

An Introduction to using WormBase ParaSite

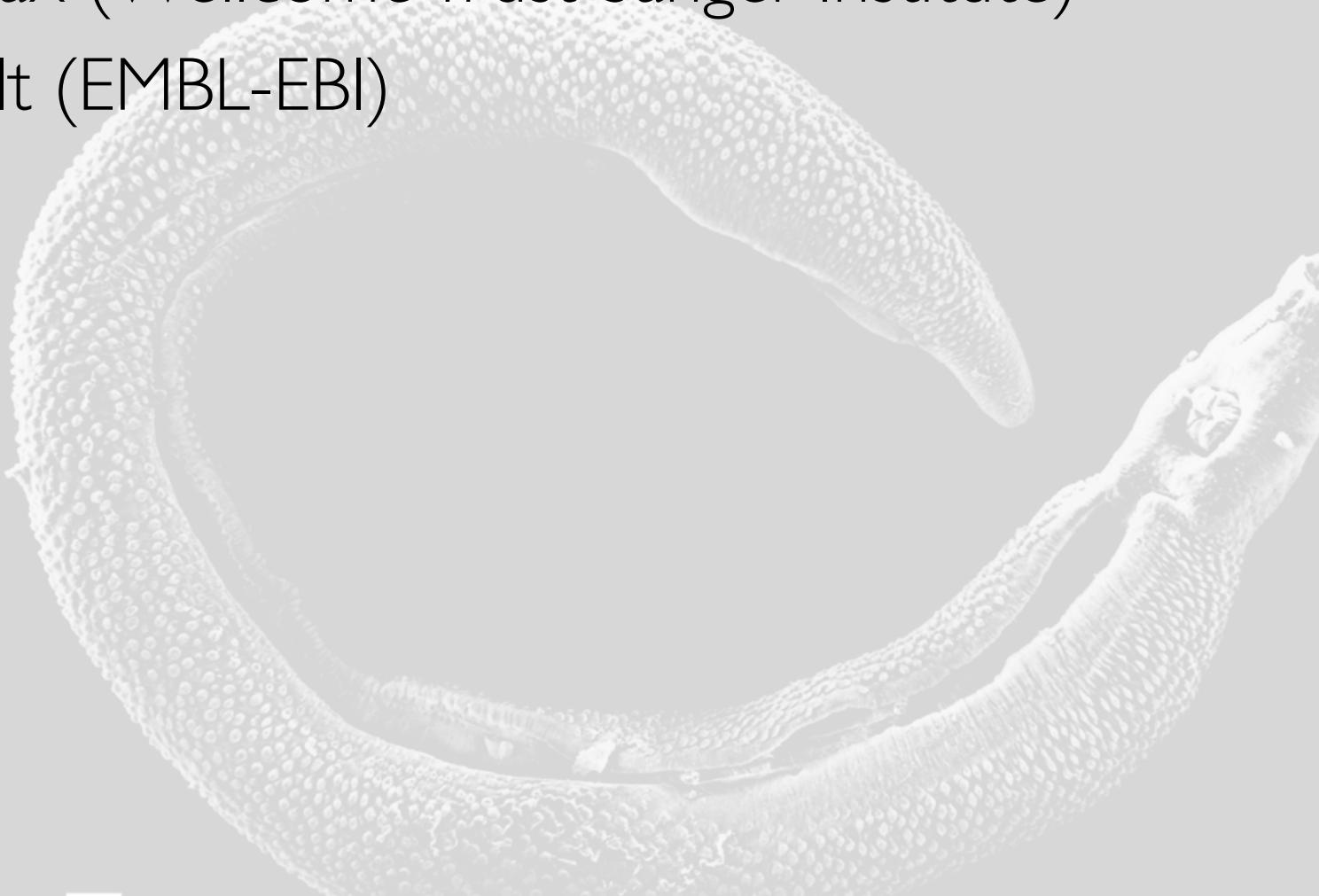
Jane Lomax, Bruce Bolt

Wellcome Trust Sanger Institute and
EMBL-EBI

Course trainers

Jane Lomax (Wellcome Trust Sanger Institute)

Bruce Bolt (EMBL-EBI)



Course Outline

10:30 - 12:00 (Jane)

Using the website: Practical exercises

12:00 - 13:00

Lunch Break

13:00 - 13:45 (Bruce)

Sequence searching with BLAST: Practical exercises

13:45 - 14:30 (Bruce)

Data export with BioMart (Part 1): Practical exercises

14:30 - 14:50

Tea & Coffee Break

14:50 - 15:20 (Jane)

Data export with BioMart (Part 2): Practical exercises

15:20 - 16:00 (Bruce)

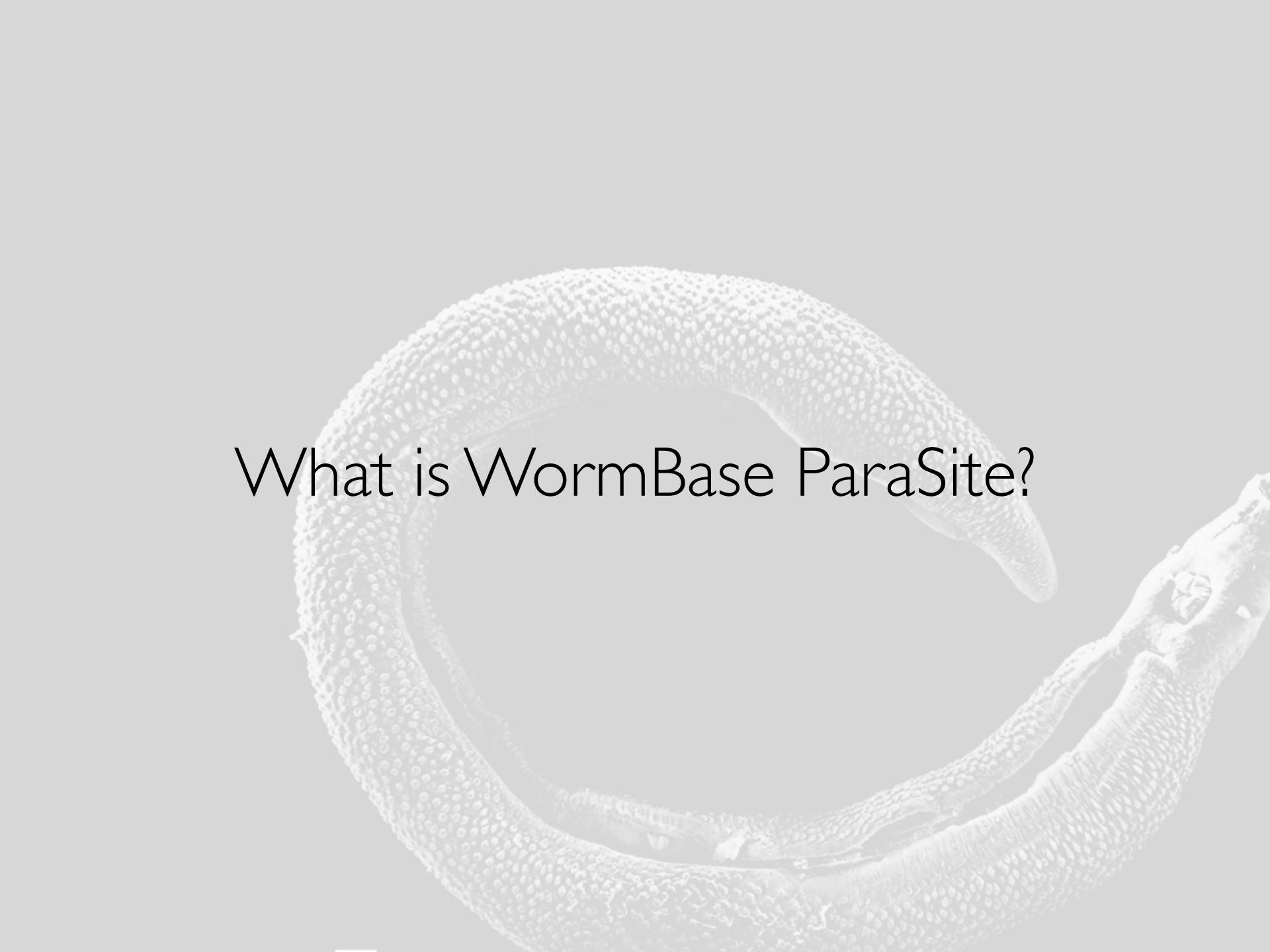
Variant Effect Predictor: Practical exercises

16:00 - 16:30

Q&A Session



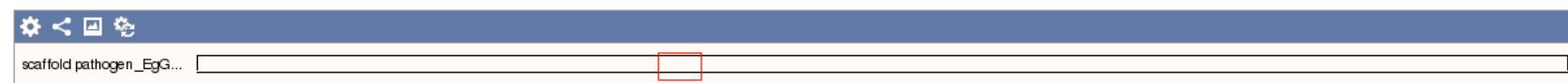
Feedback forms



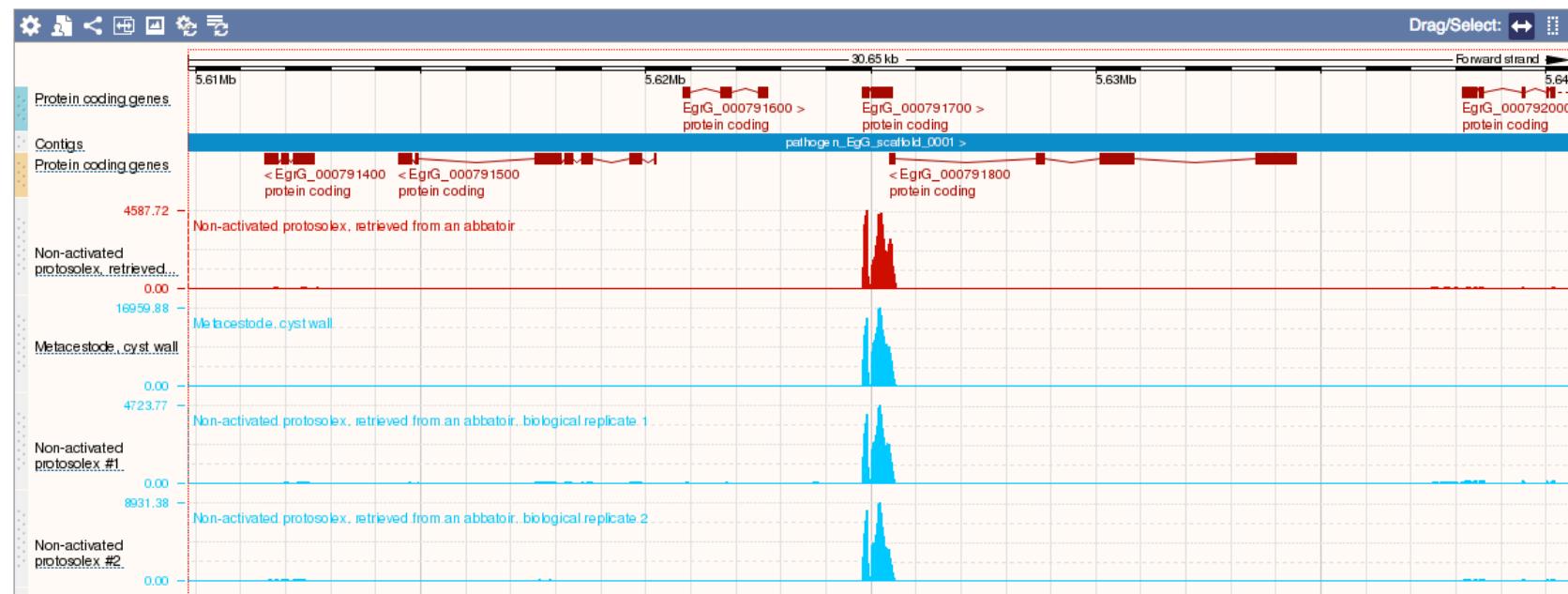
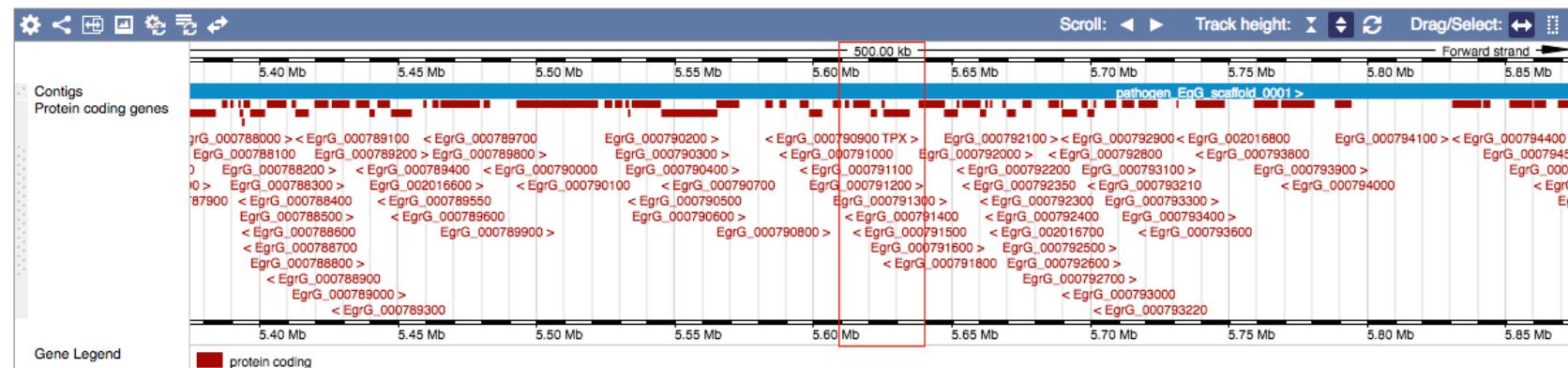
What is WormBase ParaSite?

I. Genome browser for helminth species

Scaffold pathogen_EgG_scaffold_0001: 5,609,817-5,640,467

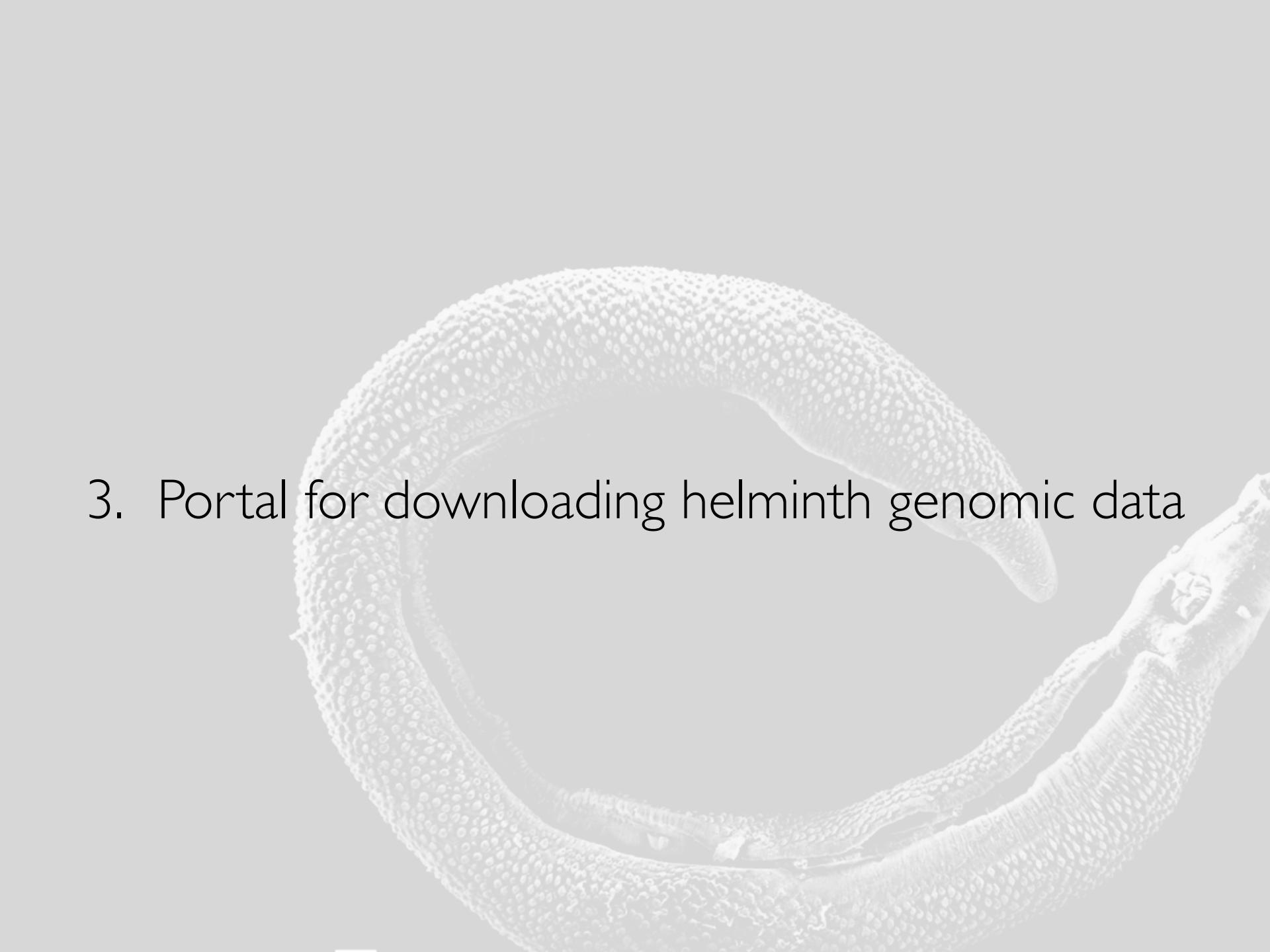


Region in detail ?





2. Set of tools for interacting with helminth genomes



3. Portal for downloading helminth genomic data



Where does the data come from?

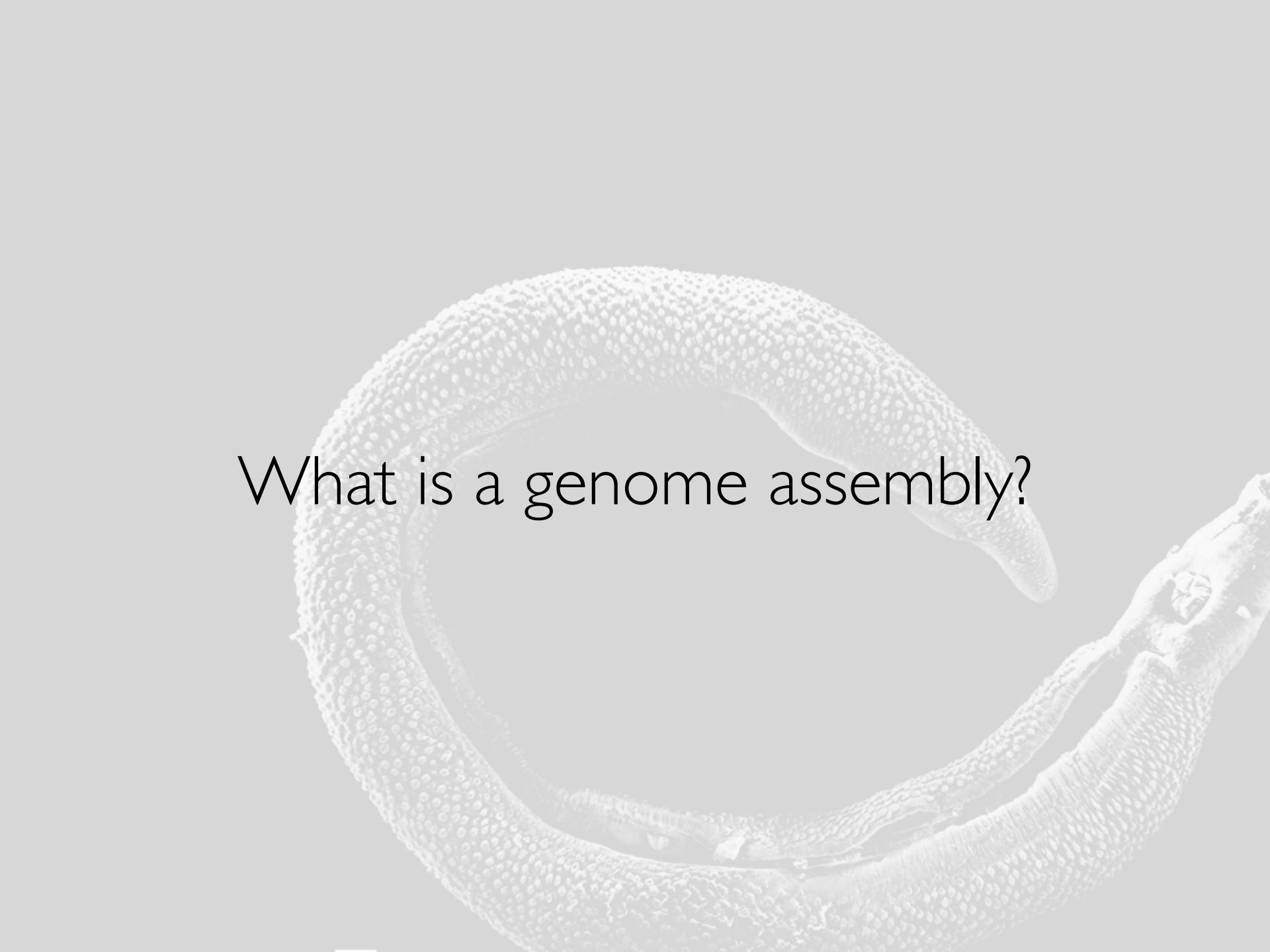


Image credit: brickdisplaycase.com



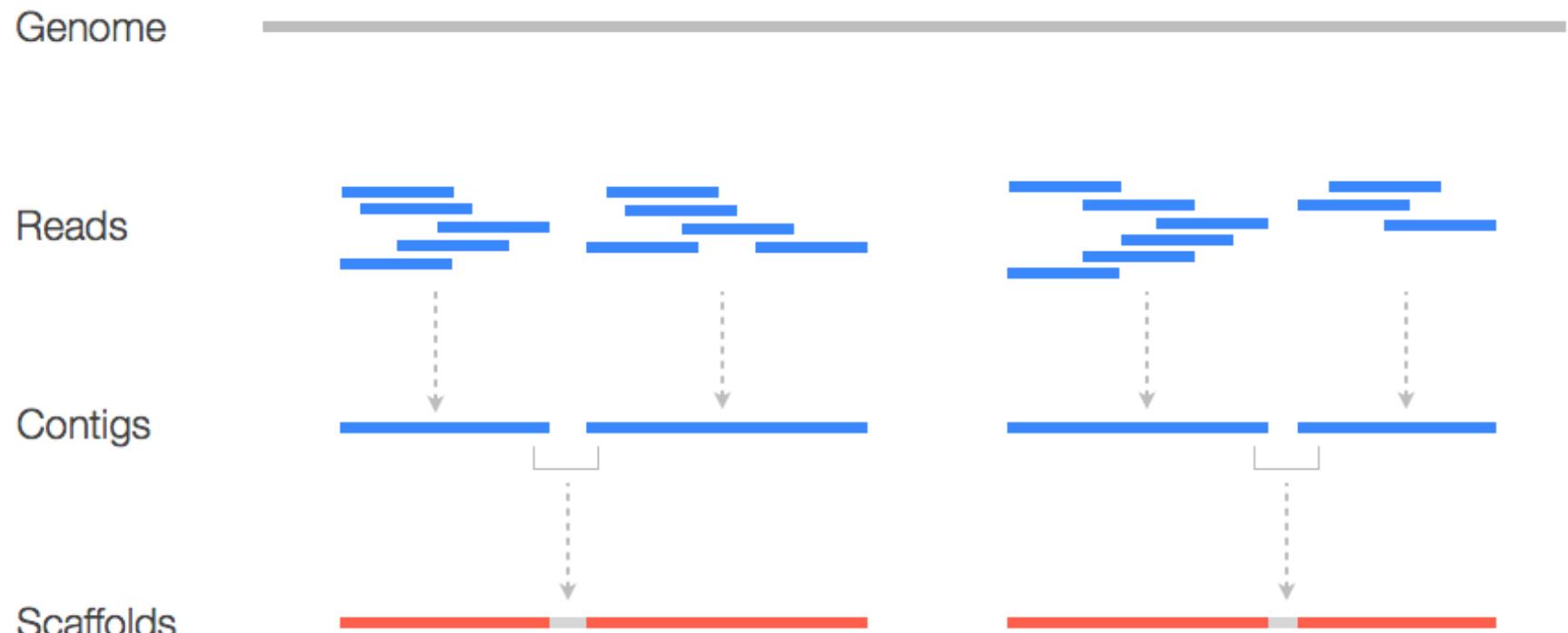
What data does WormBase ParaSite
contain?

- Genomic:
 - protein-coding genes
 - transcripts
 - proteins
- Transcriptomic
- ncRNAs
- Variation (coming soon)



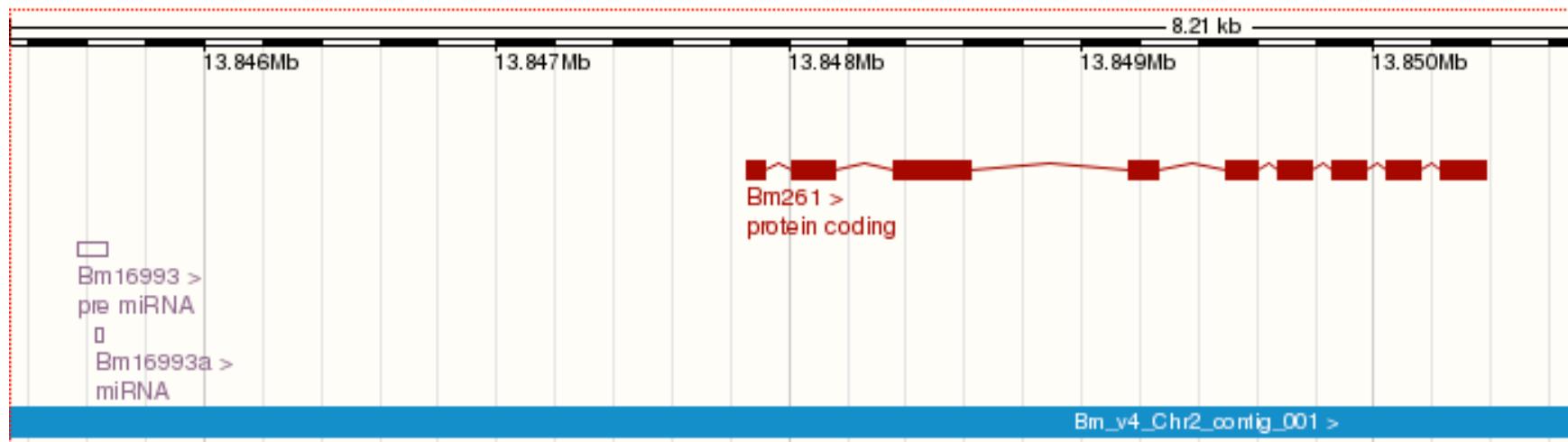
What is a genome assembly?

Genome assembly





Genome annotation





How can the data be accessed?

There are various ways for interacting with data in ParaSite:

- genome browser
- BLAST
- BioMart (data export)
- VEP
- REST API



Where is WormBase ParaSite
developed?

- Joint project between Wellcome Trust Sanger Institute and WormBase team at EMBL-EBI
- BBSRC-funded
- Based on ENSEMBL technology

Acknowledgements

- Matt Berriman
- Kevin Howe
- Paul Kersey
- Myriam Shafie

- Data providers



Using the website

Summary

1. Searching
2. Finding genomes
3. Navigating genes, transcripts and scaffolds
4. Comparative genomics
5. RNASeq data tracks
6. Adding your own data tracks
7. User accounts

I. Searching



Searching



Find a genome

[+] Nematoda (Roundworms)

[+] Platyhelminthes (Flatworms)

Statistics

- Version: WBPS6 (April 2016)
- WormBase Version: WS252
- 109 genomes, representing 100 species
- 2,189,841 genes

Release 6 - April 2016

EMBL-EBI  sanger 





WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

Search WormBase ParaSite...

e.g. *O. volvulus* PRJNA60051, WBGene00262434, *Bmz-est-4*, *est-4*, or *metalloloenidase*

Login Register Help and Documentation

Announcements

Announcing WormBase ParaSite release 6

posted 2 months ago

We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness

posted 1 month ago

We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnach, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [...]

Brugia malayi assembly update

posted 2 months ago

The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [...]

Register for a WormBase ParaSite workshop

posted 2 months ago

To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[Older]

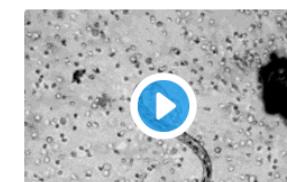
Twitter

Tweets by @WBParasite

WormBase ParaSite Retweeted



Science @scienmag
White blood cells attacking a parasite.
pic.twitter.com/zqAIUZMYDX



12 Jun

WormBase ParaSite Retweeted



eLife - the journal @eLife

Embed

View on Twitter

Information

[Help and Documentation](#)

[Citing WormBase ParaSite](#)

[About WormBase ParaSite](#)

[Data Usage](#)

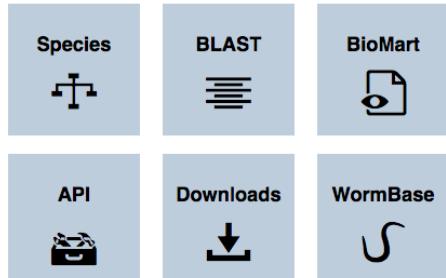
[Contact Us](#)

[Disclaimer](#)

Searching

WormBase ParaSite Version: WBPS6 (WS252)

Species List BLAST BioMart REST API VEP Downloads WormBase



Find a genome

- [+] Nematoda (Roundworms)
- [+] Platyhelminthes (Flatworms)

Statistics

- Version: WBPS6 (April 2016)
- WormBase Version: WS252
- 109 genomes, representing 100 species
- 2,189,841 genes

Release 6 - April 2016



WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

cyst

SUGGESTED TERM

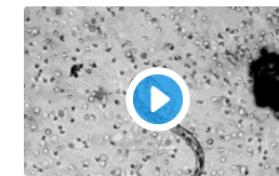
- cystathione
- cystatin
- cysteine
- cysteines
- cystinosin

Tweets by @WBParasite

WormBase ParaSite Retweeted



White blood cells attacking a parasite.
pic.twitter.com/zqAUZMYDx



12 Jun

WormBase ParaSite Retweeted



eLife - the journal
@eLife

Embed

[View on Twitter](#)

Information

- [Help and Documentation](#)
- [Citing WormBase ParaSite](#)
- [About WormBase ParaSite](#)
- [Data Usage](#)
- [Contact Us](#)
- [Disclaimer](#)

Searching

WormBase ParaSite Version: WBPS6 (WS252)

Search WormBase ParaSite... 

e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species List BLAST BioMart REST API VEP Downloads WormBase

Login Register Help and Documentation

New Search

Search WormBase ParaSite

- New Search
- Gene (485)
 - WormBase ParaSite (480)
 - WormBase (5)

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Filter by species:

Search results for 'cystatin'

Showing 1-10 of 480 Genes found in WormBase ParaSite

EEL_0000399501

Description *Cystatin* [Source:UniProtKB/TrEMBL;Acc:A0A0R3RQN0]
Gene ID [EEL_0000399501](#)
Species *Elaeophora elaphi* ([PRJEB502](#))
Location [EEL_scaffold0000034:53318-53512](#)
Gene tree [View gene tree](#)

ALUE_0002323401

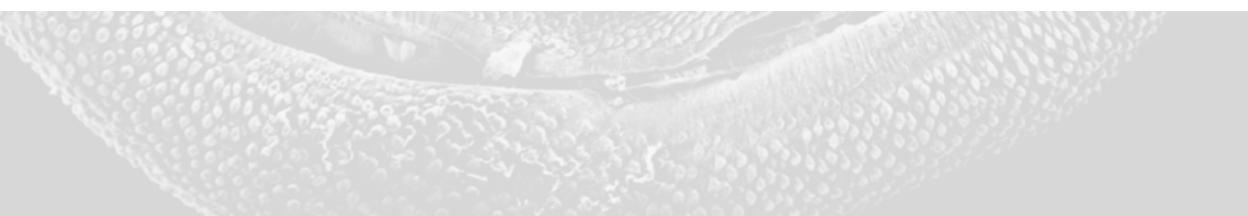
Description *Cystatin* [Source:UniProtKB/TrEMBL;Acc:A0A0M3IWV6]
Gene ID [ALUE_0002323401](#)
Species *Ascaris lumbricoides* ([PRJEB4950](#))
Location [ALUE_scaffold0008948:353-2918](#)
Gene tree [View gene tree](#)

HPLM_0001052101

Description *Cystatin* [Source:UniProtKB/TrEMBL;Acc:A0A0N4WHX7]
Gene ID [HPLM_0001052101](#)
Species *Haemonchus placei* ([PRJEB509](#))
Location [HPLM_scaffold0001226:46905-48974](#)
Gene tree [View gene tree](#)
C. elegans orthologues [cpi-2](#) [cpi-1](#)

BPAG_0000497001

Description *Cystatin* [Source:UniProtKB/TrEMBL;Acc:A0A0N4T9T3]
Gene ID [BPAG_0000497001](#)
Species *Brugia pahangi* ([PRJEB497](#))
Location [BPAG_contig0002942:40-2465](#)
Gene tree [View gene tree](#)



Searching

WormBase ParaSite

[Species List](#)
[BLAST](#)
[BioMart](#)
[REST API](#)

[New Search](#)

Search WormBase ParaSite

- New Search
- Gene (485)
- WormBase ParaSite (480)
- WormBase (5)

Search results

Showing 1-10 of 480

Configure this page

Add your data

Export data

Share this page

Bookmark this page

EEL_0000395

Description

Gene ID

Species

Location

Gene tree

ALUE_00023

Description

Gene ID

Species

Location

Gene tree

HPLM_00010

Description

Gene ID

Species

Location

Gene tree

C. elegans orthologs

BPAG_00004

Description

Gene ID

Species

Location

Gene tree

Myiostilpidae (PRJEB2244)

Hymenolepis diminuta (PRJEB507)

Hymenolepis microstoma (PRJEB124)

Hymenolepis nana (PRJEB508)

Litomosoides sigmodontis (PRJEB3075)

Loa loa (PRNA246086)

Loa loa (PRNA60051)

Meloidogyne floridensis (PRJEB6016)

Meloidogyne hapla (PRNA29083)

Mesocestoides corti (PRJEB510)

Necator americanus (PRNA72135)

Nippostrongylus brasiliensis (PRJEB511)

Oesophagostomum dentatum (PRNA72579)

Onchocerca flexuosa (PRJEB512)

Onchocerca ochengi (PRJEB1204)

Onchocerca ochengi (PRJEB1809)

Oncocerca volvulus (PRJEB513)

Opisthorchis viverrini (PRNA222628)

Panagrellus redivivus (PRNA186477)

Parastrephylloides trichosuri (PRJEB515)

Pristionchus expectatus (PRJEB6009)

Pristionchus pacificus (PRNA12644)

Protopolyystoma xenopodis (PRJEB1201)

Rhabditophanes sp. KR3021 (PRJEB1297)

Romanomermis culicivorax (PRJEB1358)

Schistocerca solidus (PRJEB527)

Schistosoma haematobium (PRNA728265)

Schistosoma japonicum (PRIEA34885)

Schistosoma mansoni (PRIEA36577)

Schistosoma margebowiei (PRJEB522)

Schistosoma rodhaini (PRJEB526)

Schmidtea mediterranea (PRNA12585)

Soboliphyme baturini (PRJEB516)

Spirometra erinaceieuropaei (PRJEB1202)

Steinernema carpocapsae (PRNA202318)

Steinernema feltiae (PRNA204661)

Steinernema glaseri (PRNA204943)

Steinernema monticolum (PRNA205067)

Steinernema scapterisci (PRNA204942)

Strongyloides papillosus (PRJEB525)

Strongyloides ratti (PRJEB125)

Strongyloides stercoralis (PRJEB528)

Strongyloides venezuelensis (PRJEB530)

Strongylus vulgaris (PRJEB531)

Syphacia muris (PRJEB524)

Taenia asiatica (PRJEB532)

Taenia solium (PRNA170813)

Teladorsagia circumincta (PRNA72569)

Thelazia callipaeda (PRJEB1205)

Toxocara canis (PRJEB533)

Trichinella nativa (PRNA179527)

Trichinella spiralis (PRNA12603)

Trichuris muris (PRJEB126)

Trichuris suis (PRNA179528)

Trichuris suis (PRNA208415)

Trichuris suis (PRNA208416)

Trichuris trichiura (PRJEB535)

Wuchereria bancrofti (PRJEB536)

Search WormBase ParaSite...

e.g. *O. volvulus*, PRJNA60051, WBGene00262434, *Bm-aat-4*, *eat-4* or metallopeptidase

[Login](#) [Register](#) [Help and Documentation](#)

Searching

WormBase ParaSite Version: WBPS6 (WS252)

Species List BLAST BioMart REST API VEP Downloads WormBase

Search WormBase ParaSite... 

e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Login Register Help and Documentation

New Search

Search WormBase ParaSite

- New Search
- Gene (485)
 - WormBase ParaSite (480)
 - WormBase (5)

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Search results for 'cystatin'

Showing 6 Genes found in WormBase ParaSite (filtered)

Filtered by species: *Strongyloides ratti* (PRJEB125) 

SRAE_2000370000 [WBGene00263925]

Description Proteinase inhibitor I25, *cystatin* domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A090LLJ9]
Gene ID WBGene00263925
Species *Strongyloides ratti* (PRJEB125)
Location SRAE_chr2:11581338-11581697
[View region in WormBase JBrowse]
Gene tree [View gene tree](#)
C. elegans orthologues cpi-2 , cpi-1 

SRAE_2000495700 [WBGene00265204]

Description Proteinase inhibitor I25, *cystatin* domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A090LQ75]
Gene ID WBGene00265204
Species *Strongyloides ratti* (PRJEB125)
Location SRAE_chr2:15634295-15634777
[View region in WormBase JBrowse]
Gene tree [View gene tree](#)

SRAE_X000107250 [WBGene00266636]

Description Proteinase inhibitor I25, *cystatin* domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A090KVQ2]
Gene ID WBGene00266636
Species *Strongyloides ratti* (PRJEB125)
Location SRAE_chRX_scaffold2:377256-379294
[View region in WormBase JBrowse]

SRAE_2000523500 [WBGene00265495]

Description Proteinase inhibitor I25, *cystatin* domain-containing protein [Source:UniProtKB/TrEMBL;Acc:A0A090LLL2]
Gene ID WBGene00265495
Species *Strongyloides ratti* (PRJEB125)
Location SRAE_chr2:16571628-16571924
[View region in WormBase JBrowse]

Searching

UniProtKB

UniProtKB Advanced Search

BLAST Align Retrieve/ID mapping Help Contact

Basket 2

UniProtKB - G4VBC8 (G4VBC8_SCHMA)

Display

Entry Feature viewer Feature table

Function Names & Taxonomy PTM / Processing Structure Family & Domains Sequence Cross-references Publications Entry information Miscellaneous Similar proteins

None

Submitted name: **Putative cystatin**

Gene: **Smp_034420.1**

Organism: *Schistosoma mansoni* (Blood fluke)

Status: Unreviewed - Annotation score: 00000 - Protein predicted

Function

GO - Molecular function: cysteine-type endopeptidase inhibitor activity (Source: InterPro)

Complete GO annotation...

Names & Taxonomy

Protein names: Submitted name: Putative cystatin Imported

Gene names: ORF Names: Smp_034420.1 Imported

Organism: Schistosoma mansoni (Blood fluke) Imported

Taxonomic identifier: 6183 [NCBI]

Taxonomic lineage: Eukaryota > Metazoa > Platyhelminthes > Trematoda > Digenea > Strigeida > Schistosomatoidea > Schistosomatidae > Schistosoma

Proteomes: UP000008854 Component: Chromosome 2

PTM / Processing

Molecule processing

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
-------------	-------------	--------	-------------	----------------	--------------------	---------

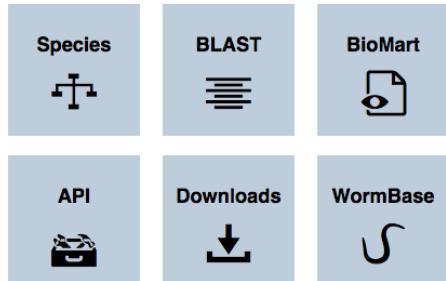
Searching

G4VBC8



ac. O. volvulus PRINA60051 WRGene00282434 Bma-act-4 act-4 or metalloendopeptidase

Login Register Help and Documentation



Announcements

Announcing WormBase ParaSite release 6

posted 2 months ago

We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness

posted 1 month ago

We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [..]

Brugia malayi assembly update

posted 2 months ago

The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [..]

Register for a WormBase ParaSite workshop

posted 2 months ago

To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[Older]

Twitter

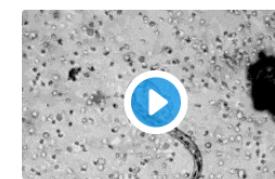
Tweets by @WBParasite

WormBase ParaSite Retweeted



Science
@scienmag

White blood cells attacking a parasite.
pic.twitter.com/zqAlUZMYDx



12 Jun

WormBase ParaSite Retweeted



eLife - the journal
@eLife

Embed

View on Twitter

Release 6 - April 2016



eEmpowered



WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

Information

[Help and Documentation](#)

[Citing WormBase ParaSite](#)

[About WormBase ParaSite](#)

[Data Usage](#)

[Contact Us](#)

[Disclaimer](#)

Searching

WormBase ParaSite Version: WBPS6 (WS252)

Species List BLAST BioMart REST API VEP Downloads WormBase

Search WormBase ParaSite... Search

e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bra-eat-4, eat-4 or metallopeptidase

Login Register Help and Documentation

New Search

Search WormBase ParaSite

- New Search
- Gene (1)
 - WormBase ParaSite (1)
 - WormBase (0)

Configure this page

Add your data

Export data

Share this page

Bookmark this page

Search results for 'G4VBC8'

Showing 1 Gene found in WormBase ParaSite

Smp_034420

Description	DIF_6; Putative cystatin [Source:UniProtKB/TrEMBL;Acc:B8Y6H4]
Gene ID	Smp_034420
Species	<i>Schistosoma mansoni</i> (PRJEA36577)
Location	Smp.Chr_2:7321609-7322171
Gene tree	View gene tree

Information

Help and Documentation

Citing WormBase ParaSite

About WormBase ParaSite

Data Usage

Contact Us

Disclaimer

Release 6 - April 2016

EMBL-EBI    

WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

2. Finding genomes

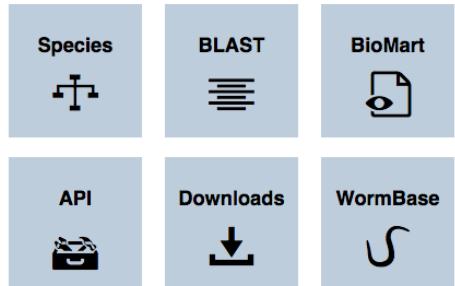
Finding a genome

Search WormBase ParaSite...



e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Login Register Help and Documentation



Find a genome

[+] Nematoda (Roundworms)

[+] Platyhelminthes (Flatworms)

Statistics

- Version: WBPS6 (April 2016)
- WormBase Version: WS252
- 109 genomes, representing 100 species
- 2,189,841 genes

Announcements

Announcing WormBase ParaSite release 6

posted 2 months ago

We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness

posted 1 month ago

We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [..]

Brugia malayi assembly update

posted 2 months ago

The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [..]

Register for a WormBase ParaSite workshop

posted 2 months ago

To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[Older]

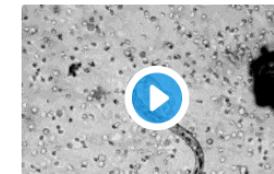
Twitter

Tweets by @WBParasite

WormBase ParaSite Retweeted

Science
@scienmag

White blood cells attacking a parasite.
pic.twitter.com/zqAIUZMYDx



12 Jun

WormBase ParaSite Retweeted

eLife - the journal
@eLife

Embed

View on Twitter

Release 6 - April 2016



WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

Information

- Help and Documentation
- Citing WormBase ParaSite
- About WormBase ParaSite
- Data Usage
- Contact Us
- Disclaimer

Finding a genome



Find a genome

[+] Nematoda (Roundworms)

[+] Platyhelminthes (Flatworms)

Statistics

- Version: WBPS6 (April 2016)
- WormBase Version: WS252
- 109 genomes, representing 100 species
- 2,189,841 genes

Announcements

Announcing WormBase ParaSite release 6

posted 2 months ago

We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness

posted 1 month ago

We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [..]

Brugia malayi assembly update

posted 2 months ago

The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [..]

Register for a WormBase ParaSite workshop

posted 2 months ago

To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[Older]

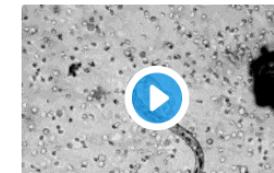
Twitter

Tweets by @WBParasite

WormBase ParaSite Retweeted

Science
@scienmag

White blood cells attacking a parasite.
pic.twitter.com/zqAIUZMYDx



12 Jun

WormBase ParaSite Retweeted

eLife - the journal
@eLife

Embed

View on Twitter

Information

- Help and Documentation
- Citing WormBase ParaSite
- About WormBase ParaSite
- Data Usage
- Contact Us
- Disclaimer

Finding a genome

WormBase ParaSite Version: WBPS6 (WS252)

[Species List](#) [BLAST](#) [BioMart](#) [REST API](#) [VEP](#) [Downloads](#) [WormBase](#)

[Search WormBase ParaSite...](#) 

[Login](#) [Register](#) [Help and Documentation](#)

Species 

BLAST 

BioMart 

API 

Downloads 

WormBase 

Announcements

[Announcing WormBase ParaSite release 6](#)
posted 2 months ago

We are pleased to announce the sixth release of WormBase ParaSite.

Blog

[Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness](#)
posted 1 month ago

We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin , Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [\[...\]](#)

[Brugia malayi assembly update](#)
posted 2 months ago

The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [\[...\]](#)

[Register for a WormBase ParaSite workshop](#)
posted 2 months ago

To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[\[Older\]](#)

Find a genome

[\[+\] Nematoda \(Roundworms\)](#)

[\[+\] Platyhelminthes \(Flatworms\)](#)

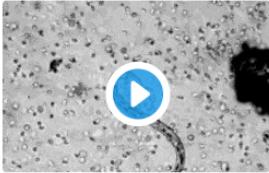
Statistics

- Version: WBPS6 (April 2016)
- WormBase Version: WS252
- 109 genomes, representing 100 species
- 2,189,841 genes

Twitter

Tweets by @WBParasite

 [Science](#) [@scienmag](#)
White blood cells attacking a parasite.
pic.twitter.com/zqAIUZMYDx



[View on Twitter](#)

 [eLife - the journal](#) [@eLife](#)

[Embed](#) [View on Twitter](#)

Release 6 - April 2016

EMBL-EBI    

WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

Information

[Help and Documentation](#) [Citing WormBase ParaSite](#) [About WormBase ParaSite](#) [Data Usage](#) [Contact Us](#) [Disclaimer](#)

Finding a genome

WormBase ParaSite Version: WBPS6 (WS252)

Species List BLAST BioMart REST API VEP Downloads WormBase

Search WormBase ParaSite... 

Login Register Help and Documentation

Species 

BLAST 

BioMart 

API 

Downloads 

WormBase 

Find a genome

[+] **Nematoda (Roundworms)**

[+] **Platyhelminthes (Flatworms)**

Statistics

- Version: WBPS6 (April 2016)
- WormBase Version: WS252
- 109 genomes, representing 100 species
- 2,189,841 genes

Announcements

Announcing WormBase ParaSite release 6
posted 2 months ago
We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness
posted 1 month ago
We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [...]

Brugia malayi assembly update
posted 2 months ago
The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [...]

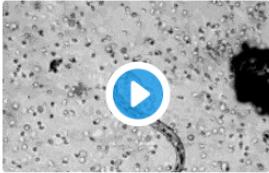
Register for a WormBase ParaSite workshop
posted 2 months ago
To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[Older]

Twitter

Tweets by @WBParasite

WormBase ParaSite Retweeted  @scienmag
White blood cells attacking a parasite. pic.twitter.com/zqAIUZMYDx



12 Jun

WormBase ParaSite Retweeted  eLife - the journal @eLife

Embed View on Twitter

Release 6 - April 2016

EMBL-EBI    

WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

Information

Help and Documentation
Citing WormBase ParaSite
About WormBase ParaSite
Data Usage
Contact Us
Disclaimer

Species List

Contents

[Nematoda \(80\)](#)
[Platyhelminthes \(29\)](#)

Nematoda

Species Name	Provider	Assembly	BioProject ID	Taxonomy ID
Acanthocheilonema viteae	University of Edinburgh	nAv.1.0	PRJEB4306	6277
Ancylostoma caninum	Genome Institute at Washington University	A_caninum_9.3.2.ec.cg.pg	PRJNA72585	29170
Ancylostoma ceylanicum	Cornell University	Acey_2013.11.30.genDNA	PRJNA231479	53326
Ancylostoma ceylanicum	Genome Institute at Washington University	A_ceylanicum1.3.ec.cg.pg	PRJNA72583	53326
Ancylostoma duodenale	Genome Institute at Washington University	A_duodenale_2.2.ec.cg.pg	PRJNA72581	51022
Angiostrongylus cantonensis	Wellcome Trust Sanger Institute	A_cantonensis_Taipei_v1_5_4	PRJEB493	6313
Angiostrongylus costaricensis	Wellcome Trust Sanger Institute	A_costaricensis_Costa_Rica_v1_5_4	PRJEB494	334426
Anisakis simplex	Wellcome Trust Sanger Institute	A_simplex_v1_5_4	PRJEB496	6269
Ascaris lumbricoides	Wellcome Trust Sanger Institute	A_lumbricoides_Ecuador_v1_5_4	PRJEB4950	6252
Ascaris suum	University of Colorado School of Medicine	ASU_2.0	PRJNA62057	6253
Ascaris suum	University of Melbourne	AscSuum_1.0_submitted	PRJNA80881	6253
Brugia malayi	WormBase	Bmali-4.0	PRJNA10729	6279
Brugia pahangi	Wellcome Trust Sanger Institute	B_pahangi_Glasgow_v1_5_4	PRJEB497	6280
Brugia timori	Wellcome Trust Sanger Institute	B_timori_Indonesia_v1_0_4	PRJEB4663	42155
Bursaphelenchus xylophilus	Wellcome Trust Sanger Institute	ASM23113v1_submitted	PRJEA64437	6326
Caenorhabditis angaria	WormBase	13mar2012	PRJNA51225	860376
Caenorhabditis brenneri	WormBase	C_brenneri-6.0.1b	PRJNA20035	135651
Caenorhabditis briggsae	WormBase	CB4	PRJNA10731	6238
Caenorhabditis elegans	WormBase	WBcel235	PRJNA13758	6239
Caenorhabditis japonica	WormBase	C_japonica-7.0.1	PRJNA12591	281687

Genome pages

WormBase ParaSite Version: WBPS6 (WS252)

Species List BLAST BioMart REST API VEP Downloads WormBase

Schistosoma mansoni (PRJEA36577) ▾

Search WormBase ParaSite...



e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Login Register Help and Documentation

Schistosoma mansoni

BioProject PRJEA36577 | Data Source Wellcome Trust Sanger Institute | Taxonomy ID 6183

Search Schistosoma mansoni (PRJEA36577)...

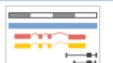


About *Schistosoma mansoni*

The trematode *Schistosoma mansoni* is one of the three major infectious agents responsible for the chronic debilitating disease schistosomiasis found throughout Africa and South America. The intermediate host for the parasite is a freshwater snail from the genus *Biomphalaria*, and the definitive host is a human. Schistosomiasis is widely considered to be second only to malaria as a global health problem and an incalculable drain on the economic development of endemic countries.

Genome assembly: ASM23792v2

The *S. mansoni* reference genome was sequenced by a consortium led by the Wellcome Trust Sanger Institute, as described by [Berriman et al \(2009\)](#). The assembly version represented here is the v5.0 freeze from [GeneDB](#) (December 2013), published in [Protasio et al \(2012\)](#) and which corresponds with INSDC assembly [ASM23792v2](#).



Example region

Gene annotation

The gene predictions were initially made by a consortium led by the Wellcome Trust Sanger Institute, as described by [Berriman et al \(2009\)](#). The gene models have since been iteratively improved, and the version represented here is the v5.0 freeze from [GeneDB](#) (December 2013) and described in [Protasio et al \(2012\)](#).

What can I find? Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.



Example gene

Statistics

Summary

Assembly:	ASM23792v2, GCA_000237925.2
Database version:	84.1
Base Pairs:	362,511,929
Genome Size:	364,538,298
Data source:	Wellcome Trust Sanger Institute
Genebuild version:	2014-05-WormBase

Gene counts

Coding genes:	10,831
Gene transcripts:	11,828

Comparative genomics

What can I find? Orthologues, paralogues, and gene trees across multiple species.



Example gene tree

Downloads

- Genomic Sequence (FASTA)
- Hard-masked Genomic Sequence (FASTA)
- Soft-masked Genomic Sequence (FASTA)
- Annotations (GFF3)
- Proteins (FASTA)
- Full-length transcripts (FASTA)
- CDS transcripts (FASTA)

Key Publications

Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, Masihiyama ST, Al-Lazikani B, Andrade LF, Ashton PD, Aslett MA, Bartholomeu DC, Blandin G, Caffrey CR, Coghlan A, Coulson R, Day TA, Delcher A, DeMarco R, Dijkeng A, Eyre T, Gamble JA, Ghedin E, Gu Y, Hertz-Fowler C, Hirai H, Hirai Y, Houston R, Ivens A, Johnston DA, Lacerda D, Macedo CD, McVeigh P, Ning Z, Oliveira G, Overington JP, Parkhill J, Pertea M, Pierce RJ, Protasio AV, Quail MA, Rajandream MA, Rogers J, Sajid M, Salzberg SL, Stanke M, Tivey AR, White O, Williams DL, Wortman J, Wu W, Zamanian M, Zerlotini A, Fraser-Liggett CM, Barrell BG, El-Sayed NM. [The genome of the blood fluke *Schistosoma mansoni*](#). *Nature*, 2009;460(7253):352-358

Protasio AV, Tsai JI, Babbage A, Nichol S, Hunt M, Aslett MA, De Silva N, Velarde GS, Anderson TJ, Clark RC, Davidson C, Dillon GP, Holroyd NE, LoVerde PT, Lloyd C, McQuillan J, Oliveira G, Otto TD, Parker-Manuel SJ, Quail MA, Wilson RA, Zerlotini A, Dunne DW, Berriman M. [A systematically improved high quality genome and transcriptome of the human blood fluke *Schistosoma mansoni*](#). *PLoS Negl Trop Dis*, 2012;6(1):e1455

Tools

- Search for sequences in the genome and proteome using BLAST
- Work with lists of data using the WormBase ParaSite BioMart data-mining tool
- Programmatically access WormBase ParaSite data using the REST API
- Predict the effects of variants using the Variant Effect Predictor

3. Navigating genes, transcripts and scaffolds



Gene pages

WormBase ParaSite

Species List BLAST BioMart REST API Downloads WormBase

Schistosoma mansoni (PRJEA365777) Location: Smp.Chr_4:31,987,438-31,988,870 Gene: SAT1 Transcript: Smp_090120.1

Search WormBase ParaSite... e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-set-4, est-4 or metallopeptidase

Help and Documentation

Gene-based displays

- Summary
- Splice variants
- Sequence
- External references
- Ontology**
 - GO: biological process
 - GO: molecular function
 - GO: cellular component
- Comparative Genomics
- Gene tree
- Orthologues
- Paralogues

Gene: SAT1 Smp_090120.1

Description Putative alpha tubulin (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VHT1]
Location Scaffold Smp.Chr_4: 31,987,438-31,988,870 reverse strand.
About this gene This gene has 2 transcripts (splice variants), 358 orthologues and 4 paralogues.
Gene type Protein coding
Annotation Method Gene models from Wellcome Trust Sanger Institute Reference Helminth Genomes project
Transcripts Hide transcript table

Show/hide columns (1 hidden) Filter

Name	Transcript ID	bp	Protein	Translation ID	Biotype	UniProt
Novel	Smp_090120.1	1356	451aa	Smp_090120.1	Protein coding	G4VHT1 Q26595
Novel	Smp_090120.2	1311	436aa	Smp_090120.2	Protein coding	G4VHT1

Summary

Go to [Region In Detail](#) for more tracks and navigation options (e.g. zooming)

Configuring the display

Tip: use the "Configure this page" link on the left to show additional data in this region.

Release 4 - November 2015

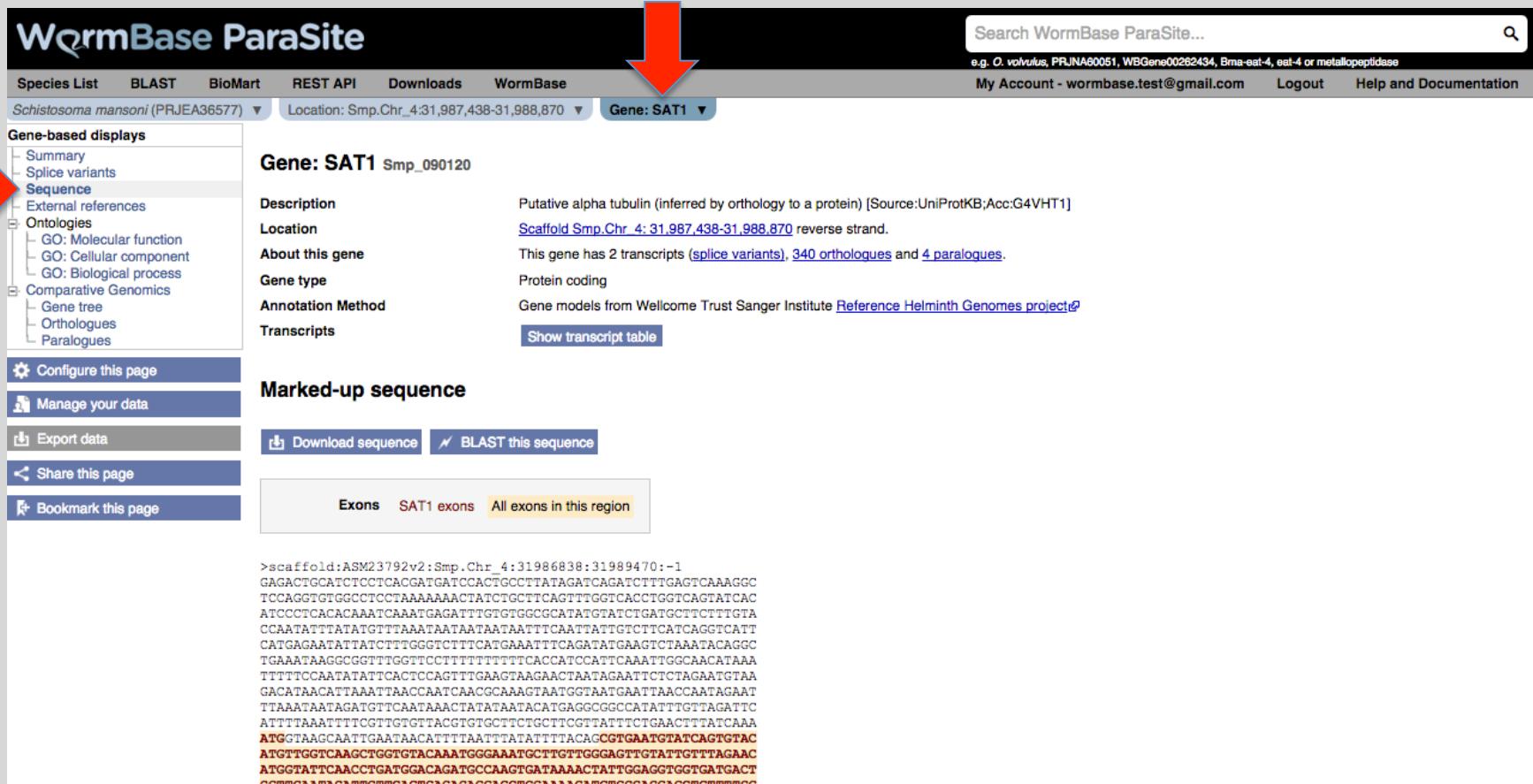


WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

Information

[Help and Documentation](#)
[About WormBase ParaSite](#)
[Data Usage](#)
[Contact Us](#)
[Disclaimer](#)

Viewing sequence



WormBase ParaSite

Species List BLAST BioMart REST API Downloads WormBase

Search WormBase ParaSite...
e.g. *O. volvulus*, PRJNA80051, WBGene00262434, Bmz-est-4, est-4 or metallopeptidase

My Account - wormbase.test@gmail.com Logout Help and Documentation

Schistosoma mansoni (PRJEA36577) Location: Smp.Chr_4:31,987,438-31,988,870 Gene: SAT1

Gene-based displays

- Summary
- Splice variants
- Sequence**
- External references
- Ontologies
 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process
- Comparative Genomics
- Gene tree
- Orthologues
- Paralogues

Configure this page

Manage your data

Export data

Share this page

Bookmark this page

Gene: SAT1 Smp_090120

Description Putative alpha tubulin (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VHT1]

Location Scaffold Smp.Chr_4: 31,987,438-31,988,870 reverse strand.

About this gene This gene has 2 transcripts ([splice variants](#)), [340 orthologues](#) and [4 paralogues](#).

Gene type Protein coding

Annotation Method Gene models from Wellcome Trust Sanger Institute [Reference Helminth Genomes project](#)

Transcripts [Show transcript table](#)

Marked-up sequence

[Download sequence](#) [BLAST this sequence](#)

Exons SAT1 exons All exons in this region

```
> scaffold:ASM23792v2:Smp.Chr_4:31986838:319894701-1
GAGACTGCATCTCCTCACGATGATCCACTGCCCTTATAGATCAGATCTTGAGTCAAAGGC
TCCAGGTGCGCTCTAAAAAAACTATCTGCTTCAGTTGGTCACCTGGTCAGTATCAC
ATCCCTACACAAATCAAAATGAGATTTGTCGGCCATATGATCTGATGCTTCTTGTAA
CCAATATTTATGTTAAATAATAATAATTTCAATTATGTCCTCATCAGGTCAATT
CATGAGAAATATTATCTTGGCTTTCATGAAAATTCAAGATATGAAGTCTAAATACAGGC
TGAATAAACCGGTTGGTTCTTTTTTCCACATCCATTCAAATTGCCAACATAAA
TTTTCCAAATATTCACCTCAGTTGAAGTAAAGACTTAAGAAATTCTGAAATGAA
GACATACATTAATAATTAACCAATCAACCCAAGAACTATGGTAATGAAATTACCAATAGAA
GACATACATTAATAATTAACCAATCAACCCAAGAACTATGGTAATGAAATTACCAATAGAA
TTAATAATAGATGTTCATATAACTATATAACATGAGGGGCCATATTGTTAGATTC
ATTTAAATTTCGTTGTTACGTTGCTGTTCTGTTGTTATTTGAACTTTATCAAA
ATGGTAAGCAATTGAATAACATTAACTTTAACAGCTGAAATGTATCACTGTCAC
ATGTTGTCAGCTGGTGTCACAAATGGAAATGGTTGGGAGTGTATTGTTAGAAC
ATGGTATTCACCTGATGGACAGATGCCAAGTGATAAAACTATGGAGGTGGTGATGACT
CGTTCAATACATTCTCAGTGAGACAGGAGCTGGAAAAACATGTGCCACGGCTGTTTCG
```

Viewing sequence

Marked-up sequence

[Download sequence](#) [BLAST this sequence](#)

[Exons](#) [SAT1 exons](#) [All exons in this region](#)

```
> scaffold:ASM23792v2:Smp.Chr_4:31986838:31989470:-1
GAGACTGCATCTCTCACGATGATCCACTGCCCTTATAGATCAGATCTTGTGACTCAAAGGC
TCCAGGTGTCGGCTCTAAAAAAACTATCTGCTCAGTTGGTCACCTGGTCAGTATCAC
ATCCCTCACACAAATGAGATTGGTGTGGCCCATATGTATCTGATGCTTCTTGT
CCAATATTATATGTTAAATAATAATAATTCAATTATGTCTCATCAGGTATT
CATGAGAAATATTATCTGGGCTTTCATGAAATTTCAGATTAAGCTAAATACAGGC
TGAATAAGCGGCTTGGCTTCTTTTTCACCATCCATTCAAATTGGCAACATAAA
TTTTCAATATATTCACTCCAGTTGAAGTAAGAACTATAAGAATTCTCTAGATGTAA
GACATAACATTAAATAACCAATCAACGCAAAGTAATGGTAATGAATTAAACCAATAGAAT
TTAAATAATAGATGTTCAATAAACATAATAATCATGAGCGGCCATTATTTGTTAGATTC
ATTTAAATTCTGGTGTGTTAGCTGTGCTCTGCTTATTCTGAACCTTATCAA
ATGCTAAGCAATTGAATAACATTAAATTATATTACAGCGTGAATGTTACGTGTAC
ATGTTGGTCAGGTGTCACAATGGAAATGCTGTTGGGAGTTGTTAGTGGAAAC
ATGGTATTCACACTGATGGACAGATCCAAGTGATAAAACTATTGGAGGTGTTGATGACT
CGTCACATACATTCTCAGTGAGACAGGGCTGGAAAACATGTGCCACGGCTTTCG
TAGATTAGAACCAACTGTAGTAGATGAAGTCCGAACCTGGAACTTATCTCAACTTTC
ATCCAGAACAAATTAACACTGGTAAGGGAGGATGCTCTAATAACATGTCCGGTCTT
ATCAATTGGGAAAGAATACTGGACTTATGTTGGATCTGATTCTCGAAGTGGTGTGACC
AGTGTACTGGACTTCAGGGTTCTGATTTCCTCATTCATTGGCGGTGTTACTGGTTCTG
GGTTCACATCTTACTGATGGAACGTTGTTGAGTTGGGATTATGCAA AAAAATCGAAATTGG
AGTTTCGCTGTGATCTGTCACACGATTGCTACAGCTGTGTTGAACCCATATAATTCCA
TTTGACTACACATAACACATTAGAACATTCTGACTGTGCTTCATGGTGAACAGGAGG
CCATTATGATATTCTGGTGAAGTTGGTCTTACTGAATTAAACATGTTTACAC
GGAATCTAGATATAGAGCGTCCAACTACACAAATTAAATCTGCTTAATTGGTCAAGATAG
TGAGTTCCATTACTGCATCACTGGCTTCGATGGTGTGCTAAATGTTGGGATTTGACGGAAAT
TCCAAACTAATCTGGTACCTTATCTCTGCTTACCTTCCGGTGGCTACCTATGCGACACAG
TAATTTCAGCTGAGAAAGCTTACCCAGGACCTGTAAGCGTTCAAGAAATCACAATGCA
GCTTCGAACCAAGCAAATCAGATGGTAAAGTGACCCCTGTCACGGCAAAATATGGGGT
GCTGTATGTTATCTGGTGTGATGTTGGGATTGGTGTGCTTACAGCTGCCCACAGGTTCAAGGTAGGTA
TAAAACCTACCAACCCAAACAGTTGGTCTGGGGTGAATTGGCAAAGATTCACCGTGTG
TCTGTATGTTGACTAATACACCGCAATAGCCGAACCTGGCGCGCTCTAGATCATAAAT
TTGATTGATGTTGACGAAACAGTGTGCTTGTACACTGGTATGTTGGTGAAGGTATGGAG
AGGGTGAATTCTGAGGGCTGTGAGATCTGGCCCACTTGAGGAAGATTACAGGGAGG
TCGGTGTGACACCGGTGATGGGGAGGGCGAAGGGGAAGGTGAAGAGTTAGGTTCAT
TTGTTTGAGTTGTTATGTTCTGGTTACTCGTCCATAATAACTTATCTTATAACTA
CTTGTGGTTTCTGGTGTGCAATGCTGTTGGAGTTTCCCTGTTGACGTCT
TAACTGTTGATAAGACTGACATTAATATAGGCTTTGATTAAGGTAAAGTTGCTTAAA
CAATTCCACTGGTGCACAGCAGCAACTAAACGTGACTAGATGCTTGGAAATGAATTA
```

Gene-based displays

- Summary
- Splice variants
- Sequence**
- External references
- Ontologies
 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process
- Comparative Genomics
 - Gene tree
 - Orthologues
 - Paralogues



GO terms

WormBase ParaSite

Species List BLAST BioMart REST API Downloads WormBase

Search WormBase ParaSite...
e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-est-4, est-4 or metallopeptidase

Login Register Help and Documentation

Schistosoma mansoni (PRJEA36577) ▾ Location: Smp.Chr_4:31,987,438-31,988,870 Gene: SAT1

Gene-based displays

- Summary
- Splice variants
- Sequence
- External references
- Ontologies**
 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process
- Comparative Genomics
- Gene tree
- Orthologues
- Paralogues

Gene: SAT1 Smp_090120

Description Putative alpha tubulin (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VHT1]

Location Scaffold Smp.Chr_4: 31,987,438-31,988,870 reverse strand.

About this gene This gene has 2 transcripts (splice variants), 340 orthologues and 4 paralogues.

Gene type Protein coding

Annotation Method Gene models from Wellcome Trust Sanger Institute [Reference Helminth Genomes project](#)

Transcripts Hide transcript table

Show/hide columns (2 hidden) Filter

Name	Transcript ID	bp	Protein	Biotype	UniProt
Novel	Smp_090120.1	1356	451aa	Protein coding	G4VHT1 Q26595
Novel	Smp_090120.2	1311	436aa	Protein coding	G4VHT1

GO: Molecular function

Accession	Term	Evidence	Annotation Source	Transcript IDs	Actions
GO:0000166	nucleotide binding	IEA	UniProtKB/TrEMBL;Q26595;SCHMA	Smp_090120.1 Smp_090120.2	Search Biomart
GO:0003924	GTPase activity	IEA	UniProtKB/TrEMBL;Q26595;SCHMA	Smp_090120.1 Smp_090120.2	Search Biomart
GO:0005200	structural constituent of cytoskeleton	IEA	UniProtKB/TrEMBL;Q26595;SCHMA	Smp_090120.1 Smp_090120.2	Search Biomart
GO:0005525	GTP binding	IEA	UniProtKB/TrEMBL;Q26595;SCHMA	Smp_090120.1 Smp_090120.2	Search Biomart



Transcripts



Brugia malayi (PRJNA10729) ▾ Location: Bmal_v3_scaffold83:109,351-111,179 Gene: Bm97 Transcript: Bm97.2 [View at WormBase central](#)

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- External References
 - General identifiers
 - Ontology
- Protein Information
 - Protein summary
 - Domains & features
- External data

[Configure this page](#)

[Add your data](#)

[Export data](#)

[Share this page](#)

[Bookmark this page](#)

Transcript: Bm97.2

Description Larval allergen [Source:UniProtKB/TrEMBL;Acc:Q8IS06]
Location SuperContig Bmal_v3_scaffold83: 109,351-111,177 reverse strand.
About this transcript This transcript has 5 exons and is annotated with 7 domains and features.
Gene This transcript is a product of gene WBGene00220358 [Hide transcript table](#)

Show/hide columns (1 hidden) Filter

Name	Transcript ID	bp	Protein	Translation ID	Biotype	UniProt
Bm97.2	Bm97.2	975	247aa	Bm97.2	Protein coding	Q8IS06
Bm97.1	Bm97.1	965	247aa	Bm97.1	Protein coding	Q8IS06

Summary

Statistics: Exons: 5 Coding exons: 4 Transcript length: 975 bps Translation length: 247 residues
Prediction Method: Protein-coding model imported from WormBase

Information

[Help and Documentation](#)
[About WormBase ParaSite](#)
[Data Usage](#)
[Contact Us](#)
[Disclaimer](#)

Release 4 - November 2015
EMBL-EBI
WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

Transcripts

Brugia malayi (PRJNA10729) ▾ Location: Bmal_v3_scaffold83:109,351-111,179 Gene: Bm97 Transcript: Bm97.2

[View at WormBase central](#)

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- External References
 - General identifiers
 - Ontology
- Protein Information
 - Protein summary
 - Domains & features
- External data

Transcript: Bm97.2

Description Larval allergen [Source:UniProtKB/TrEMBL;Acc:[Q8IS06](#)]
Location SuperContig Bmal_v3_scaffold83: 109,351-111,177 reverse strand.
About this transcript This transcript has 5 exons and is annotated with 7 domains and features.
Gene This transcript is a product of gene [WBGene00220358](#) [Hide transcript table](#)

Show/hide columns (1 hidden) Filter

Name	Transcript ID	bp	Protein	Translation ID	Biotype	UniProt
Bm97.2	Bm97.2	975	247aa	Bm97.2	Protein coding	Q8IS06
Bm97.1	Bm97.1	965	247aa	Bm97.1	Protein coding	Q8IS06

Summary

Reverse strand 1.83 kb

Statistics Exons: 5 Coding exons: 4 Transcript length: 975 bps Translation length: 247 residues
Prediction Method Protein-coding model imported from [WormBase](#)

Information

Help and Documentation
About WormBase ParaSite
Data Usage
Contact Us
Disclaimer

Release 4 - November 2015
EMBL-EBI    
WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

protein domains

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- External References
 - General identifiers
- Protein Information
 - Protein summary
 - Domains & features

Protein summary

Protein domains for Bm7483.1

Bm7483

Transmembrane helix

Low complexity (Seq.)

Superfamily

Pfam

PROSITE profiles

PANTHER

Gene3D

Scale bar: 0 40 80 120 160 200 240 280 320 360 400 475

Statistics

Ave. residue weight: 109.833 g/mol
Charge: 4.5
Isoelectric point: 7.7239
Molecular weight: 52,170.59 g/mol
Number of residues: 475 aa

Location

WormBase ParaSite 

Species List BLAST BioMart REST API Downloads WormBase

Brugia malayi (PRJNA10729) Location: Bm1_v3_scaffold83:109,351-111,179 Gene: Bm97 Transcript: Bm97.2

My Account - wormbase.test@gmail.com Logout Help and Documentation

Transcript-based displays

- Summary
- Sequence
 - Exons
 - cDNA
 - Protein
- External References
 - General identifiers
- Protein Information
 - Protein summary
 - Domains & features
- External data

[Configure this page](#)

[Manage your data](#)

[Export data](#)

[Share this page](#)

[Bookmark this page](#)

Transcript: Bm97.2 [\[View region in WormBase JBrowse\]](#) [\[View gene at WormBase central\]](#)

Location SuperContig Bm1_v3_scaffold83: 109,351-111,177 reverse strand.

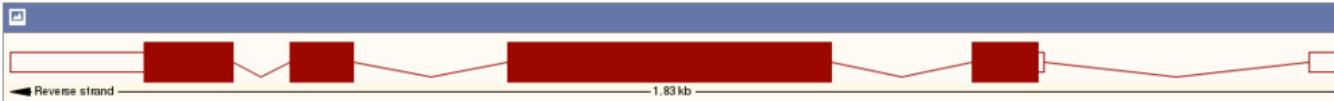
About this transcript This transcript has [5 exons](#) and is annotated with [7 domains and features](#).

Gene This transcript is a product of gene [WBGene00220358](#) [Hide transcript table](#)

Show/hide columns (2 hidden) 

Name	Transcript ID	bp	Protein	Biotype	UniProt
Bm97.2	Bm97.2	975	247aa	Protein coding	A0A0J9XM47 Q8IS06
Bm97.1	Bm97.1	965	247aa	Protein coding	A0A0J9XM47 Q8IS06

Summary



Reverse strand 1.83 kb

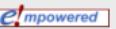
Statistics Exons: 5 Coding exons: 4 Transcript length: 975 bps Translation length: 247 residues

Prediction Method Protein-coding model imported from [WormBase](#)

Information

Help and Documentation
Citing WormBase ParaSite
About WormBase ParaSite
Data Usage
Contact Us
Disclaimer

Release 5 - January 2016

EMBL-EBI    

WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

Location view: zooming

Location: Bmal_v3_scaffold83:108,886-111,920 Gene: Bm97

SuperContig Bmal_v3_scaffold83: 74,678-76,964 [\[View region in WormBase JBrowse\]](#)

Region in detail

Contigs Protein coding genes

Gene Legend: protein coding

Location: Bmal_v3_scaffold83:108886-111920 Go Gene: Go

Contigs Protein coding genes

%GC

Gene Legend: Protein Coding protein coding

There are currently 15 tracks turned off.

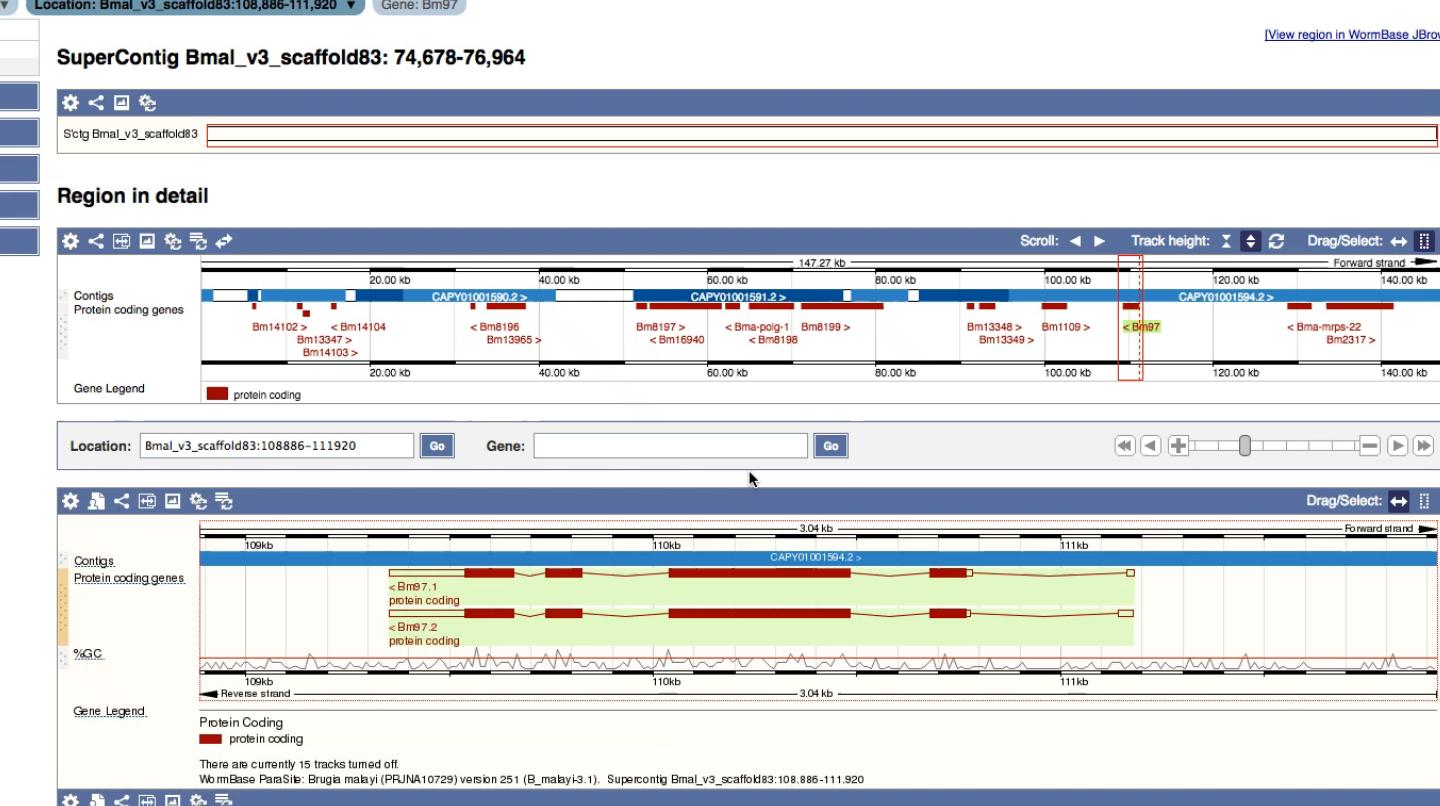
WormBase ParaSite: Brugia malayi (PRJNA10729) version 251 (B_malayi-3.1). Supercontig Bmal_v3_scaffold83:108,886-111,920

Information

Help and Documentation

Citing WormBase ParaSite

About WormBase ParaSite



Viewing gene/transcript info in browser

WormBase ParaSite

Species List BLAST BioMart REST API Downloads WormBase

Brugia malayi (PRJNA10729) Location: Bmal_v3_scaffold83:51,712-147,279 Gene: Bm97

Search WormBase ParaSite... e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bm-eat-4, eat-4 or metallopeptidase

Location-based displays
Whole genome
Region in detail

Configure this page
Add your data
Export data
Share this page
Bookmark this page

SuperContig Bmal_v3_scaffold83: 109,351-111,179

[View region in WormBase JBrowse] 

Region in detail

Contigs Protein coding genes

Location: Bmal_v3_scaffold83:51712-147279 Go Gene: Go

Drag/Select: 

Contigs Protein coding genes

Location: Bmal_v3_scaffold83:51712-147279 Go Gene: Go

Drag/Select: 

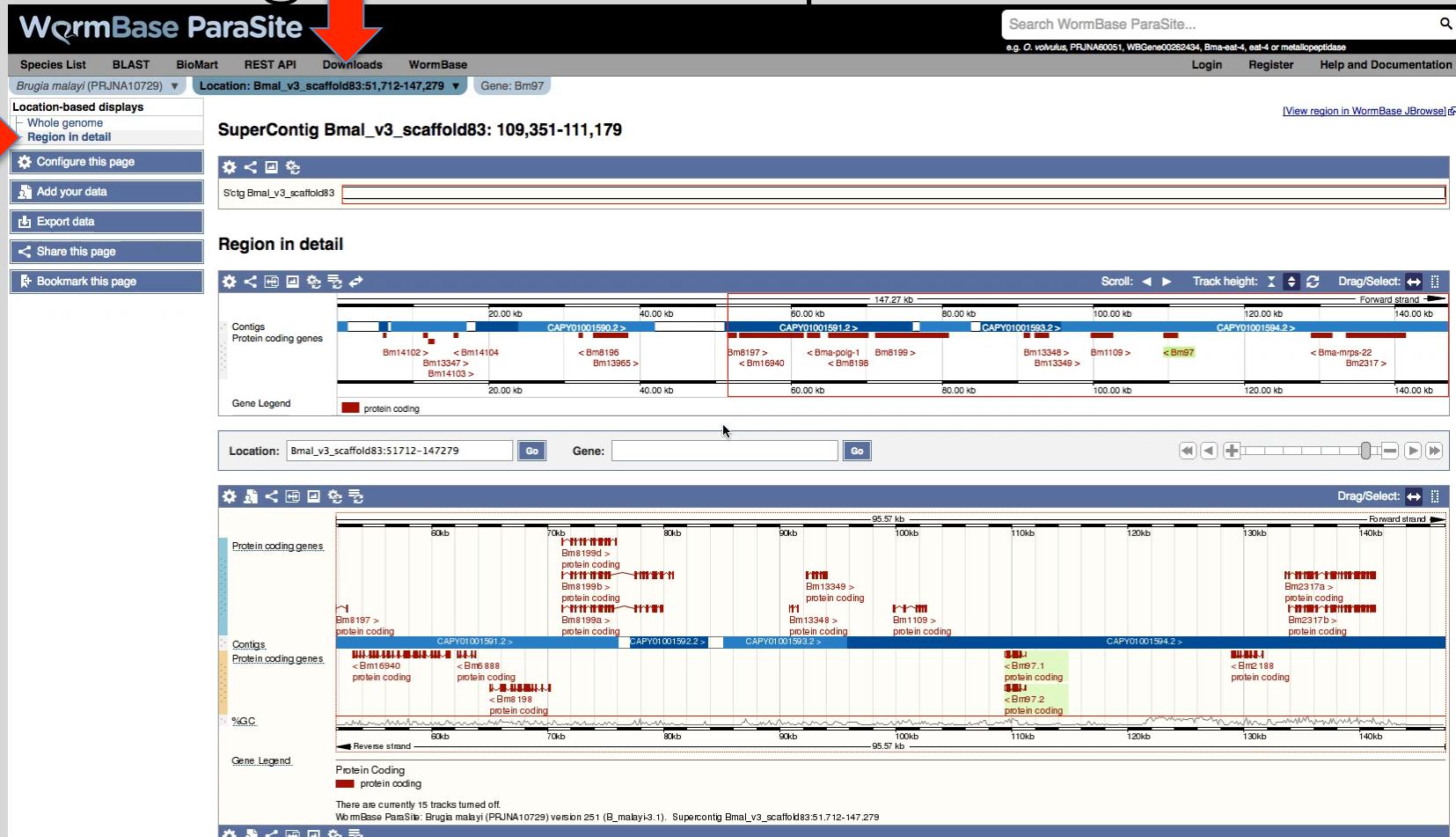
Contigs Protein coding genes

%GC

Gene Legend

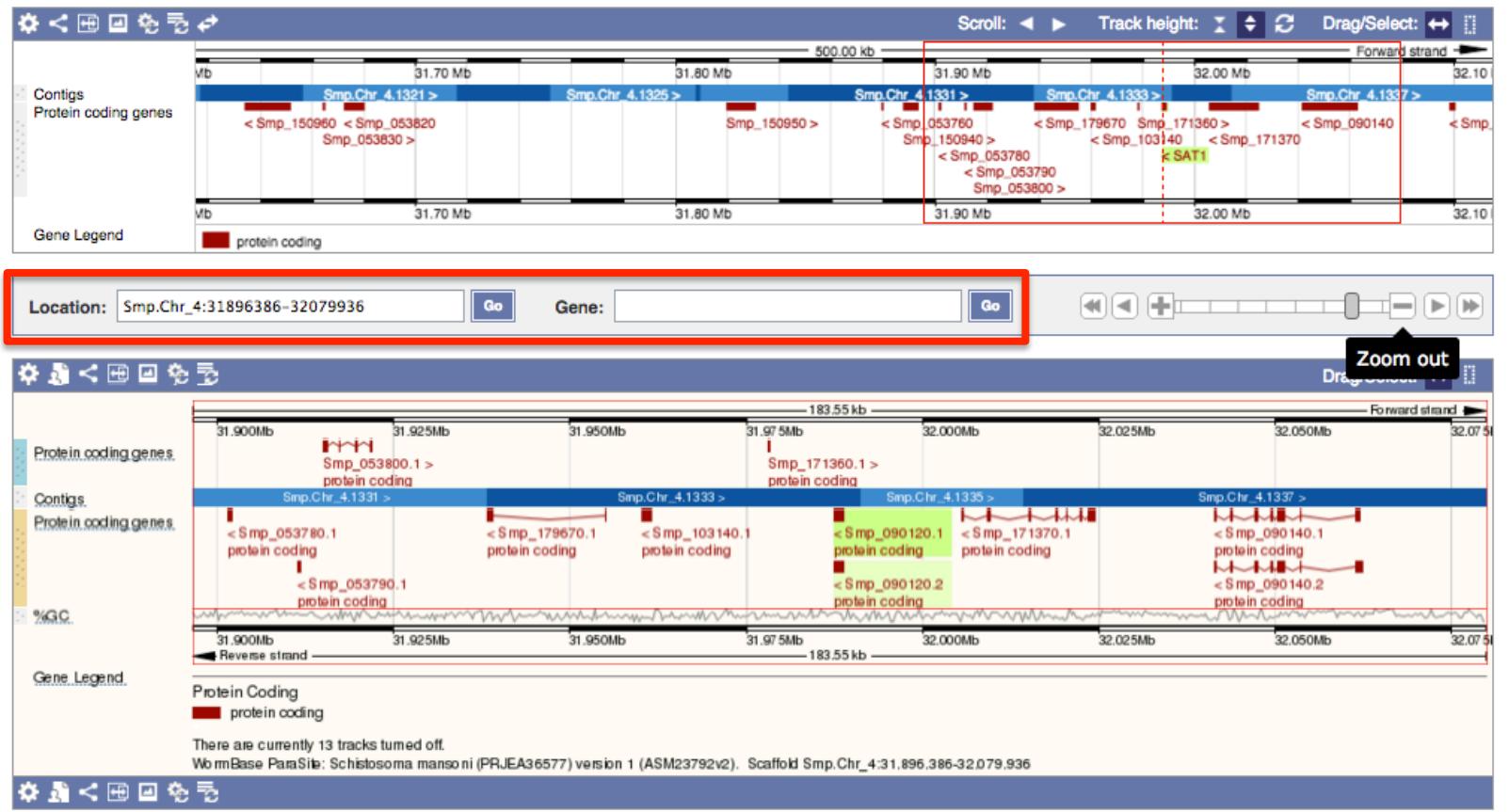
There are currently 15 tracks turned off.
WormBase ParaSite: Brugia malayi (PRJNA10729) version 251 (B_malayi-3.1). Supercontig Bmal_v3_scaffold83:51,712-147,279

There are currently 15 tracks turned off.
WormBase ParaSite: Brugia malayi (PRJNA10729) version 251 (B_malayi-3.1). Supercontig Bmal_v3_scaffold83:51,712-147,279



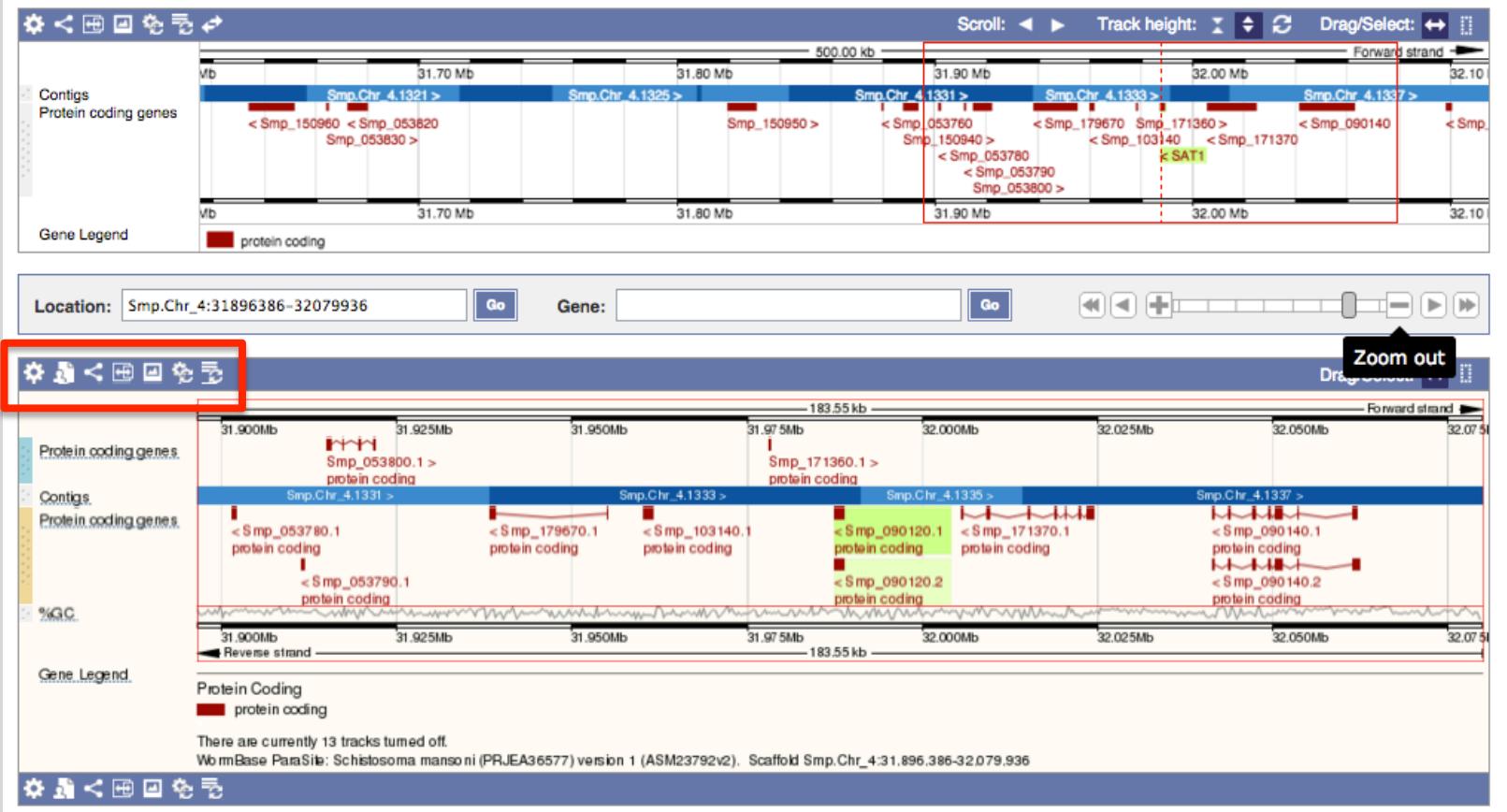
Jump to a chromosome location

Region in detail

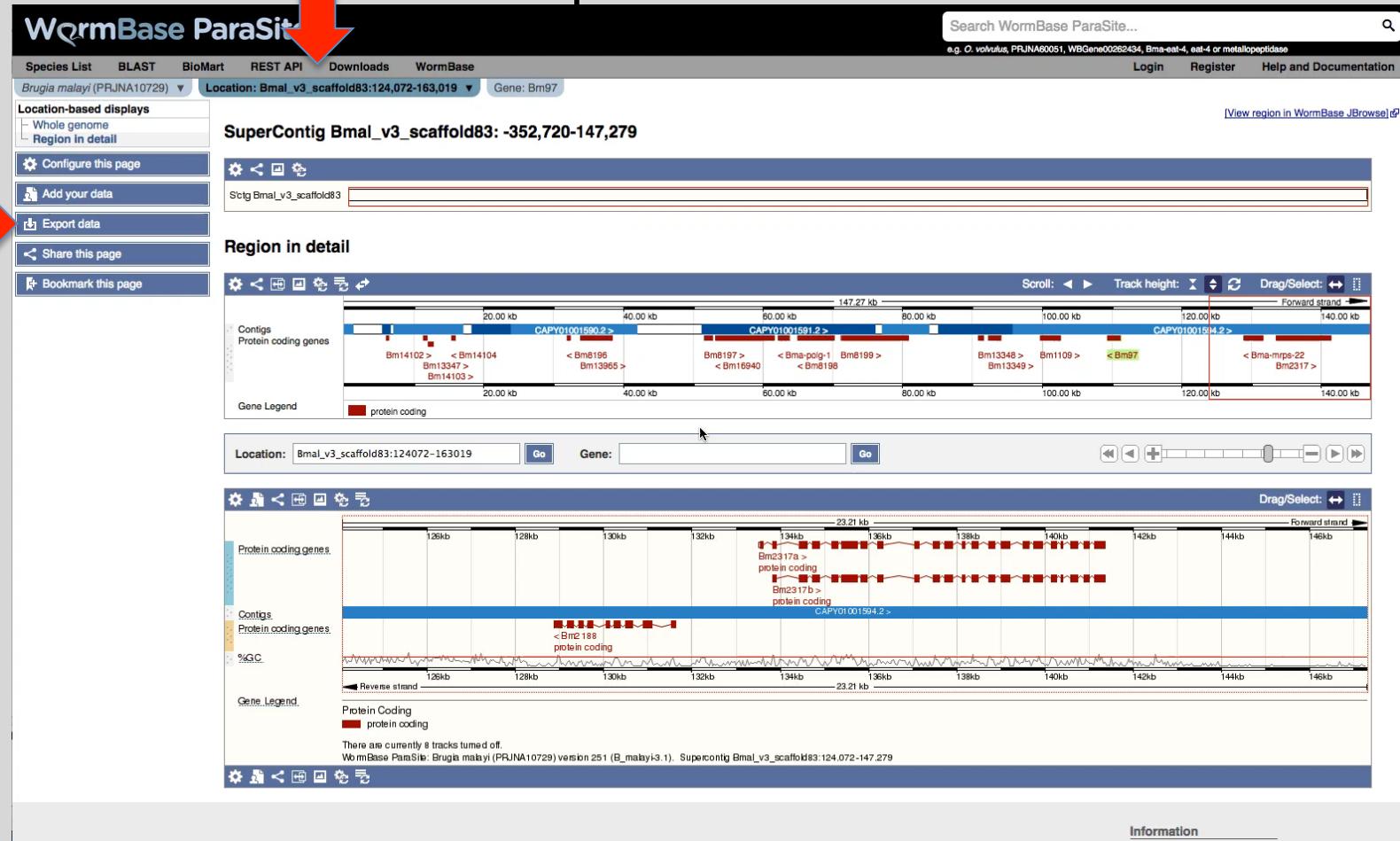


Sharing, options for browser

Region in detail



Export data



4. Comparative Genomics



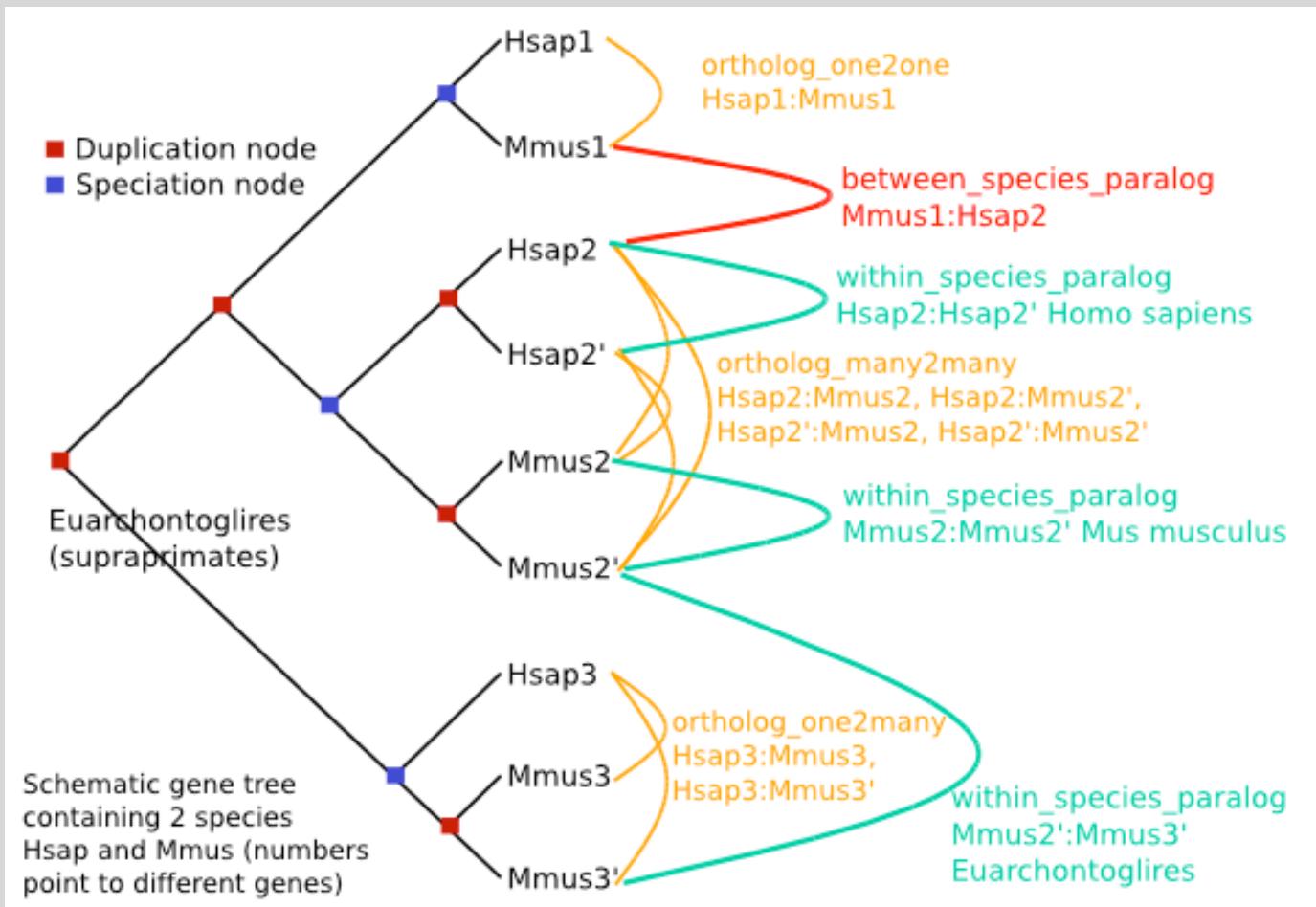
Introduction

- During each release, we compute phylogenetic trees with Compara
- 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

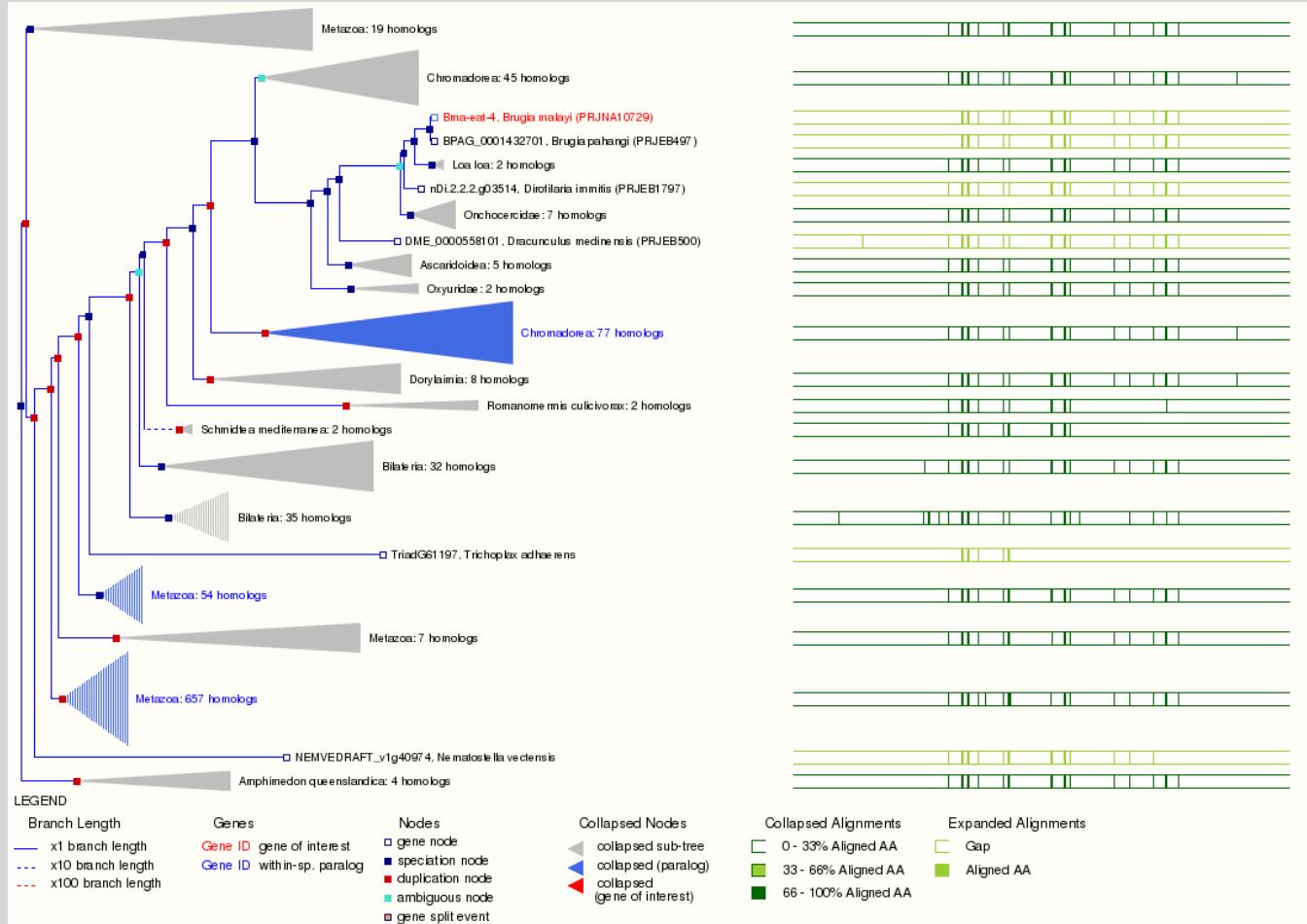
Homology types

- 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 orthologues

[View protein alignments of all orthologues](#)

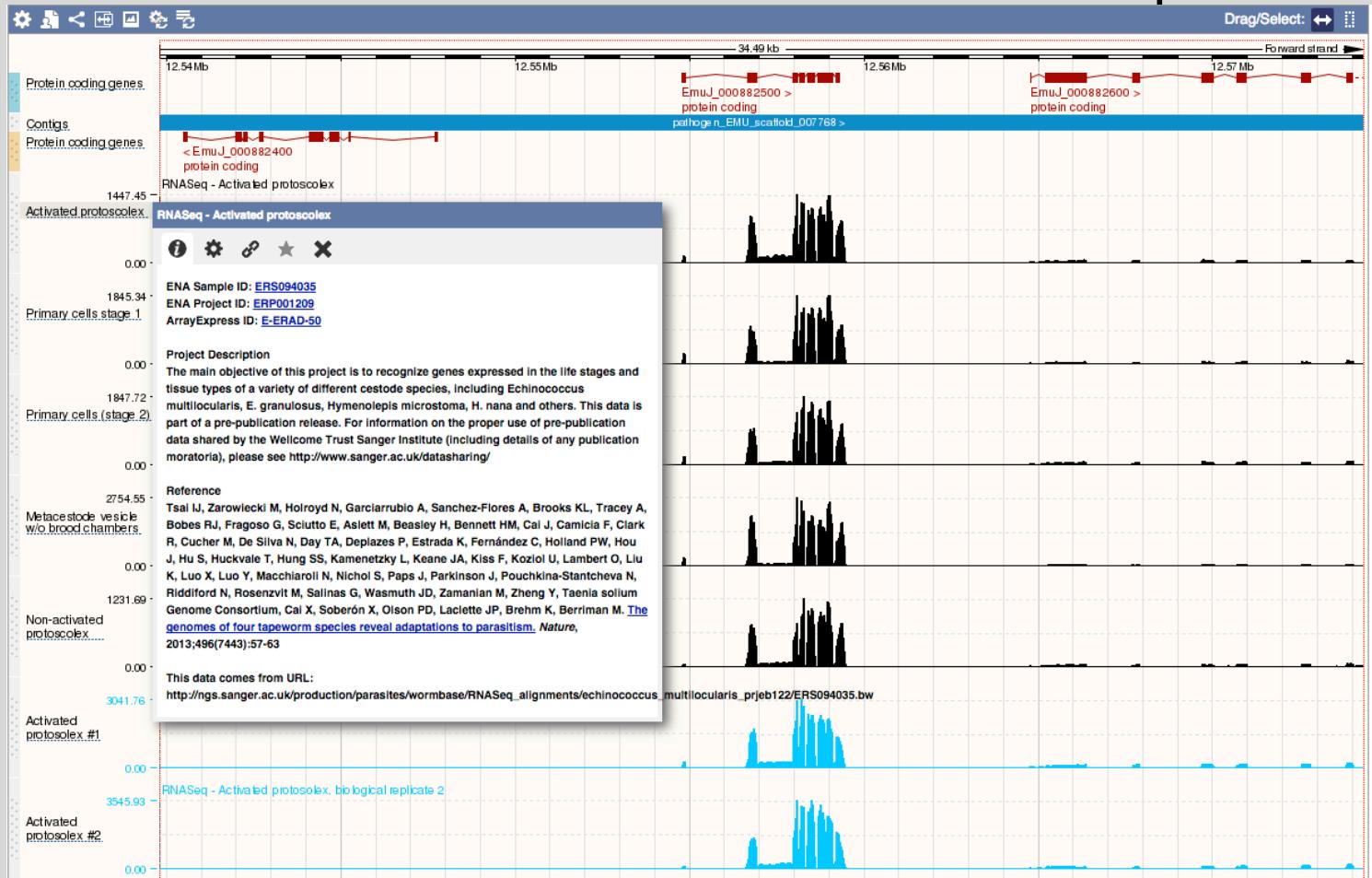
Show All <input type="button" value="entries"/>		Show/hide columns			Filter <input style="float: right;" type="button" value="X"/>		
Species	Type	dN/dS	Stable ID & gene name	Compare	Location	Target %id	Query %id
Acanthocheilonema viteae (PRJEB4306)	1-to-1	n/a	nAv.1.0.1.g03135 Probable vesicular glutamate transporter eat-4 (inferred by orthology to a <i>C. elegans</i> protein) [Source: UniProtKB; acc: P34644]	<ul style="list-style-type: none">Alignment (protein)Alignment (cDNA)Gene Tree (image)	nAv.1.0.scaf00057:63087-67685:1	73	76
Amphimedon queenslandica	Many-to-many	n/a	Aqu1.218710 No description	<ul style="list-style-type: none">Alignment (protein)Alignment (cDNA)Gene Tree (image)	Contig13347:93913-96460:1	27	28
Amphimedon queenslandica	Many-to-many	n/a	Aqu1.218711 No description	<ul style="list-style-type: none">Alignment (protein)Alignment (cDNA)Gene Tree (image)	Contig13347:96728-99289:-1	26	29
Amphimedon queenslandica	Many-to-many	n/a	Aqu1.219143 No description	<ul style="list-style-type: none">Alignment (protein)Alignment (cDNA)Gene Tree (image)	Contig13358:105162-108191:1	28	23
Amphimedon queenslandica	Many-to-many	n/a	Aqu1.219145 No description	<ul style="list-style-type: none">Alignment (protein)Alignment (cDNA)Gene Tree (image)	Contig13358:109023-112077:1	28	30
Ancylostoma caninum (PRJNA72585)	1-to-1	n/a	ANCCAN_09055 Probable vesicular glutamate transporter eat-4 (inferred by orthology to a <i>C. elegans</i> protein) [Source: UniProtKB; acc: P34644]	<ul style="list-style-type: none">Alignment (protein)Alignment (cDNA)Gene Tree (image)	ANCCANDFT_Contig116:390303-392960:1	68	60
Ancylostoma ceylanicum (PRJNA231479)	1-to-1	n/a	Acey_s0107.g3812 Probable vesicular glutamate transporter eat-4 (projected from <i>Caenorhabditis elegans</i> ortholog eat-4) [Source: UniProtKB/Swiss-Prot; acc: P34644]	<ul style="list-style-type: none">Alignment (protein)Alignment (cDNA)Gene Tree (image)	Acey_s0107_scaf:399963-412768:-1	56	67

5. RNASeq tracks



Data tracks - RNASeq

Data tracks - RNASeq



6. Adding your own data



Adding your own data

WormBase ParaSite

Species List BLAST BioMart REST API Downloads WormBase

Brugia timori (PRJEB4663) Location: BTMF_scaffold0000001:2,457-34,400

Search WormBase ParaSite... e.g. *O. volvulus*; PRJNA60051, WBGene00282434, Bm-est-4, est-4 or metallopeptidase

Help and Documentation

Location-based displays

- Whole genome
- Region in detail

Configure this page

Add your data (highlighted with a red arrow)

Export data

Share this page

Bookmark this page

Scaffold BTMF_scaffold0000001: 2,457-34,400

Region in detail

Contigs Protein coding genes

Gene Legend

Location: BTMF_scaffold0000001:2457-34400 Go Gene: Go

Contigs Protein coding genes

%GC

Gene Legend

Reverse strand

Protein Coding

Information

Help and Documentation

About WormBase ParaSite

Data Usage

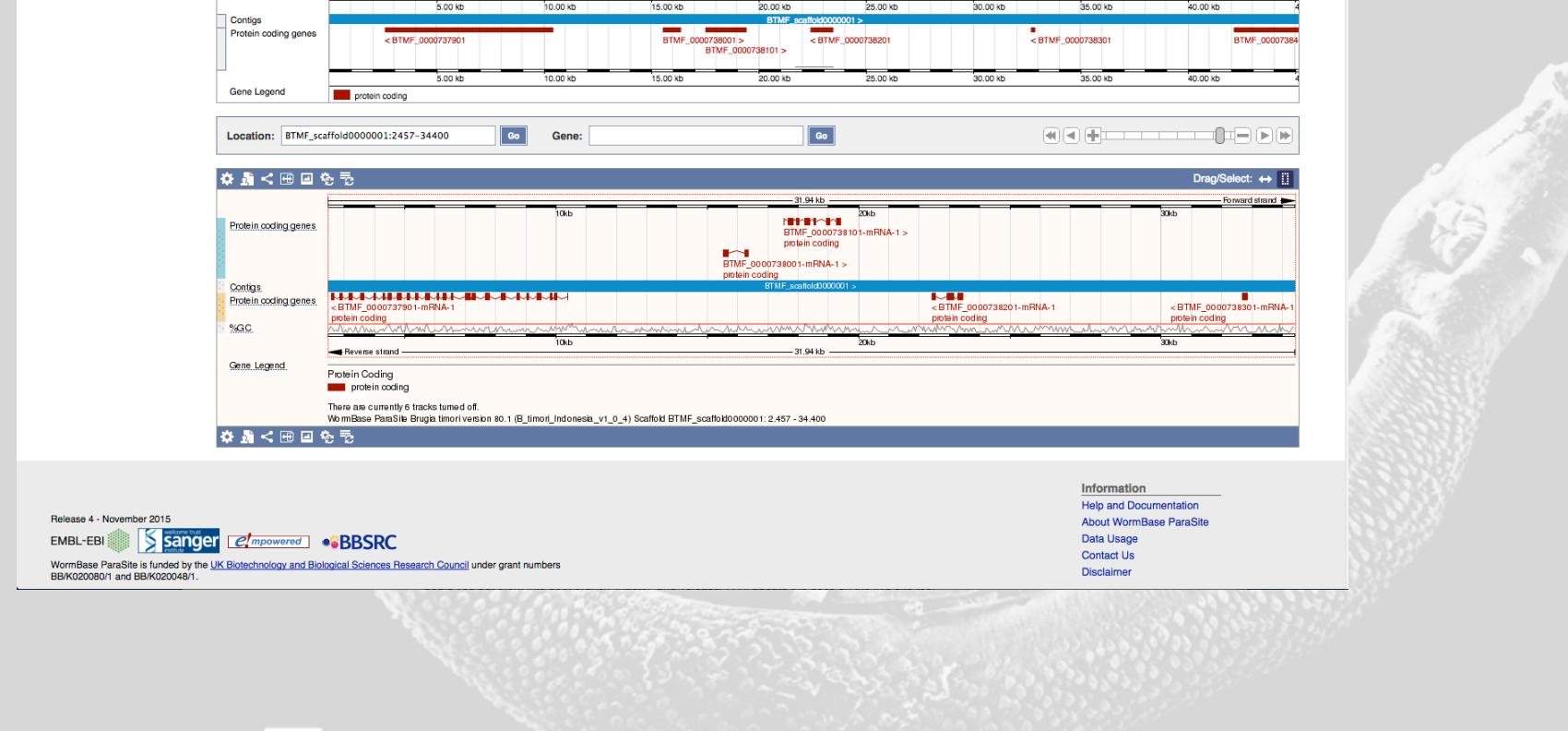
Contact Us

Disclaimer

Release 4 - November 2015

EMBL-EBI   BBSRC

WormBase ParaSite is funded by the UK Biotechnology and Biological Sciences Research Council under grant numbers BB/K020080/1 and BB/K020048/1.



Adding your own data



Account Settings

- Edit Details
- View a group
- Edit a group
- Create new group
- Join existing group
- Invite new members
- Edit bookmark
- Create new bookmark
- Share bookmark

Custom Data

- Add your data**
- Manage Data
- Features on Karyotype

Manage Configurations

- Configurations for this page
- All configurations
- Configuration sets

Logout

Add a custom track

Name for this data (optional):

Species:

Assembly:

Data:

Or upload file (max 20MB) No file chosen

[Help on supported formats, display types, etc](#)

Adding your own data

BLAST BioMart REST API Downloads WormBase My Account - ll16@sanger.ac.uk Logout Help and Documentation

Configure Region Image Configure Overview Image Configure Chromosome Image Personal Data

Add a custom track

Name for this data (optional):

Species: Trichuris muris (PRJEB126)

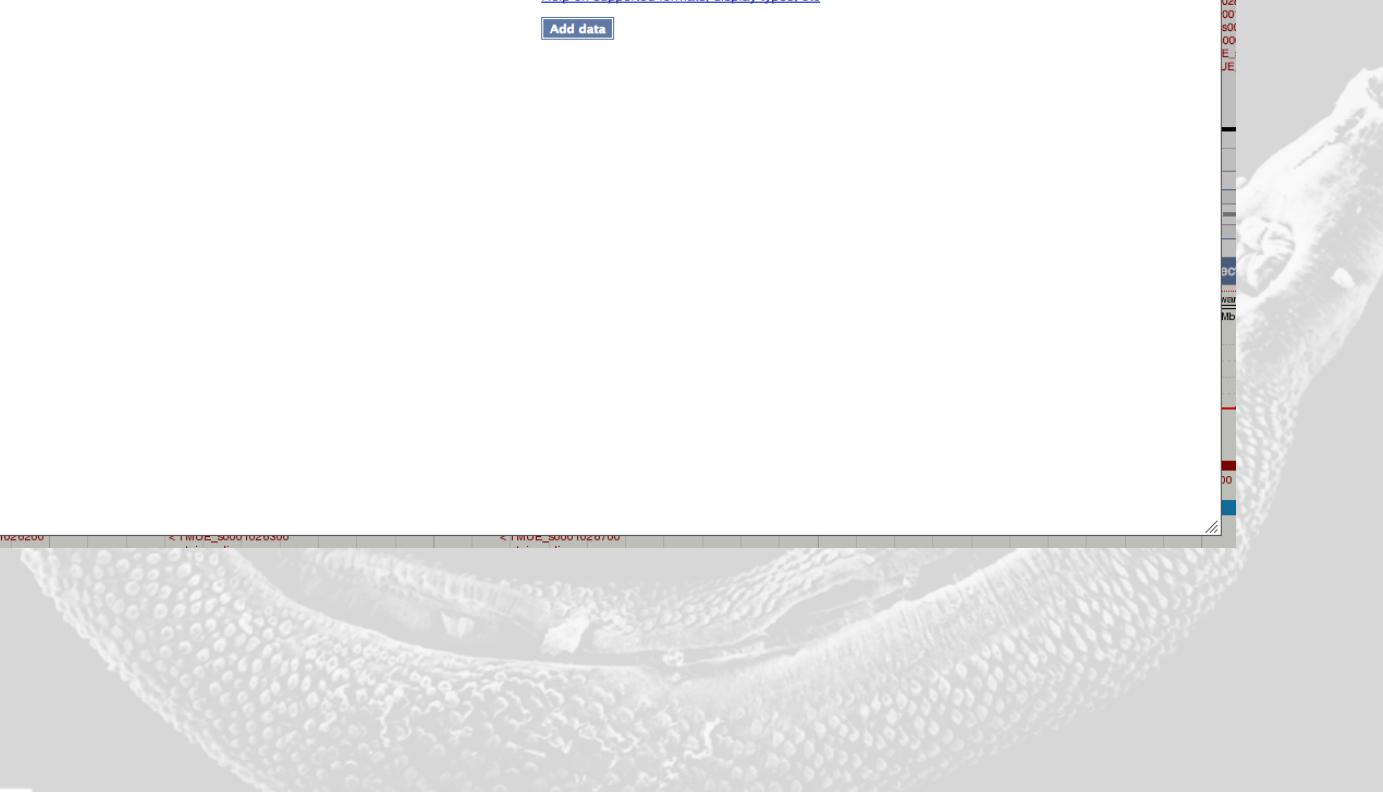
Assembly: TMUE2.2

Data: http://ngs.sanger.ac.uk/production/parasites/wormbase/RNASeq_alignments/trichuris_muris_prjeb126/ERS092077.bw

Or upload file (max 20MB) No file chosen

Data format: BigWig

[Help on supported formats, display types, etc](#)



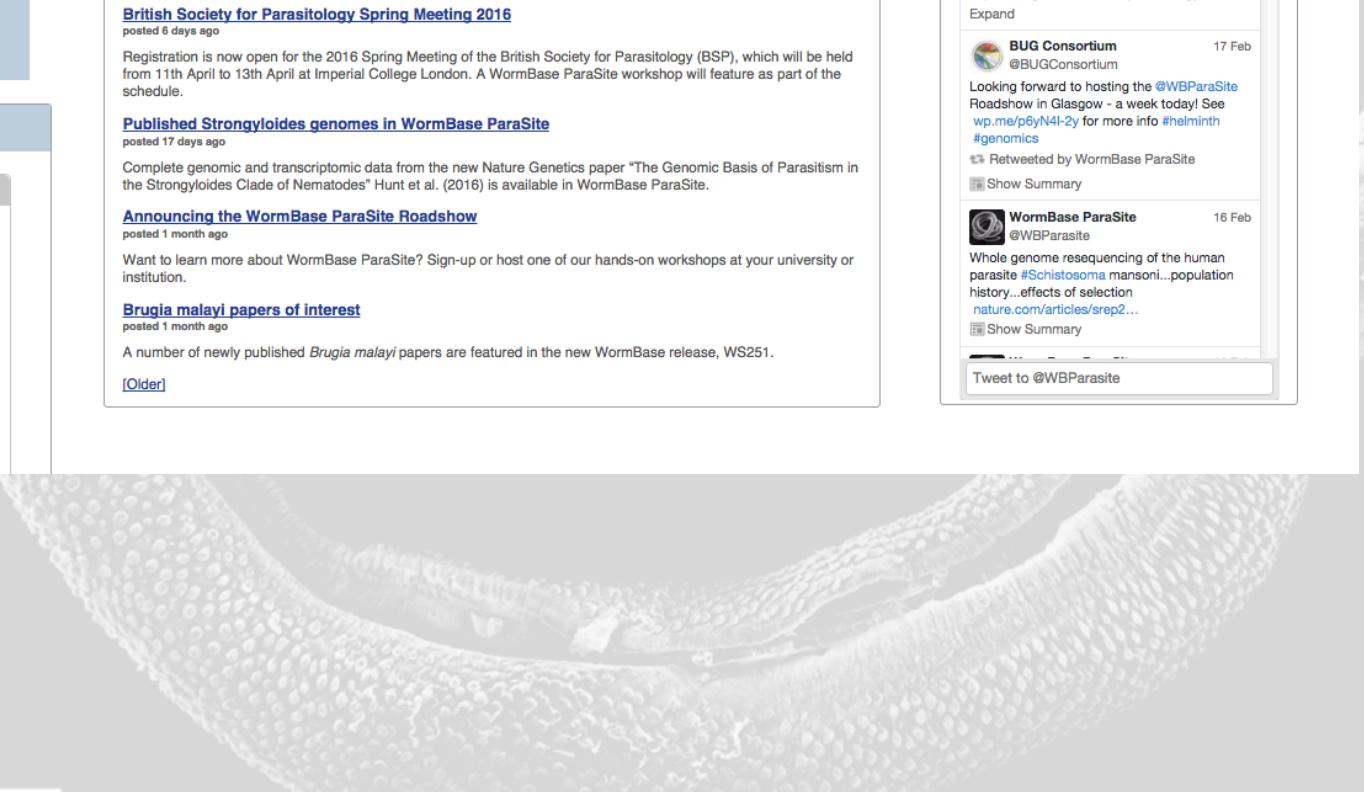
7. User Accounts



User accounts

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

User accounts



WormBase ParaSite

Species List BLAST BioMart REST API Downloads WormBase

Search WormBase ParaSite...
e.g. *O. volvulus*, PRJNA60051, WBGene0002434, Bma-eat-4, eat-4 or metal peptidase

[Login](#) [Register](#) [Help and Documentation](#)

Species **BLAST** **BioMart**

API **Downloads** **WormBase**

Find a genome

[\[-\] Nematoda \(Roundworms\)](#)

[\[-\] Clade I](#)

[Romanomermis culicivorax](#)
PRJEB1358

[Soboliphyme baturini](#)
PRJEB516

[Trichinella nativa](#)
PRJNA179527

[Trichinella spiralis](#)
PRJNA26666

Announcements

[Announcing WormBase ParaSite release 5](#)
posted 1 month ago

We are pleased to announce the fifth release of WormBase ParaSite.

Blog

[British Society for Parasitology Spring Meeting 2016](#)
posted 6 days ago

Registration is now open for the 2016 Spring Meeting of the British Society for Parasitology (BSP), which will be held from 11th April to 13th April at Imperial College London. A WormBase ParaSite workshop will feature as part of the schedule.

[Published Strongyloides genomes in WormBase ParaSite](#)
posted 17 days ago

Complete genomic and transcriptomic data from the new Nature Genetics paper "The Genomic Basis of Parasitism in the Strongyloides Clade of Nematodes" Hunt et al. (2016) is available in WormBase ParaSite.

[Announcing the WormBase ParaSite Roadshow](#)
posted 1 month ago

Want to learn more about WormBase ParaSite? Sign-up or host one of our hands-on workshops at your university or institution.

[Brugia malayi papers of interest](#)
posted 1 month ago

A number of newly published *Brugia malayi* papers are featured in the new WormBase release, WS251.

[\[Older\]](#)

Twitter

Tweets [Follow](#)

WormBase ParaSite [@WBparasite](#) 15h
Sign up for Wormbase ParaSite Workshop at the BSP Spring Meeting in April - [bsp.uk.net/news-and-events...](#) [@parasitogenomes](#) [@BSPparasitology](#)
[Expand](#)

BUG Consortium [@BUGConsortium](#) 17 Feb
Looking forward to hosting the [@WBParaSite](#) Roadshow in Glasgow - a week today! See [wp.me/p6Y4l-2y](#) for more info [#helminth](#) [#genomics](#)
[Retweeted by WormBase ParaSite](#)
[Show Summary](#)

WormBase ParaSite [@WBparasite](#) 16 Feb
Whole genome resequencing of the human parasite [#Schistosoma mansoni](#)...population history...effects of selection [nature.com/articles/srep2...](#)
[Show Summary](#)

[Tweet to @WBparasite](#)

User accounts: registering

WormBase ParaSite

Species List BLAST BioMart REST API Downloads WormBase

Search WormBase ParaSite...

e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bra-eat-4, eat-4 or metallopeptidase

Login Register Help and Documentation

Strongyloides

Register

Location-based

- Whole genome
- Region in

Configure

- Manage
- Export data
- Share this
- Bookmarks

Custom Data

- Add your data
- Manage Data
- Features on Karyotype

Manage Configurations

- Configurations for this page
- All configurations
- Configuration sets

Name:

Email Address:

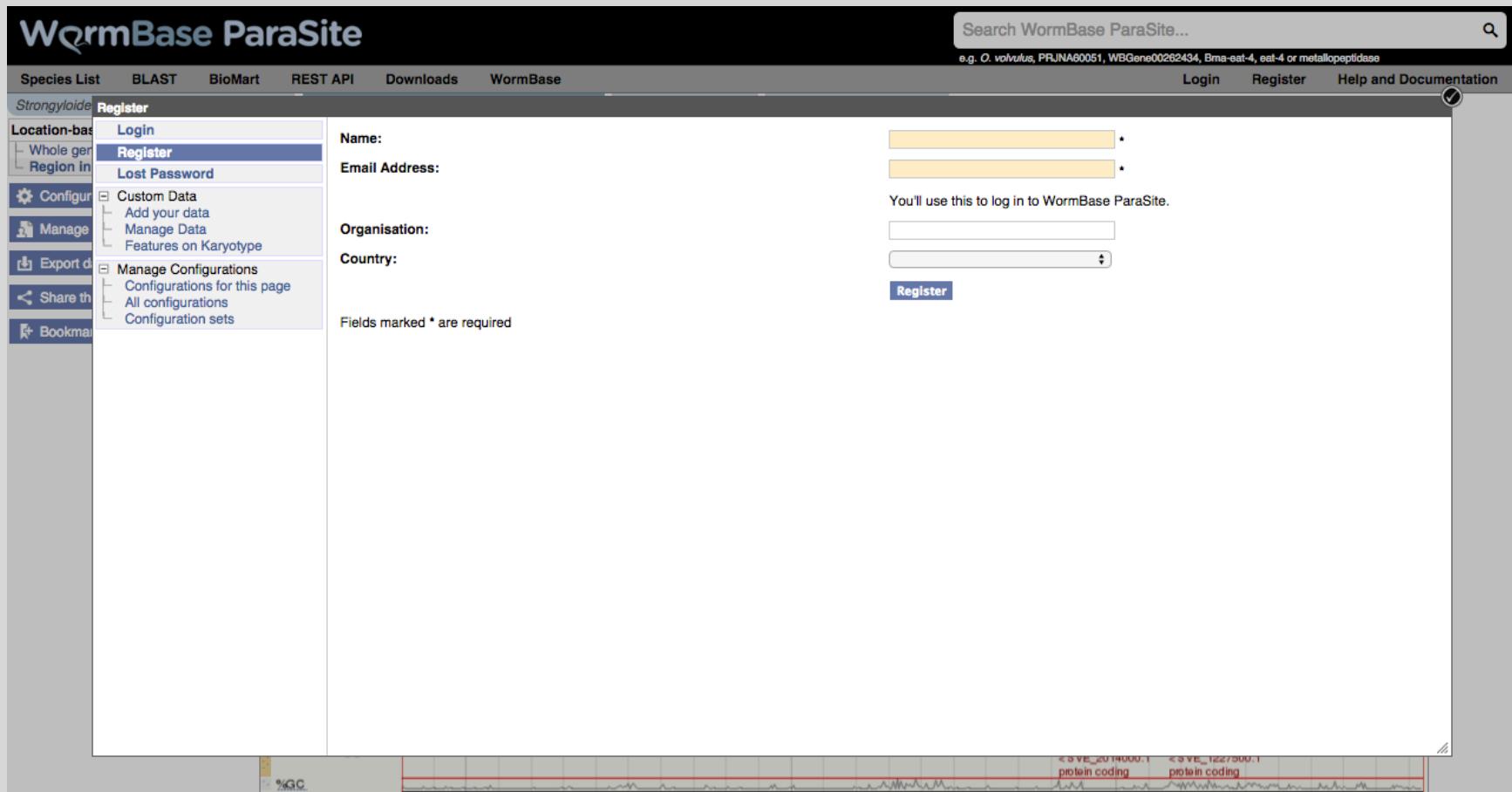
Organisation:

Country:

You'll use this to log in to WormBase ParaSite.

Fields marked * are required

Register



User accounts

WormBase ParaSite

Species List BLAST BioMart REST API Downloads WormBase

Search WormBase ParaSite...
e.g. *O. volvulus*, PRJNA60051, WBGene00282434, Brna-eat-4, eat-4 or metallopeptidase

My Account - wormbase.test@gmail.com Logout Help and Documentation

Strongyloides

My Account - wormbase.test@gmail.com

Location-based

- Whole genome
- Region in

Configure

Manage

Export data

Share this

Bookmarks

Account Settings

- Edit Details
- View a group
- Edit a group
- Create new group
- Join existing group
- Invite new members
- Edit bookmark
- Create new bookmark
- Share bookmark

Custom Data

- Add your data
- Manage Data
- Features on Karyotype

Manage Configurations

- Configurations for this page
- All configurations
- Configuration sets

Logout

User Details

Name: Test
Email: wormbase.test@gmail.com
Organisation:
Country:
Login via: WormBase ParaSite: wormbase.test@gmail.com [Change password](#)

Groups

You are not a member of any group.

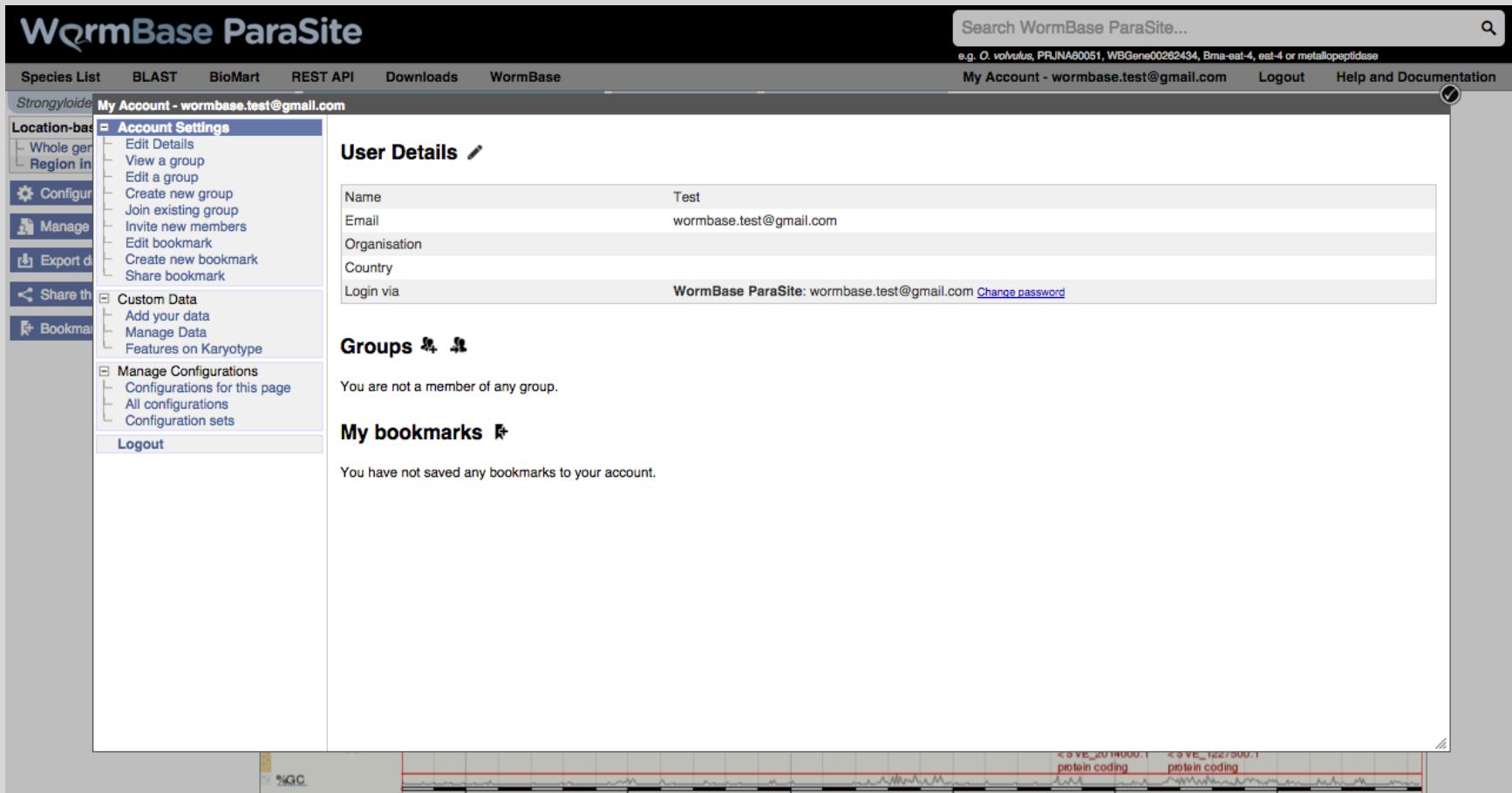
My bookmarks

You have not saved any bookmarks to your account.

GC content

protein coding

protein coding



Exercises



Answers to exercises

I. Navigate to the page for *Schistosoma mansoni*

- How many coding genes have been predicted in this genome? **10,831**
- What is the length of the genome? **364,538,298 bp**
- Which institute sequenced this genome? **Sanger**

WormBase ParaSite Version: WBPS6 (WS252)

Species List BLAST BioMart REST API VEP Downloads WormBase

Search WormBase ParaSite... 

e.g. *O. volvulus*, PFJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Login Register Help and Documentation

 Species

 BLAST

 BioMart

 API

 Downloads

 WormBase

Announcements

Announcing WormBase ParaSite release 6
posted 2 months ago

We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness
posted 2 months ago

We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [...]

Brugia malayi assembly  
posted 2 months ago

The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [...]

Register for a WormBase ParaSite workshop
posted 2 months ago

To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[\[Older\]](#)

Twitter

Tweets by @WBParasite

 WormBase ParaSite Retweeted

 British Parasitology @BSParasitology Job opportunity: Filarial Diseases Research Scientist, Northwick Park Institute for Medical Research (NPIMR). [bsp.net/jobs/](#)

 Science @scienmag White blood cells attacking a parasite. [pic.twitter.com/zqAIUZMYDx](#)

[Embed](#) [View on Twitter](#)

Release 6 - April 2016

EMBL-EBI    

WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

Information

[Help and Documentation](#) [Citing WormBase ParaSite](#) [About WormBase ParaSite](#) [Data Usage](#) [Contact Us](#) [Disclaimer](#)

Answers to exercises

2. Navigate to gene OVOC2189 from *Onchocerca volvulus*, then click on the ‘Region in detail’ link to get to the interactive browser page
 - What are the genomic coordinates of OVOC2189?
**SuperContig OVOC_OMb: 16,604,931-16,608,759
(reverse strand)**
 - Create a ‘share link’ for this display
 - Zoom out in the lower browser so that you can see more than one gene
 - Export the sequence of the region you are viewing in FASTA format (Hint: look for the ‘Export data’ button in the sidebar)

WormBase ParaSite Version: WBPS6 (WS252)

Species List BLAST BioMart REST API VEP Downloads WormBase

Search WormBase ParaSite...
e.g. *O. volvulus*, PRJNA600051, WBGene00262434, Bma-est-4, est-4 or metallopeptidase

Login Register Help and Documentation

Species

BLAST

BioMart

API

Downloads

WormBase

Announcements

Announcing WormBase ParaSite release 6
posted 2 months ago

We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness
posted 2 months ago

We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saeed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [...]

Brugia malayi assembly update
posted 2 months ago

The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [...]

Register for a WormBase ParaSite workshop
posted 2 months ago

To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[\[Older\]](#)

Twitter

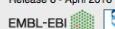
Tweets by @WBParasite

WormBase ParaSite Retweeted British Parasitology @BSPparasitology Job opportunity: Filarial Diseases Research Scientist, Northwick Park Institute for Medical Research (NPIMR). bsp.uk.net/jobs/ 15 Jun

WormBase ParaSite Retweeted Science @scienmag White blood cells attacking a parasite. pic.twitter.com/zqAIUZMYDx

Embed View on Twitter

Release 6 - April 2016



WormBase ParaSite is funded by the [UK Biotechnology and Biological Sciences Research Council](#) under grant numbers BB/K020080/1 and BB/K020048/1.

Information

- [Help and Documentation](#)
- [Citing WormBase ParaSite](#)
- [About WormBase ParaSite](#)
- [Data Usage](#)
- [Contact Us](#)
- [Disclaimer](#)

Answers to exercises

3. Scroll down the page you are on to see the RNASeq tracks aligned to this sequence

- How many studies are being displayed for this species? (Hint: studies are shown in different colours) - 2
- Identify the study ID and follow the link to see the ENA project page.
- Locate the configuration for this page and turn OFF visualization of study ERP001350 (Hint: look for the ‘Configure this page’ option in the sidebar).
- Identify the publication for study SRP056861 and navigate to the full text.

QuickTime Player File Edit View Window Help

Schisto × Smp_0 × 10 Google × www.e × WormB × WBPS × (32) Tw × https:// × http:// × wormb × School × CrossC × Jane ×

parasite.wormbase.org/Onchocerca.volvulus.prjeb513/Location/View?db=core;g=WBGene00238998;r=OVOC_OM1b:16604931-16608759;... ☆ K B

Apps Calendar QuickGO W Wikipedia webmail ENA Helix Confluence WTSI WBPS JIRA ParaSite EuropePMC UniProt Other Bookmarks

WormBase ParaSite Version: WBPS6 (WS252)

Search WormBase ParaSite... e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species List BLAST BioMart REST API VEP Downloads WormBase Login Register Help and Documentation

Onchocerca volvulus (PRJEB513) Location: OVOC_OM1b:16,604,931-16,608,759 Gene: OVOC2189 Transcript: OVOC2189 [View region in WormBase JBrowse]

Location-based displays
 Whole genome
 Region in detail

Configure this page

Add your data

Export data

Share this page

Bookmark this page

SuperContig OVOC_OM1b: 16,604,931-16,608,759

Contigs Non-coding RNA genes Protein coding genes

16.40 Mb 16.50 Mb 16.60 Mb 16.70 Mb 16.80 Mb

CBVM010000072.1 > CBVM010000072.1 >

OVOC2146 > OVOC2150 > < OVOC2157 < OVOC2165 < OVOC2174 OVOC2182 > < OVOC2147 < OVOC2152 < OVOC2151 < OVOC2168 < Ovo-hpo-21 OVOC2184 > OVOC2148 > < OVOC2153 OVOC2162 > Ovo-irr-1 > OVOC2178 > OVOC2185 > Ovo-apy-1 > Ovo-rpac-19 > OVOC2164 > OVOC2173 > < Ovo-isy-2 OVOC2186 > < Ovo-ips-41 Ovo-usp-14 > < OVOC2166 < Ovo-taf-6.2 OVOC2183 > < Ovo-dhs-29 Ovo-zhl-1 > < OVOC2169 Ovo-tbg-1 > OVOC2159 > < OVOC2167 OVOC2171 > < Ovo-tag-275 > OVOC2193 > OVOC2194 > OVOC2197 > Ovo-lcn-1 > OVOC2210 > OVOC13436 > Ovo-moma-1 > OVOC2198 > OVOC2204 > OVOC2211 > < Ovo-bal-1 < OVO2191 < Ovo-flu-2 OVOC2192 > < OVOC2196 OVOC2207 > OVOC2214 > < Ovo-dpy-7 > Ovo-gpx-6 > < Ovo-psmd-9 < OVOC2201 Ovo-gut-2 > OVOC2217 > Ovo-pqn-37 > OVOC2205 > OVOC2206 > OVOC2213 > OVOC2218 > < Ovo-osm-1

500.00 kb

Forward strand

16.40 Mb 16.50 Mb 16.60 Mb 16.70 Mb 16.80 Mb

Gene Legend protein coding RNA gene

Location: OVOC_OM1b:16604931-16608759 Go
 Gene: Go

Answers to exercises

4. Navigate to the *Trichuris muris* genome page, and click on the 'Example region' link in the Genome assembly information box:
 - Open up the 'Add your Data' window by clicking the link in the sidebar
 - Attach one of the BigWig files located at:
<http://www.ebi.ac.uk/~jane/testdata/>
by pasting the URL in to the Data box (Hint: to copy the URL, right-click file name and 'Copy link address')
 - Navigate to gene TMUE_s0016004100 and have a look at the RNASeq track. How would you judge the existing gene model? (Hint: go to the 'Region in Detail' view to see the tracks and zoom in) **(no evidence for first two exons, no evidence for intron 4)**

QuickTime Player File Edit View Window Help

WormBase ParaSite Index of /~jane WBPS Workshop Smp_034420.1 Google Calendar www.ebi.ac.uk An Integrated M Twitter Jane

parasite.wormbase.org

Apps Calendar QuickGO Wikipedia webmail ENA Helix Confluence WTSI WBPS JIRA ParaSite EuropePMC UniProt Other Bookmarks

WormBase ParaSite Version: WBPS6 (WS252)

Search WormBase ParaSite... e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species List BLAST BioMart REST API VEP Downloads WormBase Login Register Help and Documentation

Species **BLAST** **BioMart**

API **Downloads** **WormBase**

Announcements

Announcing WormBase ParaSite release 6 posted 2 months ago We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness posted 2 months ago We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in Brugia malayi Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [...]

Brugia malayi assembly update posted 2 months ago The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the Brugia malayi assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [...]

Register for a WormBase ParaSite workshop posted 2 months ago To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

Twitter

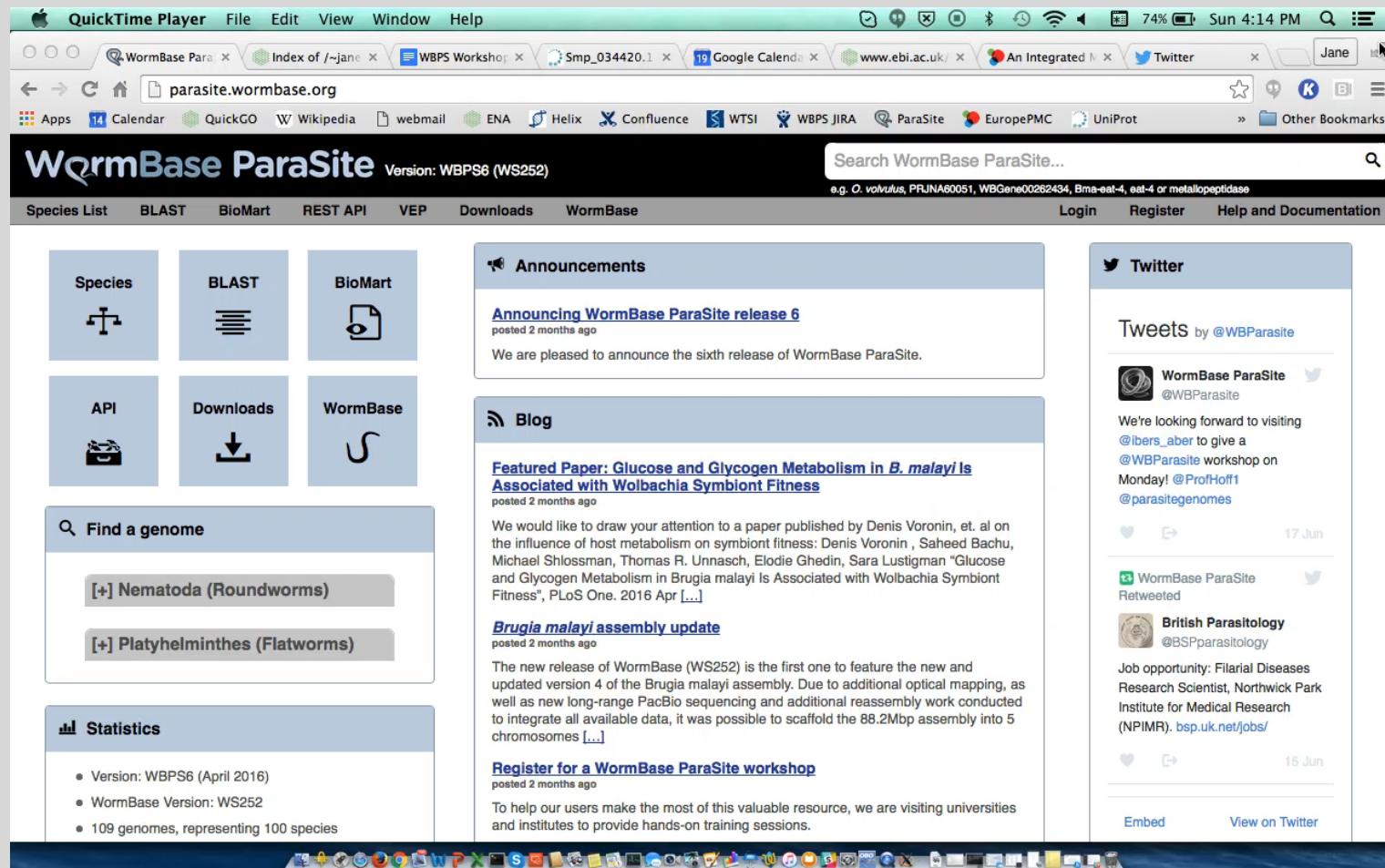
Tweets by @WBParasite

WormBase ParaSite @WBParasite We're looking forward to visiting @ibers_aber to give a @WBParasite workshop on Monday! @ProfHoff1 @parasitegenomes 17 Jun

WormBase ParaSite Retweeted

British Parasitology @BSPparasitology Job opportunity: Filarial Diseases Research Scientist, Northwick Park Institute for Medical Research (NPIMR). bsp.uk.net/jobs/ 15 Jun

Embed View on Twitter



Answers to exercises

5. Locate the gene SVE_I227300

- In which species is this gene found? *Strongyloides venezuelensis*
- What is the length of the protein product of this gene? 562aa
- How many Gene Ontology (GO) terms are assigned to this gene? 9 (3 CC, 6 BP)

QuickTime Player File Edit View Window Help

WormBase ParaSite Index of ~/jane WBPS Workshop Smp_034420.1 Google Calendar www.ebi.ac.uk An Integrated M (4) Twitter Jane

parasite.wormbase.org/?r=scaffold16:255610-270154;g=TMUE_s0016004100;db=core

Apps Calendar QuickGO Wikipedia webmail ENA Helix Confluence WTSI WBPS JIRA ParaSite EuropePMC UniProt Other Bookmarks

WormBase ParaSite Version: WBPS6 (WS252)

Search WormBase ParaSite... e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species List BLAST BioMart REST API VEP Downloads WormBase Login Register Help and Documentation

Species **BLAST** **BioMart**

API **Downloads** **WormBase**

Announcements

Announcing WormBase ParaSite release 6 posted 2 months ago We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness posted 2 months ago We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [...]

***Brugia malayi* assembly update** posted 2 months ago The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [...]

Register for a WormBase ParaSite workshop posted 2 months ago To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

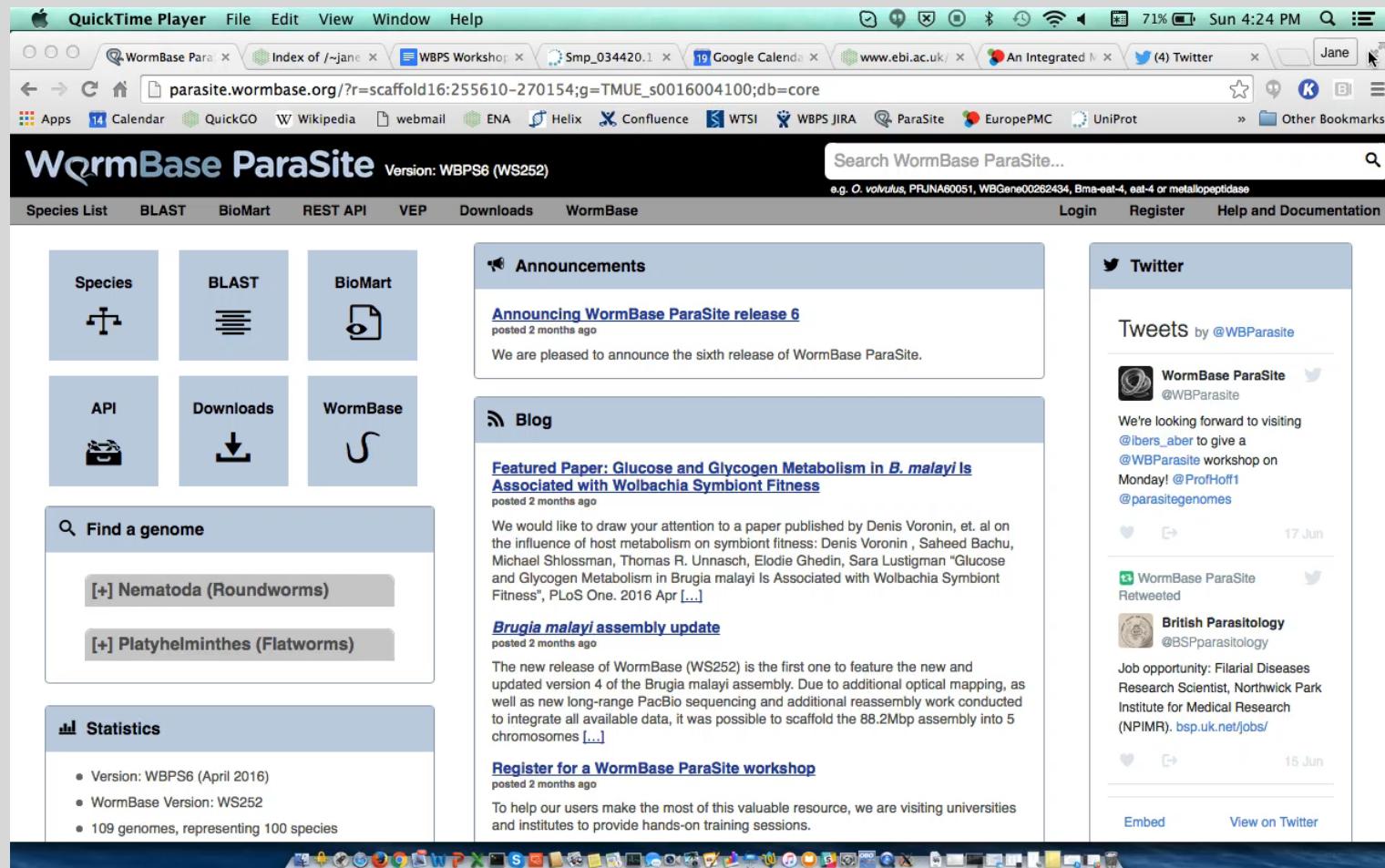
Twitter

Tweets by @WBParasite

WormBase ParaSite @WBParasite We're looking forward to visiting @ibers_aber to give a @WBParasite workshop on Monday! @ProfHoff1 @parasitegenomes 17 Jun

WormBase ParaSite Retweeted **British Parasitology** @BSPparasitology Job opportunity: Filarial Diseases Research Scientist, Northwick Park Institute for Medical Research (NPIMR). bsp.uk.net/jobs/ 15 Jun

Embed View on Twitter



Answers to exercises

6. Move onto the 'transcript' tab for SVE_I227300
 - How many exons does the single transcript of this gene have? 6
 - Which Pfam domain has been assigned to the protein product of this gene? Innexin (PF00876)

QuickTime Player File Edit View Window Help

WormBase ParaSite Index of ~/jane WBPS Workshop Smp_034420.1 Google Calendar An Integrated M Jane

parasite.wormbase.org/Strongyloides_venezuelensis_prjeb530/Gene/Summary?db=core;g=SVE_1227300;r=SVE_contig0000018:37733-396...

Apps Calendar QuickGO Wikipedia webmail ENA Helix Confluence WTSI WBPS JIRA ParaSite EuropePMC UniProt Other Bookmarks

WormBase ParaSite Version: WBPS6 (WS252)

Search WormBase ParaSite... e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species List BLAST BioMart REST API VEP Downloads WormBase Login Register Help and Documentation

Strongyloides venezuelensis (PRJEB530) Location: SVE_contig0000018:37,733-39,694 Gene: SVE_1227300 Transcript: SVE_1227300.1

Gene-based displays

- Summary
- Splice variants
- Sequence
- External references
- Ontologies
 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process
- Literature
- Comparative Genomics
 - Gene tree
 - Orthologues
 - Paralogues

Gene: SVE_1227300

Description Innixin unc-7 (projected from *Caenorhabditis elegans* ortholog unc-7) [Source:UniProtKB/Swiss-Prot;Acc:Q03412]

Location Scaffold SVE_contig0000018: 37,733-39,694 forward strand.

About this gene This gene has 1 transcript (splice variant), 89 orthologues and 16 paralogues.

Gene type Protein coding

Annotation Method Gene models from the Strongyloididae Nematode Genomes Project (unpublished) from the Parasite Genomics Group at the Wellcome Trust Sanger Institute

Transcripts Hide transcript table

Show/hide columns (2 hidden) Filter

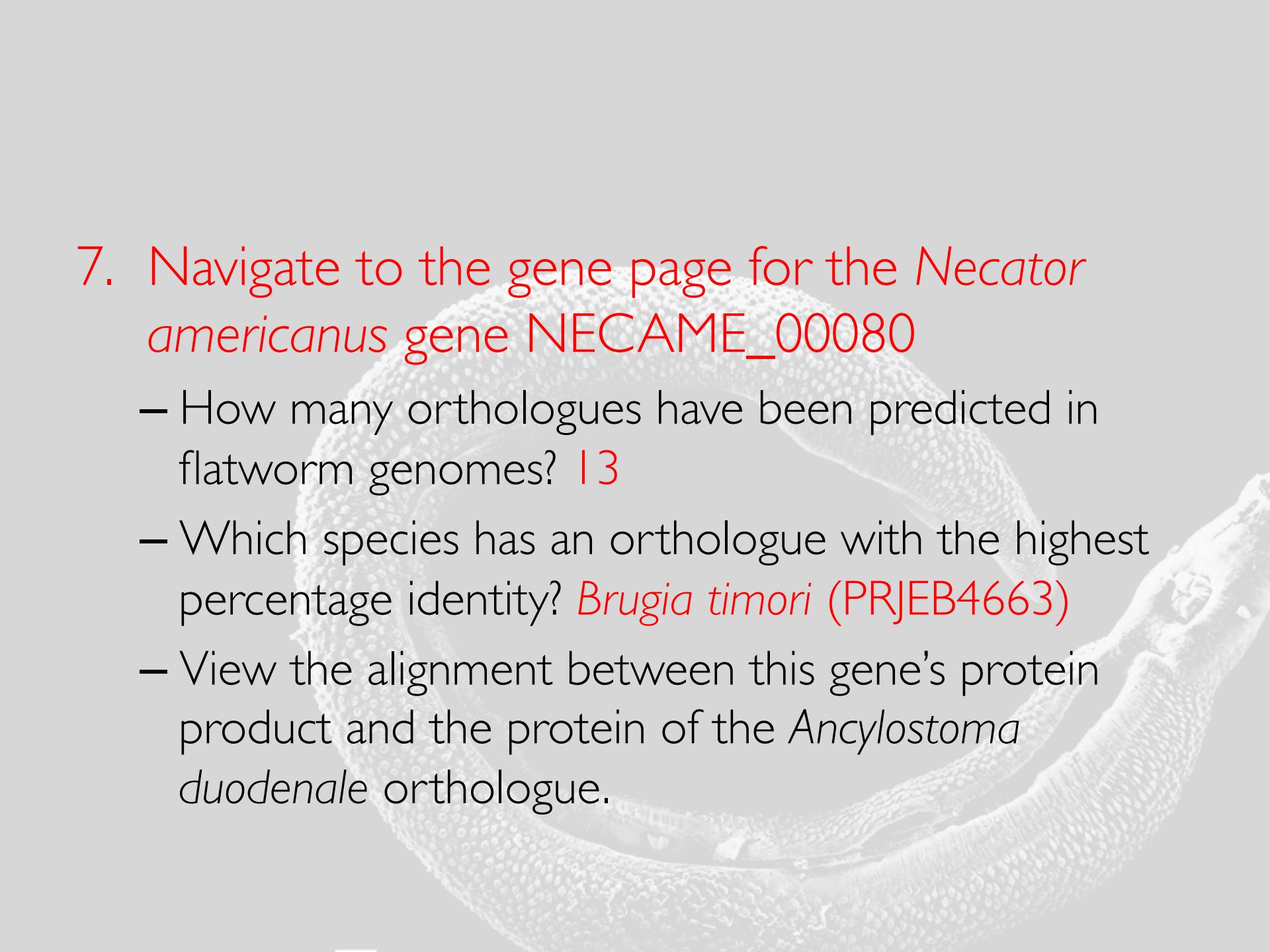
Name	Transcript ID	bp	Protein	Biotype
Novel	SVE_1227300.1	1689	562aa	Protein coding

Summary Go to Region in Detail for more tracks and navigation options (e.g. zooming)

30,000 32,500 35,000 37,500 40,000 42,500 45,000 47,500

21.96 kb Forward strand

Protein coding genes. SVE_1227200.1 > SVE_1227300.1 > SVE_1227400.1 >



7. Navigate to the gene page for the *Necator americanus* gene NECAME_00080

- How many orthologues have been predicted in flatworm genomes? 13
- Which species has an orthologue with the highest percentage identity? *Brugia timori* (PRJEB4663)
- View the alignment between this gene's protein product and the protein of the *Ancylostoma duodenale* orthologue.

QuickTime Player File Edit View Window Help

37% Sun 10:42 PM

parasite.wormbase.org

Search WormBase ParaSite... e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species List BLAST BioMart REST API VEP Downloads WormBase Login Register Help and Documentation

Species **BLAST** **BioMart**

API **Downloads** **WormBase**

Announcements

Announcing WormBase ParaSite release 6
posted 2 months ago
We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness
posted 2 months ago
We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [...]

***Brugia malayi* assembly update**
posted 2 months ago
The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [...]

Register for a WormBase ParaSite workshop
posted 2 months ago
To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

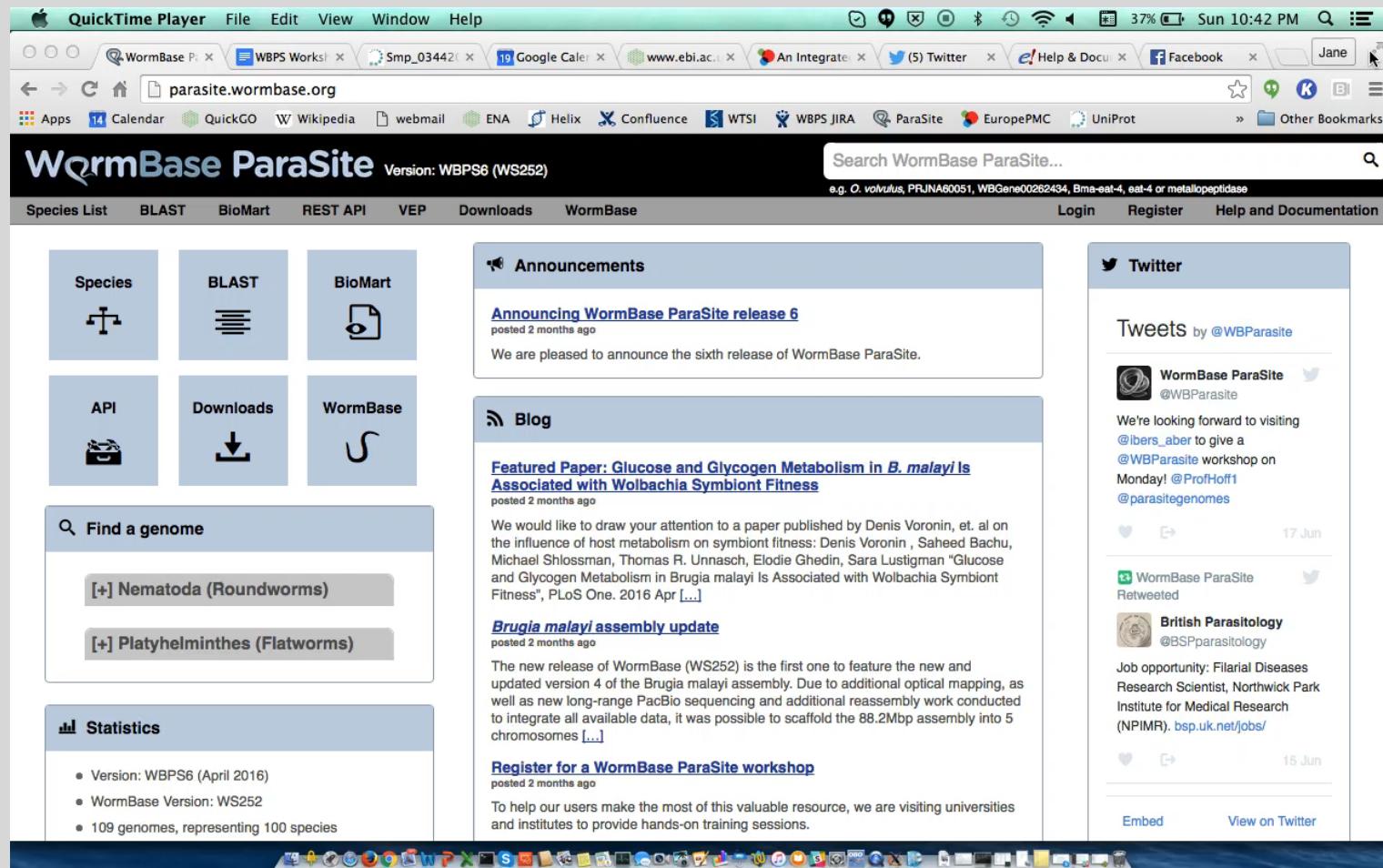
Twitter

Tweets by @WBParasite

WormBase ParaSite @WBParasite We're looking forward to visiting @ibers_aber to give a @WBParasite workshop on Monday! @ProfHoff1 @parasitegenomes 17 Jun

WormBase ParaSite Retweeted @BSPparasitology British Parasitology Job opportunity: Filarial Diseases Research Scientist, Northwick Park Institute for Medical Research (NPIMR). bsp.uk.net/jobs/ 15 Jun

Embed View on Twitter



8. Paralogues are also predicted. These are caused by duplication events.

- How many paralogues are predicted for the *Necator americanus* gene NECAME_00080?
- Look at the percent identity for this alignment - would you call this as a parologue?

QuickTime Player File Edit View Window Help

WormBase ParaSite Version: WBPS6 (WS252)

Search WormBase ParaSite... e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species List BLAST BioMart REST API VEP Downloads WormBase Login Register Help and Documentation

Necator americanus (PRJNA72135) Location: KI657455:914,690-916,914 Gene: NECAME_00080 Transcript: NECAME_00080

Gene-based displays

- Summary
- Splice variants
- Sequence
- External references
- Ontologies
 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process
- Literature
- Comparative Genomics
 - Gene tree
 - **Orthologues**
 - Paralogues

Gene: NECAME_00080

Description HMG box [Source:UniProtKB/TrEMBL;Acc:[W2U1H1](#)]

Location Scaffold KI657455: 914,690-916,914 reverse strand.

About this gene This gene has 1 transcript ([splice variant](#)), [85 orthologues](#) and [1 parologue](#).

Gene type Protein coding

Annotation Method Get models produced by the [Mitreva laboratory at the Genome Institute of Washington University](#), as described in [Tang et al \(2014\)](#)

Transcripts [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	UniProt
Novel	NECAM_00080	498	166aa	Protein coding	W2U1H1

Orthologues [?](#)

[Download orthologues](#)

Summary of orthologues of this gene
Click on 'Show' to display the orthologues for one or more groups, or click on 'Configure this page' to choose a custom list of species

Species set	Show details	1-to-1	1-to-many	many-to-many
Human	<input type="checkbox"/>	0	1	0
<i>C. elegans</i>	<input type="checkbox"/>	1	0	0

9. Locate the *Fasciola hepatica* (PRJNA179522) orthologue of the human gene BRCA2. Using the gene trees:

- How close in evolutionary history is this gene located to its orthologue?
- Are there any duplication events in the evolution of this gene and its homologues? Yes, 2

QuickTime Player File Edit View Window Help

WormBase ParaSite Version: WBPS6 (WS252)

Search WormBase ParaSite... e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species List BLAST BioMart REST API VEP Downloads WormBase Login Register Help and Documentation

Species BLAST BioMart

API Downloads WormBase

Find a genome

[+] Nematoda (Roundworms)

[+] Platyhelminthes (Flatworms)

Statistics

- Version: WBPS6 (April 2016)
- WormBase Version: WS252
- 109 genomes, representing 100 species

Announcements

Announcing WormBase ParaSite release 6
posted 2 months ago

We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness
posted 2 months ago

We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [...]

***Brugia malayi* assembly update**
posted 2 months ago

The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [...]

Register for a WormBase ParaSite workshop
posted 2 months ago

To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

Twitter

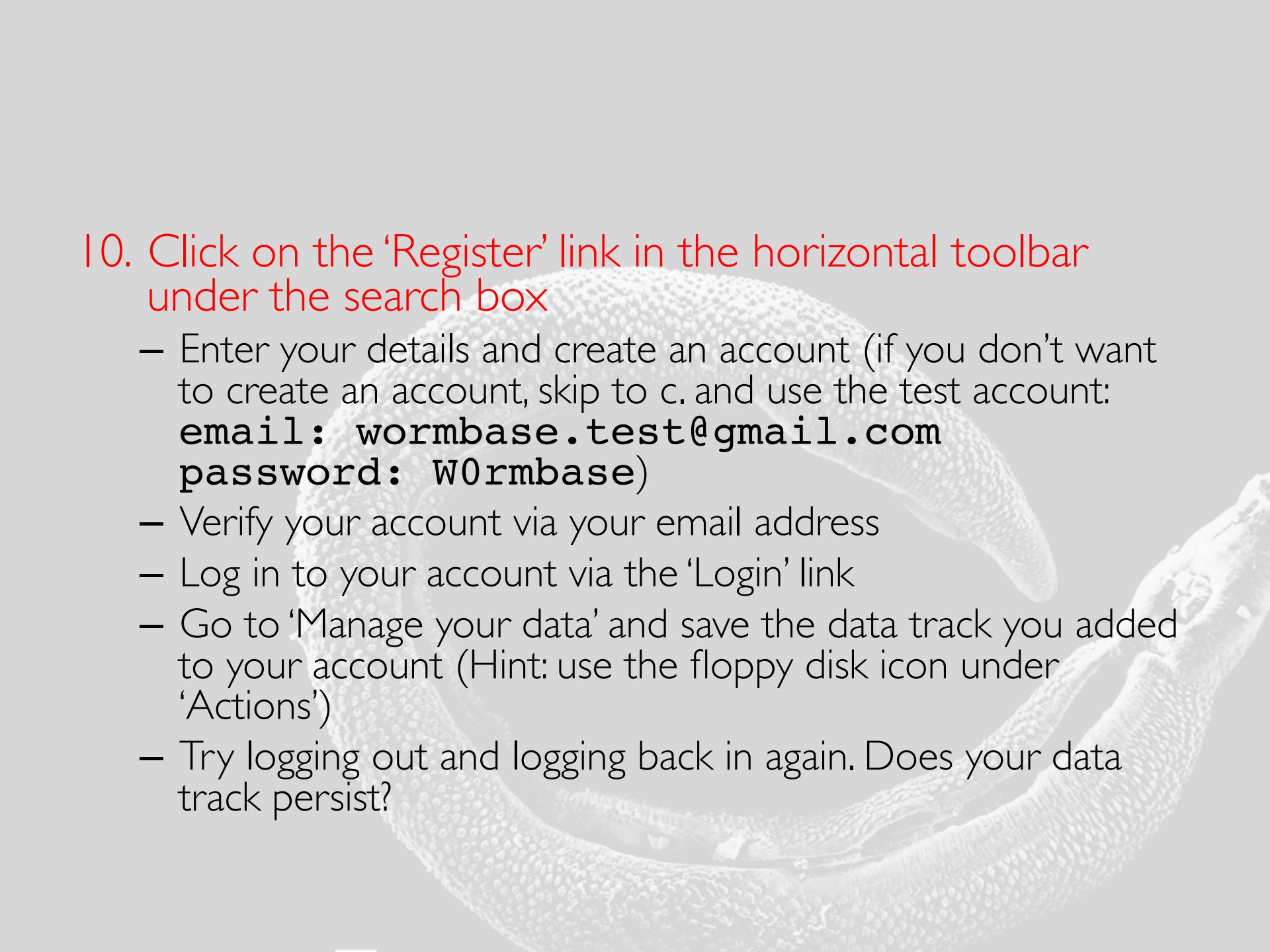
Tweets by @WBParasite

WormBase ParaSite @WBParasite We're looking forward to visiting @ibers_aber to give a @WBParasite workshop on Monday! @ProfHoff1 @parasitegenomes

WormBase ParaSite Retweeted

British Parasitology @BSPparasitology Job opportunity: Filarial Diseases Research Scientist, Northwick Park Institute for Medical Research (NPIMR). bsp.uk.net/jobs/

Embed View on Twitter



10. Click on the 'Register' link in the horizontal toolbar under the search box

- Enter your details and create an account (if you don't want to create an account, skip to c. and use the test account:
email: wormbase.test@gmail.com
password: W0rmbase)
- Verify your account via your email address
- Log in to your account via the 'Login' link
- Go to 'Manage your data' and save the data track you added to your account (Hint: use the floppy disk icon under 'Actions')
- Try logging out and logging back in again. Does your data track persist?

BioMart part II answers

- I. Using BioMart and your gene list, generate a table that contains: i. WormBase ParaSite gene ID (stable ID), ii. gene name and iii. RefSeq Protein ID.



Find a genome

[+] Nematoda (Roundworms)

[+] Platyhelminthes (Flatworms)

Statistics

- Version: WBPS6 (April 2016)
- WormBase Version: WS252
- 109 genomes, representing 100 species
- 2,189,841 genes

Announcements

Announcing WormBase ParaSite release 6

posted 2 months ago

We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness

posted 2 months ago

We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [...]

Brugia malayi assembly update

posted 2 months ago

The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [...]

Register for a WormBase ParaSite workshop

posted 2 months ago

To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[Older]

Twitter

Tweets by @WBParasite

 @WBParasite

We're looking forward to visiting @ibers_aber to give a @WBParasite workshop on Monday! @ProfHoff1 @parasitegenomes

  17 Jun

 WormBase ParaSite Retweeted

 @BSPparasitology

Job opportunity: Filarial Diseases Research Scientist, Northwick Park Institute for Medical Research (NPIMR). bsp.uk.net/jobs/

  15 Jun

[Embed](#) [View on Twitter](#)

Information

[Help and Documentation](#)

[Citing WormBase ParaSite](#)

[About WormBase ParaSite](#)

BioMart part II answers

2. Using BioMart, generate a table showing i. the WormBase ParaSite gene ID (stable ID), ii. *O. volvulus* gene name, iii. *C. elegans* orthologue gene stable ID and iv. human orthologue gene stable ID.
 - How many of these genes have an orthologue defined in both *C. elegans* and human? 4

QuickTime Player File Edit View Window Help

BioMart - M... WBPS Works! www.ebi.ac... Smp_03442... Google Calen... www.ebi.ac... An Integrate... (6) Twitter... Boden UK | Jane

parasite.wormbase.org/biomart/martview/d7eb8597e14d4d384c03180ec40db83e

Apps Calendar QuickGO Wikipedia webmail ENA Helix Confluence WTSI WBPS JIRA ParaSite EuropePMC UniProt Other Bookmarks

WormBase ParaSite

WormBase Home | ParaSite Home

Search WormBase ParaSite...

BLAST | BioMart | Downloads

New Count Results

Export all results to File TSV Unique results only Go

Email notification to

View 10 rows as HTML Unique results only

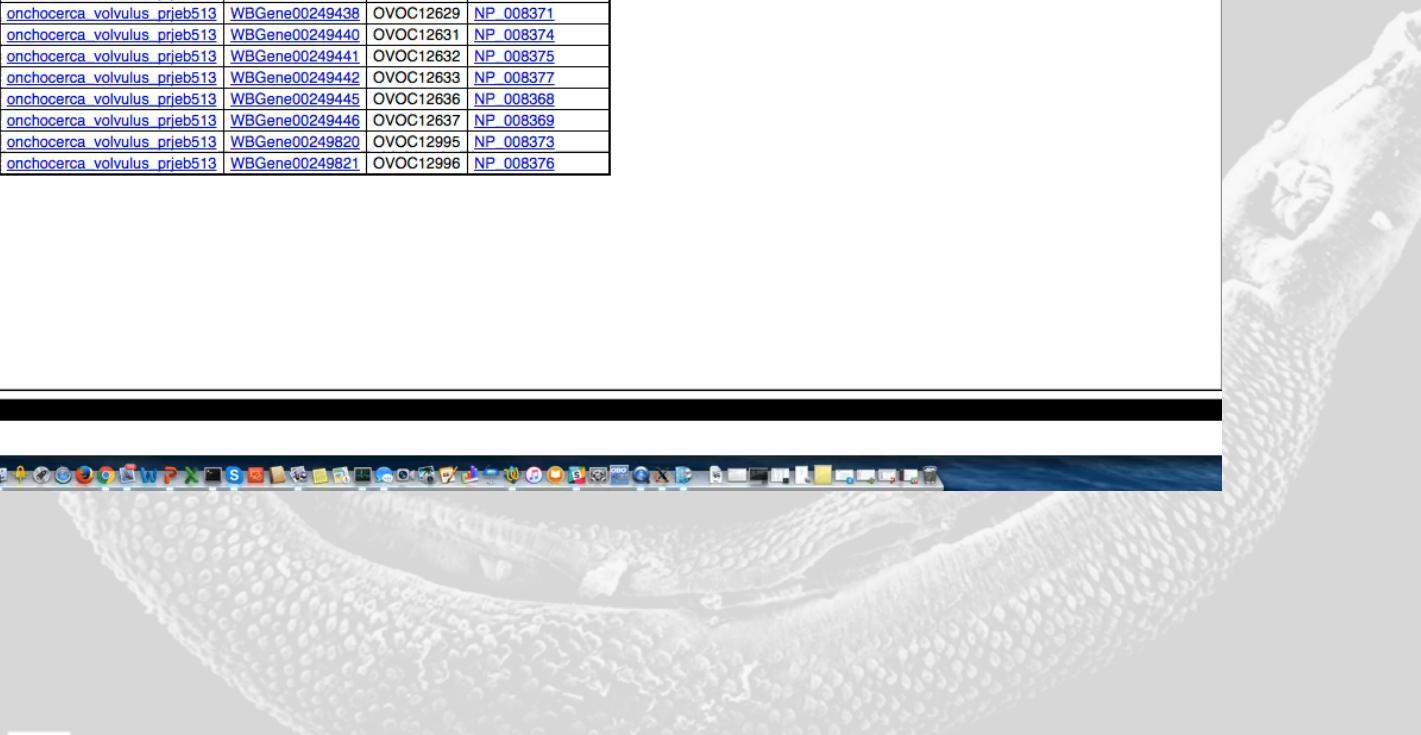
Genome project	Gene stable ID	Gene name	RefSeq protein ID
onchocerca.volvulus.prieb513	WBGene00249437	OVOC12628	NP_008370
onchocerca.volvulus.prieb513	WBGene00249438	OVOC12629	NP_008371
onchocerca.volvulus.prieb513	WBGene00249440	OVOC12631	NP_008374
onchocerca.volvulus.prieb513	WBGene00249441	OVOC12632	NP_008375
onchocerca.volvulus.prieb513	WBGene00249442	OVOC12633	NP_008377
onchocerca.volvulus.prieb513	WBGene00249445	OVOC12636	NP_008368
onchocerca.volvulus.prieb513	WBGene00249446	OVOC12637	NP_008369
onchocerca.volvulus.prieb513	WBGene00249820	OVOC12995	NP_008373
onchocerca.volvulus.prieb513	WBGene00249821	OVOC12996	NP_008376

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters
Gene stable ID(s): [ID-list specified]

2. Output Attributes
Genome project
Gene stable ID
Gene name
RefSeq protein ID

WormBase ParaSite © EBI & WTSI



BioMart part II answers

3. Using BioMart, get the sequence for the region 500bp upstream of each gene in your list. Export this as a FASTA file.

QuickTime Player File Edit View Window Help

BioMart - M... WBPS Works! www.ebi.ac... Smp_03442... Google Calen... www.ebi.ac... An Integrate... (6) Twitter... Boden UK | Jane

parasite.wormbase.org/biomart/martview/109f0407fffb4ffb112d35a294f2f352

Apps Calendar QuickGO Wikipedia webmail ENA Helix Confluence WTSI WBPS JIRA ParaSite EuropePMC UniProt Other Bookmarks

WormBase ParaSite

WormBase Home | ParaSite Home

New Count Results

Search WormBase ParaSite...

BLAST | BioMart | Downloads

URL XML Perl Help

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters
[None selected]

2. Output Attributes
Genome project
Gene stable ID

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)
For guidance, see the [relevant pages](#) of our documentation.

SPECIES

REGION (use only when selecting one species)

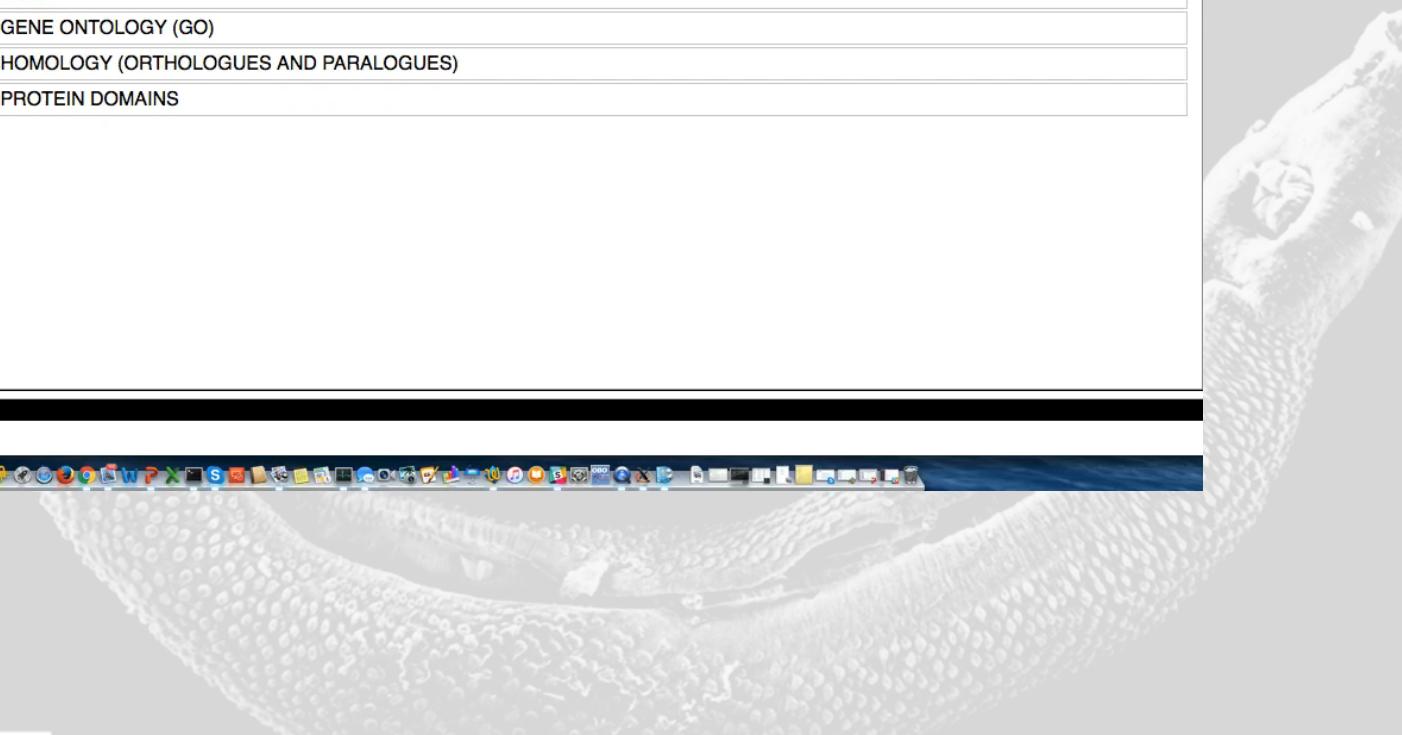
GENE

GENE ONTOLOGY (GO)

HOMOLOGY (ORTHOLOGUES AND PARALOGUES)

PROTEIN DOMAINS

WormBase ParaSite © EBI & WTSI



BioMart part II answers

4. Retrieve a table that contains:
 - i. WormBase ParaSite gene ID (stable ID),
 - ii. gene name and gene description,
 - iii. InterPro ID and
 - iv. short InterPro description.

QuickTime Player File Edit View Window Help

BioMart - M... WBPS Works! www.ebi.ac... Smp_03442... Google Calen... www.ebi.ac... An Integrate... (7) Twitter... Boden UK | Jane

parasite.wormbase.org/biomart/martview/45dc17a4137df0a954a9e3680765b8af

Apps Calendar QuickGO Wikipedia webmail ENA Helix Confluence WTSI WBPS JIRA ParaSite EuropePMC UniProt Other Bookmarks

WormBase ParaSite

WormBase Home | ParaSite Home

New Count Results

Search WormBase ParaSite...

BLAST | BioMart | Downloads

URL XML Perl Help

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters
[None selected]

2. Output Attributes
Genome project
Gene stable ID

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)
For guidance, see the [relevant pages](#) of our documentation.

SPECIES

REGION (use only when selecting one species)

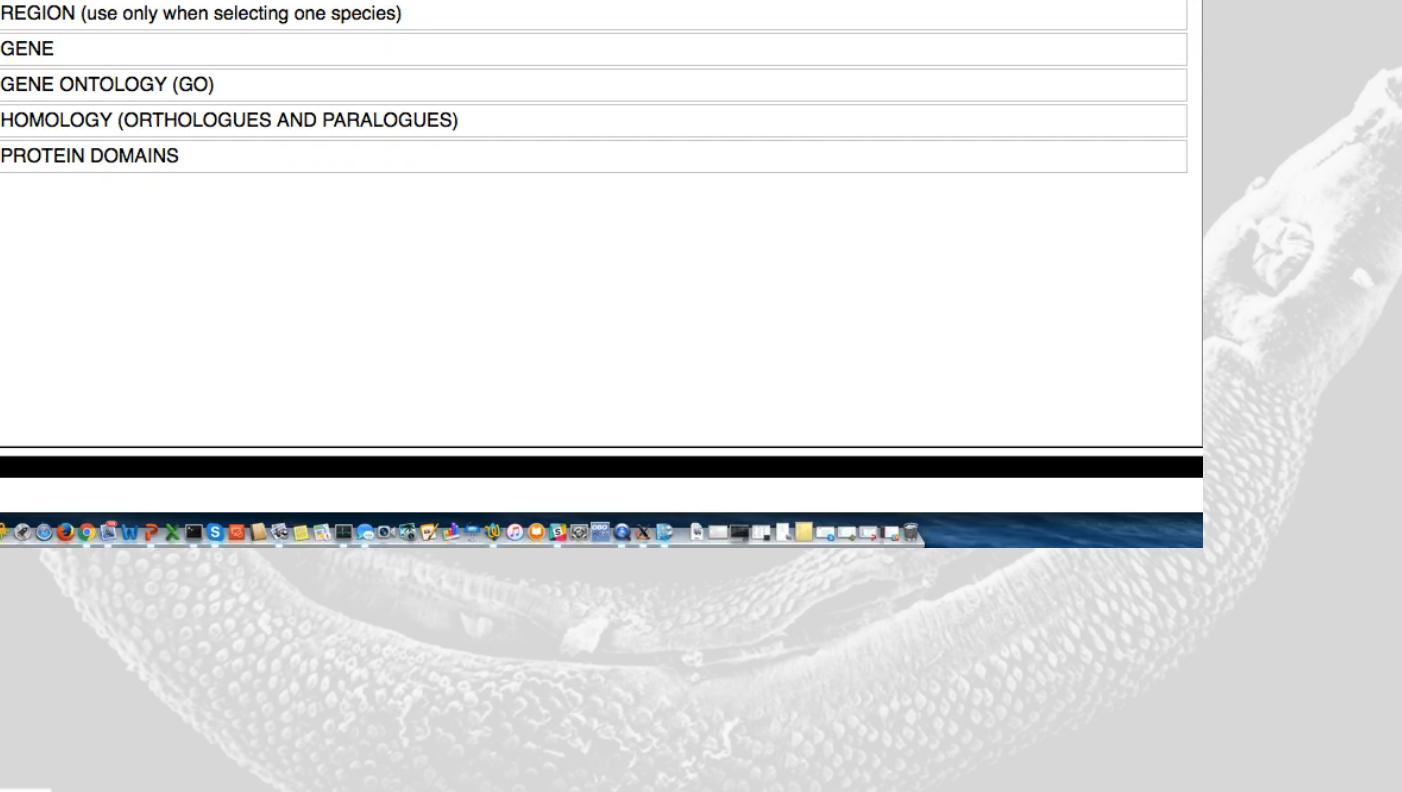
GENE

GENE ONTOLOGY (GO)

HOMOLOGY (ORTHOLOGUES AND PARALOGUES)

PROTEIN DOMAINS

WormBase ParaSite © EBI & WTSI





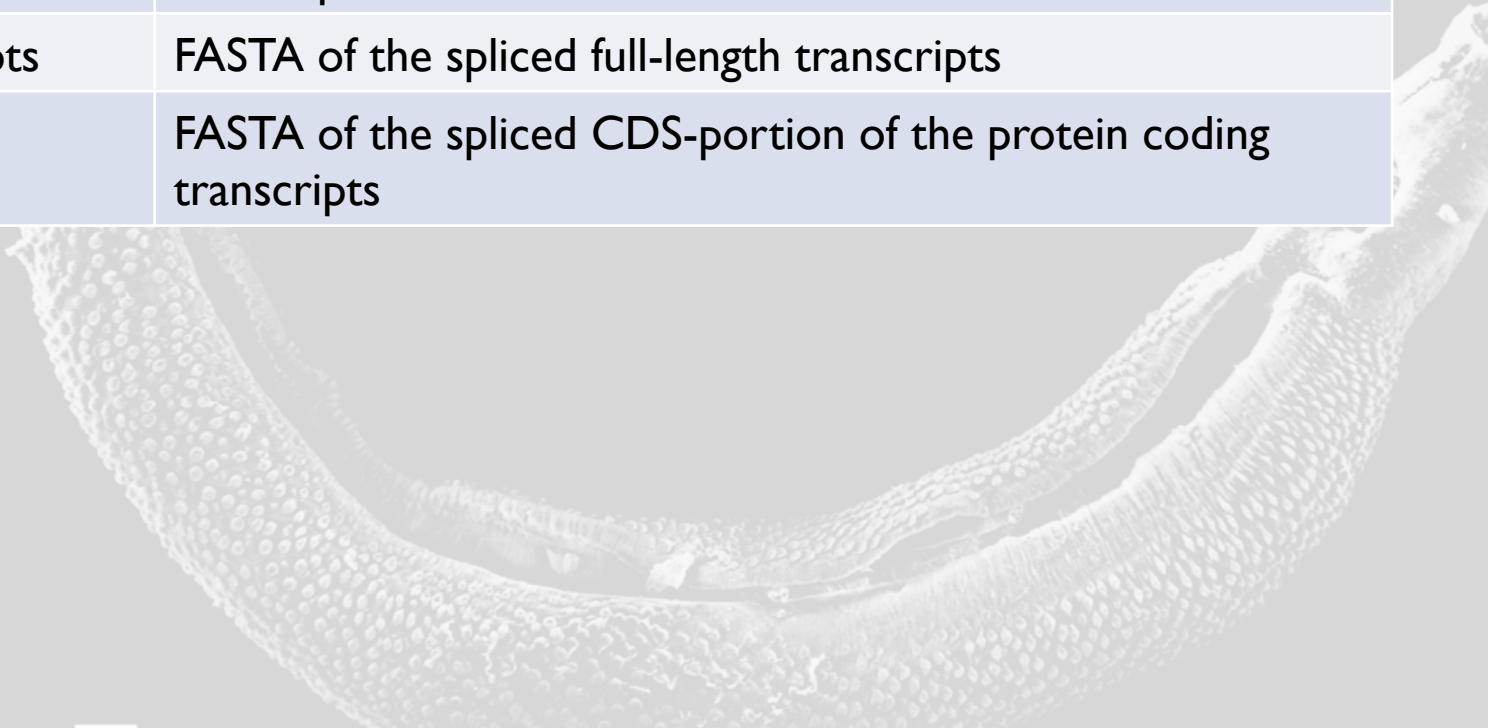
Bulk downloads and
programmatic access

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	<i>String</i>	A stable ID	-	WBGene00225050

Endpoint Examples

Example Requests

[`/rest/genetree/member/symbol/brugia_malayi_prjna10729`](#)

[`/Bma-unc-1?content-type=text/x-phyloxml%2Bxml`](#)

Example output

Perl

Python2

Python3

Ruby

Java

Curl

Wget

```
<?xml version="1.0" encoding="UTF-8"?>
<phyloxml xsi:schemaLocation="http://www.phyloxml.org http://www.phyloxml.org/1.10/phyloxml.xsd"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns="http://www.phyloxml.org">
  <phylogeny rooted="true" type="gene tree">
    <clade branch_length="0">
      <confidence type="duplication_confidence_score">0.7311</confidence>
      <taxonomy>
        <id>33208</id>
        <scientific_name>Metazoa</scientific_name>
      </taxonomy>
      <events>
        <type>speciation_or_duplication</type>
        <duplications>1</duplications>
      </events>
      <clade branch_length="0.003861">
        <confidence type="duplication_confidence_score">0.1584</confidence>
        <taxonomy>
```

Code Examples

Example Requests

[/rest/genetree/member/symbol/brugia_malayi_prjna10729](#)

[/Bma-unc-1?content-type=text/x-phyloxml%2Bxml](#)

[Example output](#)

[Perl](#)

[Python2](#)

[Python3](#)

[Ruby](#)

[Java](#)

[Curl](#)

[Wget](#)

```
1. use strict;
2. use warnings;
3.
4. use HTTP::Tiny;
5.
6. my $http = HTTP::Tiny->new();
7.
8. my $server = 'http://parasite.wormbase.org';
9. my $ext = '/genetree/member/symbol/brugia_malayi_prjna10729/Bma-unc-1?';
10. my $response = $http->get($server.$ext, {
11.     headers => { 'Content-type' => 'text/x-phyloxml+xml' }
12. });
13.
14. die "Failed!\n" unless $response->{success};
15.
16.
17. print "$response->{status} $response->{reason}\n";
18.
```

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

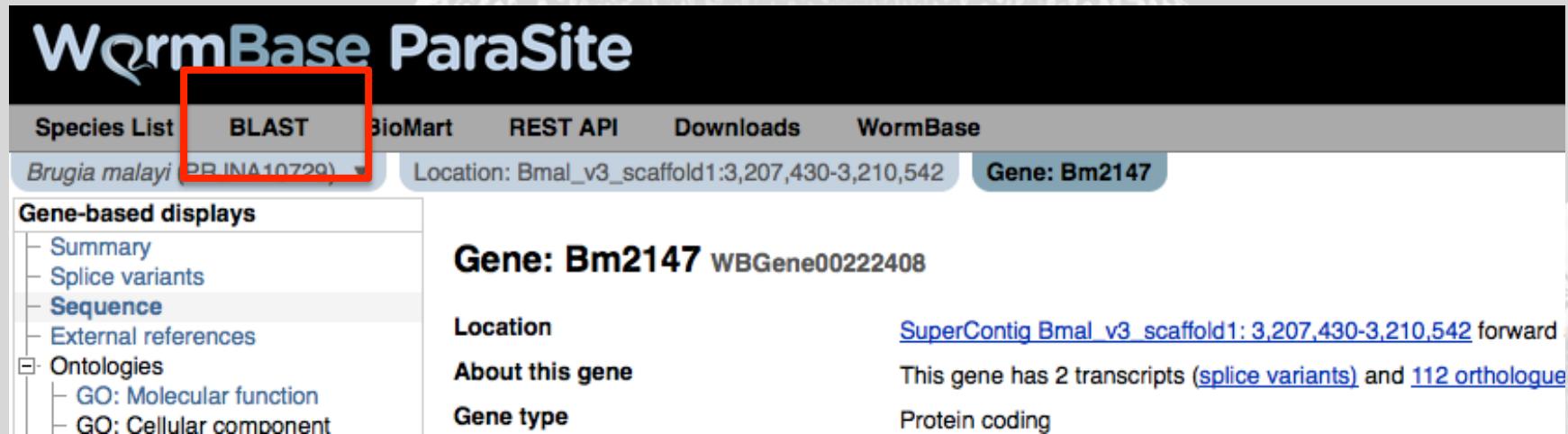
What is BLAST?

- 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



The image shows the WormBase ParaSite interface. At the top, there is a navigation bar with links for Species List, BLAST, BioMart, REST API, Downloads, and WormBase. The BLAST link is highlighted with a red box. Below the navigation bar, the species is set to *Brugia malayi* (PR INA10729). The location of the gene is Bmal_v3_scaffold1:3,207,430-3,210,542. The gene identifier is Gene: Bm2147. On the left, there is a sidebar titled "Gene-based displays" with options for Summary, Splice variants, Sequence (which is selected and highlighted in blue), External references, Ontologies (with sub-options for GO: Molecular function and GO: Cellular component), and REST API. The main content area displays the gene information for Bm2147, including its WBGene identifier, location, and gene type (Protein coding). It also mentions that the gene has 2 transcripts and 112 orthologues.

WormBase ParaSite

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

Brugia malayi (PR INA10729) Location: Bmal_v3_scaffold1:3,207,430-3,210,542 Gene: Bm2147

Gene-based displays

- Summary
- Splice variants
- Sequence**
- External references
- Ontologies
 - GO: Molecular function
 - GO: Cellular component
- REST API

Gene: Bm2147 WBGene00222408

Location SuperContig Bmal_v3_scaffold1:3,207,430-3,210,542 forward

About this gene This gene has 2 transcripts ([splice variants](#)) and [112 orthologues](#)

Gene type Protein coding

Defaults to the species you are currently browsing

Using the ParaSite BLAST

ferences
olecular function
ular component
logical process
e Genomics
ee
ques
ues

Print this page

View data

Print this page

Print this page

Location SuperContig Bmal_v3_scaffold1: 3,207,430-3,210,542 forward strand.

About this gene This gene has 2 transcripts ([splice variants](#)) and [112 orthologues](#).

Gene type Protein coding

Annotation Method Protein-coding model imported from [WormBase](#)

Transcripts [Show transcript table](#)

Marked-up sequence

[Download sequence](#) [BLAST this sequence](#)

Exons Bm2147 exons All exons in this region

```
>supercontig:B_malayi-3.1:Bmal_v3_scaffold1:3206830:3211142:1
ATTTACTTTCTGTATTTCATGATCTTATCAAACCATTTATTTGAACCTCTTTGT
CTACTCTAAGTCTACTCTACAGAGATCAGCCCTTTGTATTAGAACACCAAGCTAAAC
TCGAGAGTTACTGTATTGTTGACATTGTGTCGTTGATGCTTCTACCATGTCAGT
ATTACTAACATGCAAAATTTCTTCAGATTATAATTGGTTATTCAATAAAATCTTT
TTAAATAATTTCAGTGAATTAGACATTAAATATCAGGCAATATCTGTAAGAACATAAGTCA
TACAACATTGCTGTTCCACCGTAAACTGTCAATTTCATCTCGTTGCAATTATATTG
CAATTTCCTATCATATTGCAAGTTAAATAAAATTGATGTTGCTTCATTTCCTTCAAGA
CAGTTTACTGCAACTGTTCAAGTTAAAGCTTGGTTCAATTATGTCACATC
TGGTTGTTGAGTTGCGCAAATTATTAGAACTAGAATTGATCAGCTATTAAAGGAT
TATGTTTATAAGCAAATATAACGACAAAAGAAGTTATTGATTATAATTGTTGGAT
ATGGACGCTGGCAGTTAGTTGCTGGGTACATTACTTAGTAATGTTCTGTTGATT
TAAATGAGTTGACTTCTATTTCGATCCCTTTCTGTTATAAGGATAGTAATCCA
ATGAATGATTCAACAATTATTGTTAATTCCATTAAACTACGTAAGTGTGATA
```

Using the ParaSite BLAST

lecular function
lular component
logical process
ve Genomics
ee
ual
er

the this page

data

ata

is page

work this page

Gene type
Annotation Method
Transcripts

Protein coding

Protein-coding model imported from [WormBase](#) ✓

Show transcript table

Marked-up sequence

 Download sequence

 BLAST this sequence

Exons Bm2147 exons All exons in this region

Using the ParaSite BLAST

WormBase ParaSite

Search WormBase ParaSite... 

e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species List BLAST BioMart REST API Downloads WormBase

My Account - bbolt@ebi.ac.uk Logout Help and Documentation

Brugia malayi (PRJNA10729) ▾ Location: Bmal_v3_scaffold525:2,308-5,498 ▾ Gene: Bma-eat-4 ▾ Transcript: Bm7483 ▾

Gene-based displays

- Summary
- Splice variants
- Sequence**
- External references

Ontologies

- GO: Molecular function
- GO: Cellular component
- GO: Biological process

Comparative Genomics

- Gene tree
- Orthologues
- Paralogues

Configure this page

Manage your data

Export data

Share this page

Bookmark this page

Gene: Bma-eat-4 WBGene00227744

Location SuperContig Bmal_v3_scaffold525: 2,308-5,498 forward strand.

About this gene This gene has 1 transcript (splice variant), 126 orthologues and 7 paralogues.

Gene type Protein coding

Annotation Method Protein-coding model imported from WormBase

Transcripts [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype	UniProt
Bm7483	Bm7483	1459	475aa	Protein coding	AOA0H5SF60

Marked-up sequence

[Download sequence](#) [BLAST this sequence](#)

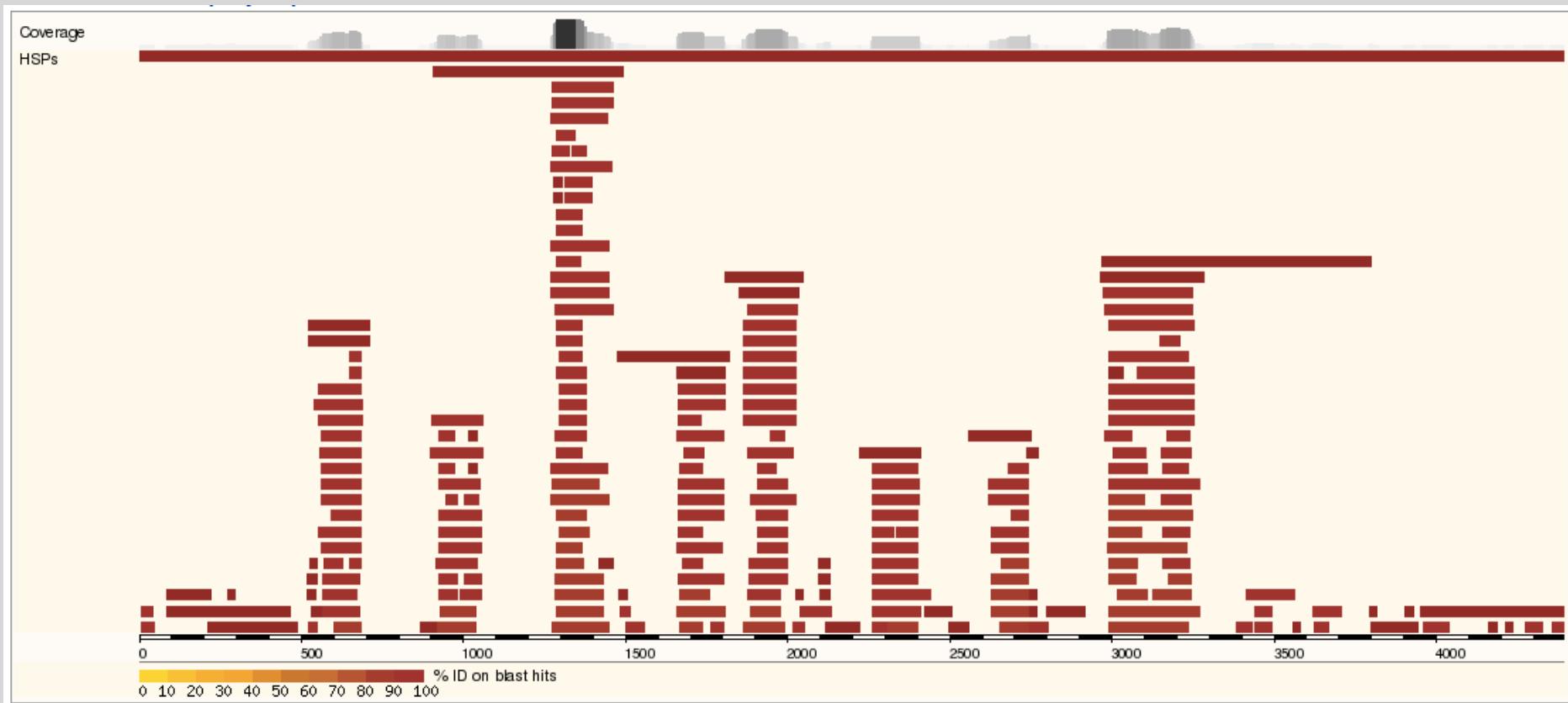
Exons Bma-eat-4 exons All exons in this region

```
>supercontig:B_malayi-3.1:Bmal_v3_scaffold525:1708:6098:1
ATTCACCTCAAAATAACTATATCATAATTGTGTTGCTAAATATAAAAGTAA
ATAATGAAAAACATTAAAAAAAAAAAAAAAAAAATGACAAAATAAAATATAA
TTCCAGATTAACGAGAAAGCAACAAATTATTAACTATAAAAGTTGATTATGT
GAATTAACTTATACTTTAGCTTAAAGGTAAAGAAAGAGGAAATCATTAA
AAAAATATTGCTCAATAATTCTTTATTTAAATTATTTGAATGAATCATAA
AAATTAGAATTGTATTACAATTAAACATTAAATTATTTAATCATATGATAAGAATAAA
TATTATATGGATAATTCCGAAAGTTAAATCTCCGGATAATTACGCTAGAAA
CAGAAACAAATTGCACTGTTTCTTTCTTTTGTGTTTATTTCTGTTTT
GATTATTTATTCATTCAATATGAGAAAAAAATTTTTTTTTAAATA
```

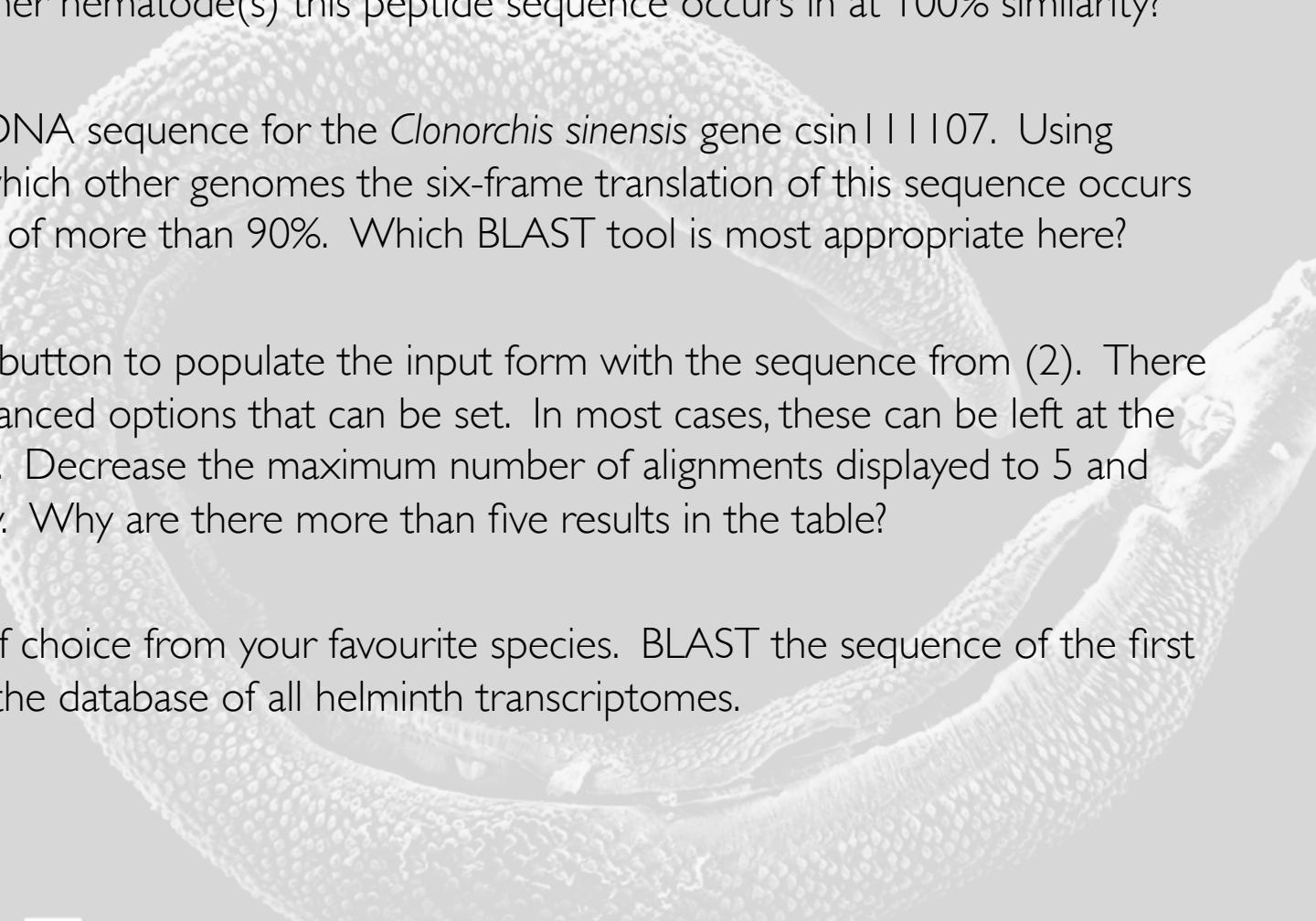
Making sense of the results

- 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

Making sense of the results

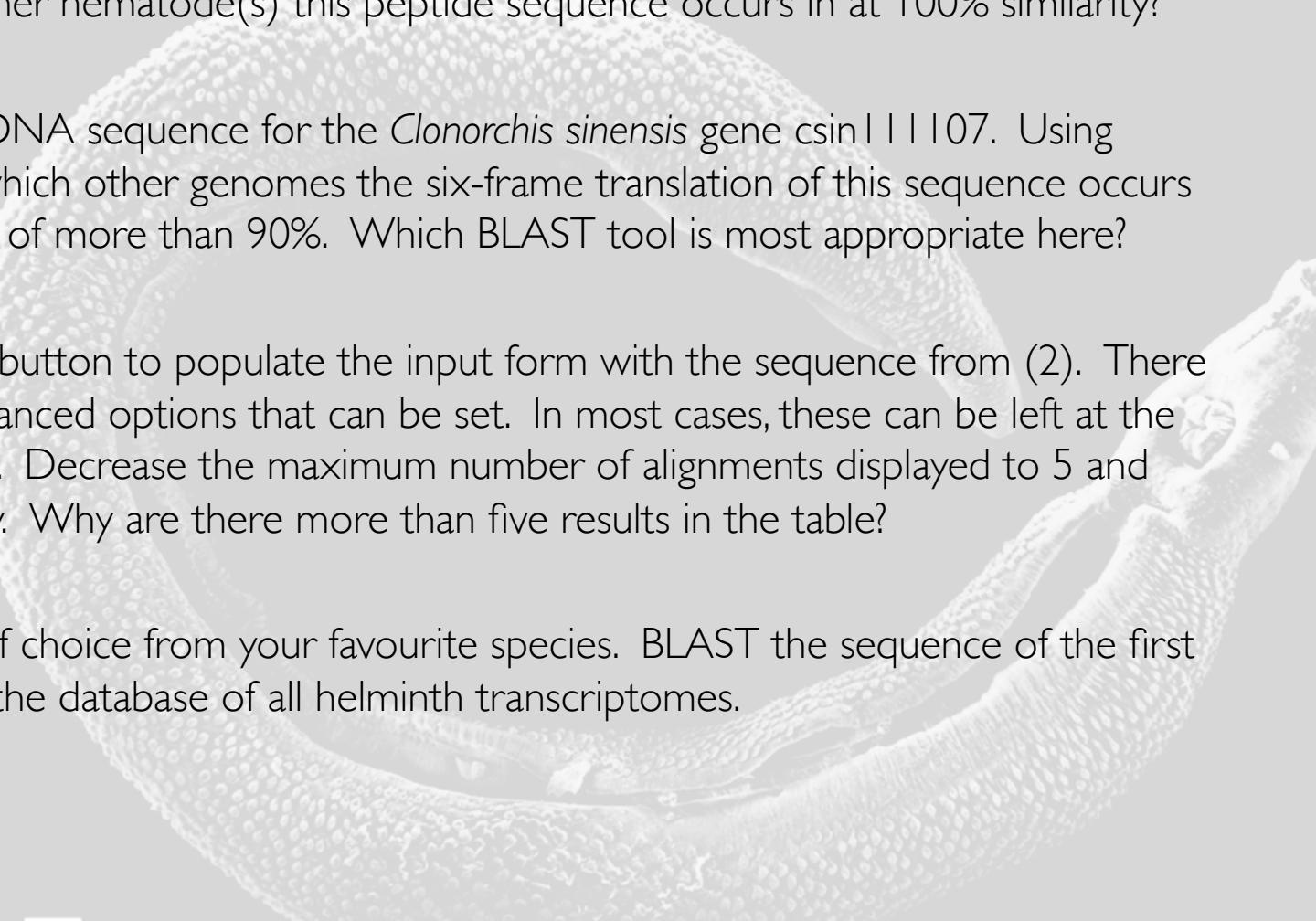


Practical Exercises



1. Locate the peptide sequence for the *Brugia malayi* gene Bma-eat-4. Using BLAST, find which other nematode(s) this peptide sequence occurs in at 100% similarity?
2. Locate the cDNA sequence for the *Clonorchis sinensis* gene csin11107. Using BLAST, find which other genomes the six-frame translation of this sequence occurs in with a %ID of more than 90%. Which BLAST tool is most appropriate here?
3. Use the 'Edit' button to populate the input form with the sequence from (2). There are many advanced options that can be set. In most cases, these can be left at the default values. Decrease the maximum number of alignments displayed to 5 and run the query. Why are there more than five results in the table?
4. Find a gene of choice from your favourite species. BLAST the sequence of the first exon against the database of all helminth transcriptomes.

Practical Exercises



1. Locate the peptide sequence for the *Brugia malayi* gene Bma-eat-4. Using BLAST, find which other nematode(s) this peptide sequence occurs in at 100% similarity?
2. Locate the cDNA sequence for the *Clonorchis sinensis* gene csin11107. Using BLAST, find which other genomes the six-frame translation of this sequence occurs in with a %ID of more than 90%. Which BLAST tool is most appropriate here?
3. Use the 'Edit' button to populate the input form with the sequence from (2). There are many advanced options that can be set. In most cases, these can be left at the default values. Decrease the maximum number of alignments displayed to 5 and run the query. Why are there more than five results in the table?
4. Find a gene of choice from your favourite species. BLAST the sequence of the first exon against the database of all helminth transcriptomes.

I. Locate the peptide sequence for the *Brugia malayi* gene Bma-eat-4. Using BLAST, find which other nematode(s) this peptide sequence occurs in at 100% similarity?

WormBase ParaSite Version: WBPS6 (WS252) Search WormBase ParaSite...

Species List BLAST BioMart REST API VEP Downloads WormBase My Account - bbolt@ebi.ac.uk Logout Help and Documentation

e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species **BLAST** **BioMart**

API **Downloads** **WormBase**

Announcements

Announcing WormBase ParaSite release 6
posted 2 months ago
We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness
posted 1 month ago
We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [...]

***Brugia malayi* assembly update**
posted 2 months ago
The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [...]

Register for a WormBase ParaSite workshop
posted 2 months ago
To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[\[Older\]](#)

Twitter

Tweets by @WBParasite

 WormBase ParaSite @WBParasite
This maintenance is now complete. Thank you for your patience. twitter.com/WBParasite/sta... 3h

 WormBase ParaSite @WBParasite
Important notice: due to essential server maintenance, WormBase ParaSite will be unavailable tomorrow (7th June) between 10:00-11:30 (BST). 23h

 WormBase ParaSite Retweeted 
 **Hayley Bennett** Embed View on Twitter



2. Locate the cDNA sequence for the *Clonorchis sinensis* gene csin111107. Using BLAST, find which other genomes the six-frame translation of this sequence occurs in with a %ID of more than 90%. Which BLAST tool is most appropriate here?

BLASTX or TBLASTX

WormBase ParaSite Version: WBPS6 (WS252)

Search WormBase ParaSite... 

e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

Species List BLAST BioMart REST API VEP Downloads WormBase Login Register Help and Documentation

Species  **BLAST**  **BioMart** 

API  **Downloads**  **WormBase** 

Find a genome

[+] Nematoda (Roundworms) 

[+] Platyhelminthes (Flatworms) 

Statistics

- Version: WBPS6 (April 2016)
- WormBase Version: WS252
- 109 genomes, representing 100 species
- 2,189,841 genes

Announcements

Announcing WormBase ParaSite release 6  posted 2 months ago
We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness  posted 1 month ago
We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [\[...\]](#)

Brugia malayi assembly update  posted 2 months ago
The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [\[...\]](#)

Register for a WormBase ParaSite workshop  posted 2 months ago
To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

[\[Older\]](#)

Twitter

Tweets by @WBParasite

 WormBase ParaSite @WBParasite This maintenance is now complete. Thank you for your patience. [twitter.com/WBParasite/sta...](#)   3h

 WormBase ParaSite @WBParasite Important notice: due to essential server maintenance, WormBase ParaSite will be unavailable tomorrow (7th June) between 10:00-11:30 (BST).   23h

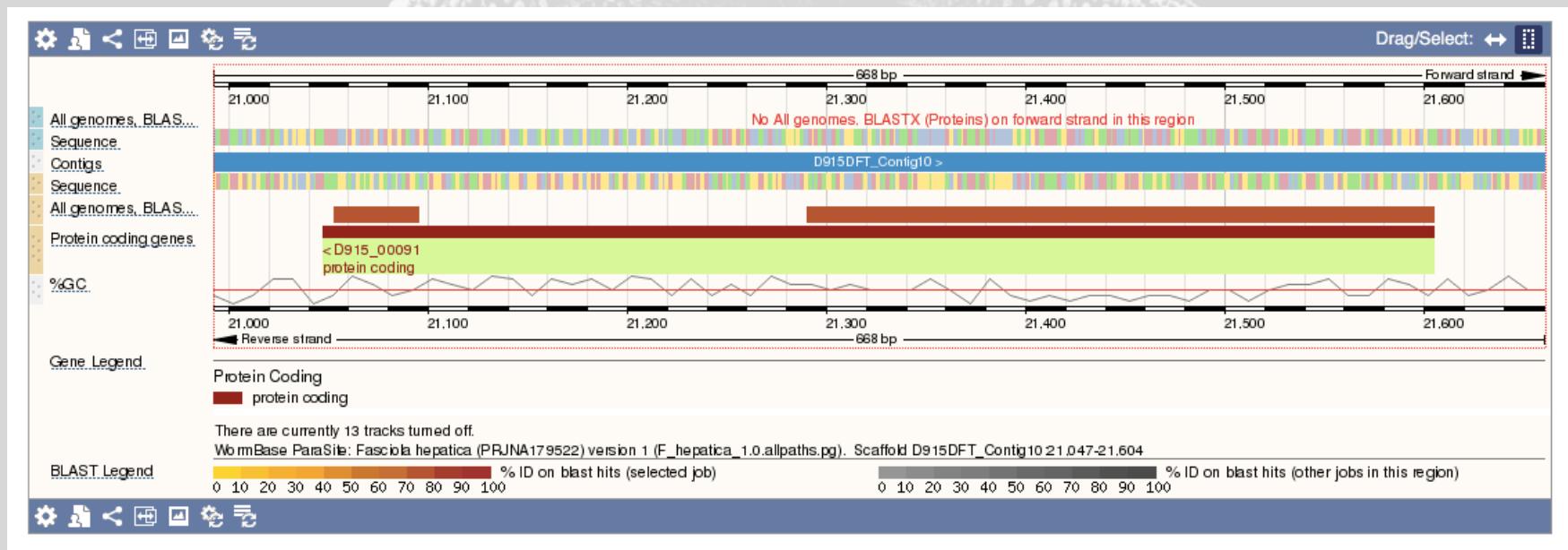
 WormBase ParaSite Retweeted  Hayley Bennett [Embed](#) [View on Twitter](#)

3. Use the 'Edit' button to populate the input form with the sequence from (2). There are many advanced options that can be set. In most cases, these can be left at the default values. Decrease the maximum number of alignments displayed to 5. Why are there more than five results in the table?

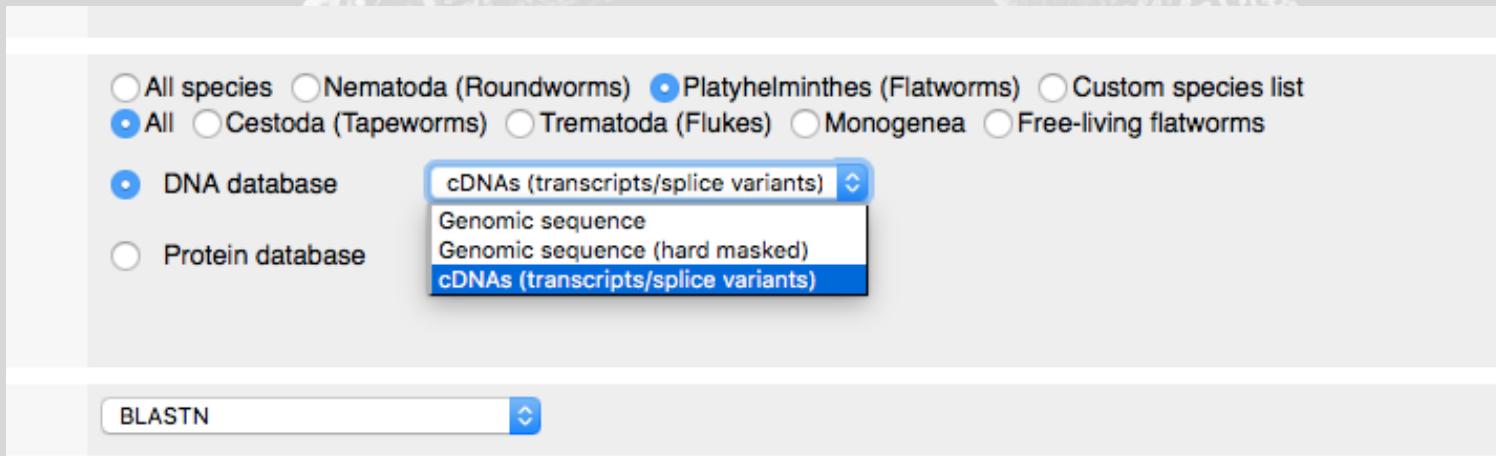
Show/hide columns (8 hidden)										Filter
Genome	Subject name	Subject description	Gene hit	Query start	Length	Score	E-val	%ID		
Clonorchis sinensis (PRJDA72781)	csin111107	DNA-binding protein A	csin111107	427	128	Sequence	677	7.1E-144	100.0	[Alignment]
Clonorchis sinensis (PRJDA72781)	csin111107	DNA-binding protein A	csin111107	1	108	Sequence	478	7.1E-144	100.0	[Alignment]
Opisthorchis viverrini (PRJNA222628)	T265_14198		T265_14198	427	17	Sequence	95	2.4E-57	100.0	[Alignment]
Opisthorchis viverrini (PRJNA222628)	T265_14198		T265_14198	1	105	Sequence	434	2.4E-57	96.2	[Alignment]
Fasciola hepatica (PRJNA179522)	D915_00091		D915_00091	1	105	Sequence	334	9.8E-38	72.4	[Alignment]
Fasciola hepatica (PRJEB6687)	BN1106_s1641B000184.mRNA-1		BN1106_s1641B000184	1	105	Sequence	330	3.6E-37	72.4	[Alignment]
Echinostoma caproni (PRJEB1207)	ECPE_0000786301-mRNA-1		ECPE_0000786301	1	105	Sequence	314	7.9E-35	72.4	[Alignment]
Fasciola hepatica (PRJEB6687)	BN1106_s1641B000184.mRNA-1		BN1106_s1641B000184	433	14	Sequence	51	3.6E-37	71.4	[Alignment]
Fasciola hepatica (PRJNA179522)	D915_00091		D915_00091	433	14	Sequence	51	9.8E-38	71.4	[Alignment]
Echinostoma caproni (PRJEB1207)	ECPE_0000786301-mRNA-1		ECPE_0000786301	427	16	Sequence	50	7.9E-35	56.3	[Alignment]

This parameter specifies the maximum number of objects in the subject database (i.e. genome or proteome). In this case, the query sequence has matched twice to each gene. This suggests there may be two conserved sequences within the query.

3. Use the ‘Edit’ button to populate the input form with the sequence from (2). There are many advanced options that can be set. In most cases, these can be left at the default values. Decrease the maximum number of alignments displayed to 5. Why are there more than five results in the table?



4. Find a gene of choice from your favourite flatworm species. BLAST the sequence of the first exon against the database of all flatworm transcriptomes.



Data Discovery & Export with BioMart



Basics of BioMart

- Advanced search and data export tool
- Produces tables of data or files containing sequence
- Table contents are entirely customisable
- Ideal for working with, or generating, lists of data

Gene stable ID	Chromosome/scaffold name	Gene start (bp)	Gene end (bp)	Exon ID	Exon region start (bp)	Exon region end (bp)	cDNA coding start	cDNA coding end
A01321	Scaffold1052	8861	10795	A01321.e1	8861	9057	1	197
A01321	Scaffold1052	8861	10795	A01321.e2	9212	9353	198	339
A01321	Scaffold1052	8861	10795	A01321.e3	10646	10795	340	489
A01322	Scaffold1052	19019	19861	A01322.e1	19019	19266	1	248
A01322	Scaffold1052	19019	19861	A01322.e2	19663	19861	249	447
A02773	Scaffold1159	9064	10222	A02773.e1	10184	10222	1	39
A02773	Scaffold1159	9064	10222	A02773.e2	9064	9354	40	330
A03307	Scaffold118	185300	190808	A03307.e1	190768	190808	1	41
A03307	Scaffold118	185300	190808	A03307.e2	190413	190656	42	285
A03307	Scaffold118	185300	190808	A03307.e3	188719	188878	286	445
A03307	Scaffold118	185300	190808	A03307.e4	188038	188180	446	588
A03307	Scaffold118	185300	190808	A03307.e5	187378	187413	589	624
A03307	Scaffold118	185300	190808	A03307.e6	185300	185401	625	726
A03308	Scaffold118	191713	191954	A03308.e1	191817	191954	1	138
A03308	Scaffold118	191713	191954	A03308.e2	191713	191775	139	201
A04991	Scaffold127	310799	318815	A04991.e1	318652	318815	1	164
A04991	Scaffold127	310799	318815	A04991.e2	315979	316050	165	236
A04991	Scaffold127	310799	318815	A04991.e3	314309	314432	237	360
A04991	Scaffold127	310799	318815	A04991.e4	313695	313832	361	498
A04991	Scaffold127	310799	318815	A04991.e5	310799	311017	499	717

Three simple steps...

- Three simple steps:
 1. Filter the entire database to include only the genes you are interested in (query can be specific or vague)
 2. Choose the data you would like to include in the output file
 3. View or download the results (including direct export to Excel, or CSV for import to R)
- No programming or database knowledge required!

Data available for export

- Sequences (genomic, cDNA, UTR, flanking, cDNA, peptide)
- Gene IDs, names and descriptions
- Identifiers for data from external databases (e.g. UniProt)
- Gene structure (e.g. exons)
- Protein domains and function (e.g. InterPro, Gene3D, PANTHER etc.)
- Gene ontology terms
- Orthologues and paralogues (in all nematodes, flatworms and a number of non-worm comparators, e.g. human, mouse and rat)

BioMart Interface

Control buttons:

New = reset form

Count = count results

Results = preview results

Query filters = search terms to restrict the query

Output attributes = select data to add into results (i.e. which columns would you like to appear in your table)

WormBase ParaSite

WormBase Home | ParaSite Home

New **Count** **Results**

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters
[None selected]

2. Output Attributes
Genome project
Gene stable ID

Please restrict your query using criteria
(If filter values are truncated in any lists, hover over the list for guidance, see the [relevant pages](#) of our documentation)

SPECIES
 REGION (use only when selecting one species)
 GENE
 GENE ONTOLOGY (GO)
 HOMOLOGY (ORTHOLOGUES AND PARALOGUES)
 PROTEIN DOMAINS

Filters and attributes appear here



BioMart Example I – Working with a list

I have a list of *Schistosoma mansoni* genes and would like to find:

1. The gene name and gene description
2. How many of these genes have a human orthologue?
3. The functional annotation of the genes with a human orthologue?

A
1 Smp_158080
2 Smp_078570
3 Smp_063300
4 Smp_204760
5 Smp_145060
6 Smp_135070
7 Smp_210640
8 Smp_160900
9 Smp_049930
10 Smp_126600
11 Smp_132740
12 Smp_139350
13 Smp_055760
14 Smp_141410
15 Smp_051410
16 Smp_175210
17 Smp_169250
18 Smp_129000
19 Smp_128010
20 Smp_079640
21 Smp_038870
22 Smp_213140
23 Smp_161540
24 Smp_151280
25 Smp_012010
26 Smp_181360
27 Smp_010260
28 Smp_199660

BioMart Example I – Working with a list



WormBase ParaSite

Species List BLAST BioMart REST API Downloads

Schistosoma mansoni (PRJEA36577) Location: Smp.Chr_1:42,206,486-42,214,230 Gene: GLUCL Transcript: Smp_104890.1

Gene-based displays

- Summary
- Splice variants
- Sequence
- External references
- Ontologies

 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process

- Comparative Genomics

 - Gene tree
 - Orthologues
 - Paralogues

Configure this page

Manage your data

Export data

Share this page

Bookmark this page

Gene: GLUCL Smp_104890

Description Location About this gene

Putative cys-loop ligand gated ion channel subunit (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V7T2] Scaffold Smp.Chr_1:42,206,486-42,214,230 reverse strand.

This gene has 1 transcript (splice variant), 375 orthologues and 3 paralogues.

WormBase ParaSite

Species List BLAST BioMart REST API Downloads

Schistosoma mansoni (PRJEA36577) Location: Smp.Chr_1:42,206,486-42,214,230 Gene: GLUCL Transcript: Smp_104890.1

Gene-based displays

- Summary
- Splice variants
- Sequence
- External references
- Ontologies

 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process

- Comparative Genomics

 - Gene tree
 - Orthologues
 - Paralogues

Configure this page

Manage your data

Export data

Share this page

Bookmark this page

Gene: GLUCL Smp_104890

Description Location About this gene

Putative cys-loop ligand gated ion channel subunit (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V7T2] Scaffold Smp.Chr_1:42,206,486-42,214,230 reverse strand.

This gene has 1 transcript (splice variant), 375 orthologues and 3 paralogues.

WormBase ParaSite

Species List BLAST BioMart REST API Downloads

Schistosoma mansoni (PRJEA36577) Location: Smp.Chr_1:42,206,486-42,214,230 Gene: GLUCL Transcript: Smp_104890.1

Gene-based displays

- Summary
- Splice variants
- Sequence
- External references
- Ontologies

 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process

- Comparative Genomics

 - Gene tree
 - Orthologues
 - Paralogues

Configure this page

Manage your data

Export data

Share this page

Bookmark this page

Gene: GLUCL Smp_104890

Description Location About this gene

Putative cys-loop ligand gated ion channel subunit (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V7T2] Scaffold Smp.Chr_1:42,206,486-42,214,230 reverse strand.

This gene has 1 transcript (splice variant), 375 orthologues and 3 paralogues.

Protein coding

Gene models from Wellcome Trust Sanger Institute Reference Helminth Genomes project

Hide transcript table

Table: Gene-based displays (2 hidden)

Filter

Transcript ID bp Protein Biotype UniProt

Level Smp_104890.1 1515 504aa Protein coding G4V7T2@T2CSW6@

GO: Molecular function

Accession	Term	Evidence	Annotation Source
GO:0005230@	extracellular ligand-gated ion channel activity	IEA	UniProtKB/Trembl

Transcript IDs Smp_104890.1

Information

Help and Documentation

Citing WormBase ParaSite

About WormBase ParaSite

Data Usage

Contact Us

Disclaimer

Save time by using BioMart!

I have a list of *Schistosoma mansoni* genes and would like to find:

I. The gene name and gene description

1	Smp_158080
2	Smp_078570
3	Smp_063300
4	Smp_204760
5	Smp_145060
6	Smp_135070
7	Smp_210640
8	Smp_160900
9	Smp_049930
10	Smp_126600
11	Smp_132740
12	Smp_139350
13	Smp_055760
14	Smp_141410
15	Smp_051410
16	Smp_175210
17	Smp_169250
18	Smp_129000
19	Smp_128010
20	Smp_079640
21	Smp_038870
22	Smp_213140
23	Smp_161540
24	Smp_151280
25	Smp_012010
26	Smp_181360
27	Smp_010260
28	Smp_199660
29	Smp_164360
30	Smp_058780
31	Smp_045420
32	Smp_136610
33	Smp_172500
34	Smp_170950

I have a list of *Schistosoma mansoni* genes and would like to find:

2. How many of these genes have a human orthologue?

WormBase ParaSite

Search WormBase ParaSite...

WormBase Home | ParaSite Home

BLAST | BioMart | Downloads

 New  Count  Results

 URL  XML  Perl  Help

100 / 2550346 Genes

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters

Gene stable ID(s): [ID-list specified]

2. Output Attributes

Genome project

Gene stable ID

Gene name

Gene description

Export all results to

File

XLS

Unique results only

Go

Email notification to

View

50

rows as

HTML

Unique results only

Genome project	Gene stable ID	Gene name	Gene description
<i>schistosoma mansoni</i> prija36577	Smp_002160		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4VLL1]
<i>schistosoma mansoni</i> prija36577	Smp_007180		Sr-related ctd associated factor (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VB15]
<i>schistosoma mansoni</i> prija36577	Smp_010260		
<i>schistosoma mansoni</i> prija36577	Smp_012010	FOXO	Forkhead transcription factor (inferred by orthology to a protein) [Source:UniProtKB;Acc:A6XDL3]
<i>schistosoma mansoni</i> prija36577	Smp_013610		
<i>schistosoma mansoni</i> prija36577	Smp_018150		Putative mannosey-oligosaccharide alpha-1,2-mannosidase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VP10]
<i>schistosoma mansoni</i> prija36577	Smp_029520		Villin, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LW89]
<i>schistosoma mansoni</i> prija36577	Smp_036550		
<i>schistosoma mansoni</i> prija36577	Smp_038870		NADH-ubiquinone oxidoreductase, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LUK7]
<i>schistosoma mansoni</i> prija36577	Smp_044010		Tropomyosin-1 [Source:UniProtKB/Swiss-Prot;Acc:P42637]
<i>schistosoma mansoni</i> prija36577	Smp_045420		Adiponectin receptor, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LXQ3]
<i>schistosoma mansoni</i> prija36577	Smp_046890		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4VCL8]
<i>schistosoma mansoni</i> prija36577	Smp_048430	TGR	Thioredoxin glutathione reductase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V8J4]
<i>schistosoma mansoni</i> prija36577	Smp_049930		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4LWN1]
<i>schistosoma mansoni</i> prija36577	Smp_051410		Septate junction protein (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LY05]
<i>schistosoma mansoni</i> prija36577	Smp_055760		ATP-dependent zinc metalloprotease YME1 homolog [Source:UniProtKB/Swiss-Prot;Acc:P54813] (projected from <i>Caenorhabditis elegans</i> ortholog ymel-1)
<i>schistosoma mansoni</i> prija36577	Smp_058780		
<i>schistosoma mansoni</i> prija36577	Smp_060480		Putative copine (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V6J0]
<i>schistosoma mansoni</i> prija36577	Smp_063300		60S ribosomal protein L36-like protein; Putative 60s ribosomal protein L36e [Source:UniProtKB/TrEMBL;Acc:Q15ET2]
<i>schistosoma mansoni</i> prija36577	Smp_069170		Putative cation efflux protein/ zinc transporter (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V1P1]
<i>schistosoma mansoni</i> prija36577	Smp_074990		Regulator of chromosome condensation-related (inferred by orthology to a protein) [Source:UniProtKB;Acc:C4QHH6]
<i>schistosoma mansoni</i> prija36577	Smp_075110		Putative rad1 DNA damage checkpoint protein (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VH02]
<i>schistosoma mansoni</i> prija36577	Smp_077790		Putative histone H2B (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VL73]
<i>schistosoma mansoni</i> prija36577	Smp_078570		Putative zinc finger protein (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VM27]
<i>schistosoma mansoni</i> prija36577	Smp_079640		Putative monocarboxylate transporter (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V7W3]
<i>schistosoma mansoni</i> prija36577	Smp_085080		
<i>schistosoma mansoni</i> prija36577	Smp_093760		Strain Puerto Rico chromosome W, complete genome [Source:UniProtKB/TrEMBL;Acc:G4VTD3]
<i>schistosoma mansoni</i> prija36577	Smp_120620		Amine GPCR, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LW20]

I have a list of *Schistosoma mansoni* genes and would like to find:

3. The functional annotation of the genes with a human orthologue?

WormBase ParaSite

Search WormBase ParaSite...

BLAST | BioMart | Downloads

 URL  XML  Perl  Help

 New  Count  Results

47 / 2550346 Genes

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters

Gene stable ID(s): [ID-list specified]

Orthologue(s) in Human: With

2. Output Attributes

Genome project

Gene stable ID

Gene name

Gene description

Human gene stable ID

Export all results to  XLS Unique results only  Go

Email notification to

View  rows as  Unique results only

Genome project	Gene stable ID	Gene name	Gene description	Human gene stable ID
schistosoma_mansoni_priea36577	Smp_002160		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4VLL1]	ENSG00000178028
schistosoma_mansoni_priea36577	Smp_007180		Sr-related ctd associated factor (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VB15]	ENSG00000085872
schistosoma_mansoni_priea36577	Smp_010260			ENSG00000136936
schistosoma_mansoni_priea36577	Smp_013610			ENSG00000170264
schistosoma_mansoni_priea36577	Smp_013610			ENSG00000156050
schistosoma_mansoni_priea36577	Smp_018150		Putative mannose-oligosaccharide alpha-1,2-mannosidase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VP10]	ENSG00000111885
schistosoma_mansoni_priea36577	Smp_018150		Putative mannose-oligosaccharide alpha-1,2-mannosidase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VP10]	ENSG00000198162
schistosoma_mansoni_priea36577	Smp_018150		Putative mannose-oligosaccharide alpha-1,2-mannosidase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VP10]	ENSG00000117643
schistosoma_mansoni_priea36577	Smp_038870		NADH-ubiquinone oxidoreductase, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LUK7]	ENSG00000023228
schistosoma_mansoni_priea36577	Smp_044010		Tropomyosin-1 [Source:UniProtKB/Swiss-Prot;Acc:P42637]	ENSG00000167460
schistosoma_mansoni_priea36577	Smp_044010		Tropomyosin-1 [Source:UniProtKB/Swiss-Prot;Acc:P42637]	ENSG00000140416
schistosoma_mansoni_priea36577	Smp_044010		Tropomyosin-1 [Source:UniProtKB/Swiss-Prot;Acc:P42637]	ENSG00000198467
schistosoma_mansoni_priea36577	Smp_044010		Tropomyosin-1 [Source:UniProtKB/Swiss-Prot;Acc:P42637]	ENSG00000143549
schistosoma_mansoni_priea36577	Smp_045420		Adiponectin receptor, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LXQ3]	ENSG0000006831
schistosoma_mansoni_priea36577	Smp_045420		Adiponectin receptor, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LXQ3]	ENSG00000159346
schistosoma_mansoni_priea36577	Smp_046890		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4VCL8]	ENSG00000196850
schistosoma_mansoni_priea36577	Smp_048430	TGR	Thioredoxin glutathione reductase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V8J4]	ENSG00000197763
schistosoma_mansoni_priea36577	Smp_048430	TGR	Thioredoxin glutathione reductase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V8J4]	ENSG00000198431
schistosoma_mansoni_priea36577	Smp_049930		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4LWN1]	ENSG00000148343
schistosoma_mansoni_priea36577	Smp_049930		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4LWN1]	ENSG00000180488
schistosoma_mansoni_priea36577	Smp_058780			ENSG00000105617
schistosoma_mansoni_priea36577	Smp_060480		Putative copine (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V6J0]	ENSG00000139117
schistosoma_mansoni_priea36577	Smp_060480		Putative copine (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V6J0]	ENSG00000144550
schistosoma_mansoni_priea36577	Smp_060480		Putative copine (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V6J0]	ENSG00000124772
schistosoma_mansoni_priea36577	Smp_063300		60S ribosomal protein L36-like protein; Putative 60s ribosomal protein L36e [Source:UniProtKB/TrEMBL;Acc:Q15FT2]	ENSG00000130255

BioMart Example 2 – Generating a list

I think a novel drug targets transmembrane signalling receptor activity in *Brugia* species. I would like to model this in *C. elegans*, so would like a list of orthologues.

Therefore I want to generate a list of *Brugia* genes, which:

- Do not have an orthologue in humans
- Are associated with transmembrane signalling receptor activity
- Annotated with the *C. elegans* orthologue, if available

I want to generate a list of *Brugia* genes, which do not have an orthologue in humans, are associated with transmembrane signalling receptor activity and annotated with the *C. elegans* orthologue, if available

WormBase ParaSite

Search WormBase ParaSite...

BLAST | BioMart | Downloads

 URL  XML  Perl  Help

WormBase Home | ParaSite Home

 New  Count  Results

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters

[None selected]

2. Output Attributes

Genome project
Gene stable ID

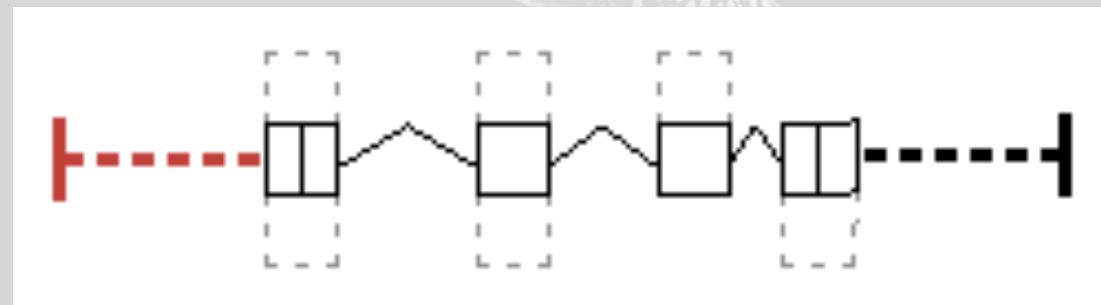
Please restrict your query using criteria below

(If filter values are truncated in any lists, hover over the list item to see the full text)
For guidance, see the [relevant pages](#) of our documentation.

- SPECIES** 
- REGION (use only when selecting one species)
- GENE
- GENE ONTOLOGY (GO)
- HOMOLOGY (ORTHOLOGUES AND PARALOGUES)
- PROTEIN DOMAINS

BioMart Example 3 – Retrieving sequences

I would like to retrieve the sequence located 500bp upstream of each *Brugia malayi* gene



I would like to retrieve the sequence located 500bp upstream of each *Brugia malayi* gene

WormBase ParaSite

Search WormBase ParaSite...

BLAST | BioMart | Downloads

 URL  XML  Perl  Help

WormBase Home | ParaSite Home

 New  Count  Results

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters

[None selected]

2. Output Attributes

Genome project

Gene stable ID

Please restrict your query using criteria below

(If filter values are truncated in any lists, hover over the list item to see the full text)
For guidance, see the [relevant pages](#) of our documentation.

SPECIES

REGION (use only when selecting one species)

GENE

GENE ONTOLOGY (GO)

HOMOLOGY (ORTHOLOGUES AND PARALOGUES)

PROTEIN DOMAINS

BioMart Summary

- Use when working with lists of data, for generating lists of genes and for retrieving sequence
- All data from the website is available
- No programming knowledge required
- Export directly to Excel or CSV for import to R
- Contact us if you are unsure how to construct a query (contact link at bottom of website)

Coming soon... (release 7 in August)

HOMOLOGY (ORTHOLOGUES AND PARALOGUES)

Restrict results to genes with orthologues in...

Caenorhabditis elegans [WS252]
Drosophila melanogaster
Human
Mouse
Saccharomyces cerevisiae
Zebrafish
Acanthocheilonema viteae (PRJEB4306)
Amphimedon queenslandica
Ancylostoma caninum (PRJNA72585)
Ancylostoma ceylanicum (PRJNA231479)
Ancylostoma ceylanicum (PRJNA72583)
Ancylostoma duodenale (PRJNA72581)
Angiostrongylus cantonensis (PRJEB493)
Angiostrongylus costaricensis (PRJEB494)
Anisakis simplex (PRJEB496)
Ascaris lumbricoides (PRJEB4950)
Ascaris suum (PRJNA62057)
Ascaris suum (PRJNA80881)
Brugia malayi (PRJNA10729) [WS252]
Brugia pahangi (PRJEB497)

Restrict results to genes without orthologues in...

Caenorhabditis elegans [WS252]
Drosophila melanogaster
Human
Mouse
Saccharomyces cerevisiae
Zebrafish
Acanthocheilonema viteae (PRJEB4306)
Amphimedon queenslandica
Ancylostoma caninum (PRJNA72585)
Ancylostoma ceylanicum (PRJNA231479)
Ancylostoma ceylanicum (PRJNA72583)
Ancylostoma duodenale (PRJNA72581)
Angiostrongylus cantonensis (PRJEB493)
Angiostrongylus costaricensis (PRJEB494)
Anisakis simplex (PRJEB496)
Ascaris lumbricoides (PRJEB4950)
Ascaris suum (PRJNA62057)
Ascaris suum (PRJNA80881)
Brugia malayi (PRJNA10729) [WS252]
Brugia pahangi (PRJEB497)

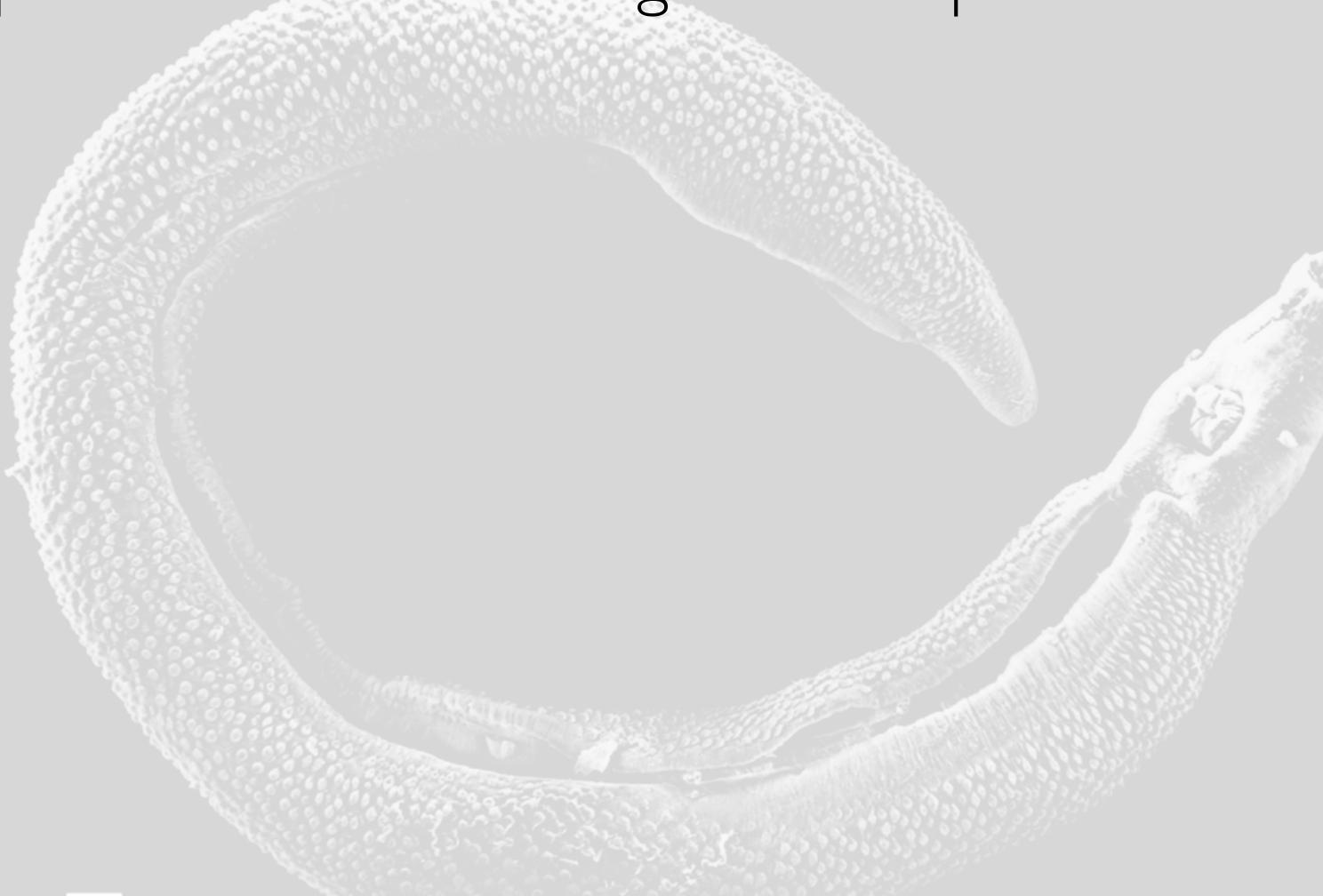
Restrict results to genes with/without...

With
 Without

PROTEIN DOMAINS

Exercises – Part I

- <http://parasite.wormbase.org/workshop>



Exercise |

WormBase ParaSite

WormBase Home | ParaSite Home

 New  Count  Results

Search WormBase ParaSite...

BLAST | BioMart | Downloads

 URL  XML  Perl  Help

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters

[None selected]

2. Output Attributes

Genome project

Gene stable ID

Please restrict your query using criteria below

(If filter values are truncated in any lists, hover over the list item to see the full text)
For guidance, see the [relevant pages](#) of our documentation.

SPECIES

REGION (use only when selecting one species)

GENE

GENE ONTOLOGY (GO)

HOMOLOGY (ORTHOLOGUES AND PARALOGUES)

PROTEIN DOMAINS

Exercise 2

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters

[None selected]

2. Output Attributes

Genome project
Gene stable ID

Please restrict your query using criteria below

(If filter values are truncated in any lists, hover over the list item to see the full text)
For guidance, see the [relevant pages](#) of our documentation.

SPECIES

REGION (use only when selecting one species)

GENE

GENE ONTOLOGY (GO)

HOMOLOGY (ORTHOLOGUES AND PARALOGUES)

PROTEIN DOMAINS

Exercise 3

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters

[None selected]

2. Output Attributes

Genome project
Gene stable ID

Please restrict your query using criteria below

(If filter values are truncated in any lists, hover over the list item to see the full text)
For guidance, see the [relevant pages](#) of our documentation.

SPECIES

REGION (use only when selecting one species)

GENE

GENE ONTOLOGY (GO)

HOMOLOGY (ORTHOLOGUES AND PARALOGUES)

PROTEIN DOMAINS

Exercise 4

WormBase ParaSite

WormBase Home | ParaSite Home

New **Count** **Results**

Search WormBase ParaSite...

BLAST | BioMart | Downloads

★ URL | ▶ XML | ⚡ Perl | ⓧ Help

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

1. Query Filters

[None selected]

2. Output Attributes

Genome project
Gene stable ID

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)
For guidance, see the [relevant pages](#) of our documentation.

- SPECIES
- REGION (use only when selecting one species)
- GENE
- GENE ONTOLOGY (GO)
- HOMOLOGY (ORTHOLOGUES AND PARALOGUES)
- PROTEIN DOMAINS

Exercise 5

WormBase ParaSite

WormBase Home | ParaSite Home

Search WormBase ParaSite...

BLAST | BioMart | Downloads

New **Count** **Results** **URL** **XML** **Perl** **Help**

BioMart allows the fast export of data tables and sequence in just a few steps:

1. Select your 'Query Filters' (these are search parameters that define your results)
2. Choose your 'Output Attributes' (what you would like to include in your results table or file)
3. Hit the 'Results' button at the top of this page

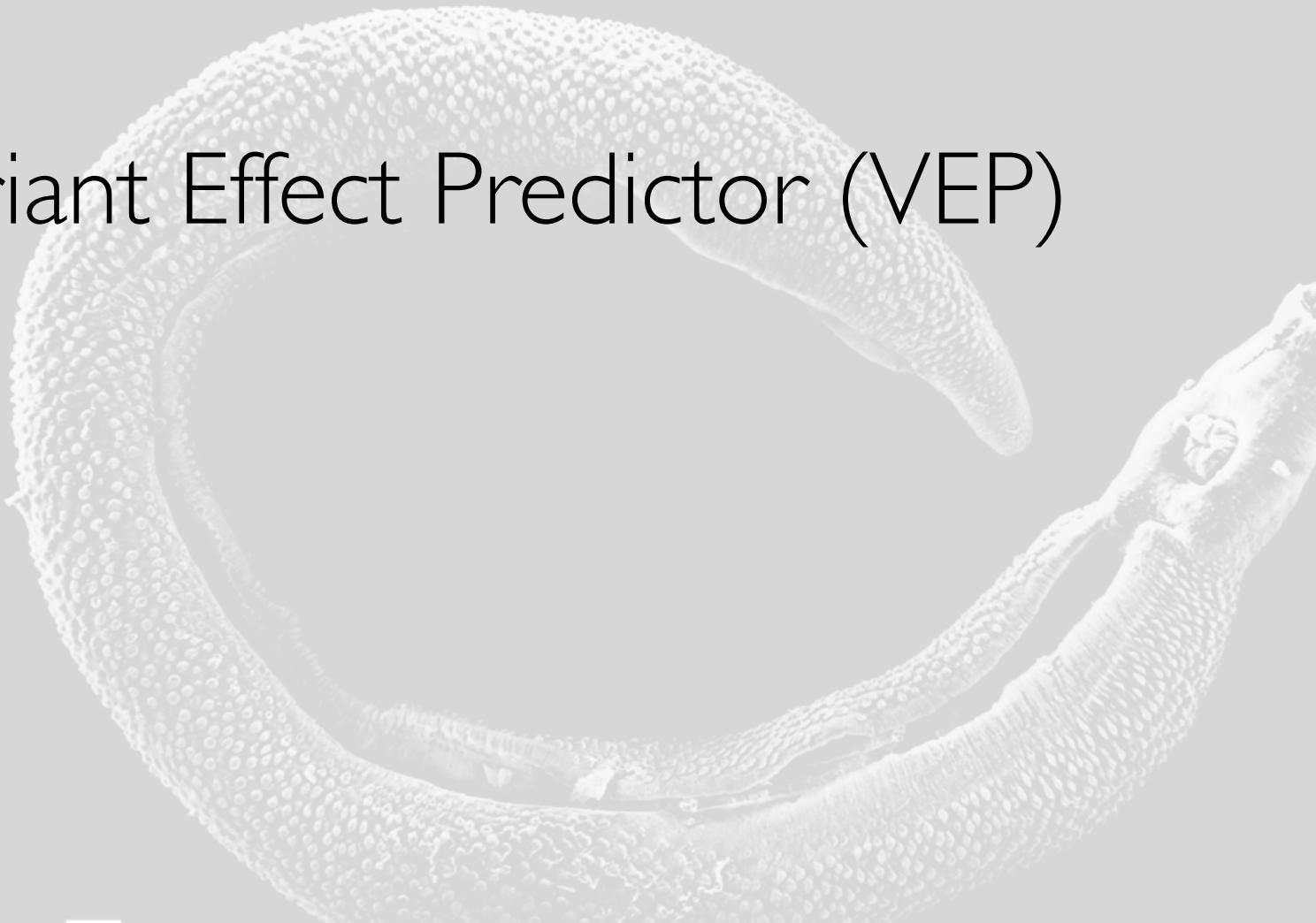
1. Query Filters
[None selected]

2. Output Attributes
Genome project
Gene stable ID

Please restrict your query using criteria below
(If filter values are truncated in any lists, hover over the list item to see the full text)
For guidance, see the [relevant pages](#) of our documentation.

- SPECIES
- REGION (use only when selecting one species)
- GENE
- GENE ONTOLOGY (GO)
- HOMOLOGY (ORTHOLOGUES AND PARALOGUES)
- PROTEIN DOMAINS

Variant Effect Predictor (VEP)



Introduction to the VEP

- Determine the effect of your variants:
 - SNPs
 - Insertions
 - Deletions
 - CNVs
 - Structural variants
- Input: coordinates of the variant and nucleotide change (or VCF file)

Introduction to the VEP

- Output:
 - Genes and transcripts affected by the variants
 - Location of the variants (e.g. upstream of a transcript, in coding sequence, in ncRNA, etc)
 - Consequence of the variant on protein sequence (e.g. stop gained/lost, missense, frameshift, synonymous, non-synonymous, etc)
- View results in table and on genome browser

Introduction to VEP

- Runs online using our servers (ideal for small number of variants)
- Possible to download and run locally for large amounts of data
 - Pre-computed “VEP caches” available for use offline

Using the VEP

WormBase ParaSite Version: WBPS6 (WS252) Search WormBase ParaSite...

Species List BLAST BioMart REST API VEP Downloads WormBase Login Register Help and Documentation

Species  **BLAST**  **BioMart** 

API  **Downloads**  **WormBase** 

Announcements

Announcing WormBase ParaSite release 6 posted 1 month ago
We are pleased to announce the sixth release of WormBase ParaSite.

Blog

Featured Paper: Glucose and Glycogen Metabolism in *B. malayi* Is Associated with Wolbachia Symbiont Fitness posted 1 month ago
We would like to draw your attention to a paper published by Denis Voronin, et. al on the influence of host metabolism on symbiont fitness: Denis Voronin, Saheed Bachu, Michael Shlossman, Thomas R. Unnasch, Elodie Ghedin, Sara Lustigman "Glucose and Glycogen Metabolism in *Brugia malayi* Is Associated with Wolbachia Symbiont Fitness", PLoS One. 2016 Apr [\[...\]](#)

Brugia malayi assembly update posted 2 months ago
The new release of WormBase (WS252) is the first one to feature the new and updated version 4 of the *Brugia malayi* assembly. Due to additional optical mapping, as well as new long-range PacBio sequencing and additional reassembly work conducted to integrate all available data, it was possible to scaffold the 88.2Mbp assembly into 5 chromosomes [\[...\]](#)

Register for a WormBase ParaSite workshop posted 2 months ago
To help our users make the most of this valuable resource, we are visiting universities and institutes to provide hands-on training sessions.

Featured Paper: Allergy the Price of Immunity posted 3 months ago
We would like to draw your attention to a paper recently published in PLOS Computational Biology: Comparisons of Allergenic and Metazoan Parasite Proteins: Allergy the Price of Immunity by Nidhi Tyagi. It is thought that part of our immune system has evolved to combat and provide immunity against infection by parasitic worms. However, in the absence [\[...\]](#)

Twitter

Tweets by @WBParasite

 WormBase ParaSite Retweeted 

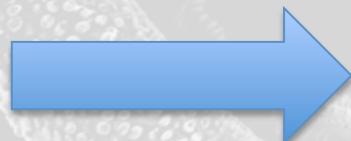
 **Hayley Bennett** @HayleyMBennett Wow @ProtocolIO publishing neat&reproducible research methods! Great parasitology example-how to culture tapeworms. [twitter.com/GigaScience/st...](#)   17h

 WormBase ParaSite Retweeted 

 **NaturalHistoryMuseum** @NHM_London You can learn more about Schistosomes and their fellow parasites at #ParasitesDay, Fri 1 Jun [nhm.ac.uk/visit/exhibiti...](#)  

Other variation data

- Starting in WBPS7 (due August 2016), we will display data from the European Variation Archive (EVA)
- Simply deposit your own variation data in EVA and it will display automatically in the next release of WormBase ParaSite



WormBase ParaSite

Gene Variation Table

WormBase ParaSite Version: WBPS31 (WS254)

Search WormBase ParaSite... 

e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

HMMER Species List BLAST BioMart REST API VEP Downloads WormBase

Login Register Help and Documentation

Sorghum bicolor Location: 1:257,765-268,116 Gene: Sb01g000355

Gene-based displays

- Summary
- Splice variants
- Sequence
- External references
- Ontologies
 - GO: Molecular function
 - GO: Cellular component
 - GO: Biological process
- Literature
- Comparative Genomics
 - Gene tree
 - Orthologues
 - Paralogues
- Variation
 - Variation Table**
 - Variation Image

Gene: Sb01g000355

Description RNA-binding ASCH domain protein [Source:Projected from *Arabidopsis thaliana* (AT2G20410) TAIR;Acc:AT2G20410]

Location Chromosome 1: 251,508-255,065 reverse strand.

About this gene This gene has 2 transcripts (splice variants).

Gene type Protein coding

Annotation Method Gene annotation by JGI through a process of automatic and manual curation

Transcripts [Hide transcript table](#)

Name	Transcript ID	bp	Protein	Biotype
Novel	Sb01g000355.1	1457	342aa	Protein coding
Novel	Sb01g000355.2	1447	318aa	Protein coding

[Show/hide columns \(2 hidden\)](#) [Filter](#)

Export data

Share this page

Ensembl Plants is produced in collaboration with Gramene

Variation Table

Show	All	entries	Show/hide columns	Filter		
Variant ID	Study	Genomic Position	Type	Alleles	Most Severe Consequence	Transcript
vcZ2KJ20E	PRJEB9507	247192	SNV	T/A	downstream_gene_variant	Sb01g000355.2
vcZ2KJ20E	PRJEB9507	247192	SNV	T/A	downstream_gene_variant	Sb01g000355.1
vcZ2KJ20F	PRJEB9507	247220	SNV	T/A	downstream_gene_variant	Sb01g000355.2
vcZ2KJ20F	PRJEB9507	247220	SNV	T/A	downstream_gene_variant	Sb01g000355.1
vcZ2KJ20G	PRJEB9507	247508	SNV	C/T	downstream_gene_variant	Sb01g000355.2
vcZ2KJ20G	PRJEB9507	247508	SNV	C/T	downstream_gene_variant	Sb01g000355.1
vcZ2KJ20H	PRJEB9507	247542	SNV	C/A	downstream_gene_variant	Sb01g000355.2
vcZ2KJ20H	PRJEB9507	247542	SNV	C/A	downstream_gene_variant	Sb01g000355.1

Variant Information Page

WormBase ParaSite Version: WBPS31 (WS254)

Search WormBase ParaSite...
e.g. *O. volvulus*, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase

HMMER Species List BLAST BioMart REST API VEP Downloads WormBase Login Register Help and Documentation

Sorghum bicolor ▾ Location: 1:257,765-268,116

Location-based displays

- Whole genome
- Region in detail
- Other genome browsers
 - Phytozome

Variant Information

Variant ID	Scaffold/Chromosome	Start	End	Reference Allele	Alternative Allele
vcZ2KJ26l	1	260344	-	T	C

Consequences
This variant affects 3 transcripts

Show/hide columns											Filter
Gene ID	Transcript ID	Strand	Biotype	cDNA Position	CDS Position	AA Position	AA Change	Codon Change	SO Term(s)		
Sb01g000365	Sb01g000365.1	-	protein_coding	-	-	-	-	-	downstream_gene_variant		
Sb01g000360	Sb01g000360.1	+	protein_coding	142	142	48	S/P	Tct/Cct	missense_variant		
Sb01g000357	Sb01g000357.1	+	protein_coding	-	-	-	-	-	downstream_gene_variant		

Study PRJEB9507_ERZ115806

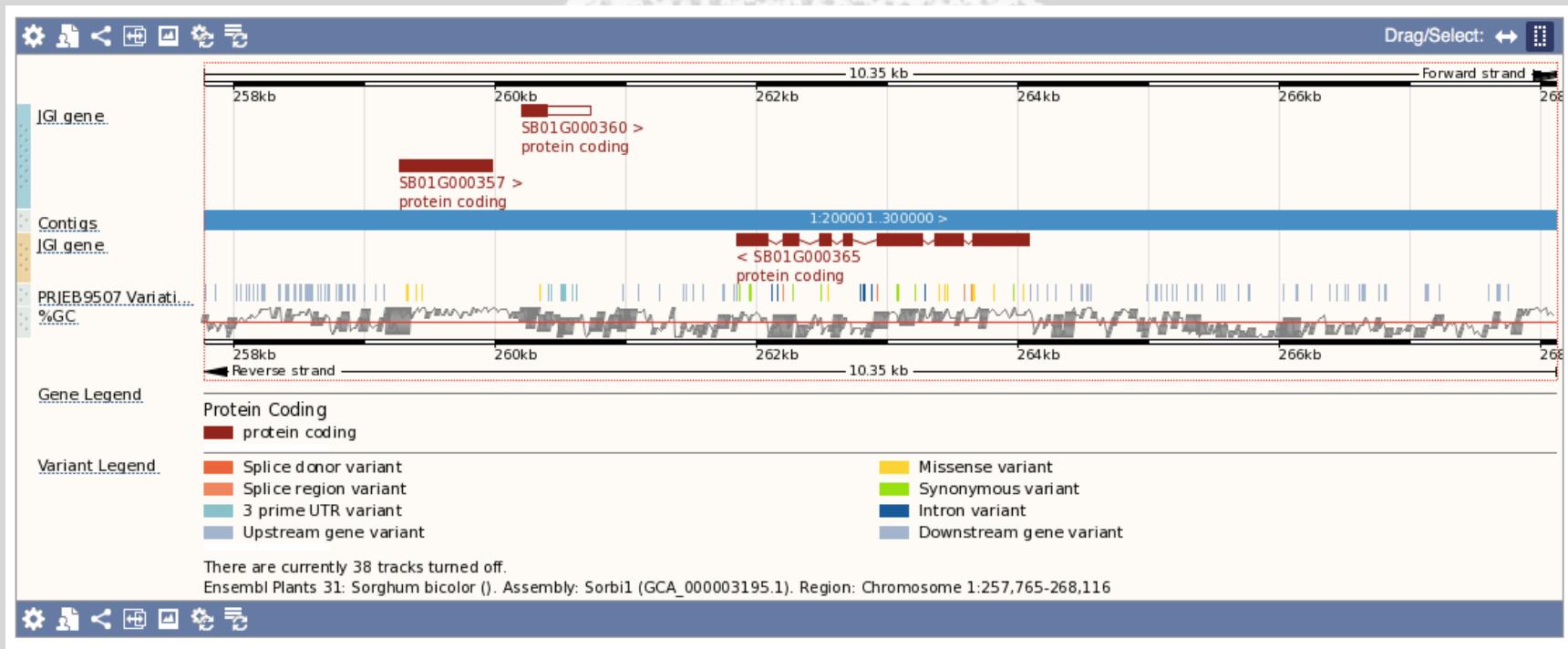
Quality Overview

QUAL	DP	FILTER	AF	NS
33.0	278	PASS	0.0645	47

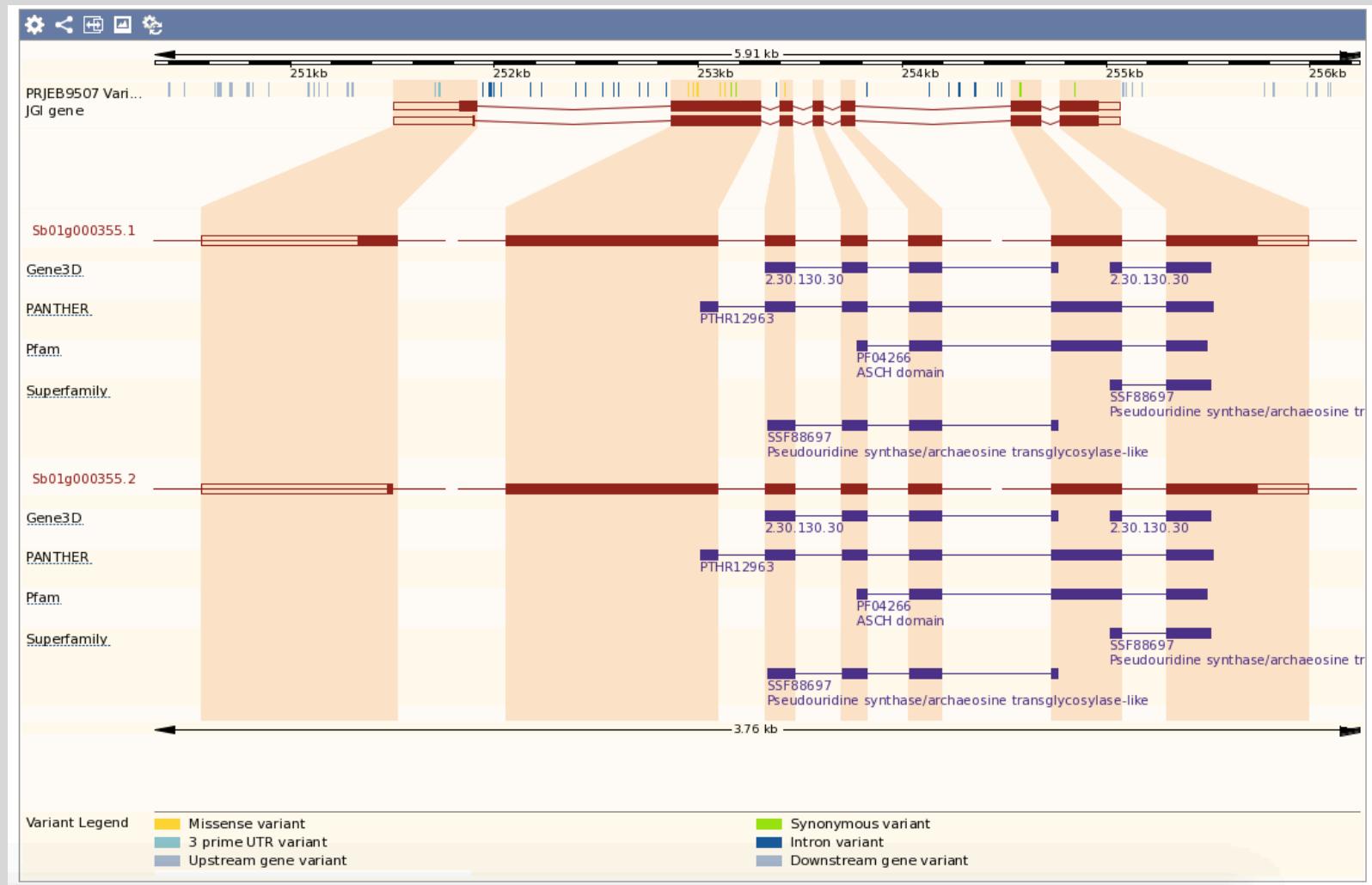
Genotypes
This study included 47 individuals. The genotype for each is shown in the table below.

Show All entries	Show/hide columns	Filter
Sample Name	Genotype	
B35	C/C	
SC23	C/C	
SC326-6	T/T	
IS8525	T/T	
B923296	T/T	
Greenleaf	T/T	
PI586430	T/T	
Kilo	T/T	
SC56-14E	T/T	
Malisor84-7	T/T	
SC62C	T/T	
Rio	T/T	
M35-1	T/T	
S. bicolor subsp. <i>Verticilliflorum</i> (PI300119)	T/T	

Genome Browser Tracks

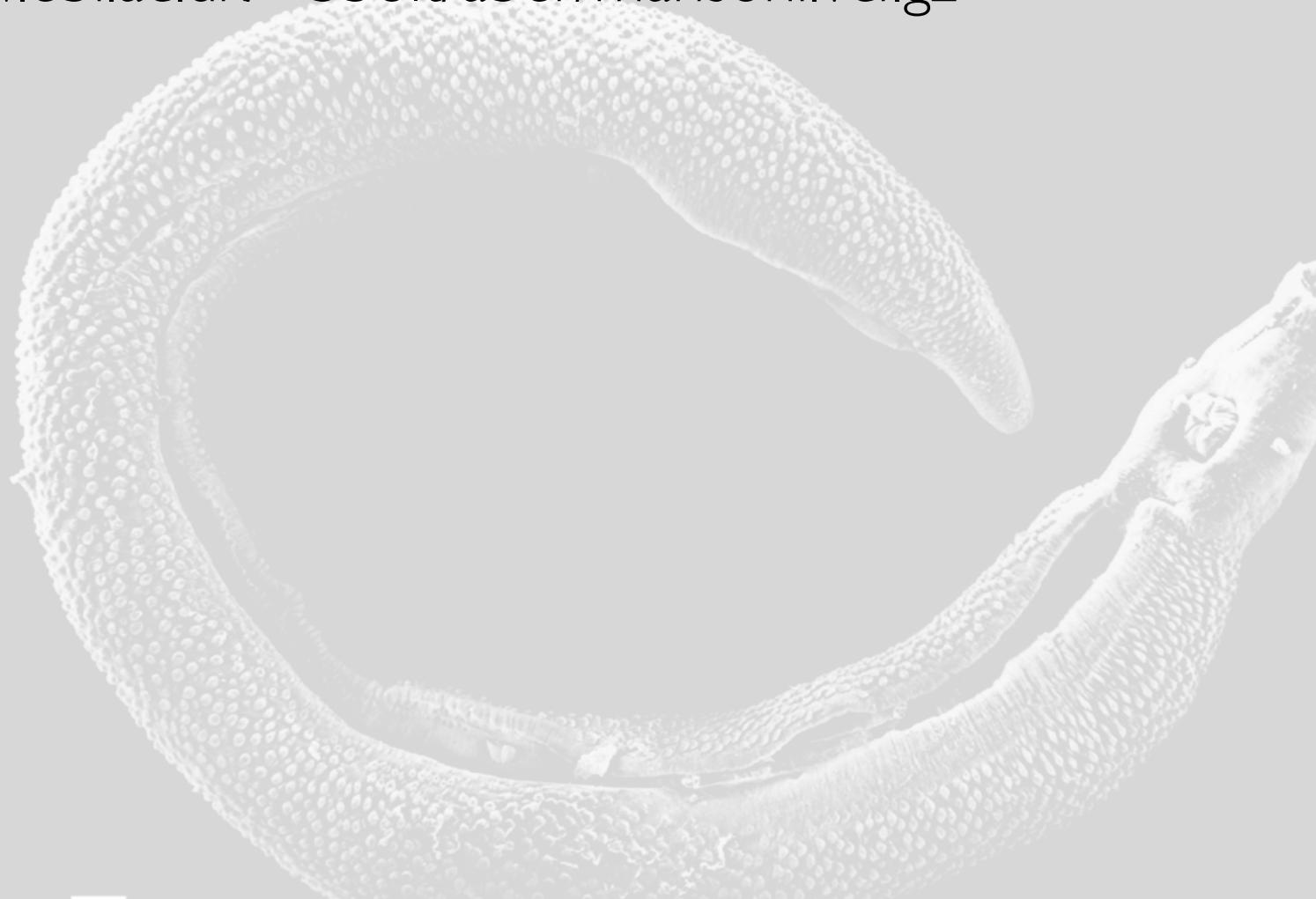


Variant/Transcript Image



Exercises

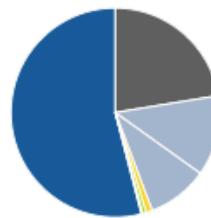
<http://www.ebi.ac.uk/~bbolt/aber/mansoni.vcf.gz>



Exercise Solutions

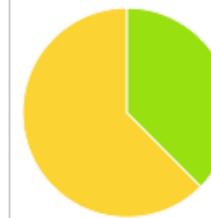
Category	Count
Variants processed	499
Variants remaining after filtering	499
Novel / existing variants	-
Overlapped genes	5
Overlapped transcripts	5
Overlapped regulatory features	-

Consequences (all)



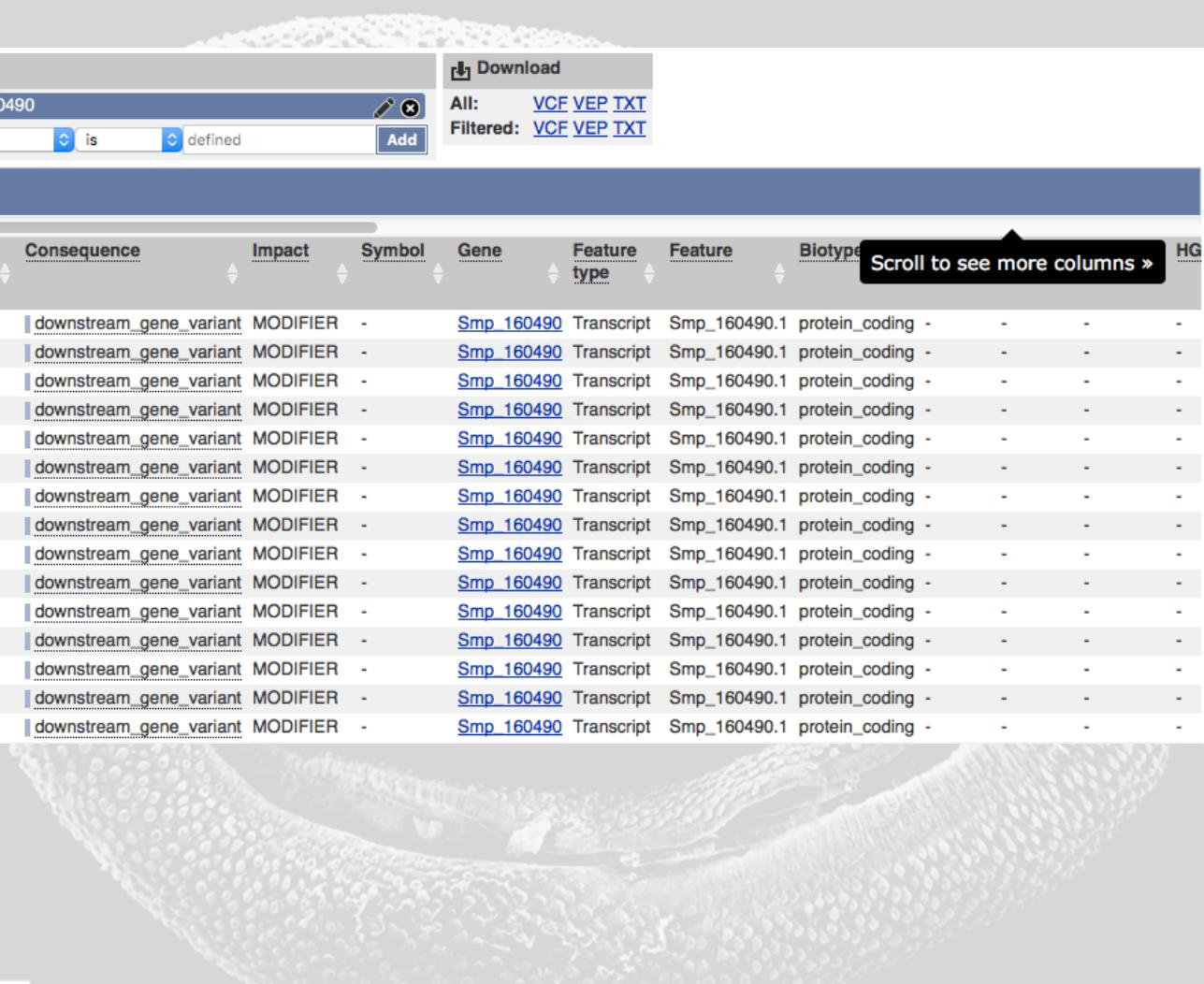
- intron_variant: 54%
- intergenic_variant: 22%
- upstream_gene_variant: 12%
- downstream_gene_variant: 9%
- missense_variant: 1%
- synonymous_variant: 1%
- splice_region_variant: 0%

Coding consequences



- missense_variant: 63%
- synonymous_variant: 38%

Exercise Solutions



Navigation: Show: 1 5 10 50 All variants

Filters: Gene is Smp_160490

Download: All: VCF VEP TXT Filtered: VCF VEP TXT

Show/hide columns

Uploaded variant Location Allele Consequence Impact Symbol Gene Feature type Feature Biotype

Scroll to see more columns » HG

Uploaded variant	Location	Allele	Consequence	Impact	Symbol	Gene	Feature type	Feature	Biotype	HG
.	Smp.Chr 1:118335-118335	G	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:118771-118771	G	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:118772-118772	T	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:118973-118973	C	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:119104-119104	T	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:119118-119118	G	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:119483-119483	T	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:119484-119484	G	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:119485-119485	G	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:119573-119573	C	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:119738-119738	T	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:119775-119775	C	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:119946-119946	A	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:120098-120098	A	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-
.	Smp.Chr 1:120380-120380	T	downstream_gene_variant	MODIFIER	-	Smp_160490	Transcript	Smp_160490.1	protein_coding	-