasitology (BSP), which will be held from 19th to 21st April at Imperial College London. A Wormbase ParaSite workshop will take place on the 19th April.

Published Strongyloides genomes in Wormbase ParaSite

Coronavirus genomic and transcriptomic data from the new Nature Genetics paper "The Genome Basis of Parasitism in the Strongyloides Caste of Nematomorphs" (Hart et al. 2016) is available in Wormbase ParaSite.

Dataset

Information Help and Documentation Citing Wormbase ParaSite About Wormbase ParaSite Data Usage Contact Us Disclaimer

Release 1 January 2016

EMAIL: [EMAIL](mailto:EMAIL@Wormbase.org) Sanger BBSRC

Wormbase ParaSite is funded by the UK Biotechnology and Biological Sciences Research Council under grant numbers BB/N008552/1 and BB/N008553/1.

Front page

WormBase ParaSite

Species List BLAST BioMart PEST API Downloads Wormbase

Search Wormbase ParaSite...

Login Register Help and Documentation

Announcements

Announcing Wormbase ParaSite release 3

We are pleased to announce the 10th release of Wormbase ParaSite.

Blog

Published Paper: Allergy: the Price of Immunity

We would like to draw your attention to a paper recently published in PLOS Computational Biology. Correspondence of Allergen and Mite Parasite Proteins. This paper describes how the Wormbase ParaSite team has used the Wormbase ParaSite Perl module to support the analysis of the data presented in the paper.

Allergen

The Wormbase ParaSite Perl module supports the perl module package for P. This enables the fast retrieval of data directly from the source.

British Society for Parasitology Spring Meeting 2016

Registration is now open for the 2016 Spring Meeting of the British Society for Parasitology (BSP), which will be held from 19th to 21st April at Imperial College London. A Wormbase ParaSite workshop will take place on the 19th April.

Published Strongyloides genomes in Wormbase ParaSite

Coronavirus genomic and transcriptomic data from the new Nature Genetics paper "The Genome Basis of Parasitism in the Strongyloides Caste of Nematomorphs" (Hart et al. 2016) is available in Wormbase ParaSite.

Dataset

Information Help and Documentation Citing Wormbase ParaSite About Wormbase ParaSite Data Usage Contact Us Disclaimer

Release 1 January 2016

EMAIL: [EMAIL](mailto:EMAIL@Wormbase.org) Sanger BBSRC

Wormbase ParaSite is funded by the UK Biotechnology and Biological Sciences Research Council under grant numbers BB/N008552/1 and BB/N008553/1.

2. Locating genomes

Front page

WormBase ParaSite

Species List BLAST BioMart PEST API Downloads Wormbase

Search Wormbase ParaSite...

Login Register Help and Documentation

Announcements

Announcing Wormbase ParaSite release 3

We are pleased to announce the 10th release of Wormbase ParaSite.

Blog

Published Paper: Allergy: the Price of Immunity

We would like to draw your attention to a paper recently published in PLOS Computational Biology. Correspondence of Allergen and Mite Parasite Proteins. Allergy: the Price of Immunity (Tang et al. 2016). We would like to draw your attention to a paper recently published in PLOS Computational Biology. Correspondence of Allergen and Mite Parasite Proteins. Allergy: the Price of Immunity (Tang et al. 2016).

Allergen

The Wormbase ParaSite Perl module supports the perl module package for P. This enables the fast retrieval of data directly from the source.

British Society for Parasitology Spring Meeting 2016

Registration is now open for the 2016 Spring Meeting of the British Society for Parasitology (BSP), which will be held from 19th to 21st April at Imperial College London. A Wormbase ParaSite workshop will take place on the 19th April.

Published Strongyloides genomes in Wormbase ParaSite

Coronavirus genomic and transcriptomic data from the new Nature Genetics paper "The Genome Basis of Parasitism in the Strongyloides Caste of Nematomorphs" (Hart et al. 2016) is available in Wormbase ParaSite.

Dataset

Information Help and Documentation Citing Wormbase ParaSite About Wormbase ParaSite Data Usage Contact Us Disclaimer

Release 1 January 2016

EMAIL: [EMAIL](mailto:EMAIL@Wormbase.org) Sanger BBSRC

Wormbase ParaSite is funded by the UK Biotechnology and Biological Sciences Research Council under grant numbers BB/N008552/1 and BB/N008553/1.

Front page: find genomes

WormBase ParaSite

Species List BLAST BioMart PEST API Downloads Wormbase

Search Wormbase ParaSite...

Login Register Help and Documentation

Announcements

Announcing Wormbase ParaSite release 3

We are pleased to announce the 10th release of Wormbase ParaSite.

Blog

Published Paper: Allergy: the Price of Immunity

We would like to draw your attention to a paper recently published in PLOS Computational Biology. Correspondence of Allergen and Mite Parasite Proteins. Allergy: the Price of Immunity (Tang et al. 2016). We would like to draw your attention to a paper recently published in PLOS Computational Biology. Correspondence of Allergen and Mite Parasite Proteins. Allergy: the Price of Immunity (Tang et al. 2016).

Allergen

The Wormbase ParaSite Perl module supports the perl module package for P. This enables the fast retrieval of data directly from the source.

British Society for Parasitology Spring Meeting 2016

Registration is now open for the 2016 Spring Meeting of the British Society for Parasitology (BSP), which will be held from 19th to 21st April at Imperial College London. A Wormbase ParaSite workshop will take place on the 19th April.

Published Strongyloides genomes in Wormbase ParaSite

Coronavirus genomic and transcriptomic data from the new Nature Genetics paper "The Genome Basis of Parasitism in the Strongyloides Caste of Nematomorphs" (Hart et al. 2016) is available in Wormbase ParaSite.

Dataset

Information Help and Documentation Citing Wormbase ParaSite About Wormbase ParaSite Data Usage Contact Us Disclaimer

Release 1 January 2016

EMAIL: [EMAIL](mailto:EMAIL@Wormbase.org) Sanger BBSRC

Wormbase ParaSite is funded by the UK Biotechnology and Biological Sciences Research Council under grant numbers BB/N008552/1 and BB/N008553/1.

Locating genomes

Locating genomes

WormBase ParaSite

Species List BLAST BioMart PEST API Downloads Wormbase

Search Wormbase ParaSite...

Login Register Help and Documentation

Announcements

Announcing Wormbase ParaSite release 3

We are pleased to announce the 10th release of Wormbase ParaSite.

Blog

Published Paper: Allergy: the Price of Immunity

We would like to draw your attention to a paper recently published in PLOS Computational Biology. Correspondence of Allergen and Mite Parasite Proteins. Allergy: the Price of Immunity (Tang et al. 2016). We would like to draw your attention to a paper recently published in PLOS Computational Biology. Correspondence of Allergen and Mite Parasite Proteins. Allergy: the Price of Immunity (Tang et al. 2016).

Allergen

The Wormbase ParaSite Perl module supports the perl module package for P. This enables the fast retrieval of data directly from the source.

British Society for Parasitology Spring Meeting 2016

Registration is now open for the 2016 Spring Meeting of the British Society for Parasitology (BSP), which will be held from 19th to 21st April at Imperial College London. A Wormbase ParaSite workshop will take place on the 19th April.

Published Strongyloides genomes in Wormbase ParaSite

Coronavirus genomic and transcriptomic data from the new Nature Genetics paper "The Genome Basis of Parasitism in the Strongyloides Caste of Nematomorphs" (Hart et al. 2016) is available in Wormbase ParaSite.

Dataset

Information Help and Documentation Citing Wormbase ParaSite About Wormbase ParaSite Data Usage Contact Us Disclaimer

Release 1 January 2016

EMAIL: [EMAIL](mailto:EMAIL@Wormbase.org) Sanger BBSRC

Wormbase ParaSite is funded by the UK Biotechnology and Biological Sciences Research Council under grant numbers BB/N008552/1 and BB/N008553/1.

Locating genomes

Locating genomes

WormBase ParaSite

Species List BLAST BioMart PEST API Downloads Wormbase

Search Wormbase ParaSite...

Login Register Help and Documentation

Announcements

Announcing Wormbase ParaSite release 3

We are pleased to announce the 10th release of Wormbase ParaSite.

Blog

British Society for Parasitology Spring Meeting 2016

Registration is now open for the 2016 Spring Meeting of the British Society for Parasitology (BSP), which will be held from 19th to 21st April at Imperial College London. A Wormbase ParaSite workshop will take place on the 19th April.

Published Strongyloides genomes in Wormbase ParaSite

Coronavirus genomic and transcriptomic data from the new Nature Genetics paper "The Genome Basis of Parasitism in the Strongyloides Caste of Nematomorphs" (Hart et al. 2016) is available in Wormbase ParaSite.

Dataset

Information Help and Documentation Citing Wormbase ParaSite About Wormbase ParaSite Data Usage Contact Us Disclaimer

Release 1 January 2016

EMAIL: [EMAIL](mailto:EMAIL@Wormbase.org) Sanger BBSRC

Wormbase ParaSite is funded by the UK Biotechnology and Biological Sciences Research Council under grant numbers BB/N008552/1 and BB/N008553/1.

Genomes list					
Contents	Provider	Assembly	Bioproject ID	Taxonomy ID	
Nematoda (1)					
Phylum nematodes (26)					
Nematoda					
<i>Acanthocephala virens</i>	University of Edinburgh	Mh-1.0	PRJNA2636	8521	
<i>Ankylostoma caninum</i>	Genome Institute at Washington University	A_caniun_1.0	PRJNA25966	93170	
<i>Ankylostoma oxyuris</i>	Genome Institute at Washington University	A_oxyuris_1.0	PRJNA25967	93171	
<i>Ankylostoma pseudacanthicum</i>	Genome Institute at Washington University	A_pseuacanth_1.0	PRJNA25968	93172	
<i>Ankylostoma duodenale</i>	Genome Institute at Washington University	A_duodenale_2.2.0	PRJNA2581	91622	
<i>Angiostrongylus cantonensis</i>	Wellcome Trust Sanger Institute	A_cantoniensis_1.0	PRJEB1000	93145	
<i>Angiostrongylus costaricensis</i>	Wellcome Trust Sanger Institute	A_costaricensis_1.0	PRJEB1001	93146	
<i>Anisakis simplex</i>	Wellcome Trust Sanger Institute	A_simplex_1.0	PRJEB1002	93147	
<i>Anisakis setifera</i>	Wellcome Trust Sanger Institute	A_setifera_1.0	PRJEB1003	93148	
<i>Ascaris suum</i>	University of California School of Medicine	ABU_1.0	PRJNA2632	8525	
<i>Ascaris suum</i>	University of Michigan	A_suum_1.0	PRJNA2633	8526	
<i>Biprora metely</i>	Wellcome Trust Sanger Institute	B_metely_1.0	PRJNA25929	8228	
<i>Bugula heinrichi</i>	Wellcome Trust Sanger Institute	B_heinrichi_1.0	PRJEB1004	93149	
<i>Buaphelomys xylophilus</i>	Wellcome Trust Sanger Institute	ABM001_1.0	PRJNA2634	8226	
<i>Cyathostoma cestorum</i>	Wellcome Trust Sanger Institute	C_cestorum_1.0	PRJEB1005	93150	
<i>Dirofilaria vivipara</i>	University of Edinburgh	DOV_1.0	PRJEB1018	91712	
<i>Dirofilaria vivipara</i>	Genome Institute at Washington University	D_vivipara_3.0.1	PRJNA2597	93152	
<i>Dirofilaria repens</i>	Genome Institute at Washington University	D_repens_3.0.1	PRJNA2598	93153	
<i>Dirofilaria medenensis</i>	Wellcome Trust Sanger Institute	D_medenensis_1.0	PRJEB1003	93149	
<i>Elaeophora sibirica</i>	Wellcome Trust Sanger Institute	E_sibirica_1.0	PRJEB1002	93148	
<i>Enterogobius vermicularis</i>	Wellcome Trust Sanger Institute	E_vermicularis_1.0	PRJEB1004	93150	
<i>Globovittula palmae</i>	Wellcome Trust Sanger Institute	GPAL1_0	PRJEB1131	36500	
<i>Gymnophorus eichhorniae</i>	Wellcome Trust Sanger Institute	GEICH_1.0	PRJEB1132	36501	
<i>Hemachirus concolor</i>	Wellcome Trust Sanger Institute	H.concolor_1.0	PRJEB1005	93151	
<i>Hemachirus concolor</i>	University of Melbourne	H.concolor_1.0	PRJNA2630	85188	
<i>Hemachirus concolor</i>	Wellcome Trust Sanger Institute	H.concolor_1.0	PRJEB1006	93152	
<i>Heligmosomoides polygyrus</i>	Wellcome Trust Sanger Institute	H.polygyrus_1.0	PRJEB1003	93149	

Genome pages

WormBase ParaSite

Species List BLAST BrAPI ANI Downloads Wormbase

Submit new species [View Species](#) [View BLAST](#)

Echinococcus multilocularis

BarcodeDB [Data Sources](#) [Submit Data](#) [Help](#) [Feedback](#) [Log in](#) [Register](#) [Help and Documentation](#)

Echinococcus multilocularis

BarcodeDB [Data Sources](#) [Submit Data](#) [Help](#) [Feedback](#) [Log in](#) [Register](#) [Help and Documentation](#)

Search Echinococcus multilocularis (PRJEB124)

About Echinococcus multilocularis

The *Echinococcus* multilocularis is a member of the Cyclophyllidae, which comprise the majority of tapeworms that are of medical importance. The disease arachnoiditis echinococcosis, which is caused by the larval stage of *E. multilocularis*, is considered to be one of the most dangerous worm infections in the world. *E. multilocularis* has recently spread in Europe, Japan, and North America areas that were previously free of the parasite leading to the reclassification of several echinococcosis as arachnoiditis echinococcosis.

Genome assembly: PRJEB124

PRJEB124 (Echinococcus multilocularis) was submitted by the Parasite Genomics group at the Max-Planck-Institut für Infection Biologie, Berlin, Berlin, Germany in collaboration with Olaf Reichhardt University of Heidelberg, The Institute of Parasitology, University of Bern, Switzerland, and the Institute of Parasitology, University of Bern, Switzerland. The genome was released, and the genome represented here since PrjRelease 3 (January 2016) in HCB Data Bank.

Example region

Comparative genomics

What can I do? Homologs, para trees, and whole genome alignments across multiple species

Example gene tree

Tools

- Search for annotations in the genome and proteome using BLAST
- From any site, click the Wormbase Protein Master page linking tool
- Proteome search using the Wormbase Protein Master page
- Predict the effects of variants using the Variant Effect Predictor

Example gene

Gene annotation

The original gene predictions were made by the Parasite Genomics group at the Max-Planck-Institut für Infection Biologie, Berlin, Berlin, Germany. Instead of describing the genes from scratch, they have been taken from the genome assembly PRJEB124. The genes have since been refined by the Parasite Genomics group at the Max-Planck-Institut für Infection Biologie, Berlin, Berlin, Germany.

Example gene

Downloads

- Genomic Data (FASTA)
- Genomic Data (GFF3)
- Orthologous Genes (FASTA)
- Protein (FASTA)
- Protein (GFF3)
- CDB annotations (FASTA)

Example protein

3. Navigating genes, transcripts and scaffolds

The figure shows a screenshot of the WormBase ParaSite Gene pages. The main title is "Gene pages". Below it, the specific gene page for "Gene SAT-1 (4402)" is displayed. The page includes a search bar, a navigation menu, and a detailed gene information panel. The panel contains sections for "Basic Information", "Gene Structure", "Protein Domains", "Gene Expression", "Gene Function", "Gene Ontology", "Gene Interactions", "Gene Variants", "Gene Phenotype", "Gene Disease", "Gene Pathway", and "Gene References". A "Summary" section at the bottom provides a brief overview of the gene's characteristics.

Gene pages: exons

The figure shows a screenshot of a bioinformatics tool interface. At the top, the title "Gene pages: exons" is displayed above a sub-header "Marked-up sequence". Below this are two buttons: "Download sequence" and "BLAST this sequence". A navigation bar below the buttons includes links for "Exons", "DAT1 exons", and "All exons in this region". The main content area displays a sequence of DNA bases (A, T, C, G) with specific regions highlighted in blue and red. A legend at the bottom left defines the color coding: blue for "Coding sequence", red for "UTR", green for "Intron", and black for "Exon". The sequence itself is a long string of nucleotides.

GO terms

WormBase ParaSite

Gene: SAT1 (eng. Bm0018)

Description: Protein with unknown function. It encodes a protein (Source: UniProtKB/Proteins)

About this gene: This gene has a UniProtKB/Swiss-Prot entry.

Gene ID: Bm0018

Annotation Method: Gene models from Wormbase Trust, Ganger, Genome annotation

Transcripts:

GO: Molecular function

Annotation	Term	Evidence	Associated Model	UniProt ID
GO:0003650	Protein binding	Experimental	Bm0018	Q8BQD2
GO:0003650	GTPase activity	Experimental	Bm0018	Q8BQD2
GO:0003650	Structure constuction of polypeptide	Experimental	Bm0018	Q8BQD2
GO:0003650	GTP binding	Experimental	Bm0018	Q8BQD2

Transcript pages: summary

Transcript: Bm072

Description: Larval antigen (Source: UniProtKB/Proteins; Acc: Q8BQD2)

Location: SuperContig_Bm1_v2_scaf0013779_108_301-11172

About this transcript: This transcript is a product of gene Bm0018 (UniProtKB/Swiss-Prot: Q8BQD2).

Transcript columns (1 hidden):

Name	Transcript ID	Re	Protein	Translation ID	Start	End	Motif	UniProt
Bm072	Bm072	975	2420	Bm072	Protein coding	Q8BQD2		
Bm071	Bm071	966	2524	Bm072	Protein coding	Q8BQD2		

Summary:

Statistics: Exons: 1 Coding exons: 1 Translational length: 275 lps Translation length: 247 residues

Prediction Method: Protein-coding model imported from WormBase

Information: Help and Documentation — About WormBase ParaSite — Gene Search — Contact Us — Disclaimer

Transcript pages: navigating

Transcript: Bm072

Description: Larval antigen (Source: UniProtKB/Proteins; Acc: Q8BQD2)

Location: SuperContig_Bm1_v2_scaf0013779_108_301-11172

About this transcript: This transcript is a product of gene Bm0018 (UniProtKB/Swiss-Prot: Q8BQD2).

Summary:

Statistics: Exons: 1 Coding exons: 1 Translational length: 275 lps Translation length: 247 residues

Prediction Method: Protein-coding model imported from WormBase

Information: Help and Documentation — About WormBase ParaSite — Gene Search — Contact Us — Disclaimer

Transcript pages: protein domains

Protein summary:

Protein domains for Bm7483.1

Statistics:

- Avg. residue weight: 109.83 g/mol
- Charge: 4.5
- Isoelectric point: 7.7259
- Molecular weight: 52,170.59 g/mol
- Number of residues: 470 aa

Navigating: tabs

WormBase ParaSite

Gene: Bm072

Description: Larval antigen (Source: UniProtKB/Proteins; Acc: Q8BQD2)

About this transcript: This transcript is a product of gene Bm0018 (UniProtKB/Swiss-Prot: Q8BQD2).

Transcripts:

Summary:

Statistics: Exons: 1 Coding exons: 1 Translational length: 275 lps Translation length: 247 residues

Prediction Method: Protein-coding model imported from WormBase

Information: Help and Documentation — About WormBase ParaSite — Gene Search — Contact Us — Disclaimer

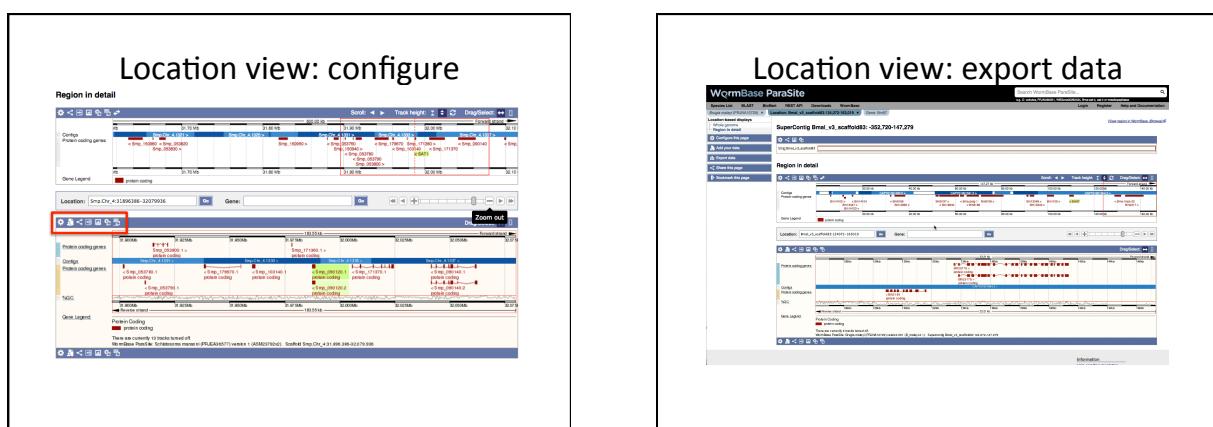
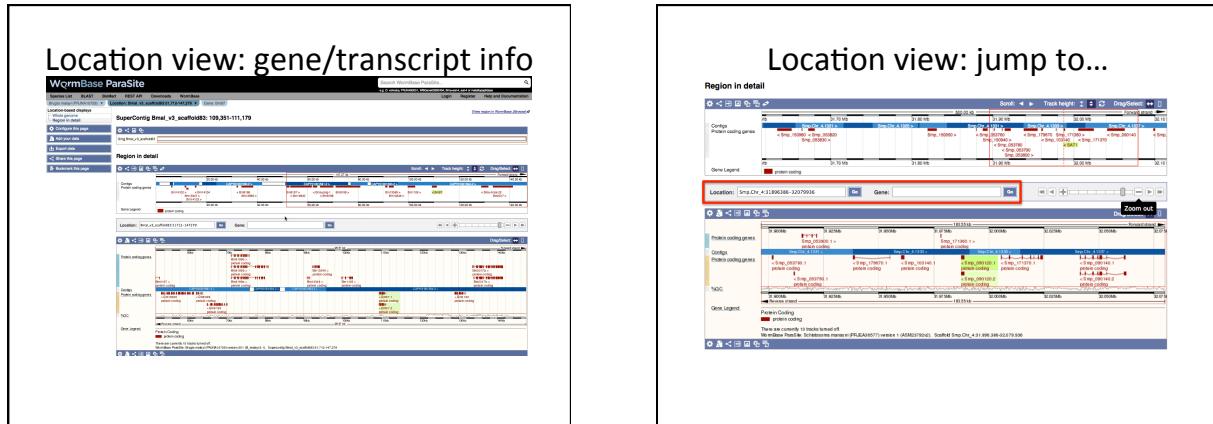
Location view: zooming

Region in detail:

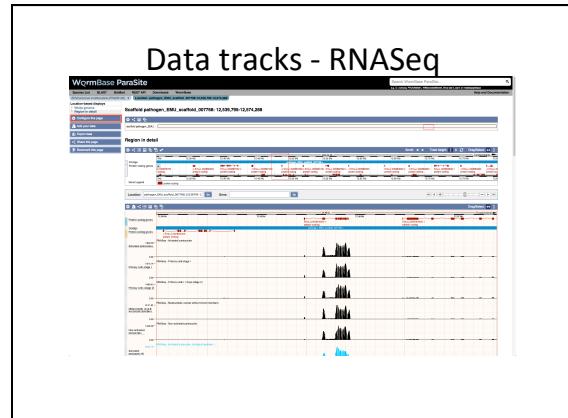
SuperContig_Bm1_v2_scaf0013779

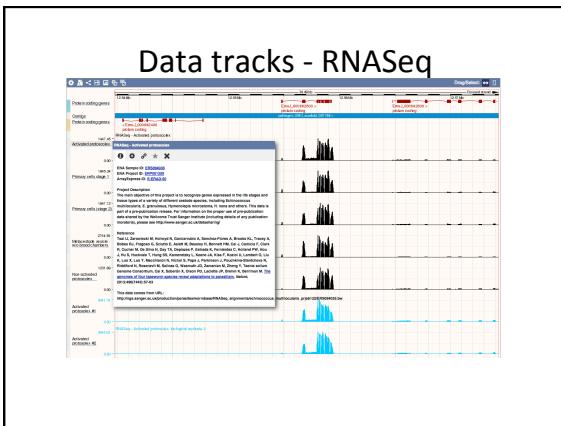
Location: Bm072

Information: Help and Documentation — About WormBase ParaSite — Gene Search — Contact Us — Disclaimer

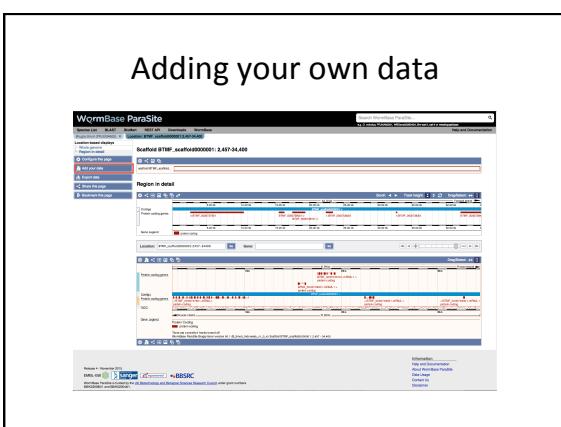


4. RNASeq tracks

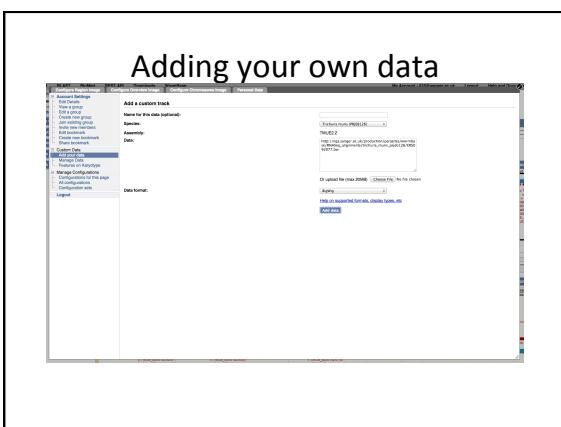




5. Adding your own data



Adding your own data



Part 1b: Browsing the website

Searching the website
Comparative genomics
User accounts



WormBase ParaSite

Species List BLAST BLURB REST API Downloads WormBase

[Species List](#)

[BLAST](#)

[BLURB](#)

[REST API](#)

[Downloads](#)

[WormBase](#)

[Find a gene](#)
[Statistics](#)
[About](#)
[Contact](#)

Searching

[Search WormBase ParaSite](#)

Annotations

Annotating WormBase ParaSite release 3

We are pleased to announce the 11th release of WormBase ParaSite.

Bug

Feature Paper: Always at the Price of Integrity

We would like to draw your attention to a paper published in PLOS Computational Biology, which highlights the potential risks of using automated annotation tools. The paper shows that it is possible that our own annotation system has worked so hard and provides seemingly accurate information by introducing errors.

Access WormBase ParaSite data using R

The WormBase ParaSite R library supports the [WormBase package](#) for R. This enables the fast retrieval of WormBase data from within R. The WormBase package is available on CRAN and can be installed via `install.packages("WormBase")`. We hope that you will find it useful for your work.

Publish! Shareomics: generate a WormBase ParaSite genome

Complete genomic and transcriptomic data from the new *Near Genotype* paper ("The Genome Basis of Resistance to Metabolite Stress in *Caenorhabditis elegans*," 2018) is available in WormBase ParaSite.

[View all news items](#)

Statistics

- Version: WMP15 (Lokur, 2018)
- 100 species
- 1000+ genes
- 100+ gene families
- 100+ orthologous groups

About

WormBase ParaSite is funded by the US National Institutes of Health and Wellcome Trust.

WormBase ParaSite is a service of the US National Bioinformatics and Behavioral Research Council under grant numbers

Contact

Support: [support@wormbase.org](#)
Feedback: [feedback@wormbase.org](#)

Information

Help with Documentation
About WormBase ParaSite
About WormBase

The screenshot shows the WormBase ParaSite search results for the query "eat-4". The results are displayed in two main sections: "WormBase Parasite" and "C. elegans".

WormBase Parasite Results

- Showing 1-10 of 82 Genes Found in WormBase Parasite
- Filter by species: Search & sort: eat-4

Gene ID	Description	Species	Location
eat-4 (Wormbase[1])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[2])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[3])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[4])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[5])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[6])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[7])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[8])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[9])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[10])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]

C. elegans Results

- Showing 1-10 of 10 Genes Found in C. elegans
- Filter by species: Search & sort: eat-4

Gene ID	Description	Species	Location
eat-4 (Wormbase[1])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[2])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[3])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[4])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[5])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[6])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[7])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[8])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[9])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]
eat-4 (Wormbase[10])	Probable vesicular glutamate transporter homolog eat-4 [Source:UniProtKB/Swiss-Prot;Acc:P54644]	<i>Wormbase</i>	Genomic location 8.03915-9.11382 C. elegans [Bioscale_id:2]

Filtering search results

areaSite

Site (83)

Filter by species: Select a species... Select a species... [eat-4 \[Source:UniProtKB\]](#)

eat-4 | WB

Description *Ascaris suum* (PRJNA1366)

Gene ID *PRNA243174*

Species *Ascaris suum* (PRJNA1366)

Location *Brugia malayi* (PRJNA10729)

Cbn-eat-4 | WB

Description *Caenorhabditis elegans* (PRJEA4437)

Gene ID *PRJEB16981*

Species *Dicyclonus viviparus* (PRJEB5116)

Location *Dicyclonus viviparus* (PRJNA72587)

[Cite | Config | View region in WormBase | Download](#)

Comparative Genomics

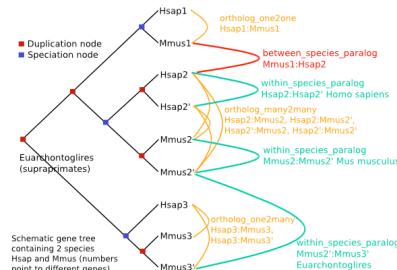
Introduction

- During each release, we compute phylogenetic trees
 - Every gene is included from 120 species:
 - 99 helminths
 - 9 free-living nematodes
 - 12 comparator species (e.g. human, mouse, etc)
 - Determine orthologues and paralogues

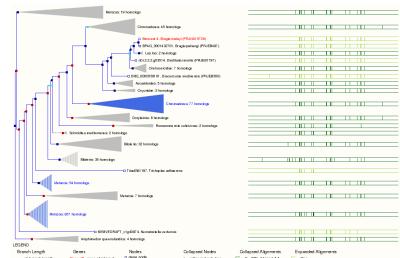
Homology types

- Orthologues: any gene pairwise relation where the ancestor node is a speciation event
 - 1-to-1 orthologue
 - 1-to-many orthologue
 - Many-to-many orthologue
 - Paralogues: any pairwise relation where the ancestor node is a duplication event

Understanding the gene tree



Visual access to the trees



Tabular access to tree data

Selected orthologs						
	Type	dN/dS	Showhide columns	Compare	Location	Filter
Arachnididae Mites (PUK42426)	1-to-1	n/a	oMe1.0_1.g0135	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	oMe1.0_00007_00007_07990_1	79 79
Probable insect-like protostomes (PUK42426)	1-to-1	n/a	oMe1.0_1.g0135	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	oMe1.0_00007_00007_07990_1	79 79
Arthropoda Crustacea (PUK42426)	Many-to-many	n/a	Asy1_210710_P	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	Contig13047_00019_00000_1.g	27 28
Probable insect-like protostomes (PUK42426)	Many-to-many	n/a	Asy1_210711_P	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	Contig13047_00020_00000_1.g	26 26
Probable insect-like protostomes (PUK42426)	Many-to-many	n/a	Asy1_210712_P	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	Contig13047_00019_00001_1.g	28 29
Probable insect-like protostomes (PUK42426)	Many-to-many	n/a	Asy1_210713_P	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	Contig13047_00019_00001_1.g	28 29
Probable insect-like protostomes (PUK42426)	Many-to-many	n/a	Asy1_210714_P	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	Contig13047_00019_00001_1.g	28 30
Probable insect-like protostomes (PUK42426)	Many-to-many	n/a	Asy1_210715_P	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	Contig13047_00019_00001_1.g	28 30
Probable insect-like protostomes (PUK42426)	Many-to-many	n/a	ANCCAN00PT_Contig116_000003_302980_1	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	ANCCAN00PT_Contig116_000003_302980_1	68 60
Probable insect-like protostomes (PUK42426)	1-to-1	n/a	Asy1_210716_P	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	Asy1_210716_P0f_000003_412798_1	66 67
Probable insect-like protostomes (PUK42426)	1-to-1	n/a	Asy1_210717_P	• Alignment (protein) • Alignment (DNA) • Gene Tree (protein) • Gene Tree (DNA)	Asy1_210717_P0f_000003_412798_1	66 67

User Accounts

User accounts

- Saving and sharing attached data tracks
- Saving configuration settings
- Saving and sharing BLAST results

User accounts

User accounts: registering

User accounts

Part 2: Sequence Similarity Search using BLAST

What is BLAST?

- BLAST = Basic Local Alignment Search Tool
- Sequence similarity tool
- Allows comparison of a **query** sequence, against a **database** of sequences
- Query = your nucleotide or protein sequence
- Database = the genome or proteome of any species

What is BLAST?

- Input:
Nucleotide or protein sequence
Search Parameters
- Output:
List of all hits ranked in order of statistical significance

Types of BLAST

BLAST Type	Query Sequence	Target Database
BLASTN	Nucleotide	Genome (nucleotide)
BLASTP	Peptide	Proteome (peptide)
BLASTX	Six frame translation of a nucleotide sequence	Proteome (peptide)
TBLASTX (slowest)	Six frame translation of a nucleotide sequence	Six frame translation of genome
TBLASTN	Peptide	Six frame translation of genome

Using the ParaSite BLAST

WormBase ParaSite

Species List **BLAST** BioMart REST API Downloads WormBase

Brugia malayi [Bm2147](#) [Bm2147]

Gen-based displays

- Species
- Splice variants
- Sequence
- External references

Ontologies

- GO: Molecular function
- GO: Cellular component

Gene: Bm2147 WBGene00222408

Location: SuperContig Bm1_v3_scaffold1[3,207,430-3,210,542] Gene: Bm2147

Location

About this gene

Gene type

SuperContig Bm1_v3_scaffold1[3,207,430-3,210,542] forward

This gene has 2 transcripts (splice variants) and 112 orthologues

Protein coding

Defaults to the species you are currently browsing

The figure shows a screenshot of the ParaSite BLAST interface. At the top, the title "Using the ParaSite BLAST" is displayed. Below it, a search result is shown for a query sequence. The result includes the following details:

- Gene type:** Protein coding
- Annotation Method:** Protein-coding model imported from WormBase®
- Transcripts:** A link to "Show transcript table".

Below this, a section titled "Marked-up sequence" contains the query sequence and its alignment with a target sequence. The target sequence is highlighted with a red box. The alignment shows matches in green and mismatches in red. The target sequence is as follows:

```

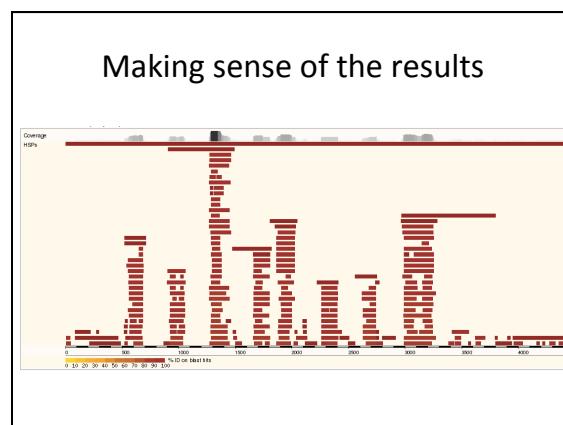
>WBGene00145163_Balmy-3_1-bal3_v3_startorf1d|2264850|21114451
AATTCCTCTGATTTCTGGCTCTTACCAACNTTAACTTGATCTTCTTGT
CIACTCTAGCTACTCTACAGAGAGATCAGCCCTTTTGAATTGAAACCGAGCTAAC
TGAGAGCTCTATTAATTTTGATCTTGCTGTTGATCTCTTCTACACCTGCTACAT
TAAATAATTTAGGATGTTGAGCTTATTAATGAGCTTATGTTAGAGCATAAAGCT
CTTACCTTCTGATCTACGGCTAAATTAATATGAGTGTGTTCTAACTT
CAGTTTCTGATGAGCTTCTAGTATTCTGAACTTCTGTTGTTCTAACTT
ATGTTCTGATCTACGGCTAAATTAATATGAGTGTGTTCTAACTT
TAAATGAGTTGACTCTTATTGATCTTCTGTTGTTGTTCTAACTT
ATGAAATGAGCTTATTAATGATCTTCTGTTGTTGTTCTAACTT

```

Below the target sequence, a red box highlights the first few lines of the sequence: "TAAATAATTTAGGATGTTGAGCTTATTAATGAGCTTATGTTAGAGCATAAAGCT". To the right of this box, the text "Part of aligned sequence" is displayed.

Making sense of the results

- Score
Used to assess the biological relevance by describing the alignment quality
Higher score = higher similarity
- E-value
Similar to (but not the same as) a *p*-value that has been corrected for multiple testing - decreases exponentially as the score increases
Lower *E*-value = more significant result
- %ID
Percentage of your query sequence that matches the genome/proteome database



Part 4: Data-mining with BioMart

Data-mining with BioMart

The screenshot shows the WormBase ParaSite BioMart interface. On the left, a sidebar titled 'Dataset' lists options: 'All Species (WBPS4)' (selected), 'None selected', 'Attributes', 'Genome project', and 'Gene stable ID'. On the right, a search bar says 'Search WormBase Parasite...' and a 'BLAST' button is visible. Below the search bar is a section titled 'Please restrict your query using criteria below' with a note '(If filter values are truncated in any lists, hover over the list item to see the full text)'. A list of filters includes: SPECIES, REGION, ID, GENE, GENE ONTOLOGY, MULTI-SPECIES COMPARISONS, and PROTEIN DOMAINS.

Setting filters

The screenshot shows the WormBase ParaSite BioMart interface. On the left, a sidebar lists 'Dataset' options: 'All Species (WBPS4)' (selected), 'None selected', 'Attributes', 'Genome project', and 'Gene stable ID'. On the right, a search bar says 'Search WormBase Parasite...' and a 'BLAST' button is visible. Below the search bar is a section titled 'Please restrict your query using criteria below' with a note '(If filter values are truncated in any lists, hover over the list item to see the full text)'. A list of filters includes: SPECIES, REGION, ID, GENE, GENE ONTOLOGY, MULTI-SPECIES COMPARISONS, and PROTEIN DOMAINS. The 'SPECIES' filter is highlighted with a red border.

- **SPECIES:** Use this filter to select either individual genomes or nematode clades.
 - Multiple genomes can be selected by holding down the ctrl key or the option key on a Mac.

- **REGION:** Restrict to a particular genomic region.
 - Should only be used where a single genome has been selected, as it is possible that a particular region is present in multiple genomes.
 - If start/end co-ordinates are being specified, a scaffold or chromosome id is always required.
 - Where multiple regions are specified, the format is 'Scaffold/Chr:Start:End:Strand' e.g. AG00032:411187:446321:1.
 - If no strand is specified, both strands are selected.
 - Regions should be separated by a comma or new line.

- **GENE:** Specify a list of genes with WormBase IDs, or one of the other ID types listed.
 - IDs should be separated by a new line.

- **GENE ONTOLOGY:** Restrict by one or more Gene Ontology (GO) terms for functional descriptions.
 - Paste or upload a list of GO IDs or use the autocomplete box to populate the list.
 - Alternatively restrict to a particular GO evidence type e.g. Inferred by Electronic Annotation (IEA).
 - Multiple codes can be selected by holding down the ctrl key, or option key on a Mac.

- **PROTEIN DOMAINS:** Allows you to restrict your query based on the presence or absence of protein domains.
 - **Limit to genes**...lets you choose a particular database feature set in include or exclude e.g. "restrict to all proteins containing any feature found in Pfam".
 - **Limit to genes with these family or domain IDs**:, allows you to restrict to one or more protein domains/families.
 - Accepts IDs from several databases including InterPro, Pfam and Panther. IDs should be separated by a new line.

BioMart output

Setting Attributes (output): features

<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> WormBase ParaSite Home Search Help Logout </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> Detailed View Edit Delete Filters None Advanced Sort None Alphabetical Descending Genome ID None </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> Features <input checked="" type="checkbox"/> Structures <input type="checkbox"/> Sequences <input type="checkbox"/> Genes <input type="checkbox"/> Genes & Transcripts <input type="checkbox"/> Genomic project <input type="checkbox"/> Genomic region <input type="checkbox"/> Taxonomy ID </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> Gene Attributes <input checked="" type="checkbox"/> Gene ID <input type="checkbox"/> Organism ID <input type="checkbox"/> Transcript stable ID <input type="checkbox"/> Product name <input type="checkbox"/> Chromosome/Contig ID <input type="checkbox"/> Description <input type="checkbox"/> Synonyms <input type="checkbox"/> Shared <input type="checkbox"/> Translated (tbl) <input type="checkbox"/> Translated (tblp) </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> External References <input type="checkbox"/> Chado gene ID <input type="checkbox"/> Chado transcript ID <input type="checkbox"/> Chado contig ID <input type="checkbox"/> RefSeq ID <input type="checkbox"/> UniProtKB/Swiss-Prot ID <input type="checkbox"/> UniProtKB/TrEMBL ID </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> GO ID <input type="checkbox"/> None <input type="checkbox"/> All <input type="checkbox"/> GO term name <input type="checkbox"/> GO term ID <input type="checkbox"/> GO domain </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> GOSim ID <input type="checkbox"/> None </div>	<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> Please select columns to be included in the output and 'Info' when ready. </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> <input type="checkbox"/> Assembly information <input type="checkbox"/> RefSeq ID <input type="checkbox"/> WormBase ID </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> <input type="checkbox"/> Gene name <input type="checkbox"/> Source of gene name <input type="checkbox"/> Description <input type="checkbox"/> Gene family <input type="checkbox"/> Chromosome <input type="checkbox"/> Taxon ID <input type="checkbox"/> Shared <input type="checkbox"/> Translated (tbl) <input type="checkbox"/> Translated (tblp) </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> <input type="checkbox"/> Wormbase gene ID <input type="checkbox"/> Wormbase transcript ID <input type="checkbox"/> Wormbase contig ID <input type="checkbox"/> Wormbase taxon ID <input type="checkbox"/> Wormbase allele ID <input type="checkbox"/> Wormbase ID </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> <input type="checkbox"/> GO term evidence code <input type="checkbox"/> GO domain </div>
--	---

Setting Attributes (output): structures

WormBase Parasite

Database Name: **Parasite**

Search **Logout** **Help** **Submit** **Download**

Dataset WormBase (WPDB) Filters Gene identifier Gene symbol Gene stable ID Attributes Species Organism Genome name Transliteration ID +GENE Gene Attributes Chromosome ID Protein code ID Chromosome location name Gene start (bp) Gene end (bp) Strand Transliteration start (bp) Transliteration end (bp)	Features → Wormbase → Structure → Sequences → Genes → Genes → Genes → Assembly accessions DBProject ID Remarks table	Gene name Gene description Gene location Gene description Transliteration count Chromosome ID Strand UTR start UTR end	+EXON Gene Information Gene ID Gene name (bp) Gene length (bp) Gene start (bp) Gene end (bp) Gene start in transcript Gene end in transcript	+CDS CDS coding start CDS coding end CDS location CDS description CDS start (bp) CDS end (bp) CDS length (bp)
--	---	--	--	--

Please select column(s) to be included in the output and hit **Results** when ready.

Setting Attributes (output):

<div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> WormBase Parasite Home Parasite Home </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> Dataset <input type="checkbox"/> WBSTC <input type="checkbox"/> Filters <input type="checkbox"/> Search <input type="checkbox"/> Help </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> Species <input type="checkbox"/> All species <input type="checkbox"/> H. contortus <input type="checkbox"/> H. meleagridis <input type="checkbox"/> H. nelsoni <input type="checkbox"/> H. rostratus <input type="checkbox"/> H. vermicularis <input type="checkbox"/> N. brasiliensis <input type="checkbox"/> S. haemolyticum <input type="checkbox"/> S. stercoralis <input type="checkbox"/> T. suis <input type="checkbox"/> T. trichiura <input type="checkbox"/> Unclassified </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> Gene Attribute <input type="checkbox"/> All genes <input type="checkbox"/> Protein coding <input type="checkbox"/> Protein isoform <input type="checkbox"/> Protein state <input type="checkbox"/> Protein isoform state <input type="checkbox"/> Gene <input type="checkbox"/> Gene part <input type="checkbox"/> Gene alias <input type="checkbox"/> Gene name </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> Orthologues <input type="checkbox"/> Conservation <input type="checkbox"/> Homology <input type="checkbox"/> Protein <input type="checkbox"/> Protein isoform <input type="checkbox"/> Protein state <input type="checkbox"/> Protein isoform state <input type="checkbox"/> Gene <input type="checkbox"/> Gene part <input type="checkbox"/> Gene alias <input type="checkbox"/> Gene name </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> Disease <input type="checkbox"/> Diseases <input type="checkbox"/> Pathogen <input type="checkbox"/> Phenotype <input type="checkbox"/> Phenotype </div> <div style="border: 1px solid #ccc; padding: 5px; margin-bottom: 10px;"> Human Orthologues <input type="checkbox"/> All orthologs <input type="checkbox"/> Protein coding <input type="checkbox"/> Protein isoform <input type="checkbox"/> Protein state <input type="checkbox"/> Protein isoform state <input type="checkbox"/> Gene <input type="checkbox"/> Gene part <input type="checkbox"/> Gene alias <input type="checkbox"/> Gene name </div>	<p>Please select columns to be included in the output and hit 'Submit' when ready.</p> <table border="0"> <tr> <td style="vertical-align: top; width: 30%;"> <input type="checkbox"/> Features <input type="checkbox"/> Homologs <input type="checkbox"/> Structures <input type="checkbox"/> Taxonomy <input type="checkbox"/> Sequences </td> <td style="vertical-align: top; width: 30%;"> <input type="checkbox"/> Assembly <input type="checkbox"/> Genomic ID <input type="checkbox"/> Genomic Name <input type="checkbox"/> Genomic Type </td> <td style="vertical-align: top; width: 30%;"> <input type="checkbox"/> Assembly accessible <input type="checkbox"/> BioMart ID <input type="checkbox"/> BioMart name </td> </tr> <tr> <td colspan="3"> <input type="checkbox"/> Gene name <input type="checkbox"/> Gene alias <input type="checkbox"/> Gene description <input type="checkbox"/> Gene ID <input type="checkbox"/> NCBI ID <input type="checkbox"/> UniProt ID </td> </tr> <tr> <td colspan="3"> <input type="checkbox"/> Homology type <input type="checkbox"/> Identity <input type="checkbox"/> Length <input type="checkbox"/> Log-odds <input type="checkbox"/> P-value <input type="checkbox"/> Score </td> </tr> <tr> <td colspan="3"> <input type="checkbox"/> Homology type <input type="checkbox"/> Homologous regions <input type="checkbox"/> Homology <input type="checkbox"/> Identity <input type="checkbox"/> Length <input type="checkbox"/> Log-odds <input type="checkbox"/> P-value <input type="checkbox"/> Score </td> </tr> </table>	<input type="checkbox"/> Features <input type="checkbox"/> Homologs <input type="checkbox"/> Structures <input type="checkbox"/> Taxonomy <input type="checkbox"/> Sequences	<input type="checkbox"/> Assembly <input type="checkbox"/> Genomic ID <input type="checkbox"/> Genomic Name <input type="checkbox"/> Genomic Type	<input type="checkbox"/> Assembly accessible <input type="checkbox"/> BioMart ID <input type="checkbox"/> BioMart name	<input type="checkbox"/> Gene name <input type="checkbox"/> Gene alias <input type="checkbox"/> Gene description <input type="checkbox"/> Gene ID <input type="checkbox"/> NCBI ID <input type="checkbox"/> UniProt ID			<input type="checkbox"/> Homology type <input type="checkbox"/> Identity <input type="checkbox"/> Length <input type="checkbox"/> Log-odds <input type="checkbox"/> P-value <input type="checkbox"/> Score			<input type="checkbox"/> Homology type <input type="checkbox"/> Homologous regions <input type="checkbox"/> Homology <input type="checkbox"/> Identity <input type="checkbox"/> Length <input type="checkbox"/> Log-odds <input type="checkbox"/> P-value <input type="checkbox"/> Score		
<input type="checkbox"/> Features <input type="checkbox"/> Homologs <input type="checkbox"/> Structures <input type="checkbox"/> Taxonomy <input type="checkbox"/> Sequences	<input type="checkbox"/> Assembly <input type="checkbox"/> Genomic ID <input type="checkbox"/> Genomic Name <input type="checkbox"/> Genomic Type	<input type="checkbox"/> Assembly accessible <input type="checkbox"/> BioMart ID <input type="checkbox"/> BioMart name											
<input type="checkbox"/> Gene name <input type="checkbox"/> Gene alias <input type="checkbox"/> Gene description <input type="checkbox"/> Gene ID <input type="checkbox"/> NCBI ID <input type="checkbox"/> UniProt ID													
<input type="checkbox"/> Homology type <input type="checkbox"/> Identity <input type="checkbox"/> Length <input type="checkbox"/> Log-odds <input type="checkbox"/> P-value <input type="checkbox"/> Score													
<input type="checkbox"/> Homology type <input type="checkbox"/> Homologous regions <input type="checkbox"/> Homology <input type="checkbox"/> Identity <input type="checkbox"/> Length <input type="checkbox"/> Log-odds <input type="checkbox"/> P-value <input type="checkbox"/> Score													

Practical exercises: part 1

"I'd like to extract all *C. elegans* orthologs for *Nippostrongylus* genes involved in a particular process."

1. In the SPECIES menu select *Nippostrongylus*
 2. In the MULTI-SPECIES COMPARISONS menu select **Orthologous *C. elegans* genes -> Only**
 3. Further refine this list by function, process or location by choosing one or more categories from the GENE ONTOLOGY list.
 - Start typing in the upper box and choose your terms of interest from the autocomplete, they will be added to the box beneath.
 4. Click the **Results** button (top left) to see your results. By default a two-column file is returned that contains gene ID and Genome Project. To configure different options for the output, select **Attributes** in the left menu.

"I have a list of genes from *Ascaris suum* and would like to know which ones have orthologs in humans and mammals and which ones might be nematode-specific."

- In the GENE menu paste in your gene list
- in the MULTI-SPECIES COMPARISONS select **Orthologous human genes -> Excluded**
- You can also run this query against mouse orthologs by selecting **Orthologous mouse genes -> Excluded** (the results are the same in this case)
- Click the Results button (top left) to see your results. By default a two-column file is returned that contains gene ID and Genome Project. To configure different options for the output, select **Attributes** in the left menu.

The screenshot shows the WormBase ParaSite search interface. On the left, there is a sidebar with various filters like 'Dataset', 'Species', 'Filters', 'Regions', 'Sequences', and 'Attributes'. The 'Species' section is expanded, showing 'Orthologous Human genes' selected under 'Multi-species comparisons' and 'Excluded' selected under 'Orthologous genes'. The main search area has a 'Please enter your query using genome names' input field. Below it, there are sections for 'SPECIES' (with 'Orthologous genes' checked), 'GENE TYPES' (with 'Excluded' checked), and 'SEQUENCES' (with 'Excluded' checked). At the bottom right of the search area, there is a 'Results' button.

"I need the sequences for a set of *Schistosoma mansoni* genes. I have the chromosome, start, and stop for each."

- From the **SPECIES** filter choose *Schistosoma mansoni*.
- Open the **REGION** section and enter the list of co-ordinates under 'Multiple regions' separated by commas or new lines.
- In **Attributes**, check the **Sequences** option, then in the **SEQUENCES** section choose **Unspliced (genes)**.
- Click the **Results** button

The screenshot shows the WormBase ParaSite search interface. The 'SEQUENCES' section in the sidebar is expanded, showing 'Unspliced (genes)' selected. The main search area has a 'Please enter your query using genome names' input field. Below it, there are sections for 'SPECIES' (with 'Schistosoma mansoni' selected), 'GENE TYPES' (unchecked), and 'SEQUENCES' (with 'Unspliced (genes)' checked). At the bottom right of the search area, there is a 'Results' button.

"I need a list of genes with predicted signal peptide that are present in *Brugia malayi* a given organism but not present in *C. elegans*."

- In the **SPECIES** section choose *Brugia malayi*, then in the **MULTI-SPECIES COMPARISONS** select **Orthologous C. elegans genes -> Excluded**
- In the **PROTEIN DOMAINS** section check **Limit to genes...**
- From the menu select **with signal P protein features -> Only**
- Click the **Results** button (top left) to see your results. By default a two-column file is returned that contains gene ID and Genome Project. To configure different options for the output, select **Attributes** in the left menu.

The screenshot shows the WormBase ParaSite search interface. In the 'SPECIES' dropdown, 'C. elegans' is selected. Under 'MULTI-SPECIES COMPARISONS', 'Orthologous C. elegans genes -> Excluded' is chosen. In the 'PROTEIN DOMAINS' section, the 'Limit to genes...' checkbox is checked. The search results are displayed in a table with columns for 'Gene ID' and 'Genome Project'. At the bottom, there are options for 'Attributes' and 'Output'.

Part 4: Bulk downloads and programmatic access

Downloads

- All genomes, proteomes and annotations available to download as compressed flat files
- Ideal for use with alignment software, etc.
- Data from all previous releases available to download
- Please remember to cite the genome provider and WormBase ParaSite

Downloads – File Formats

Genomic	Raw FASTA genome file
Masked Genomic	Genome FASTA with repeat regions hard-masked
Soft-masked Genomic	Genome FASTA with repeat regions soft-masked
Annotations	GFF3 file containing all annotations
Proteins	FASTA protein file
mRNA Transcripts	FASTA of the spliced full-length transcripts
CDS Transcripts	FASTA of the spliced CDS-portion of the protein coding transcripts

Access using R

- Access our database directly from R, via the biomaRt package
- Syntax identical to Ensembl
- Very quick access to large amounts of data
- Please don't use excessively (i.e. download the results once then store them locally for processing)

WormBase ParaSite in R

- Install the biomaRt package:

```
source("http://bioconductor.org/biocLite.R")
biocLite("biomaRt")
```

- Install the biomaRt package:

```
library(biomaRt)
```

WormBase ParaSite in R

- Establish a connection to WormBase ParaSite

```
mart <- useMart("parasite_mart",
                 dataset = "wbps_eg_gene",
                 host = "parasite.wormbase.org")
```

WormBase ParaSite in R

- Example: get all the *Schistosoma mansoni* genes with a *C. elegans* orthologue:

```
genes <- getBM(mart = mart,
                filters = c("species_id 1010",
                           "with_celegans_eg_homologue"),
                value = list("prjea36577", TRUE),
                attributes = c("ensembl_gene_id",
                              "celegans_eg_gene"))
head(genes)

  ensembl_gene_id celegans_eg_gene
1   Smp_078570    WBGene00009448
2   Smp_063300    WBGene00004450
3   Smp_210640    WBGene00009305
4   Smp_049930    WBGene00010465
5   Smp_132740    WBGene00001395
6   Smp_132740    WBGene00001396
```

Language neutral queries

- REST API allows access using any programming language
- For processing large amounts of data: consider whether making one query to BioMart may be more suitable
- Examples provided in Perl, Python, Ruby, Java, Curl and Wget

Endpoint Catalogue

Comparative Genomics

Resource	Description
GET /rest/genetree/id/:id	Retrieves a gene tree dump for a gene tree stable identifier
GET /rest/genetree/member/id/:id	Retrieves a gene tree that contains the stable identifier
GET /rest/genetree/member/:symbol/:species/:symbol	Retrieves a gene tree containing the gene identified by a symbol
GET /rest/homology/id/:id	Retrieves homology information (orthologues) by gene id
GET /rest/homology/:symbol/:species/:symbol	Retrieves homology information (orthologues) by symbol

Endpoint Specifics

GET genetree/member/id/:id

Retrieves a gene tree that contains the stable identifier

Parameters

Required

Name	Type	Description	Default	Example Values
id	String	A stable ID	-	WBGene00225050

Endpoint Examples

Example Requests

/rest/genetree/member/symbol/brugia_malayi_prjna10729
/Bma-unc-1?content-type=text/x-phylxml%2Bxml

Example output | Perl | Python2 | Python3 | Ruby | Java | Curi | Wget

```
<?xml version="1.0" encoding="UTF-8"?>
<xphylxml xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.phyloxml.org http://www.phyloxml.org/1.10/phylxml.xsd"
  xmlns="http://www.phyloxml.org/1.10/phylxml">
  <phylogeny rooted="true" type="phylogenetic tree">
    <clade branch_length="0">
      <confidence type="duplication_confidence_score">0.731</confidence>
      <taxon>
        <id>3324</id>
        <scientific_name>brugia_malayi</scientific_name>
      </taxon>
      <events>
        <speciation_or_duplication type="speciation_or_duplication">
          < duplication>
            <branch_length>0.003661</branch_length>
            <confidence type="duplication_confidence_score">0.1584</confidence>
          </ duplication>
        </speciation_or_duplication>
      </events>
    </clade>
  </phylogeny>
</xphylxml>
```

Code Examples

Example Requests

/rest/genetree/member/symbol/brugia_malayi_prjna10729
/Bma-unc-1?content-type=text/x-phylxml%2Bxml

Example output | Perl | Python2 | Python3 | Ruby | Java | Curi | Wget

```
use strict;
use warnings;
use HTTP::Tiny;
my $http = HTTP::Tiny->new();
my $server = "http://openbionetworks.org";
my $path = "/rest/genetree/member/symbol/brugia_malayi_prjna10729/Bma-unc-1?";
my $response = $http->get($server . $path, {
  headers => { "Content-type" => 'text/x-phylxml+xml' }
});
die "Failed!\n" unless $response->(success);
print "$response->[status] $response->[reason]\n";
```