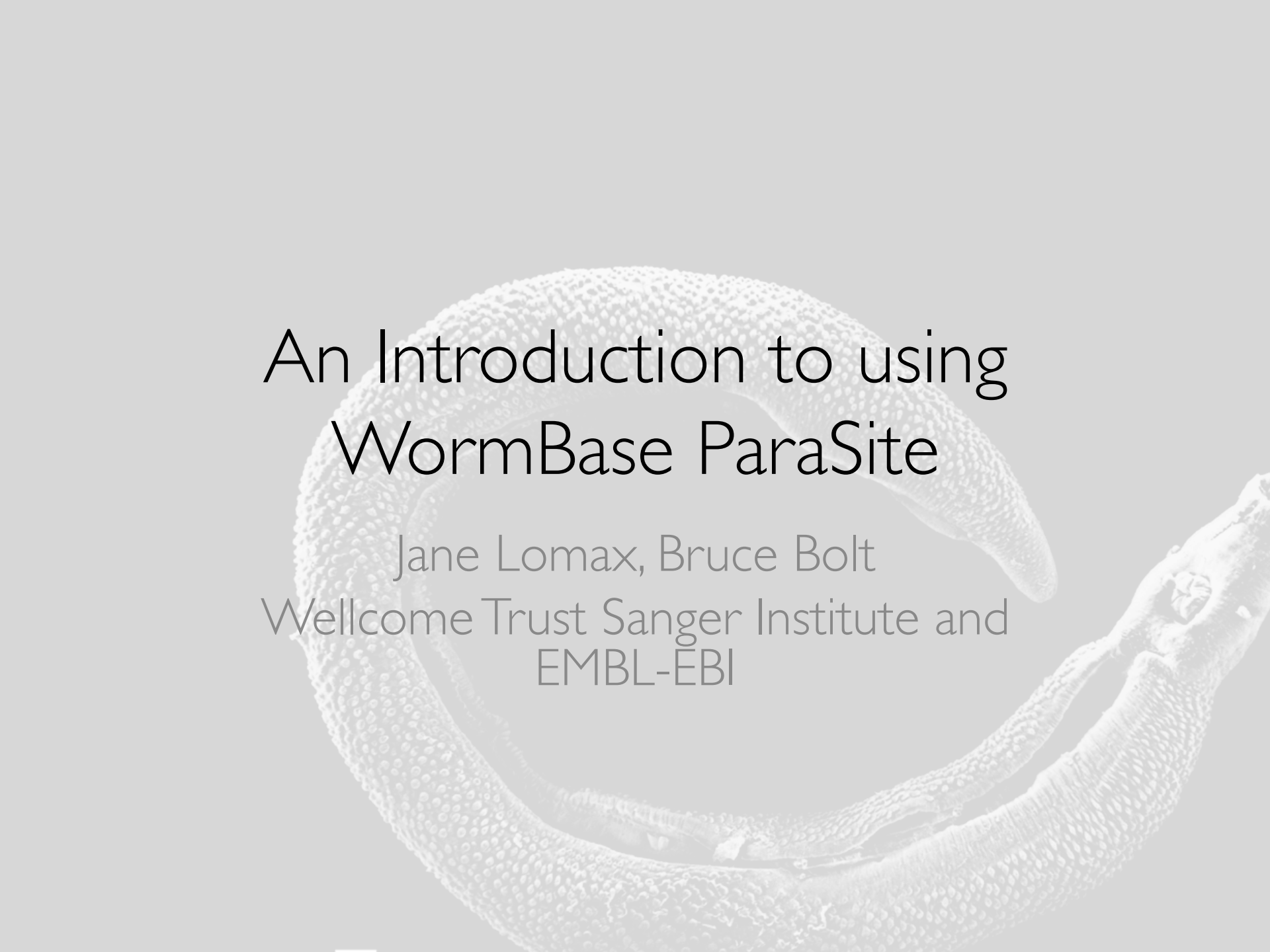
An Introduction to using WormBase ParaSite

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

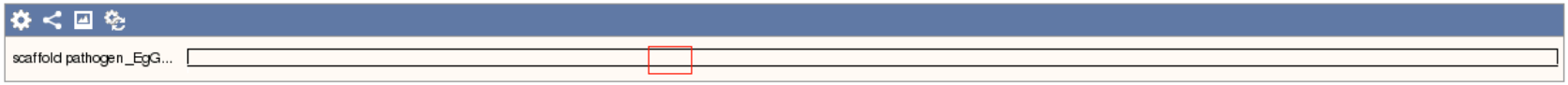
16:00 - 16:30
Q&A Session

Feedback forms

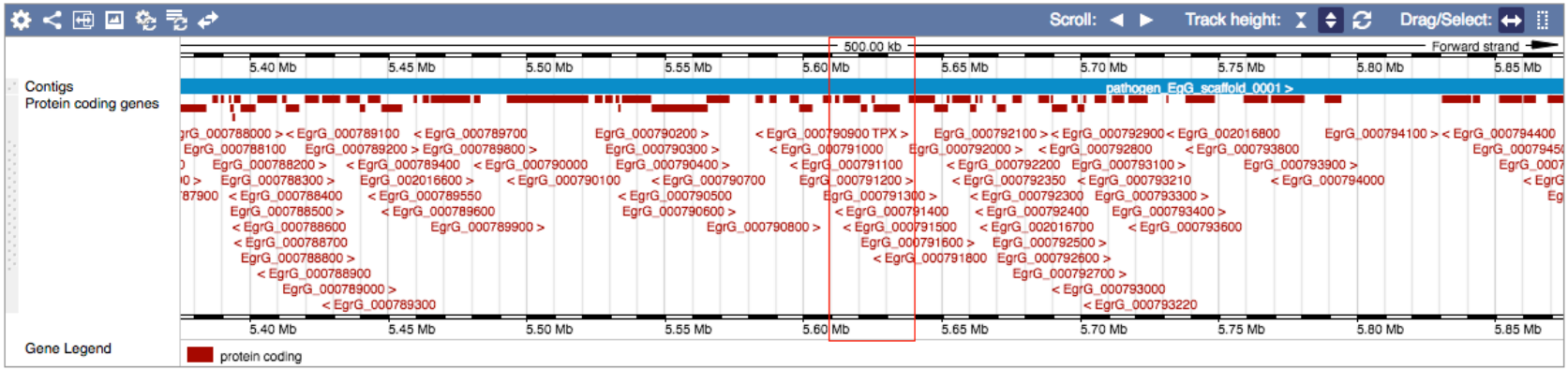
What is WormBase ParaSite?

1. Genome browser for helminth species

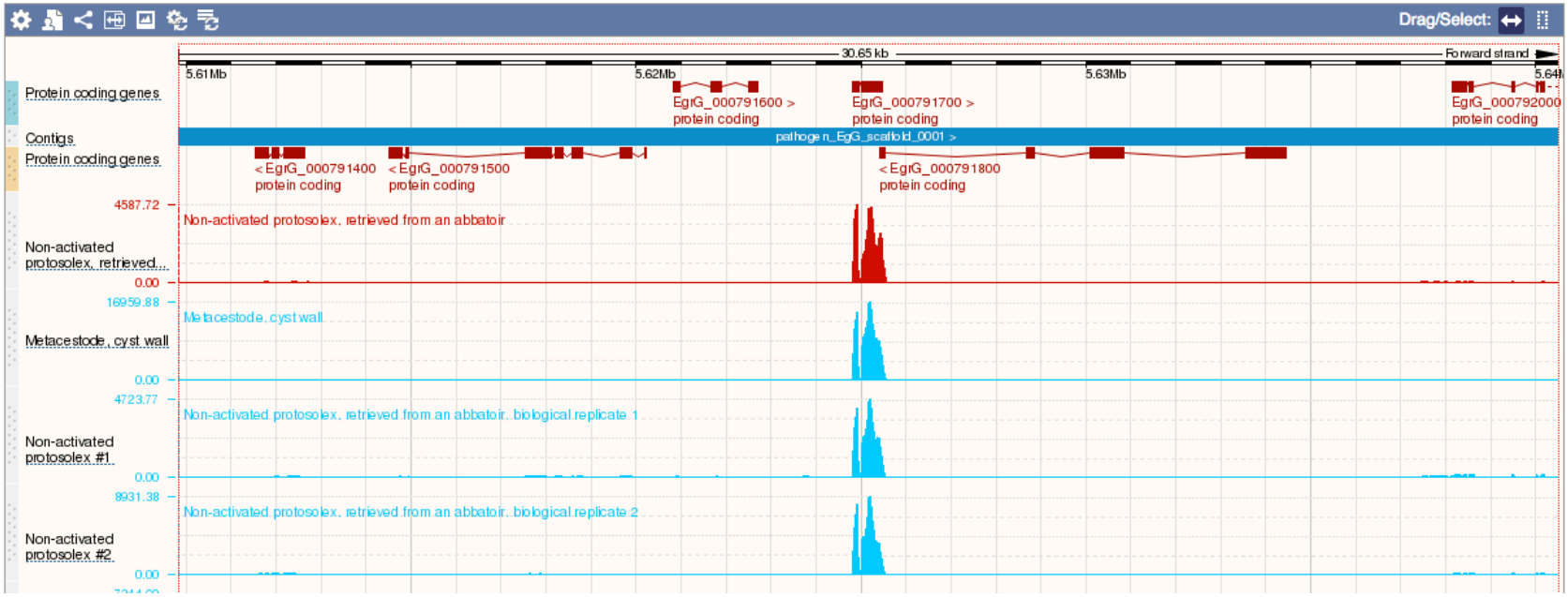
Scaffold pathogen_EgG_scaffold_0001: 5,609,817-5,640,467



Region in detail



Location: pathogen_EgG_scaffold_0001:5609817-56404 Go Gene: Go



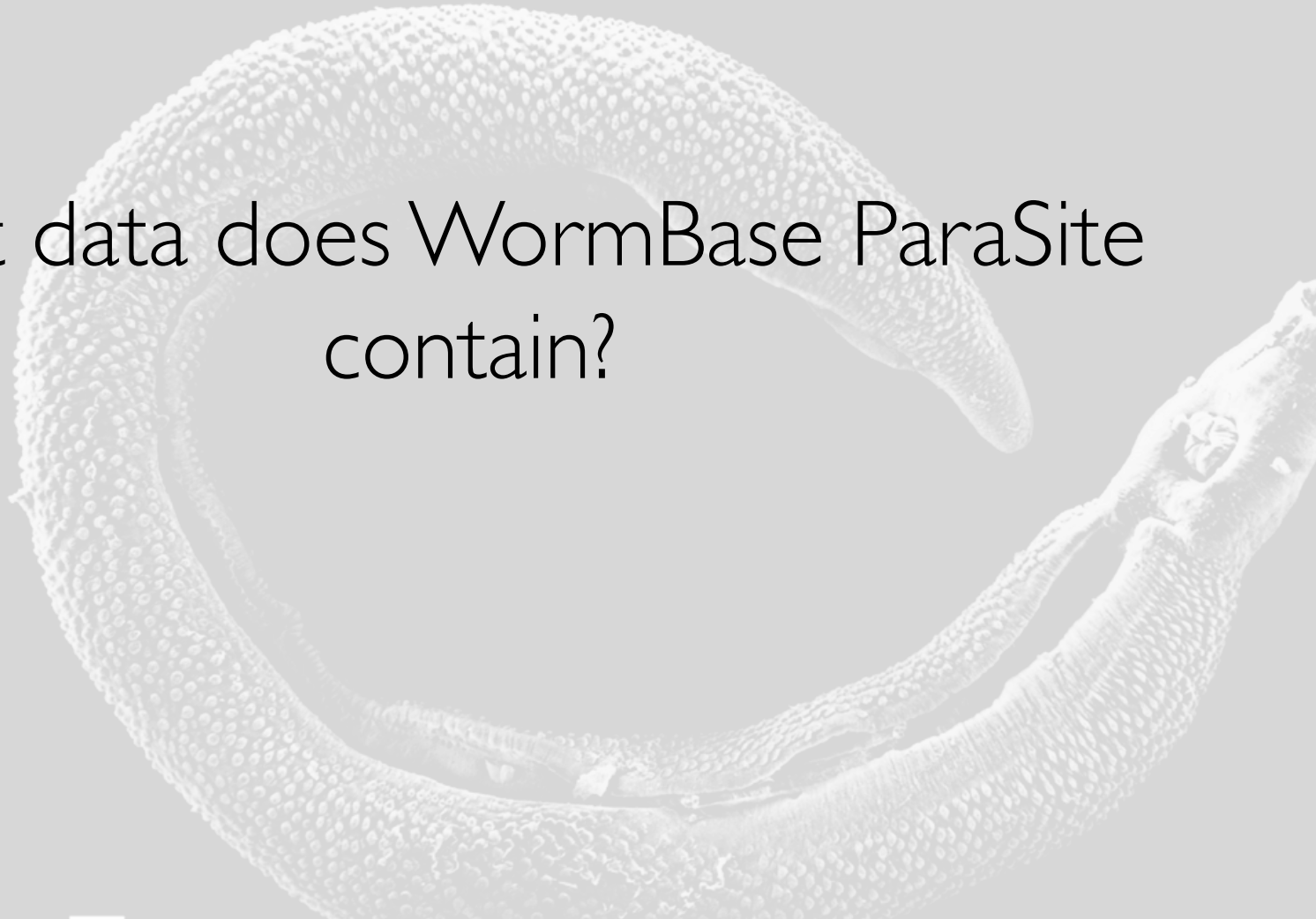
2. Set of tools for interacting with helminth genomes

3. Portal for downloading helminth genomic data

Where does the data come from?

Image credit: brickdisplaycase.com

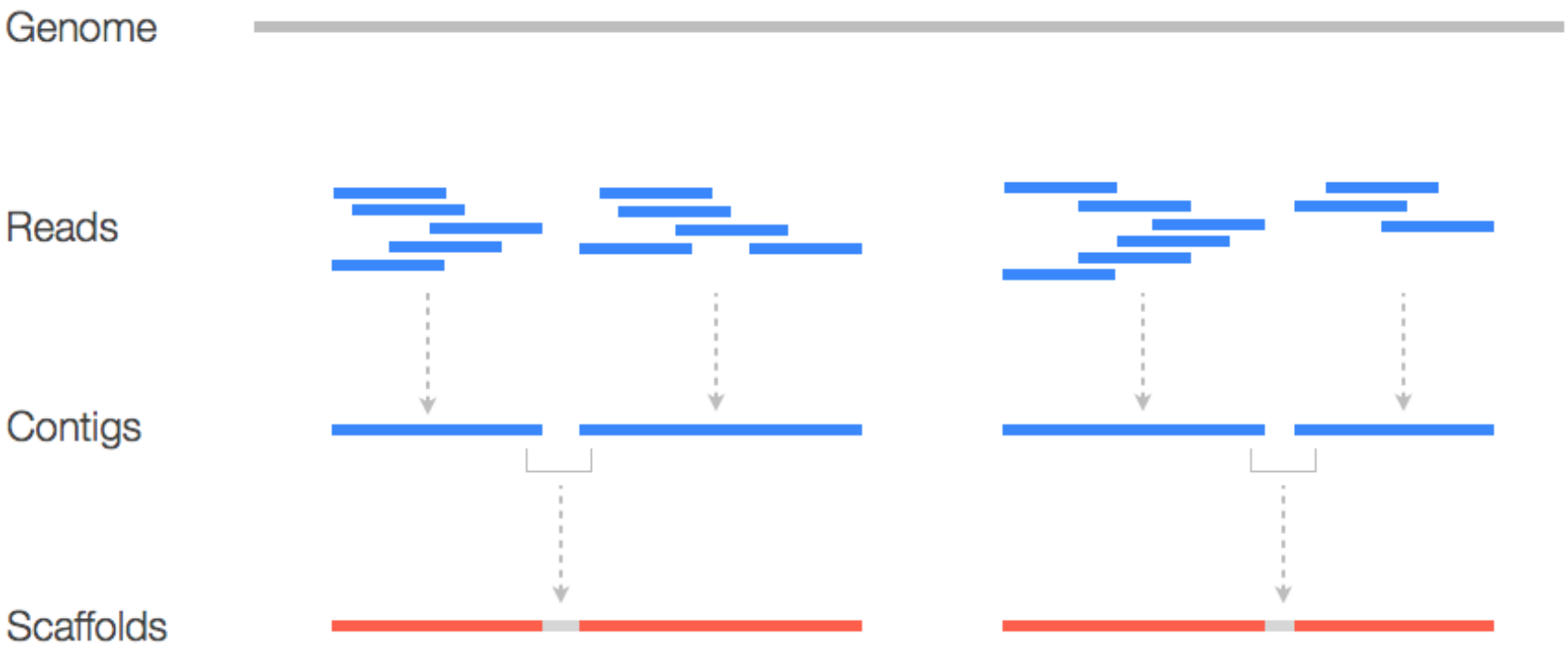
What data does WormBase ParaSite contain?



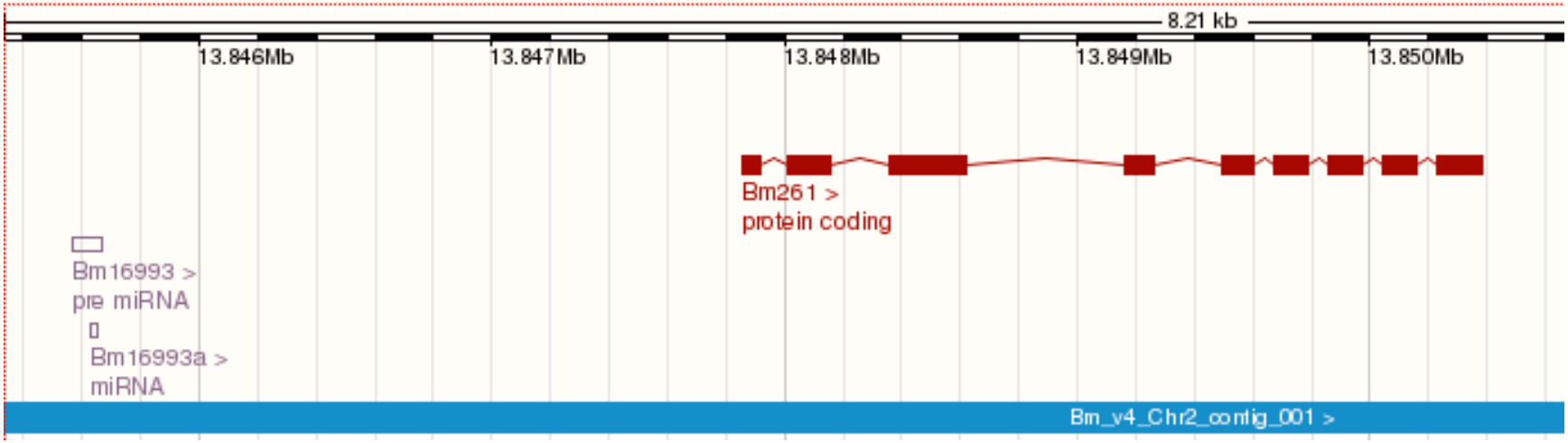
- Genomic:
  - protein-coding genes
  - transcripts
  - proteins
- Transcriptomic
- ncRNAs
- Variation (coming soon)

What is a genome assembly?

Genome assembly



Genome annotation


8.21 kb
13.846Mb
13.847Mb
13.848Mb
13.849Mb
13.850Mb
Bm261 >
protein coding
Bm16993 >
pre miRNA
Bm16993a >
miRNA
Bm_v4_Chr2_contig_001 >

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

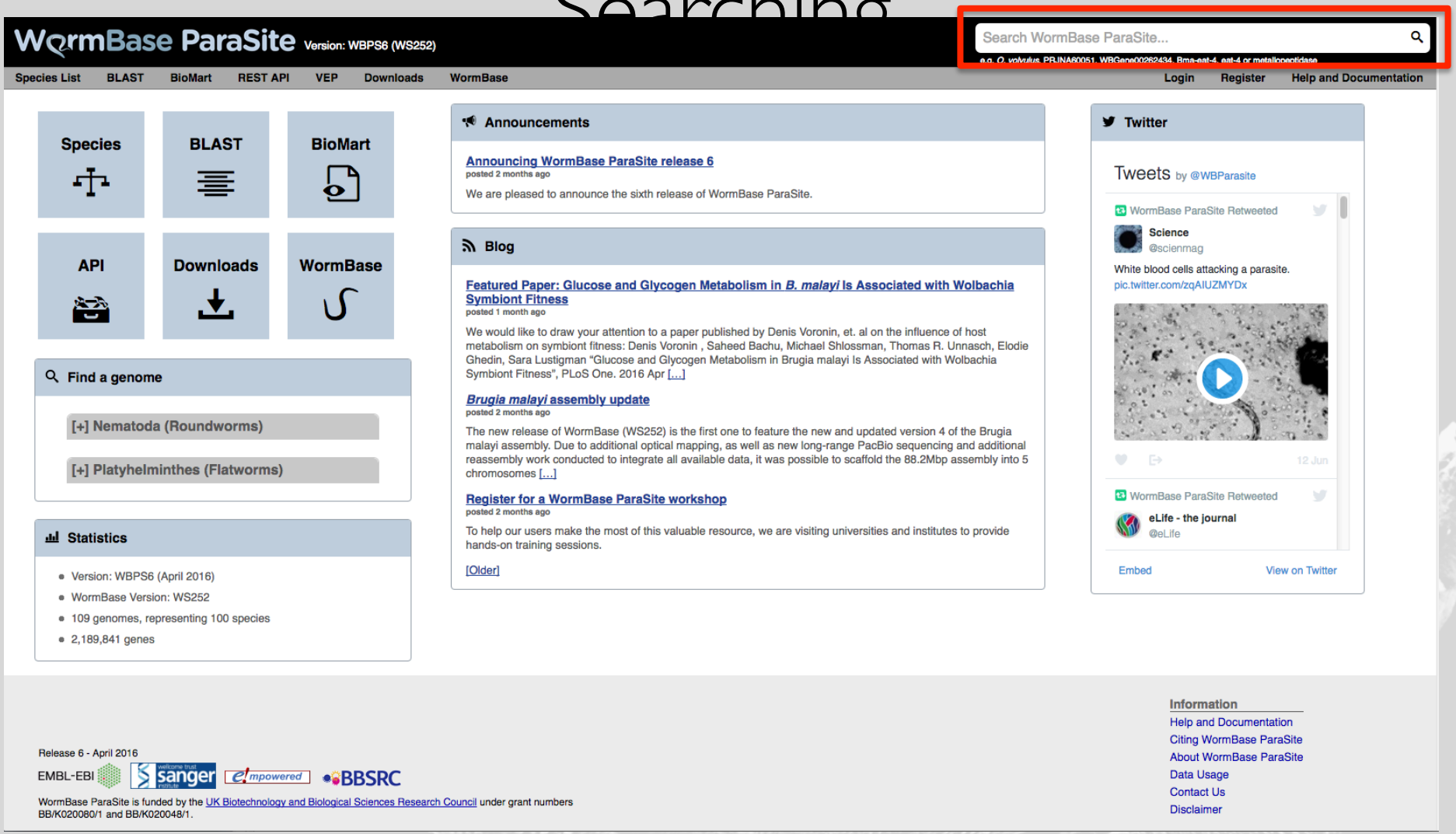
Using the website

Summary

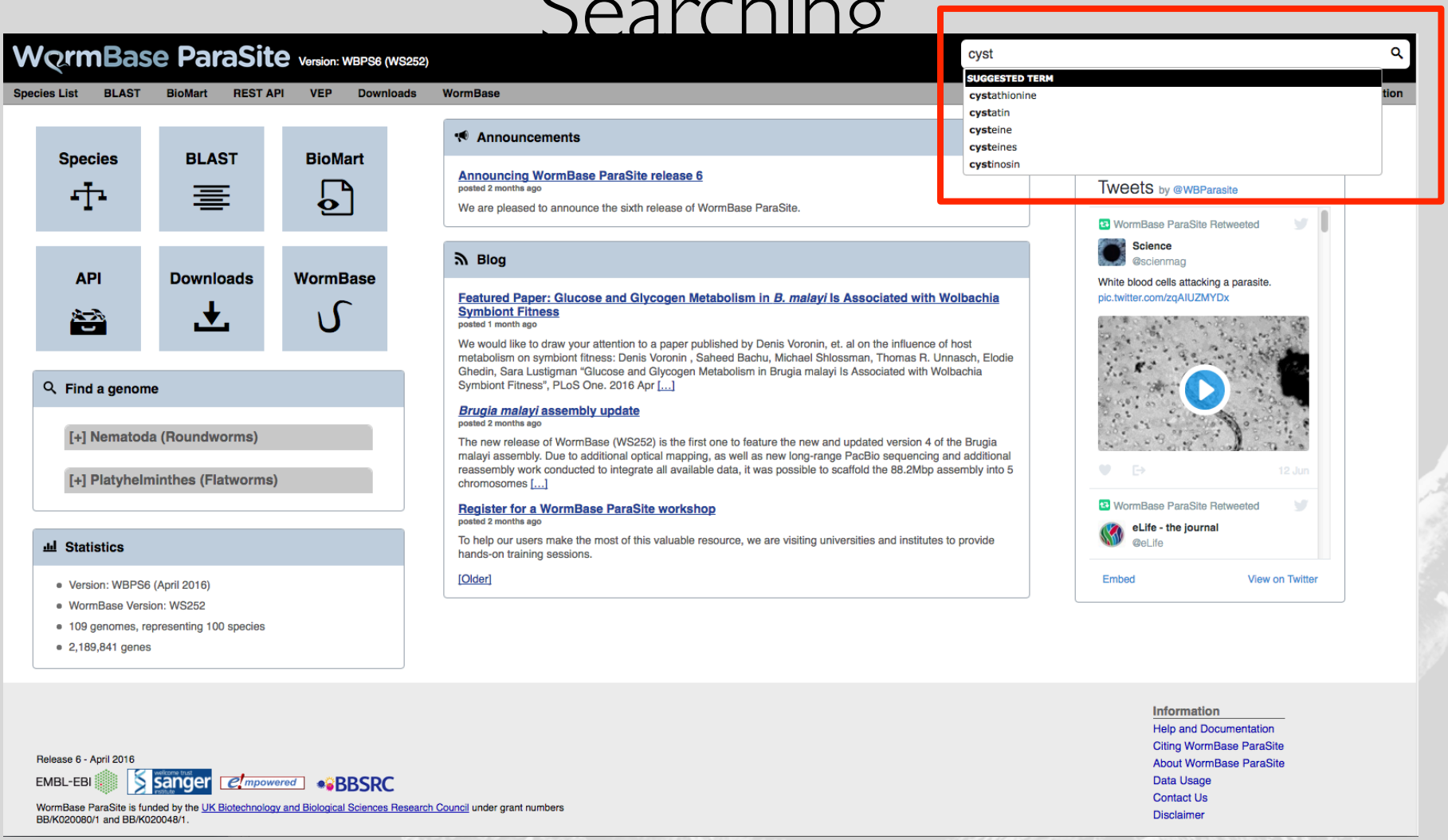
1. Searching
2. Finding genomes
3. Navigating genes, transcripts and scaffolds
4. Comparative genomics
5. RNASeq data tracks
6. Adding your own data tracks
7. User accounts

1. Searching

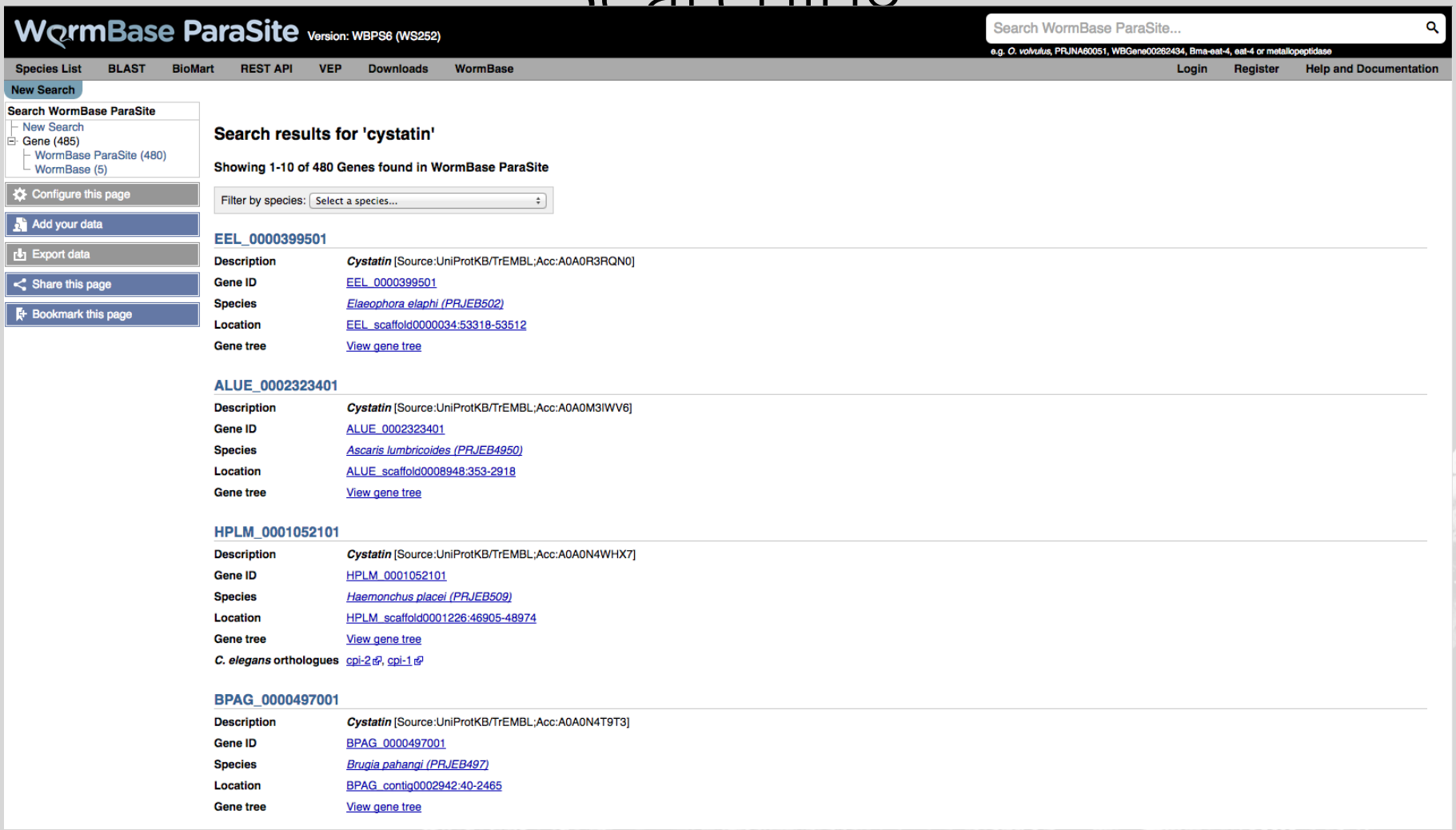
Searching


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
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
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
EMBL-EBI
sanger
empowered
BBSRC
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

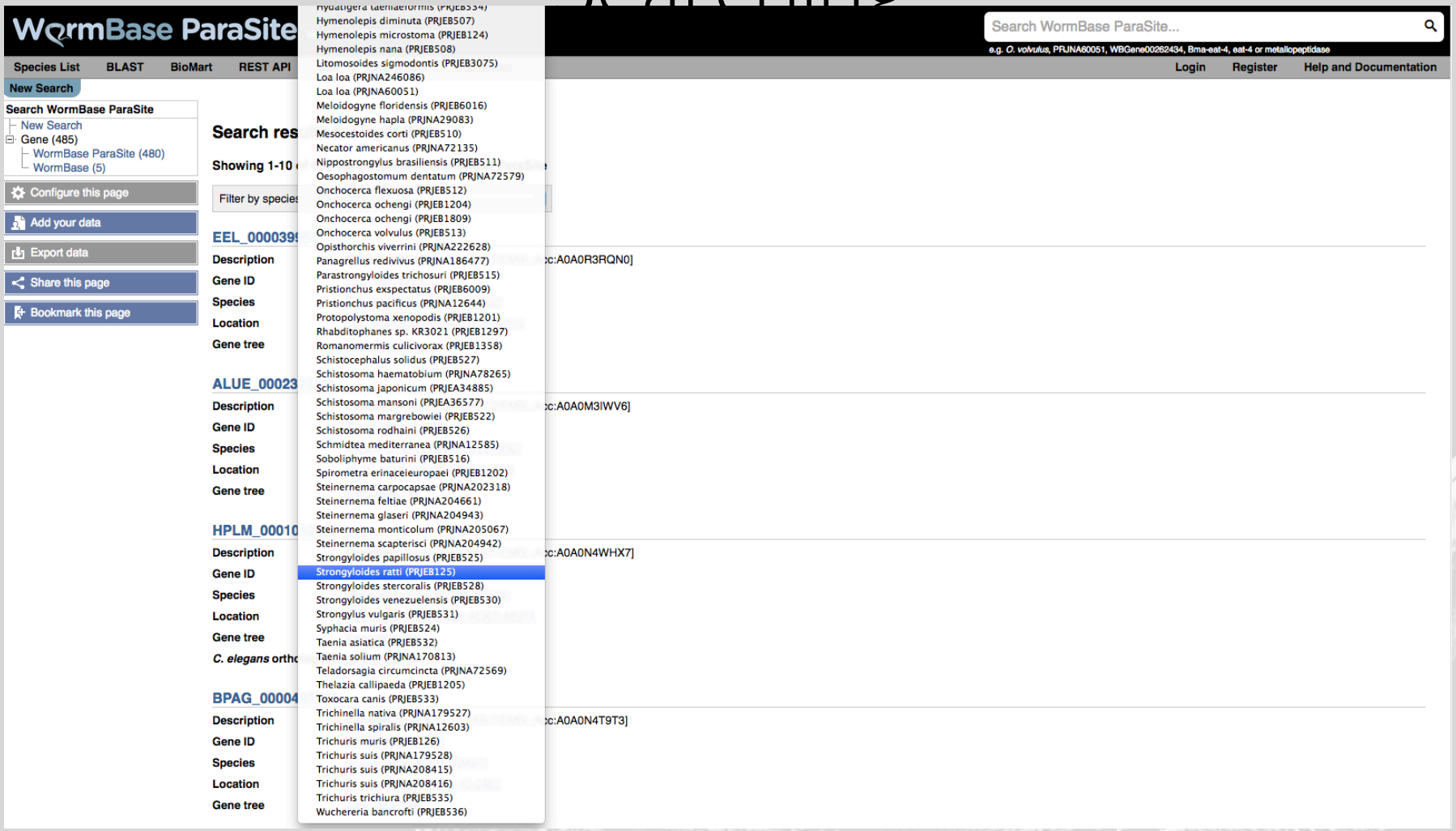
Searching


WormBase ParaSite Version: WBPS6 (WS252)
Species List
BLAST
BioMart
REST API
VEP
Downloads
WormBase
cyst
SUGGESTED TERM
cystathionine
cystatin
cysteine
cysteines
cystinosin
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]
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
EMBL-EBI
sanger
empowered
BBSRC
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

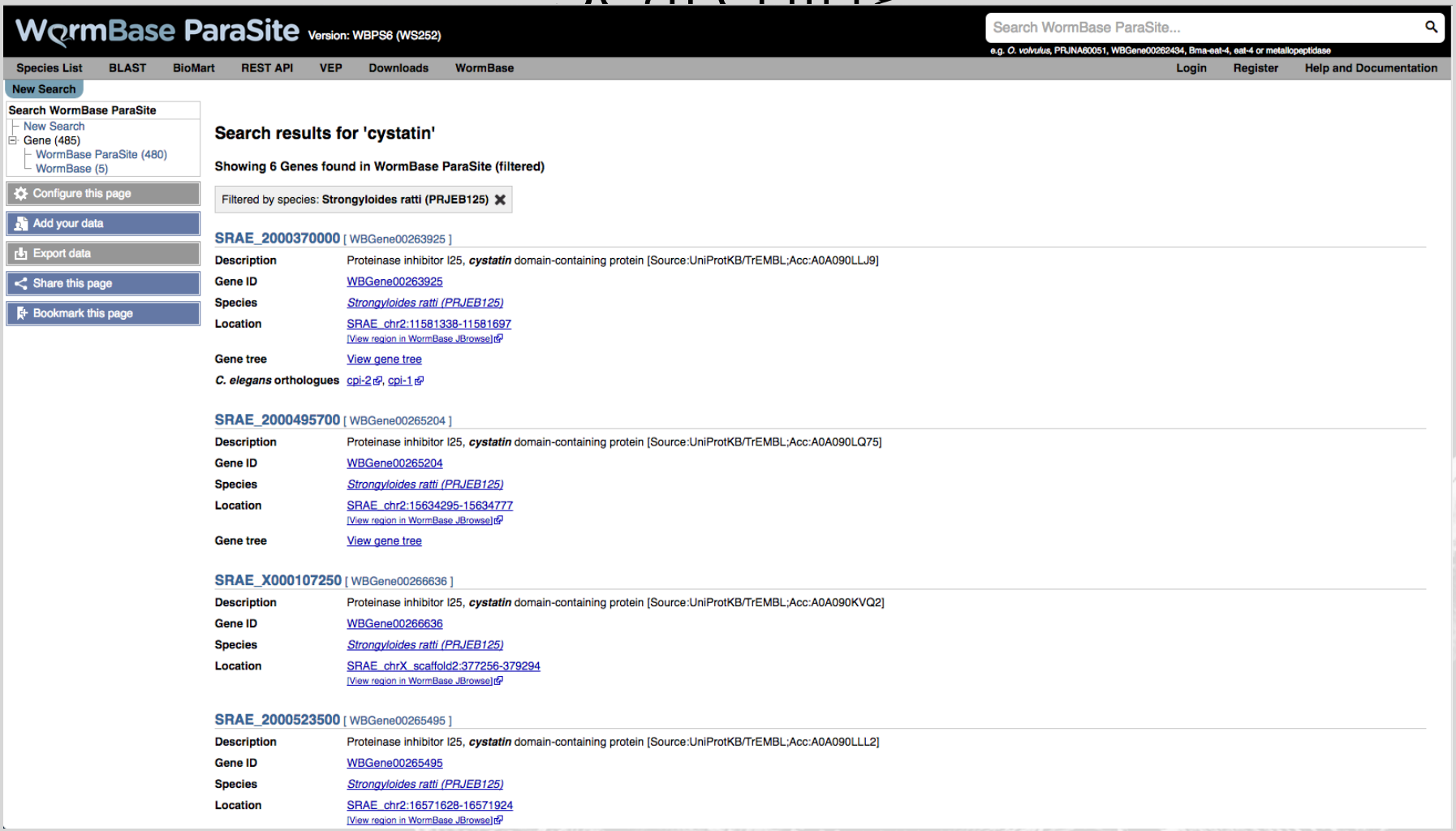
Searching


WormBase ParaSite Version: WBPS6 (WS252)
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
Search results for 'cystatin'
Showing 1-10 of 480 Genes found in WormBase ParaSite
Filter by species: Select a species...
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

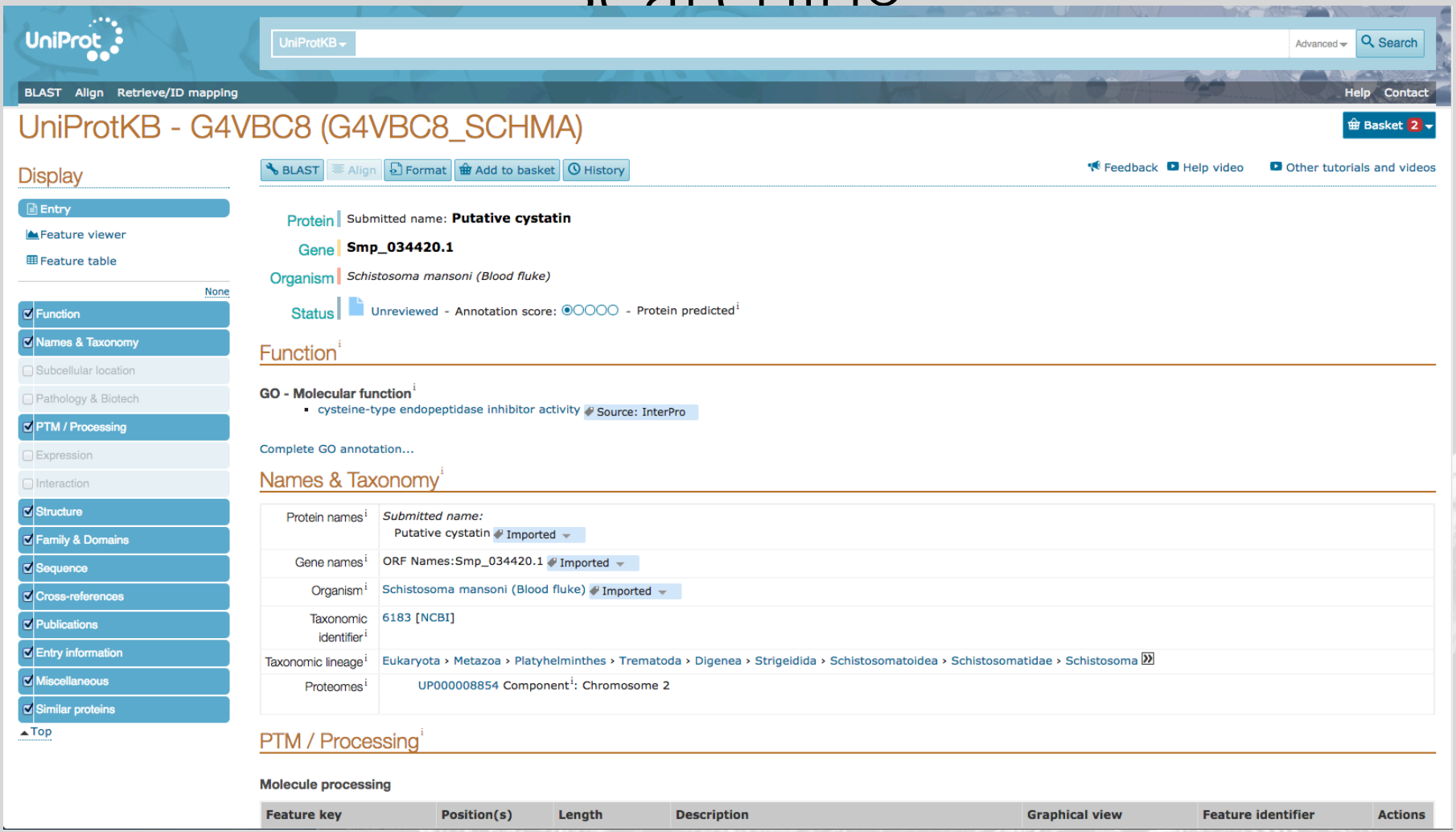
Searching


WormBase ParaSite
Search WormBase ParaSite...
e.g. C. volvulus, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase
Species List
BLAST
BioMart
REST API
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 res
Showing 1-10
Filter by species
EEL_000039
Description
Gene ID
Species
Location
Gene tree
ALUE_00023
HPLM_0001
C. elegans ortho
BPAG_00004
Hymenolepis diminuta (PRJEB507)
Hymenolepis microstoma (PRJEB124)
Hymenolepis nana (PRJEB508)
Litomosoides sigmodontis (PRJEB3075)
Loa loa (PRJNA246086)
Loa loa (PRJNA60051)
Meloidogyne floridensis (PRJEB6016)
Meloidogyne hapla (PRJNA29083)
Mesocestoides corti (PRJEB510)
Necator americanus (PRJNA72135)
Nippostrongylus brasiliensis (PRJEB511)
Oesophagostomum dentatum (PRJNA72579)
Onchocerca flexuosa (PRJEB512)
Onchocerca ochengi (PRJEB1204)
Onchocerca ochengi (PRJEB1809)
Onchocerca volvulus (PRJEB513)
Opisthorchis viverrini (PRJNA222628)
Panagrellus redivivus (PRJNA186477)
Parastrongyloides trichosuri (PRJEB515)
Pristionchus exspectatus (PRJEB6009)
Pristionchus pacificus (PRJNA12644)
Protopolystoma xenopodis (PRJEB1201)
Rhabditophanes sp. KR3021 (PRJEB1297)
Romanomermis culicivorax (PRJEB1358)
Schistocephalus solidus (PRJEB527)
Schistosoma haematobium (PRJNA78265)
Schistosoma japonicum (PRJEA34885)
Schistosoma mansoni (PRJEA36577)
Schistosoma margrebowiei (PRJEB522)
Schistosoma rodhaini (PRJEB526)
Schmidtea mediterranea (PRJNA12585)
Soboliphyme baturini (PRJEB516)
Spirometra erinaceieuropaei (PRJEB1202)
Steinernema carpocapsae (PRJNA202318)
Steinernema feltiae (PRJNA204661)
Steinernema glaseri (PRJNA204943)
Steinernema monticolum (PRJNA205067)
Steinernema scapterisci (PRJNA204942)
Strongyloides papillosus (PRJEB525)
Strongyloides ratti (PRJEB125)
Strongyloides stercoralis (PRJEB528)
Strongyloides venezuelensis (PRJEB530)
Strongylus vulgaris (PRJEB531)
Syphacia muris (PRJEB524)
Taenia asiatica (PRJEB532)
Taenia solium (PRJNA170813)
Teladorsagia circumcincta (PRJNA72569)
Thelazia callipaeda (PRJEB1205)
Toxocara canis (PRJEB533)
Trichinella nativa (PRJNA179527)
Trichinella spiralis (PRJNA12603)
Trichuris muris (PRJEB126)
Trichuris suis (PRJNA179528)
Trichuris suis (PRJNA208415)
Trichuris suis (PRJNA208416)
Trichuris trichiura (PRJEB535)
Wuchereria bancrofti (PRJEB536)
cc:A0A0R3RQN0]
cc:A0A0M3IWV6]
cc:A0A0N4WHX7]
cc:A0A0N4T9T3]

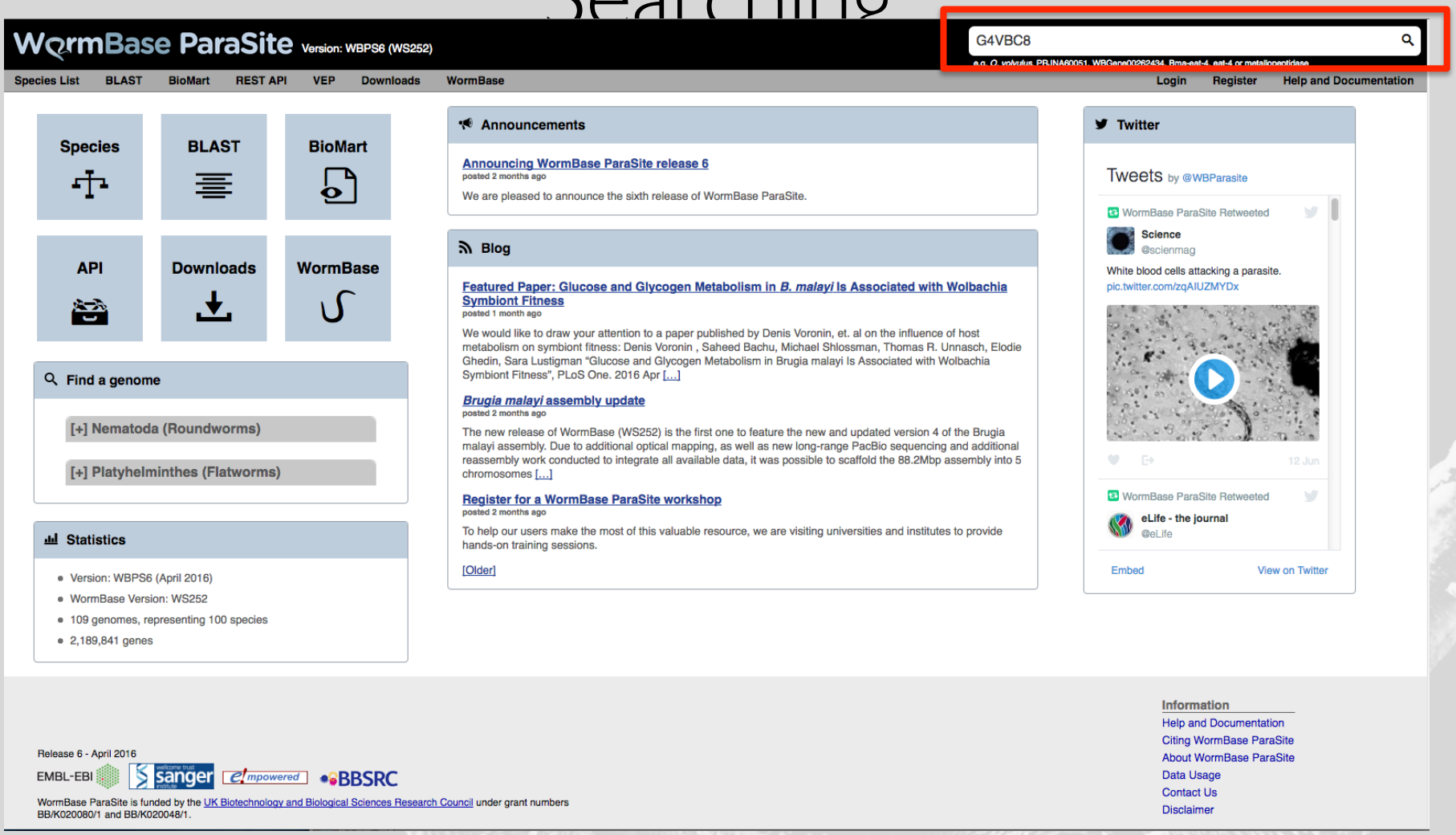
Searching


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
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]

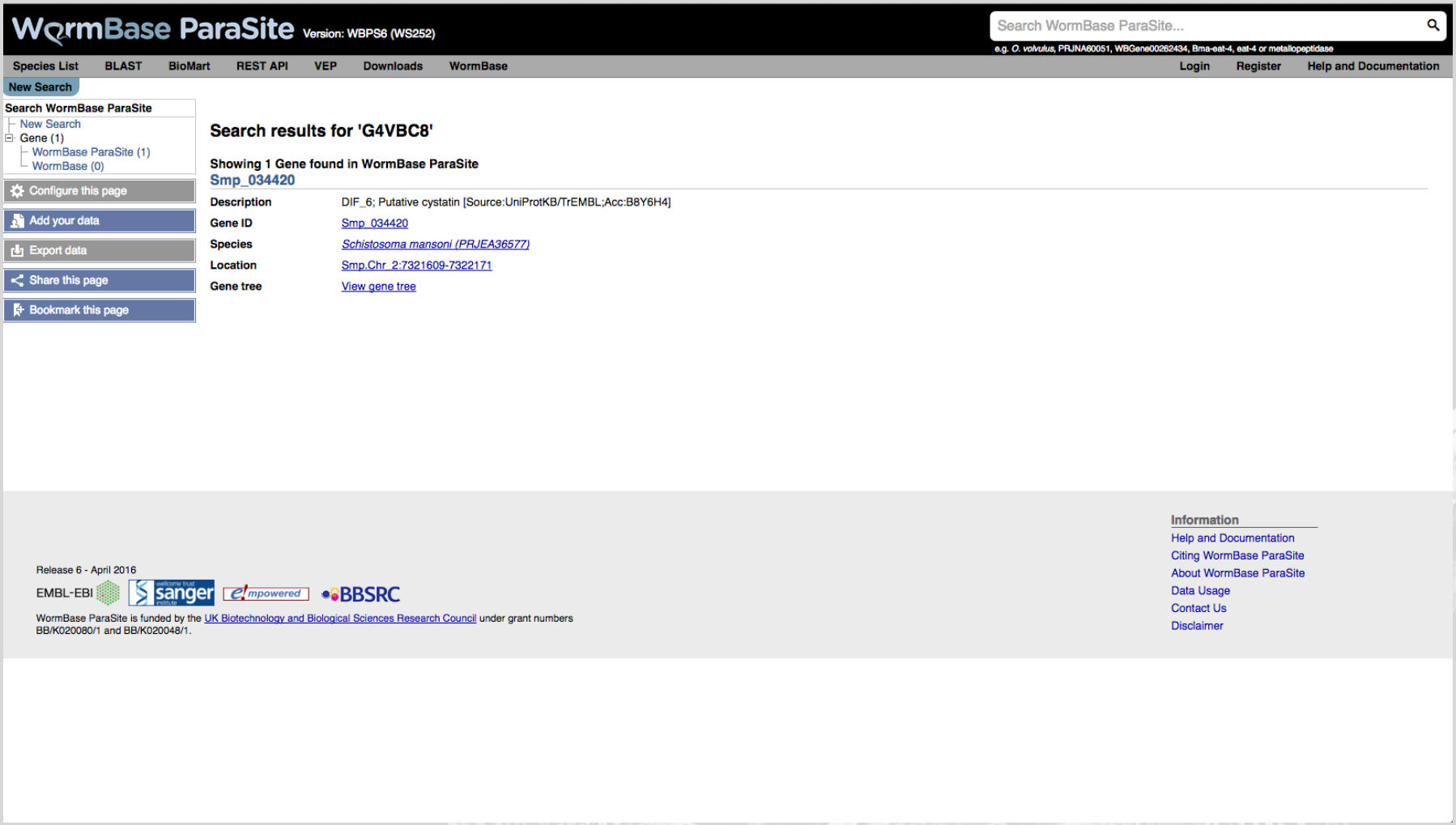
Searching


UniProt
UniProtKB
Advanced
Search
BLAST Align Retrieve/ID mapping
Help Contact
UniProtKB - G4VBC8 (G4VBC8_SCHMA)
Basket 2
Display
Entry
Feature viewer
Feature table
None
Function
Names & Taxonomy
Subcellular location
Pathology & Biotech
PTM / Processing
Expression
Interaction
Structure
Family & Domains
Sequence
Cross-references
Publications
Entry information
Miscellaneous
Similar proteins
Top
BLAST Align Format Add to basket History
Feedback Help video Other tutorials and videos
Protein Submitted name: Putative cystatin
Gene Smp_034420.1
Organism Schistosoma mansoni (Blood fluke)
Status Unreviewed - Annotation score: - 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 › Strigeidida › Schistosomatoidea › Schistosomatidae › Schistosoma
Proteomes UP000008854 Component: Chromosome 2
PTM / Processing
Molecule processing
Feature key
Position(s)
Length
Description
Graphical view
Feature identifier
Actions

Searching

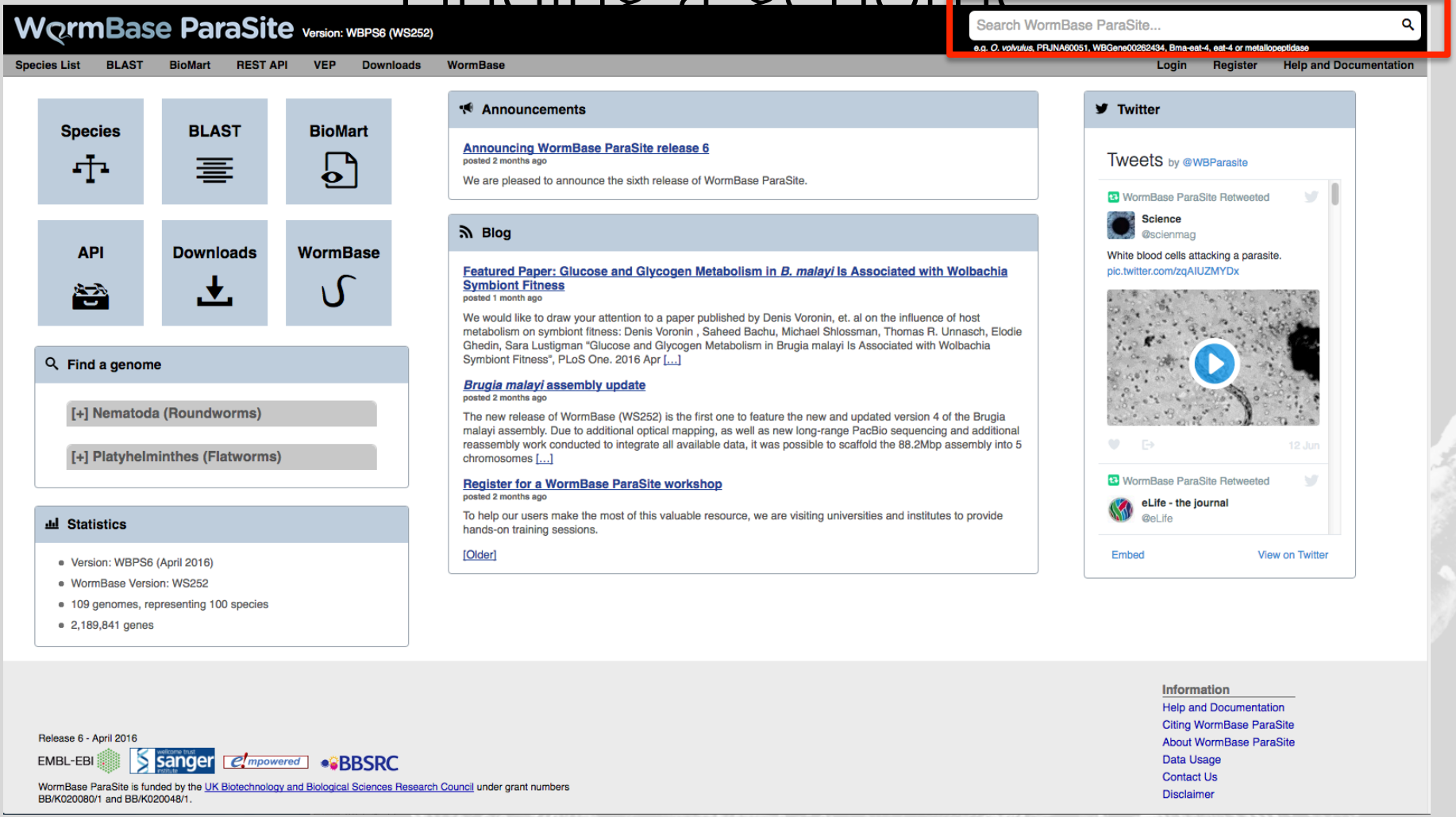

WormBase ParaSite Version: WBPS6 (WS252)
G4VBC8
Species List
BLAST
BioMart
REST API
VEP
Downloads
WormBase
Login
Register
Help and Documentation
Species
BLAST
BioMart
API
Downloads
WormBase
Find a genome
[+] Nematoda (Roundworms)
[+] Platyhelminthes (Flatworms)
Statistics
Version: WBPS6 (April 2016)
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 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
Release 6 - April 2016
EMBL-EBI
sanger
empowered
BBSRC
WormBase ParaSite is funded by the UK Biotechnology and Biological Sciences Research Council under grant numbers BB/K020080/1 and BB/K020048/1.

Searching

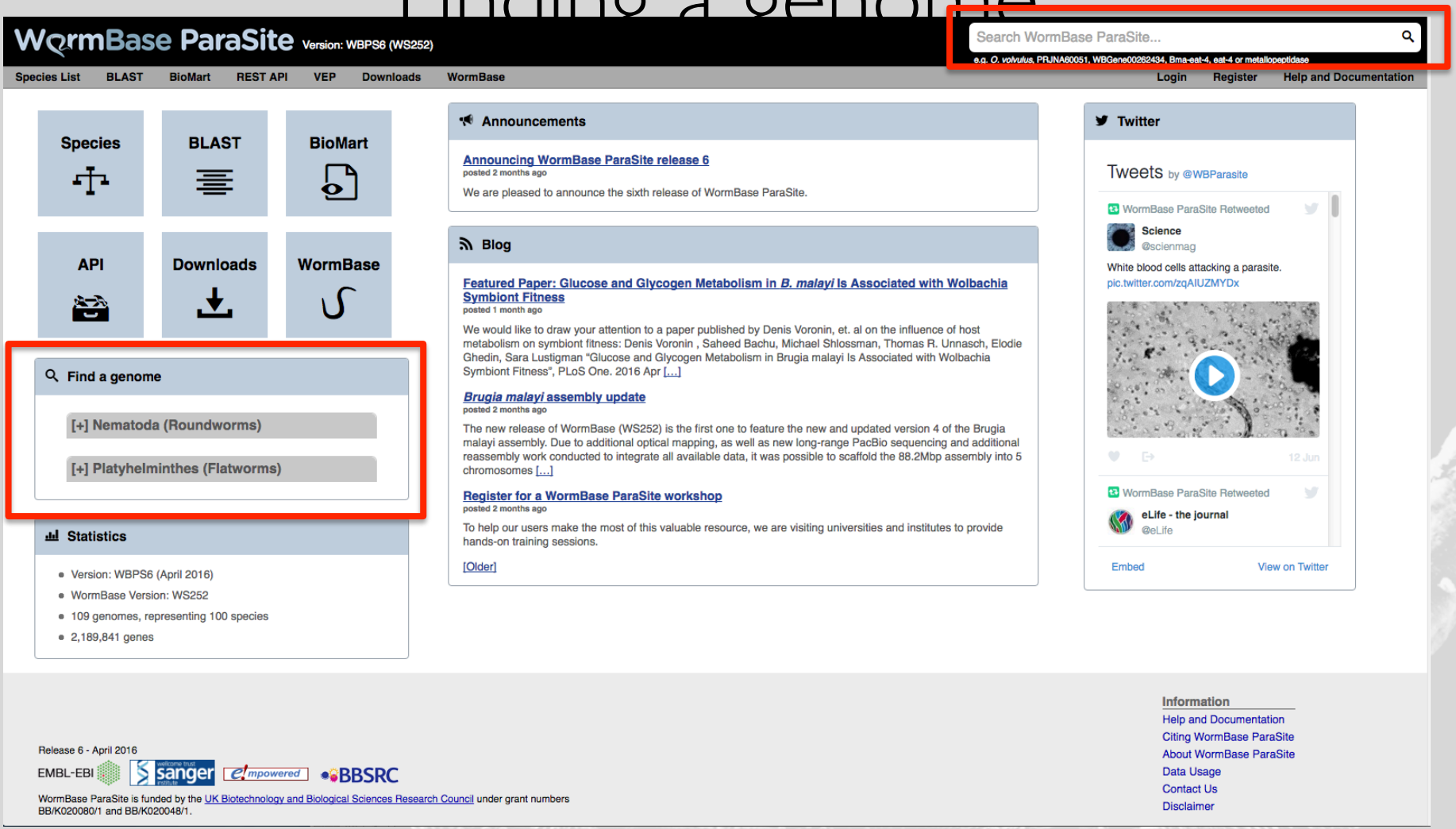

WormBase ParaSite Version: WBPS6 (WS252)
Search WormBase ParaSite...
e.g. O. volvulus, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase
Species List BLAST BioMart REST API VEP Downloads WormBase
Login Register Help and Documentation
New Search
Search WormBase ParaSite
New Search
Gene (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 Schistosoma mansoni (PRJEA36577)
Location Smp.Chr_2:7321609-7322171
Gene tree View gene tree
Release 6 - April 2016
EMBL-EBI
sanger
empowered
BBSRC
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

2. Finding genomes

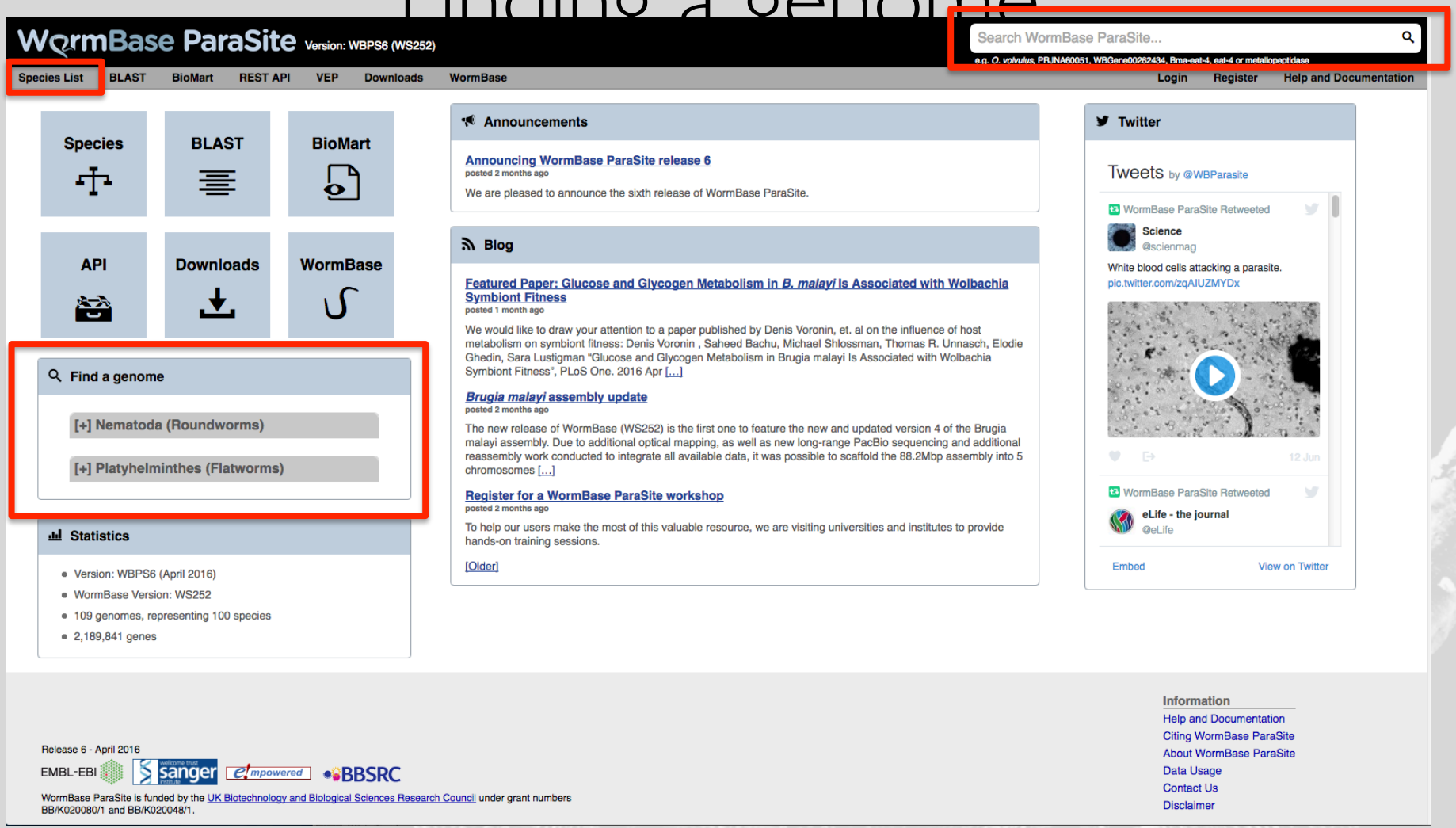
Finding a genome



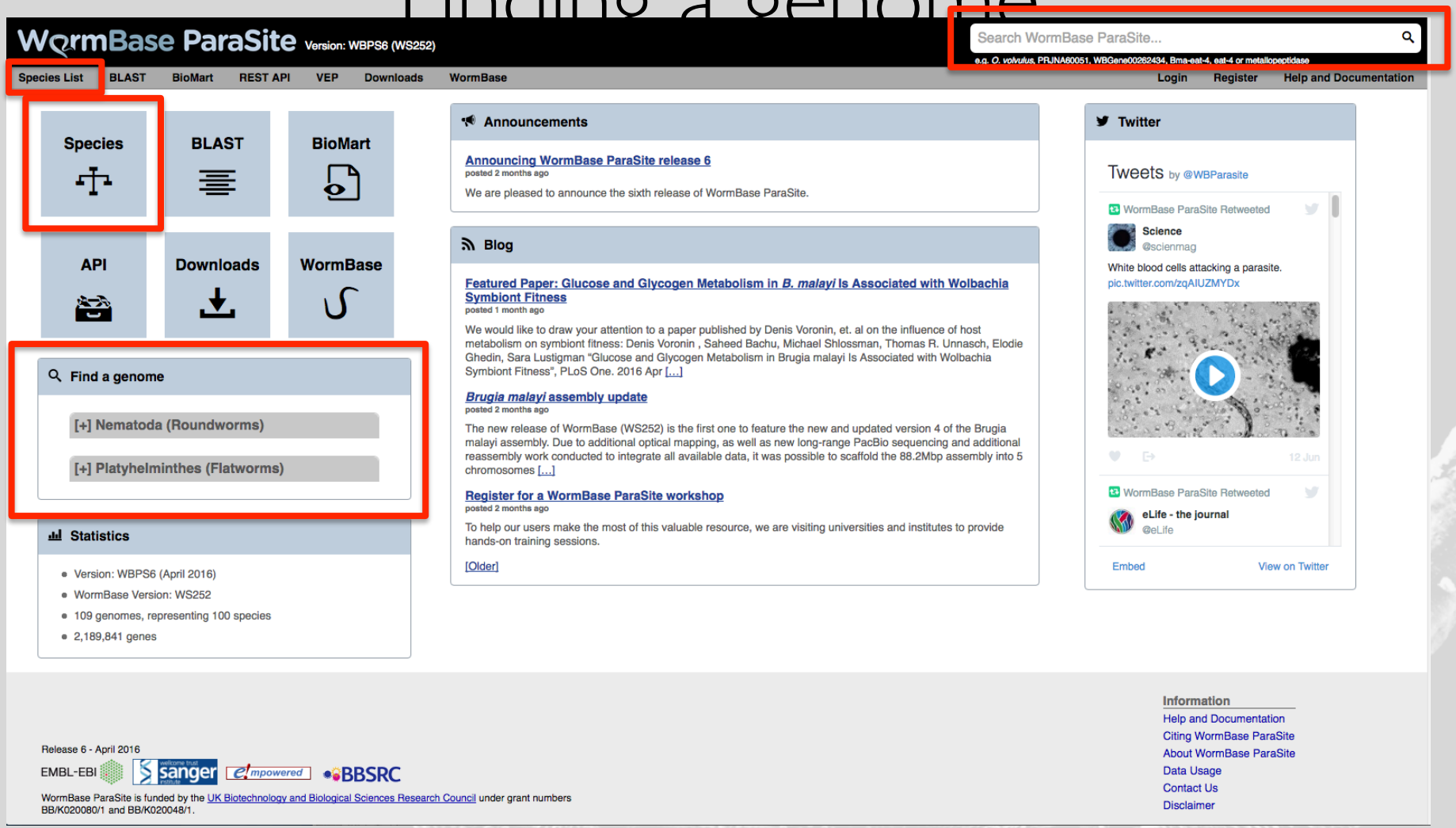
Finding a genome



Finding a genome



Finding a genome



Genomes list

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 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	Bmal-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)

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

Schistosoma mansoni (PRJEA36577) ▼

Schistosoma mansoni

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

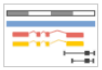
Search Schistosoma mansoni (PRJEA36577)...

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 Biomphlaria, 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 intially 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

Comparative genomics

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

More information and statistics

Example gene tree

Key Publications

Berriman M, Haas BJ, LoVerde PT, Wilson RA, Dillon GP, Cerqueira GC, Mashiyama 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, Djikeng 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 IJ, 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

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

Downloads

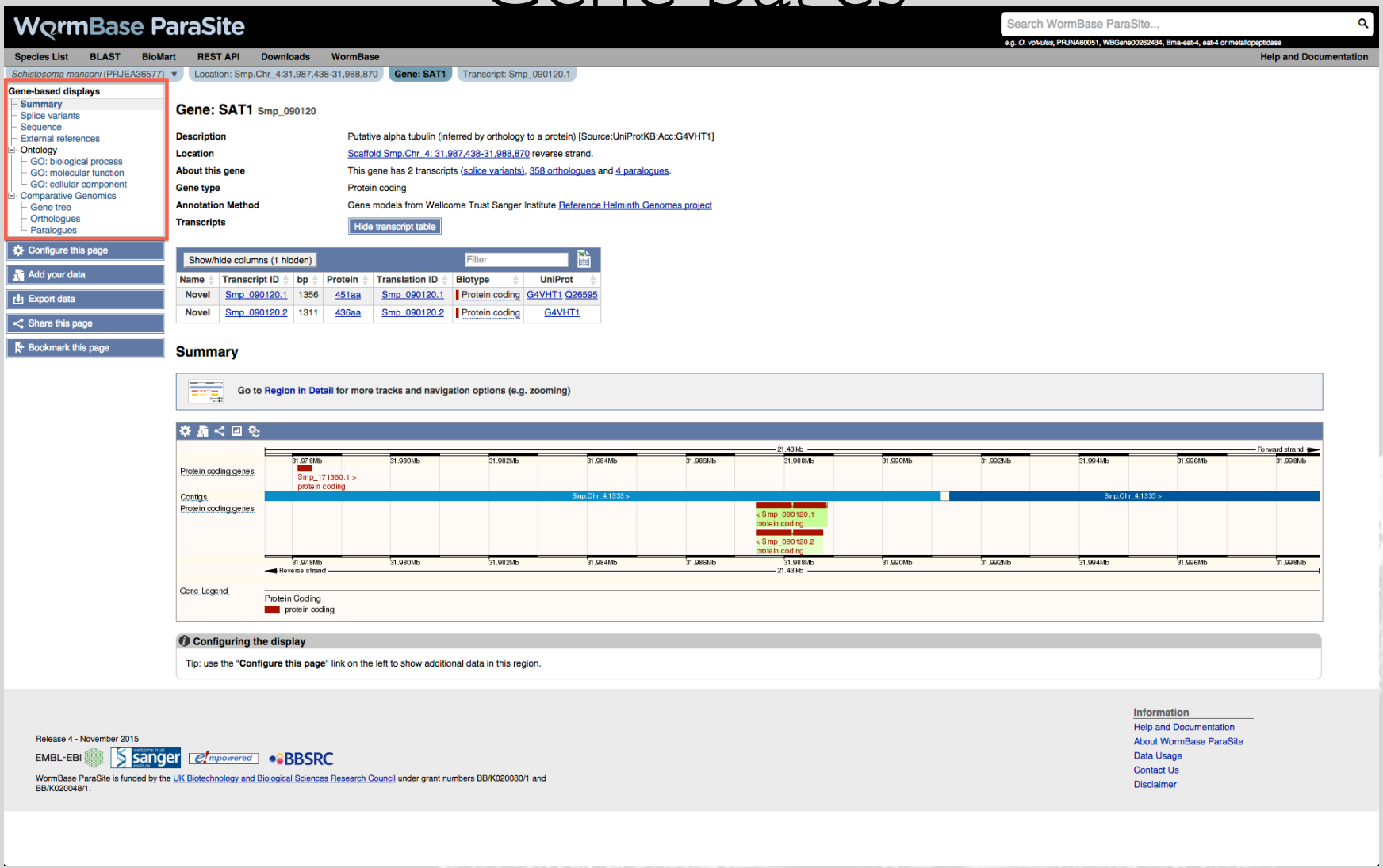
- Genomic Sequence (FASTA)
- Hard-masked Genomic Sequence (FASTA)
- Soft-masked Genomic Sequence (FASTA)
- Annotations (GFF3)
- Proteins (FASTA)
- Full-length transcripts (FASTA)
- CDS transcripts (FASTA)

Tools

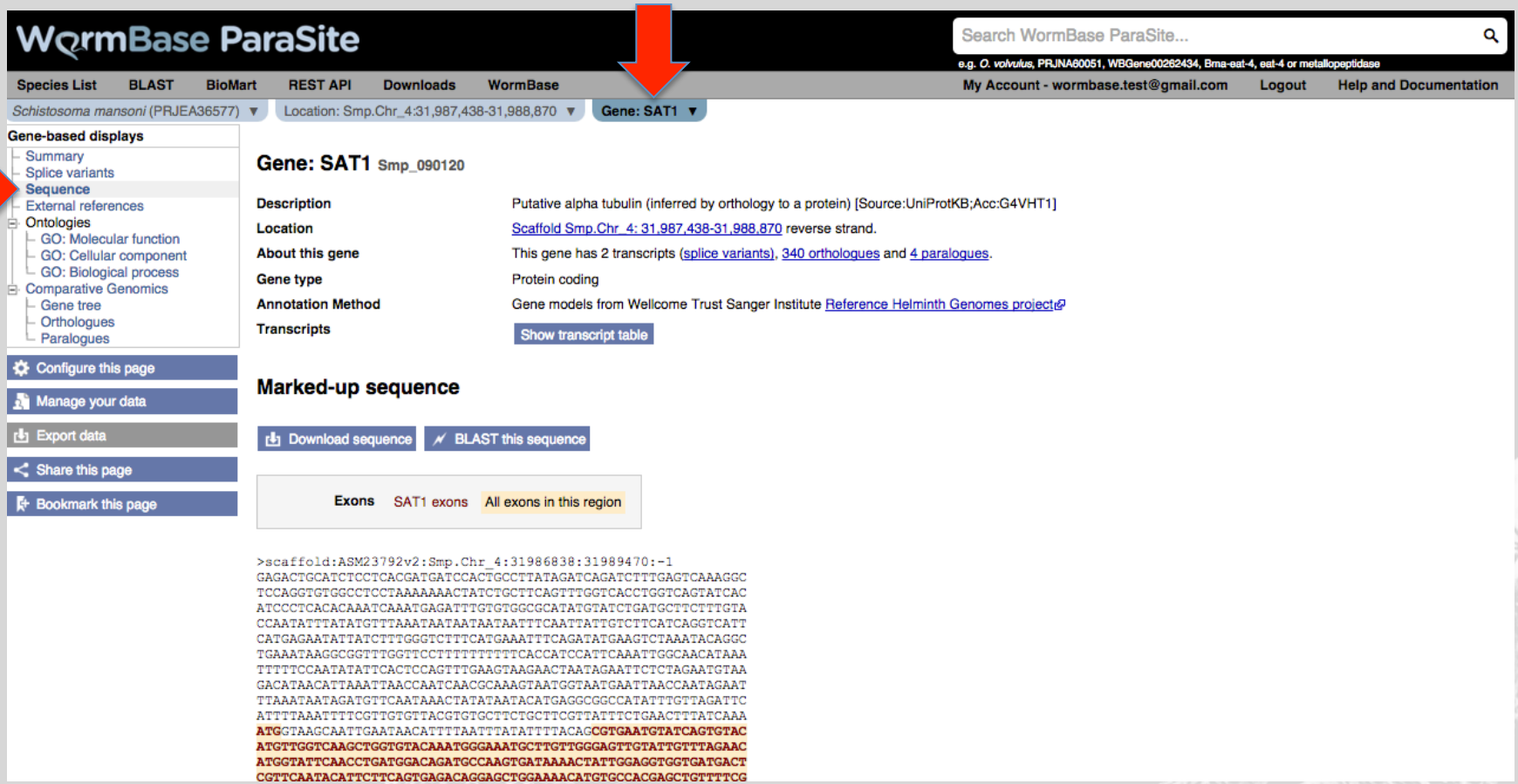
- Search for sequences in the genome and proteome using BLAST
- Work with lists of data using the WormBase ParaSite BioMart data-mining tool
- Programatically access WormBase ParaSite data using the REST API
- Predict the effects of variants using the Variant Effect Predictor

3. Navigating genes, transcripts and scaffolds

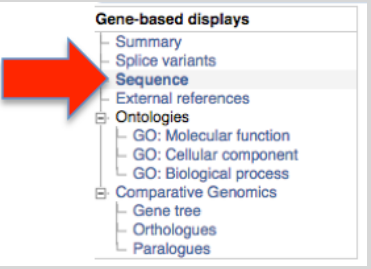
Gene pages


WormBase ParaSite
Search WormBase ParaSite...
e.g. O. volvulus, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase
Species List
BLAST
BioMart
REST API
Downloads
WormBase
Help and Documentation
Schistosoma mansoni (PRJEA36577)
Location: Smp.Chr_4:31,987,438-31,988,870
Gene: SAT1
Transcript: Smp_090120.1
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
Configure this page
Add 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), 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)
Protein coding genes
Smp_171360.1 > protein coding
Contigs
Smp.Chr_4:1333 >
Smp.Chr_4:1335 >
< Smp_090120.1 protein coding
< Smp_090120.2 protein coding
Gene Legend
Protein Coding
protein coding
Configuring the display
Tip: use the "Configure this page" link on the left to show additional data in this region.
Release 4 - November 2015
EMBL-EBI
sanger
empowered
BBSRC
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



Viewing sequence



Marked-up sequence

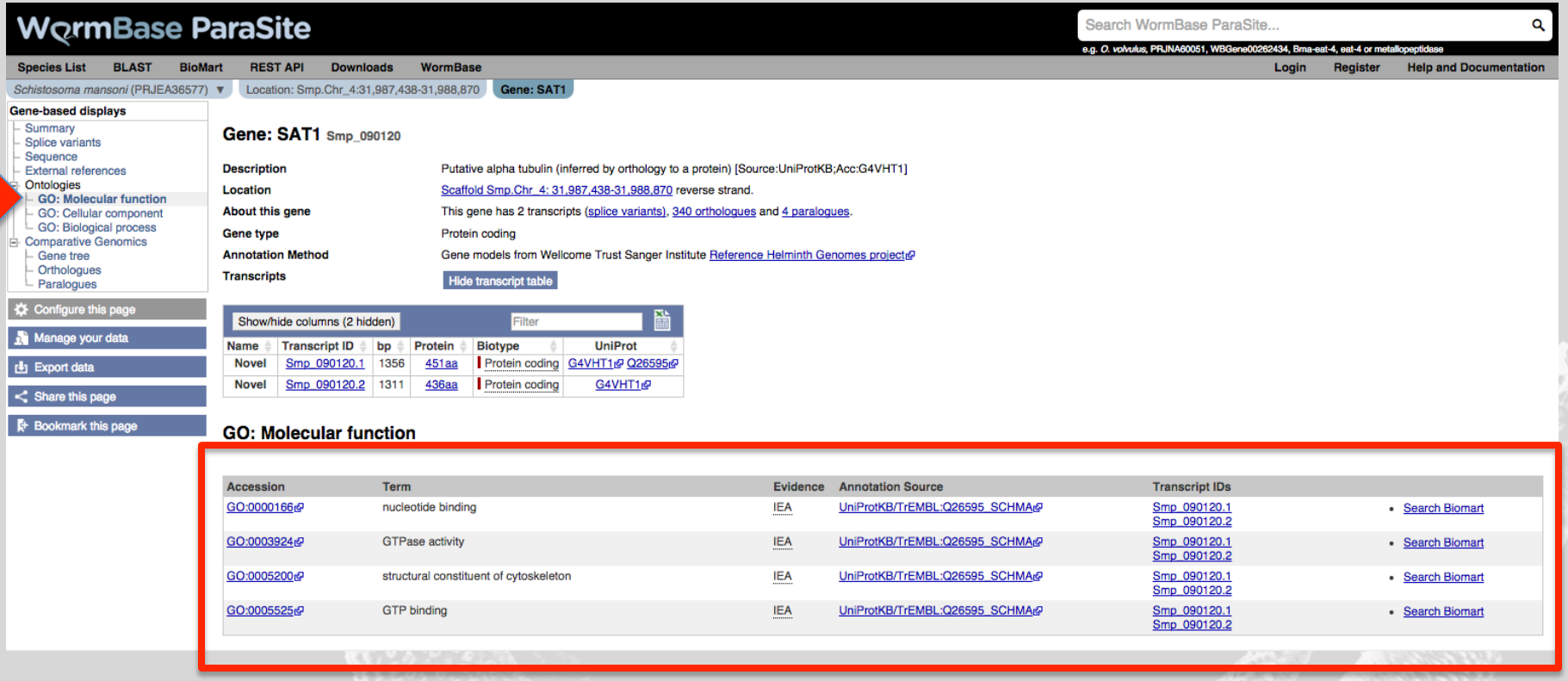
Download sequence | BLAST this sequence

Exons SAT1 exons All exons in this region

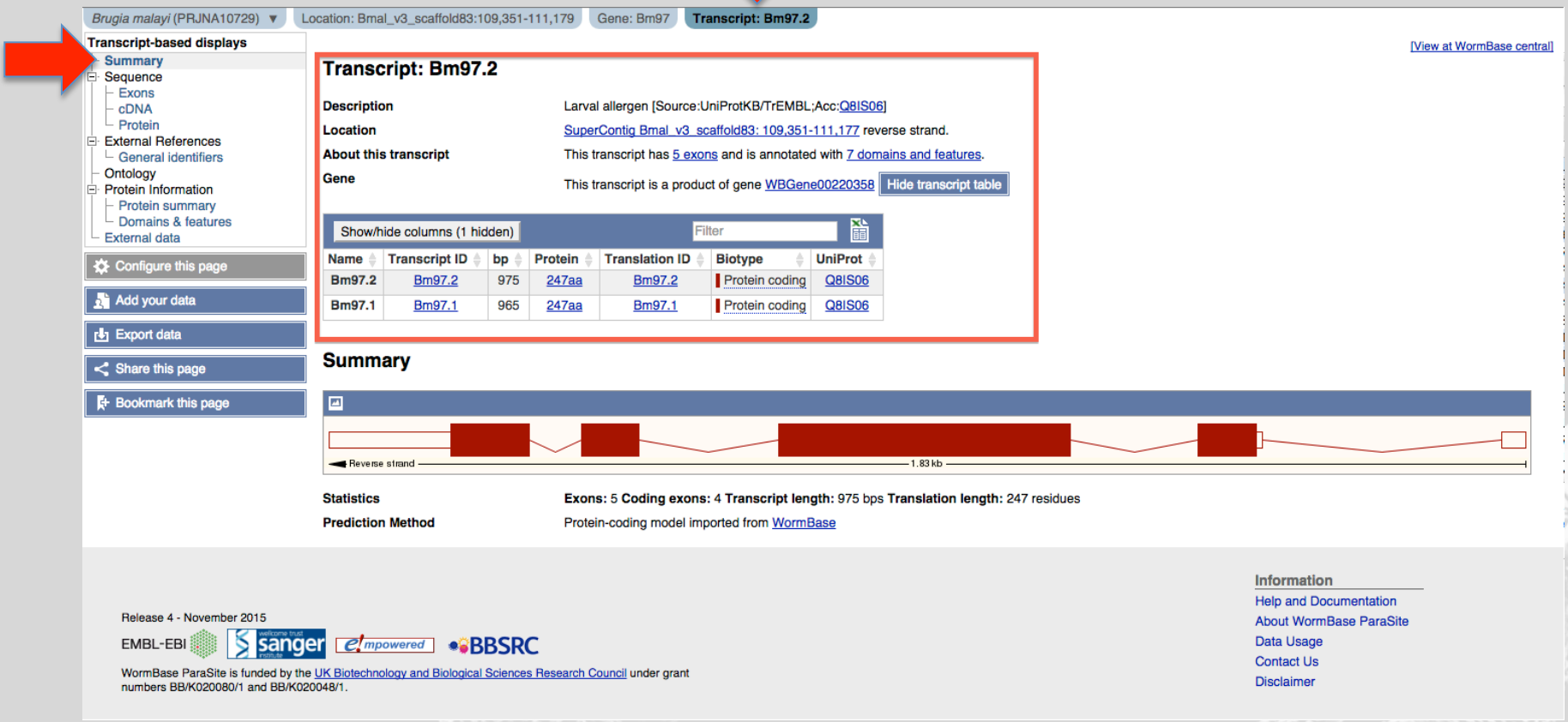
>scaffold:ASM23792v2:Smp.Chr_4:31986838:31989470:-1

GAGACTGCATCTCCTCACGATGATCCACTGCCTTATAGATCAGATCTTTGAGTCAAAGGC
TCCAGGTGTGGCCTCCTAAAAAAACTATCTGCTTCAGTTTGGTCACCTGGTCAGTATCAC
ATCCCTCACACAAATCAAATGAGATTTGTGTGGCGCATATGTATCTGATGCTTCTTTGTA
CCAATATTTATATGTTTAAATAATAATAATAATTTCAATTATTGTCTTCATCAGGTCATT
CATGAGAATATTATCTTTGGGTCTTTCATGAAATTTCAGATATGAAGTCTAAATACAGGC
TGAAATAAGGCGGTTTGGTTCCTTTTTTTTTTCACCATCCATTCAAATTGGCAACATAAA
TTTTTCCAATATATTCACTCCAGTTTGAAGTAAGAACTAATAGAATTCTCTAGAATGTAA
GACATAACATTAAATTAACCAATCAACGCAAAGTAATGGTAATGAATTAACCAATAGAAT
TTAAATAATAGATGTTCAATAAACTATATAATACATGAGGCGGCCATATTTGTTAGATTC
ATTTTAAATTTTCGTTGTGTTACGTGTGCTTCTGCTTCGTTATTTCTGAACTTTATCAAA
ATGGTAAGCAATTGAATAACATTTTAATTTATATTTTACAG**CGTGAATGTATCAGTGTAC**
ATGTTGGTCAAGCTGGTGTACAAATGGGAAATGCTTGTTGGGAGTTGTATTGTTTAGAAC
ATGGTATTCAACCTGATGGACAGATGCCAAGTGATAAAACTATTGGAGGTGGTGATGACT
CGTTCAATACATTCTTCAGTGAGACAGGAGCTGGAAAACATGTGCCACGAGCTGTTTTCG
TAGATTTAGAACCAACTGTAGTAGATGAAGTGCGAACTGGAACTTATCGTCAACTTTTCC
ATCCAGAACAATTAATAACTGGTAAGGAGGATGCTGCTAATAACTATGCTCGCGGTCATT
ATACAATTGGGAAAGAAATAGTGGACTTAGTTTTGGATCGTATTCGCAAGTTGGCTGACC
AGTGTACTGGACTTCAAGGGTTTCTGATTTTCCATTCATTTGGCGGTGGTACTGGTTCTG
GGTTCACATCTTTACTGATGGAACGTTTGAGTGTGGATTATGGCAAAAAATCGAAATTGG
AGTTTGCTGTGTATCCTGCACCACAGATTGCTACAGCTGTGGTTGAACCATATAATTCCA
TTTTGACTACACATACAACATTAGAACATTCTGACTGTGCTTTCATGGTTGACAACGAGG
CCATTTATGATATTTGTCGGTAAGTTGCTGTTTCAGTGAATTAAACAATGTTTTTTAG**AC**
GGAATCTAGATATAGAGCGTCCAACCTACACAAATTTAAATCGTCTAATTGGTCAGATAG
TGAGTTCCATTACTGCATCACTGCGCTTCGATGGTGCGCTAAATGTGGATTTGACGGAAT
TCCAAACTAATCTGGTACCTTATCCTCGTATCCACTTCCCGTTGGCTACCTATGCACCAG
TAATTTCAGCTGAGAAGGCTTACCACGAGCAGTTAAGCGTTGCAGAAATCACAAATGCAT
GCTTCGAACCAGCAAATCAGATGGTGAAATGTGACCCTCGTCACGGCAAATATATGGCGT
GCTGTATGTTATATCGTGGTGATGTTGTCCCCAAGGATGTCAACGCTGCCATAGCAACAA
TAAAAACTAAACGTACTATCCAGTTTGTGGATTGGTGTCCAACAGGTTTCAAGGTAGGTA
TAAACTACCAACCACCAACAGTTGTTCCTGGCGGTGATTTGGCAAAAGTTCAACGTGCTG
TCTGTATGTTGAGTAATACAACCGCAATAGCCGAAGCTTGGGCGCGTCTAGATCATAAAT
TTGATTTGATGTACGCAAAACGTGCGTTTGTACACTGGTATGTTGGTGAAGGTATGGAAG
AGGGTGAATTCTCTGAGGCTCGTGAAGATCTGGCCGCACTTGAGAAAGATTACGAGGAGG
TCGGTGTTGACAGCGTTGATGGCGAGGGCGAAGGGGAAGGTGAAGAGTATTAGGTTTCAT
TTCTTTTGAGTGTTATGTTTCTGGTTGTTACTCGTTCCTAATATACTTATCTTATAACTA
CTTGTGGTTTTTCCGTGTGCCATTGTGTCAATGCTGTGGAGTTTTCCCTGTGTAGCGTCT
TAACTGTTGATAAGACTAGCATCTAATATAGGCTTTTTGATTAAGGTAAGTTGCTTAAAA
CATTTCCACTGGGTGCCAACGCACTACTAAACGTGACTAGATCGTCTTGGAAATGAATTA

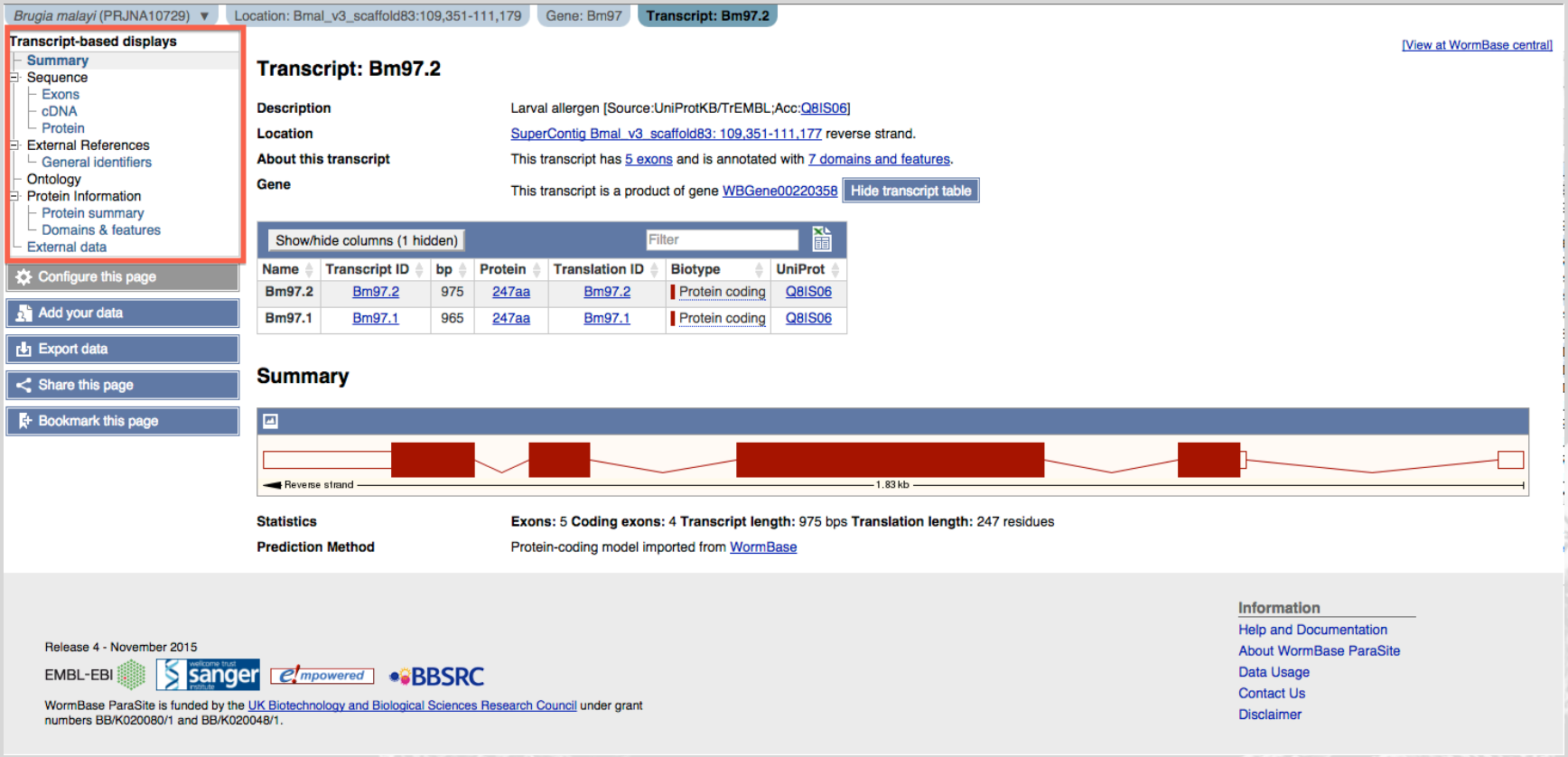
GO terms


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
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
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
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
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

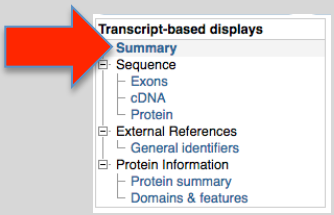
Transcripts



Transcripts

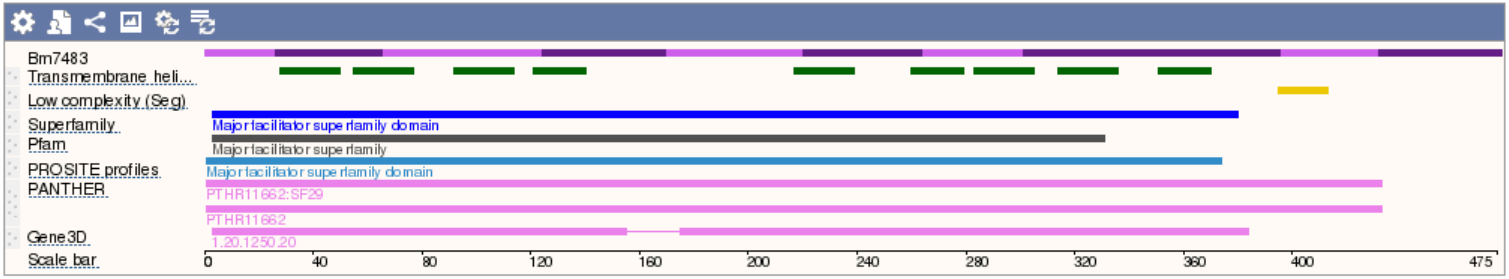


protein domains



Protein summary

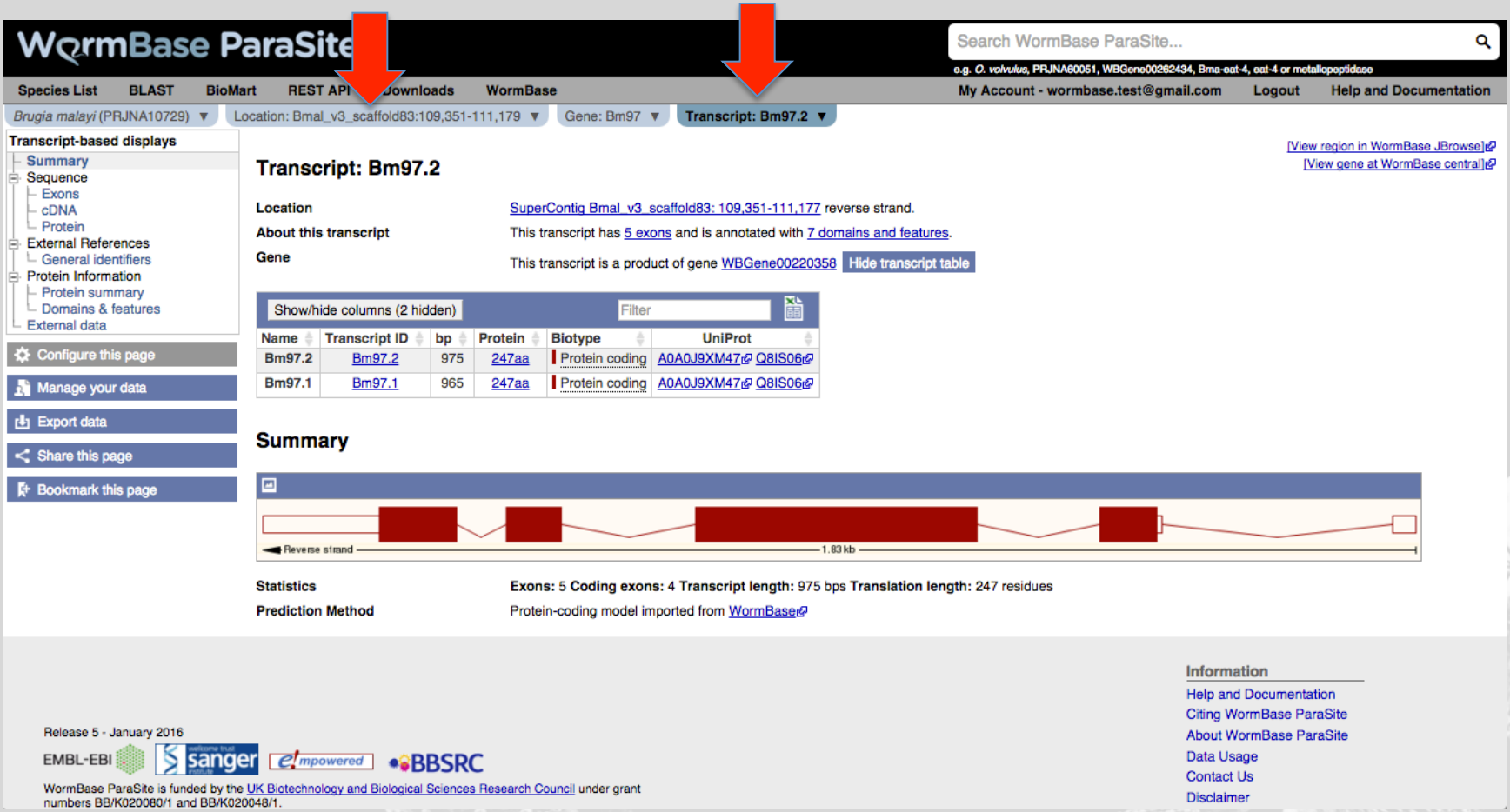
Protein domains for Bm7483.1



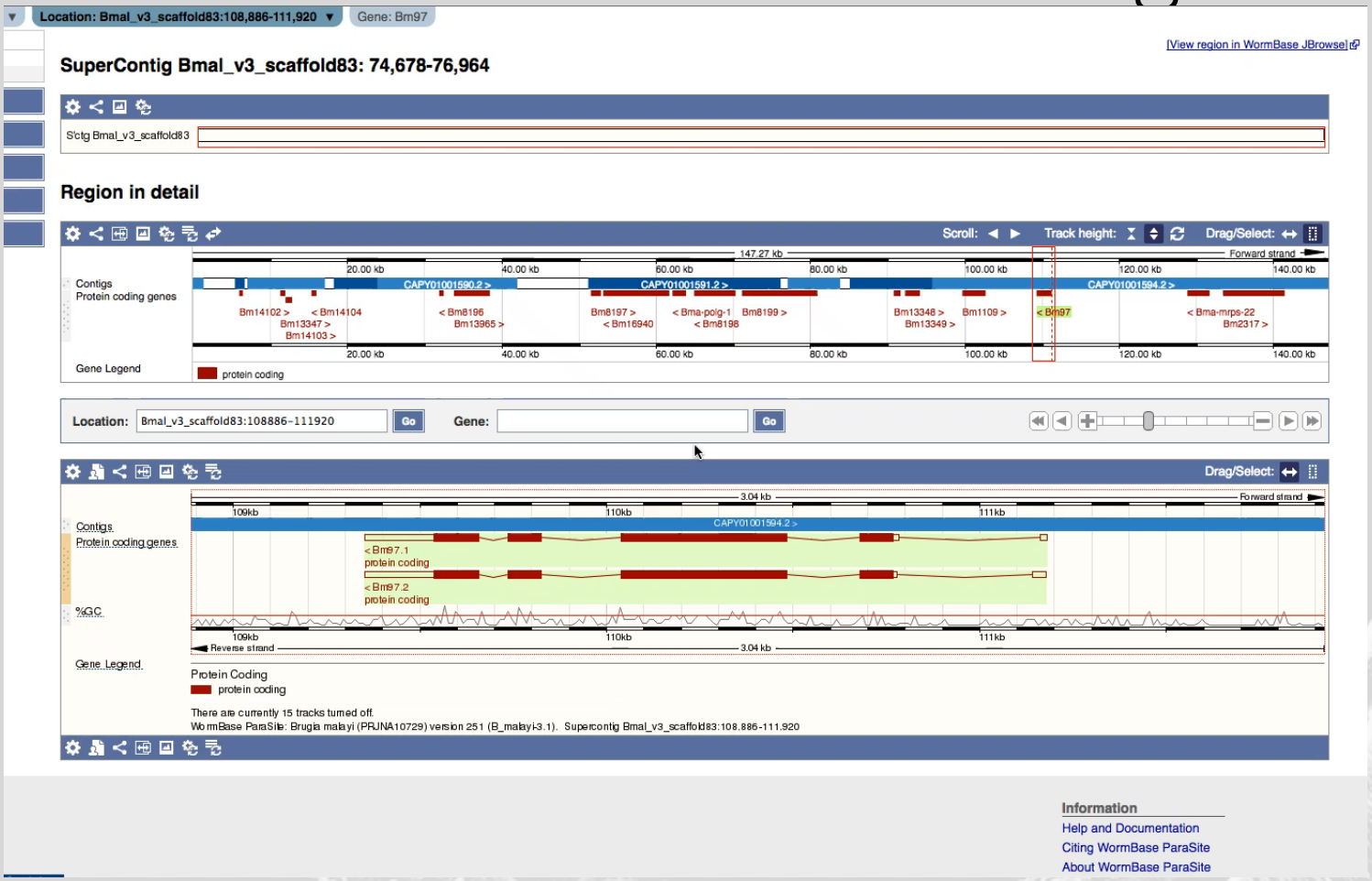
Statistics

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

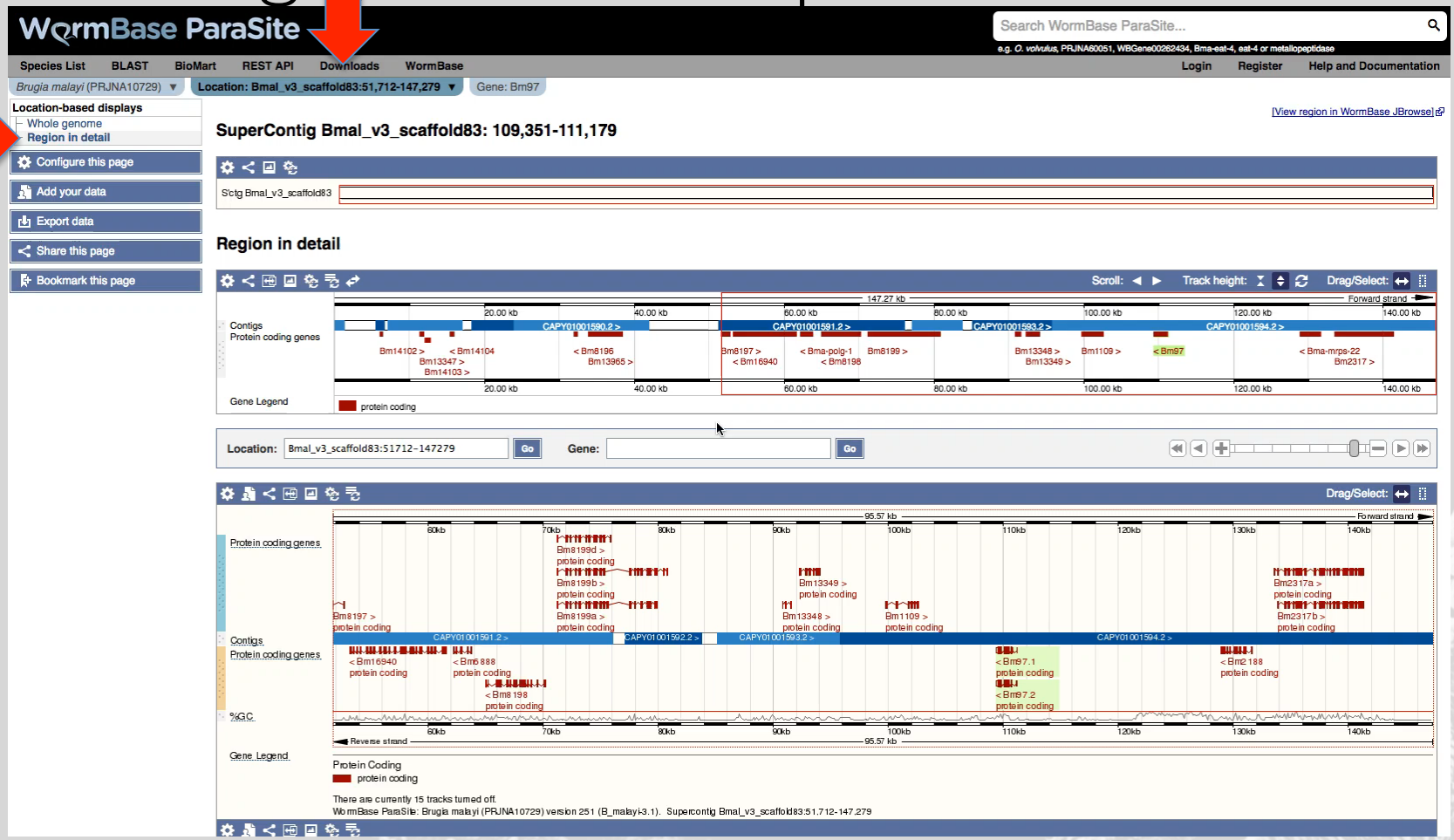
Location


WormBase ParaSite
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 - wormbase.test@gmail.com
Logout
Help and Documentation
Brugia malayi (PRJNA10729)
Location: Bmal_v3_scaffold83:109,351-111,179
Gene: Bm97
Transcript: Bm97.2
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
[View region in WormBase JBrowse]
[View gene at WormBase central]
Transcript: Bm97.2
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 (2 hidden)
Filter
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
sanger
BBSRC
WormBase ParaSite is funded by the UK Biotechnology and Biological Sciences Research Council under grant numbers BB/K020080/1 and BB/K020048/1.

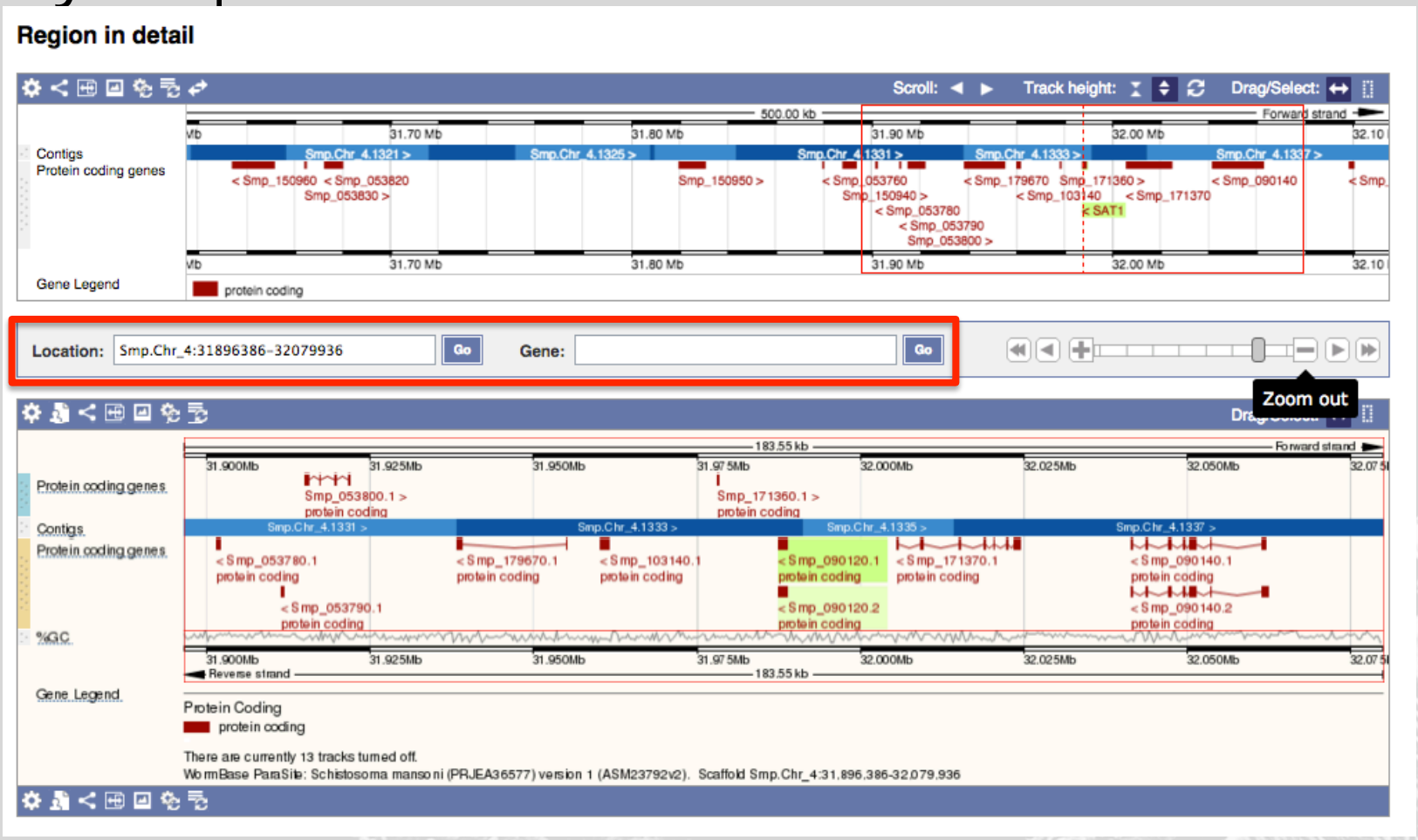
Location view: zooming



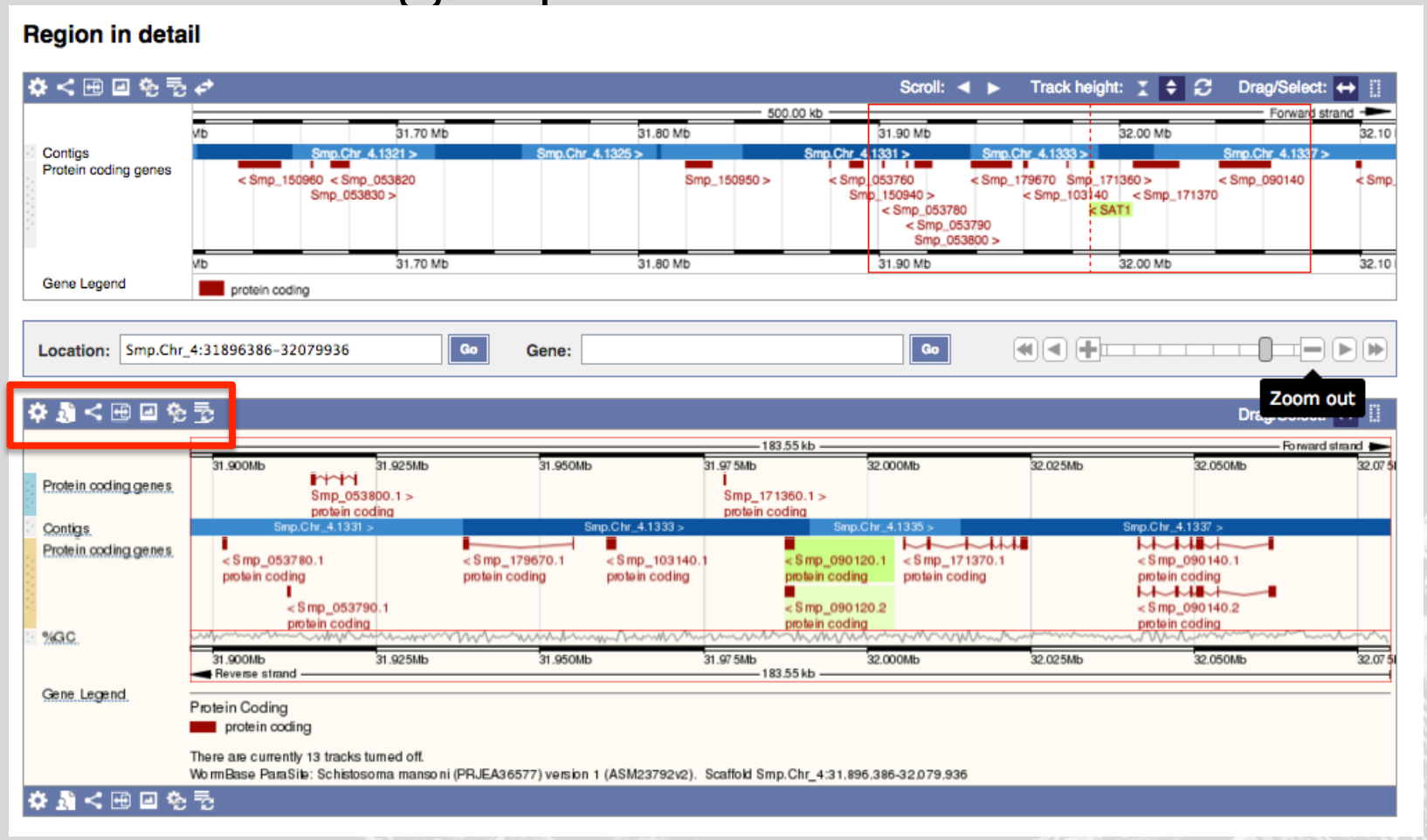
Viewing gene/transcript info in browser



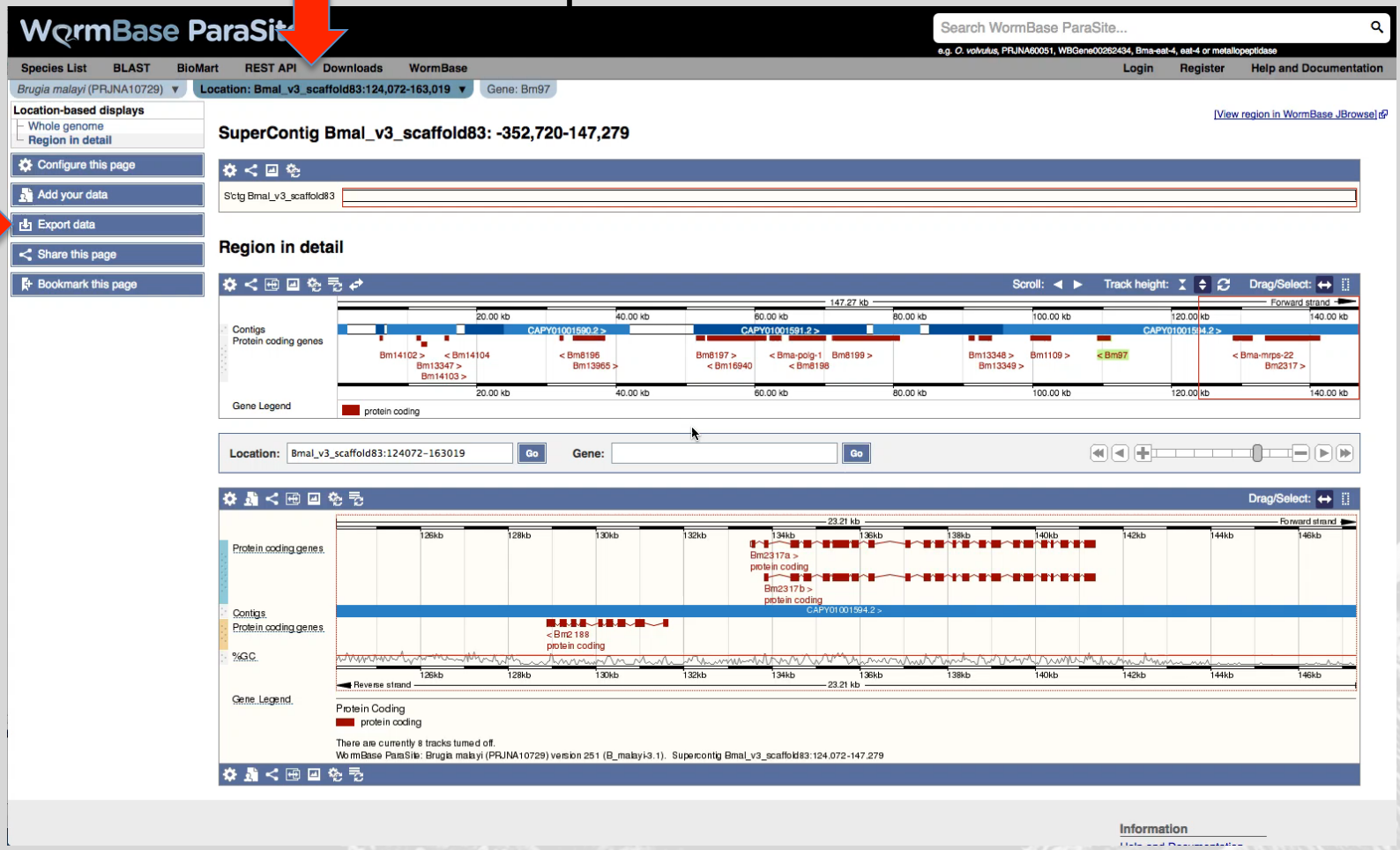
Jump to a chromosome location



Sharing, options for browser



Export data


WormBase ParaSite
Search WormBase ParaSite...
Species List
BLAST
BioMart
REST API
Downloads
WormBase
Login
Register
Help and Documentation
Brugia malayi (PRJNA10729)
Location: Bmal_v3_scaffold83:124,072-163,019
Gene: Bm97
Location-based displays
Whole genome
Region in detail
Configure this page
Add your data
Export data
Share this page
Bookmark this page
[View region in WormBase JBrowse]
SuperContig Bmal_v3_scaffold83: -352,720-147,279
S'ctg Bmal_v3_scaffold83
Region in detail
Scroll:
Track height:
Drag/Select:
Contigs
Protein coding genes
Gene Legend
protein coding
Location:
Bmal_v3_scaffold83:124072-163019
Go
Gene:
Go
Protein coding genes
Contigs
Protein coding genes
%GC
Gene Legend
Protein Coding
protein coding
There are currently 8 tracks turned off.
WormBase ParaSite: Brugia malayi (PRJNA10729) version 251 (B_malayi-3.1). Supercontig Bmal_v3_scaffold83:124,072-147,279
Information

4. Comparative Genomics

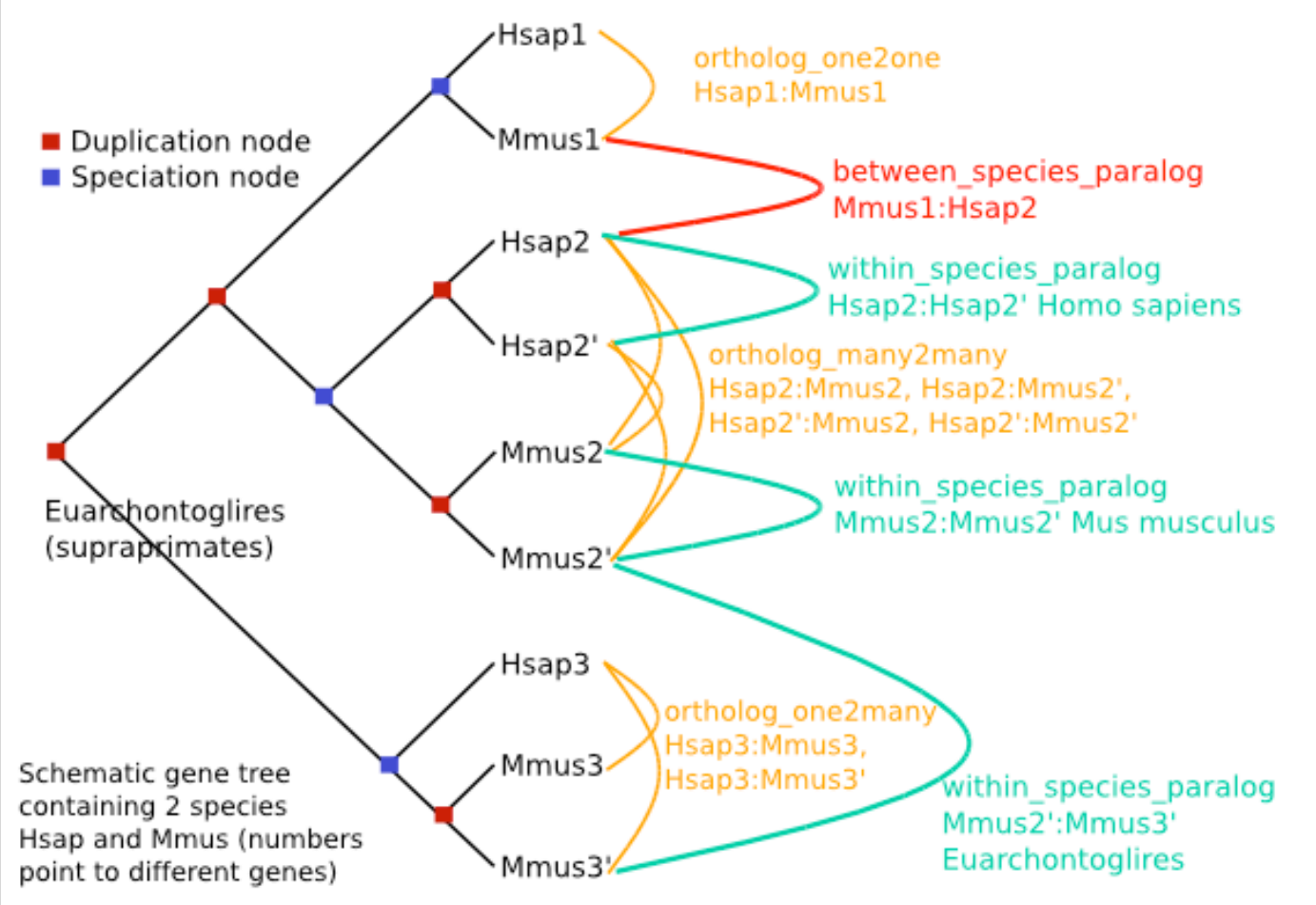
Introduction

- 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

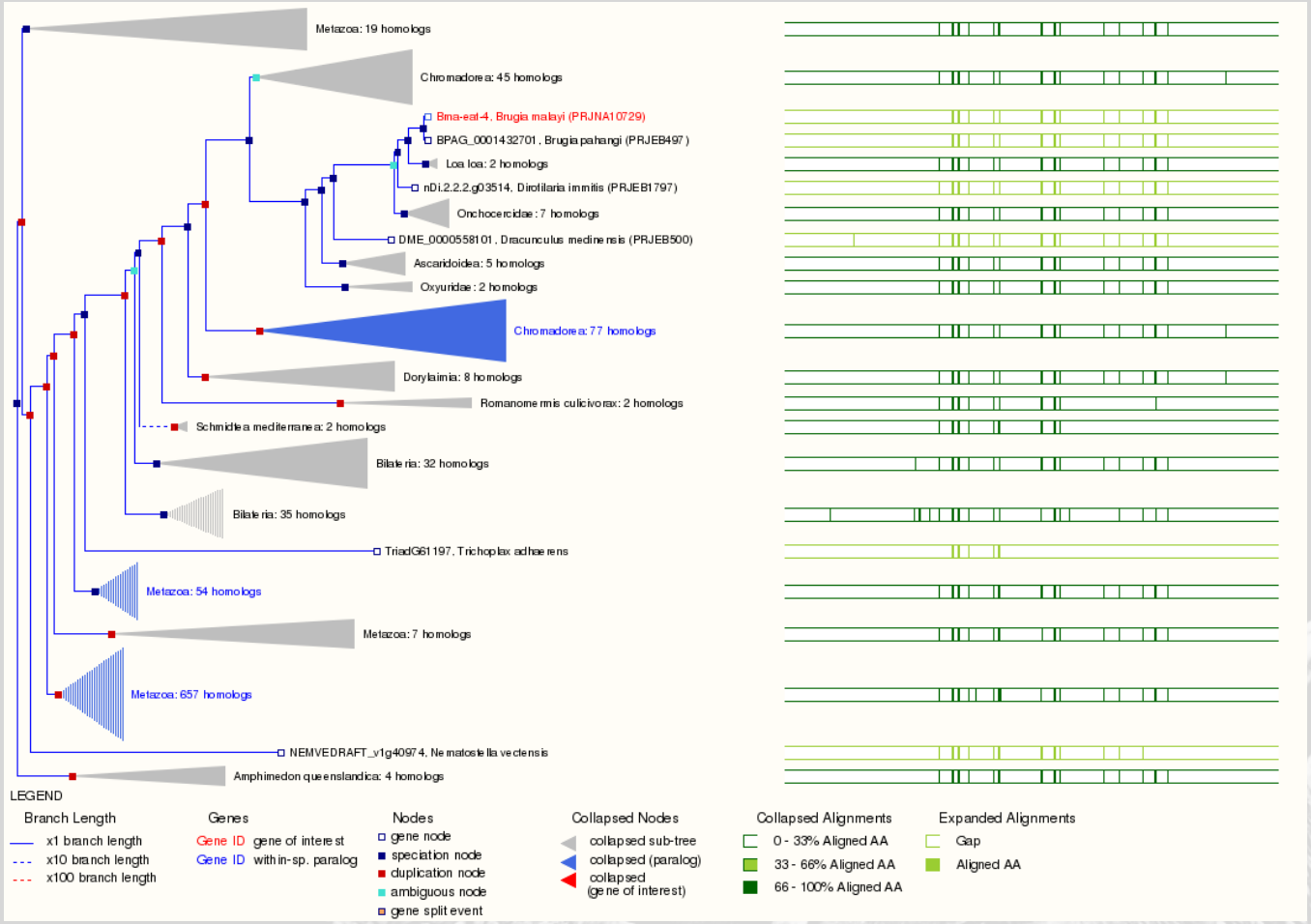
Homology types

- 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

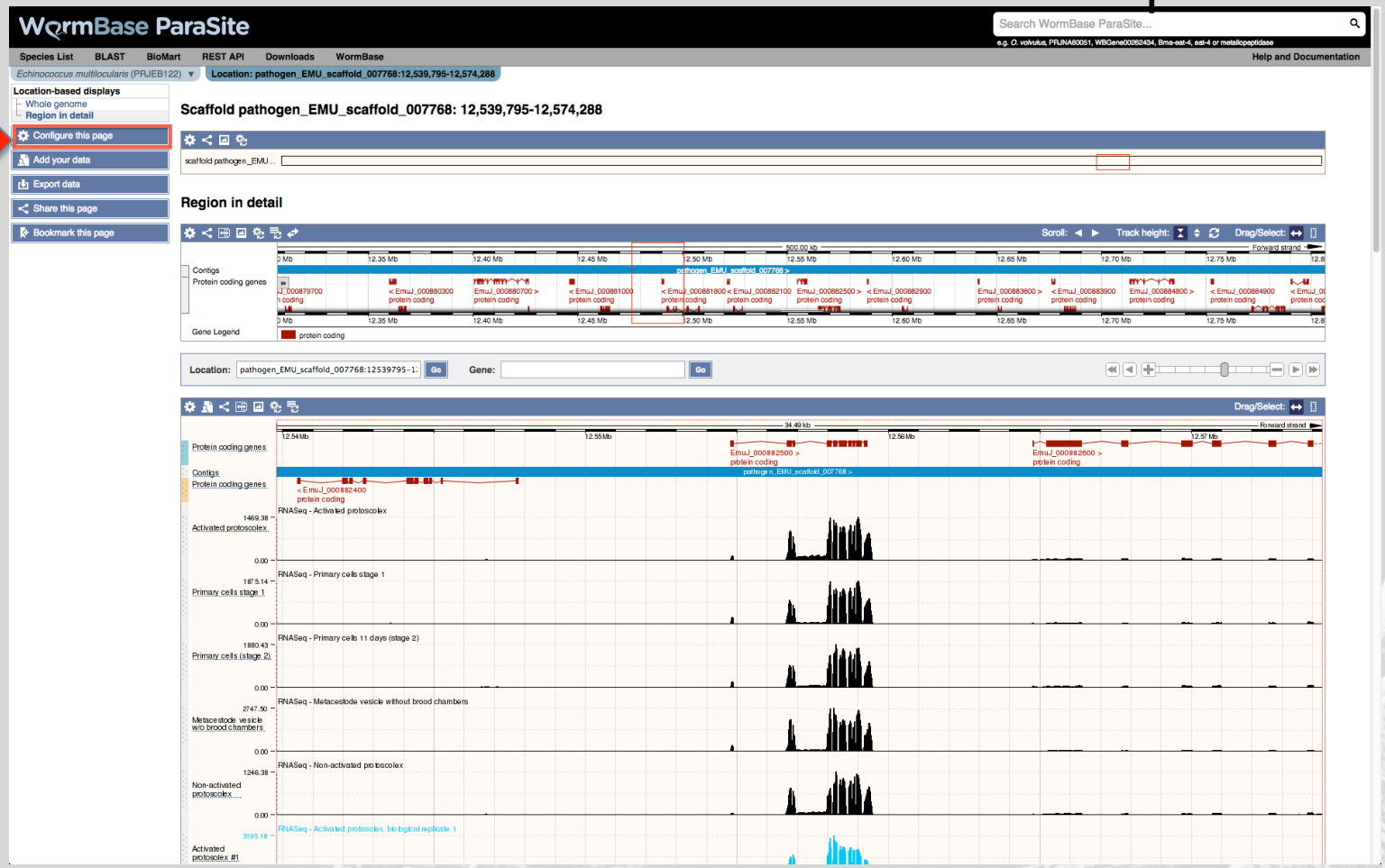
View protein alignments of all orthologues

Show All entries | Show/hide columns | Filter

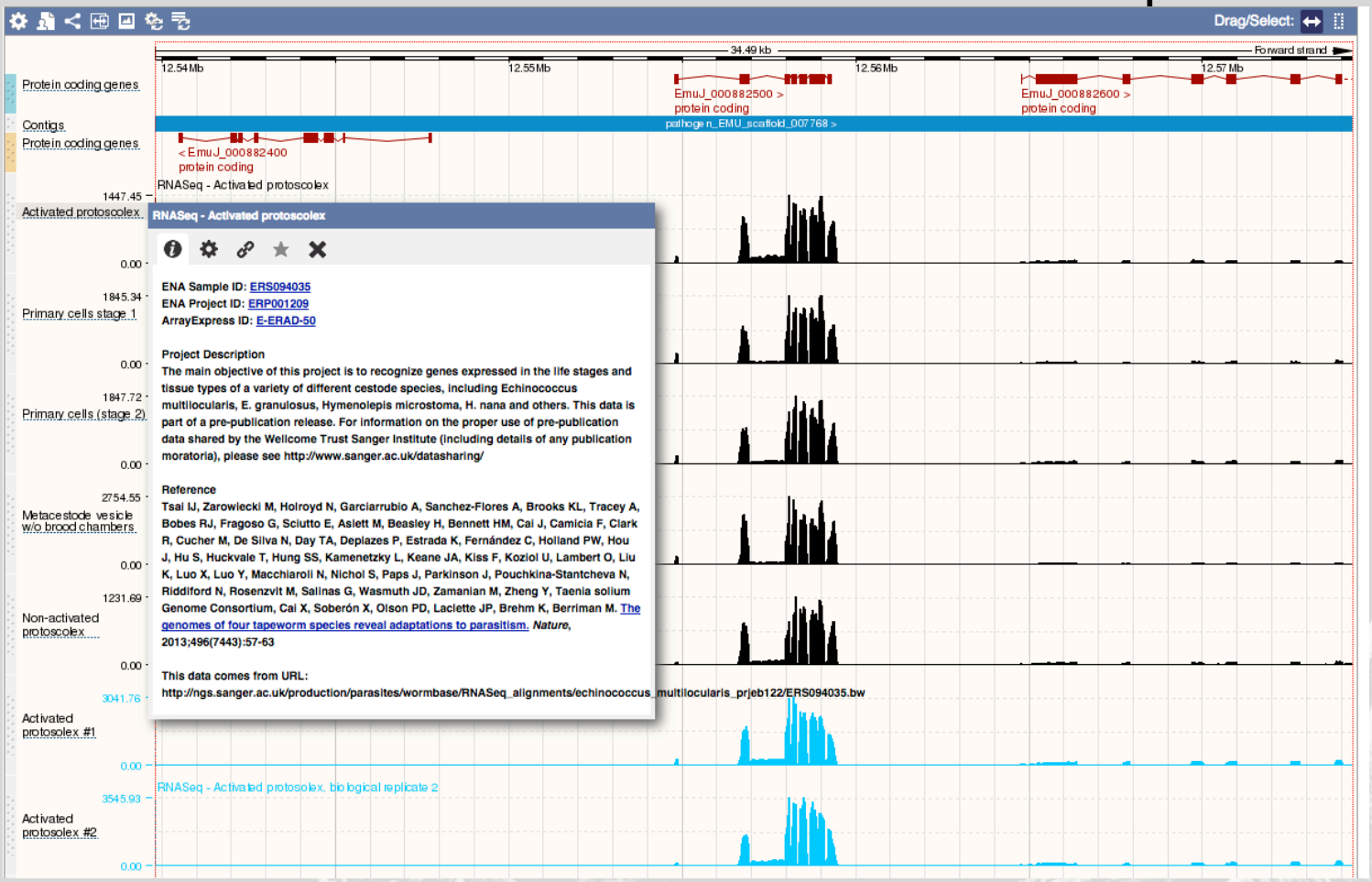
Species	Type	dN/dS	Stable ID & gene name	Compare	Location	Target %id	Query %id
Acanthocheilonema viteae (PRJEB4306)	1-to-1	n/a	nAv.1.0.1.g03135 Probable vesicular glutamate transporter eat-4 (inferred by orthology to a C. elegans protein) [Source: UniProtKB; acc: P34644]	• Alignment (protein) • Alignment (cDNA) • Gene Tree (image)	nAv.1.0.scaf00057:63087-67685:1	73	76
Amphimedon queenslandica	Many-to-many	n/a	Aqu1.218710 No description	• Alignment (protein) • Alignment (cDNA) • Gene Tree (image)	Contig13347:93913-96460:1	27	28
Amphimedon queenslandica	Many-to-many	n/a	Aqu1.218711 No description	• Alignment (protein) • Alignment (cDNA) • Gene Tree (image)	Contig13347:96728-99289:-1	26	29
Amphimedon queenslandica	Many-to-many	n/a	Aqu1.219143 No description	• Alignment (protein) • Alignment (cDNA) • Gene Tree (image)	Contig13358:105162-108191:1	28	23
Amphimedon queenslandica	Many-to-many	n/a	Aqu1.219145 No description	• 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 C. elegans protein) [Source: UniProtKB; acc: P34644]	• 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 Caenorhabditis elegans ortholog eat-4)[Source: UniProtKB/Swiss-Prot; acc: P34644]	• Alignment (protein) • Alignment (cDNA) • Gene Tree (image)	Acey_s0107_scaf:399963-412768:-1	56	67

5. RNASeq tracks

Data tracks - RNASeq

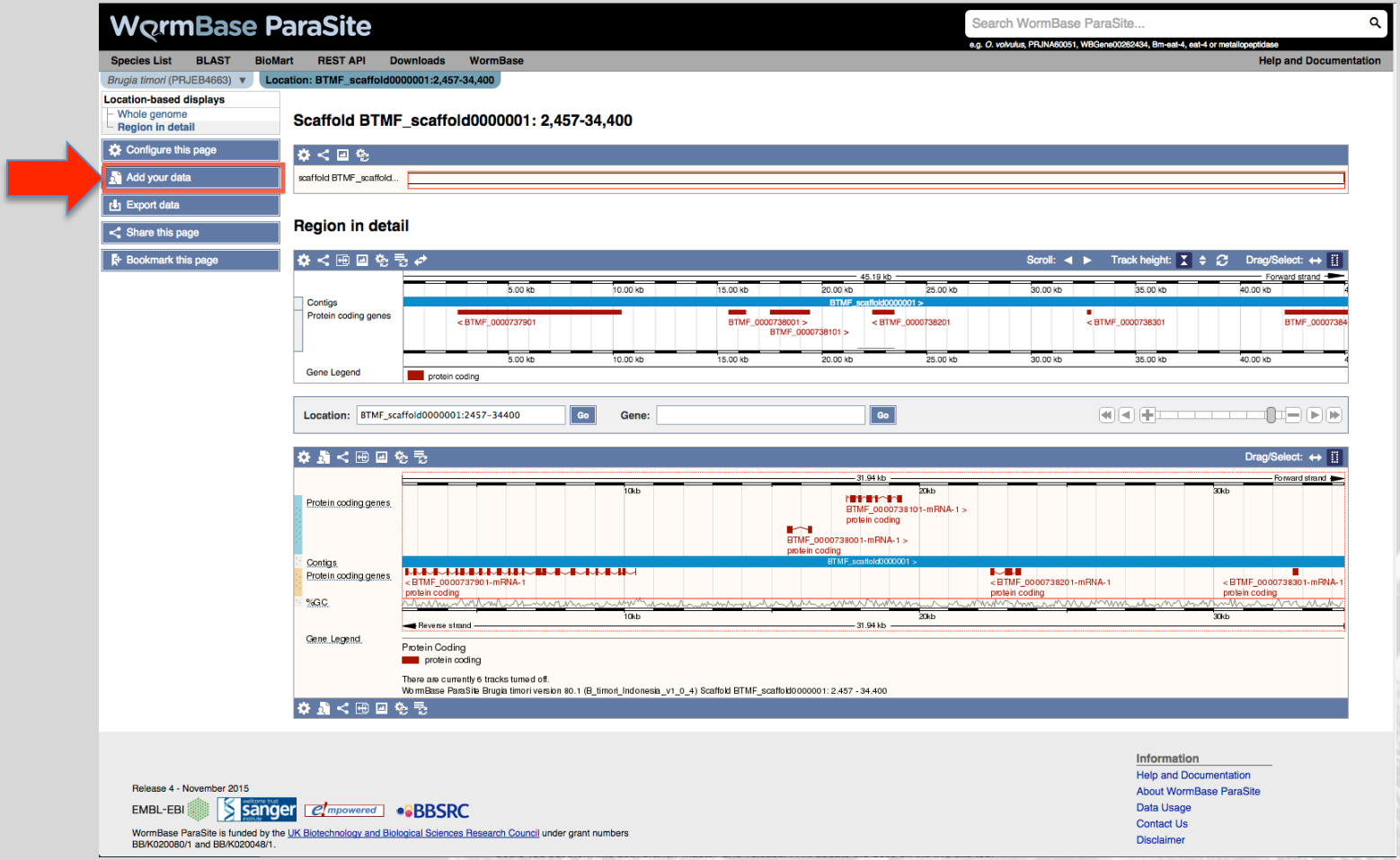


Data tracks - RNASeq

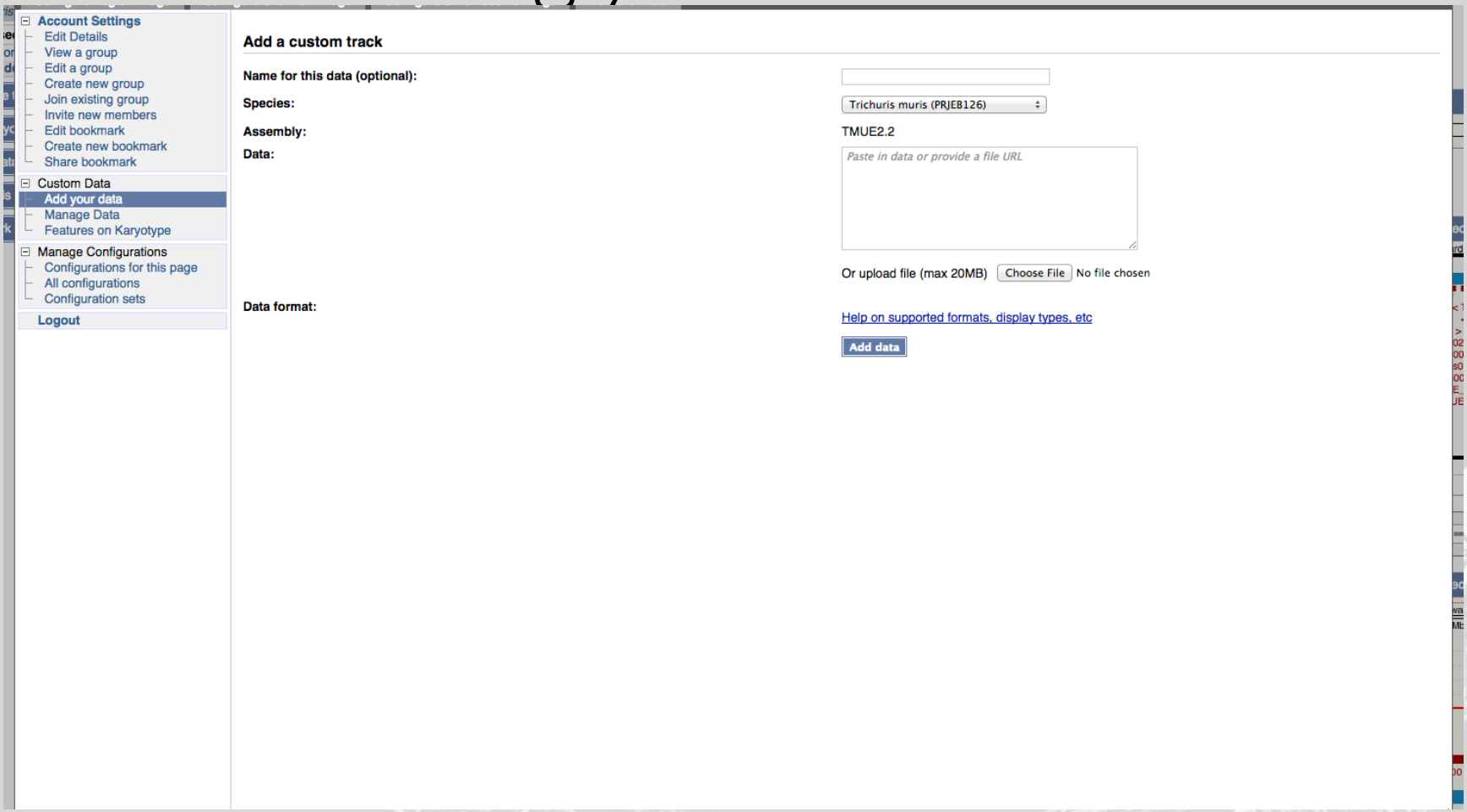


6. Adding your own data

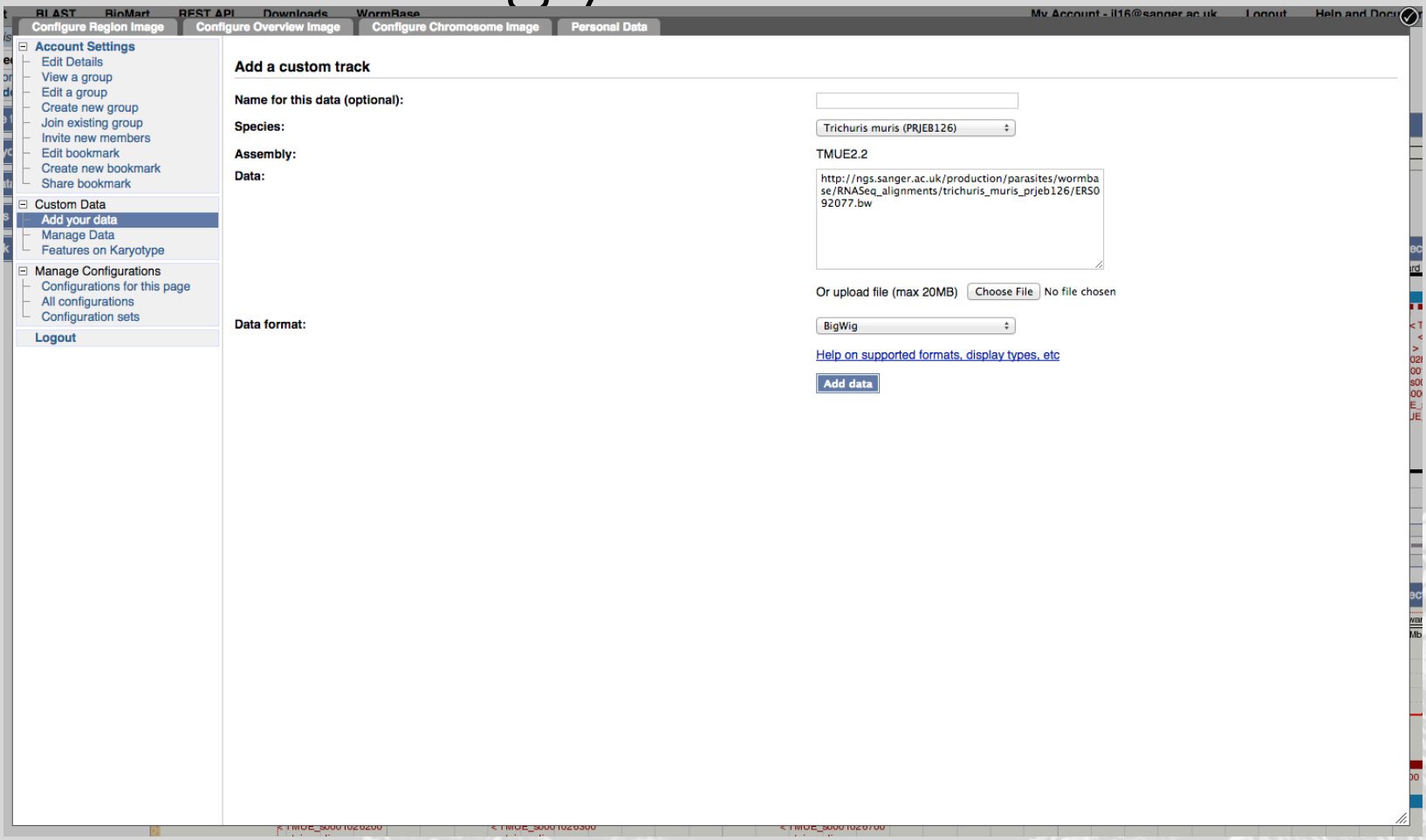
Adding your own data



Adding your own data



Adding your own data

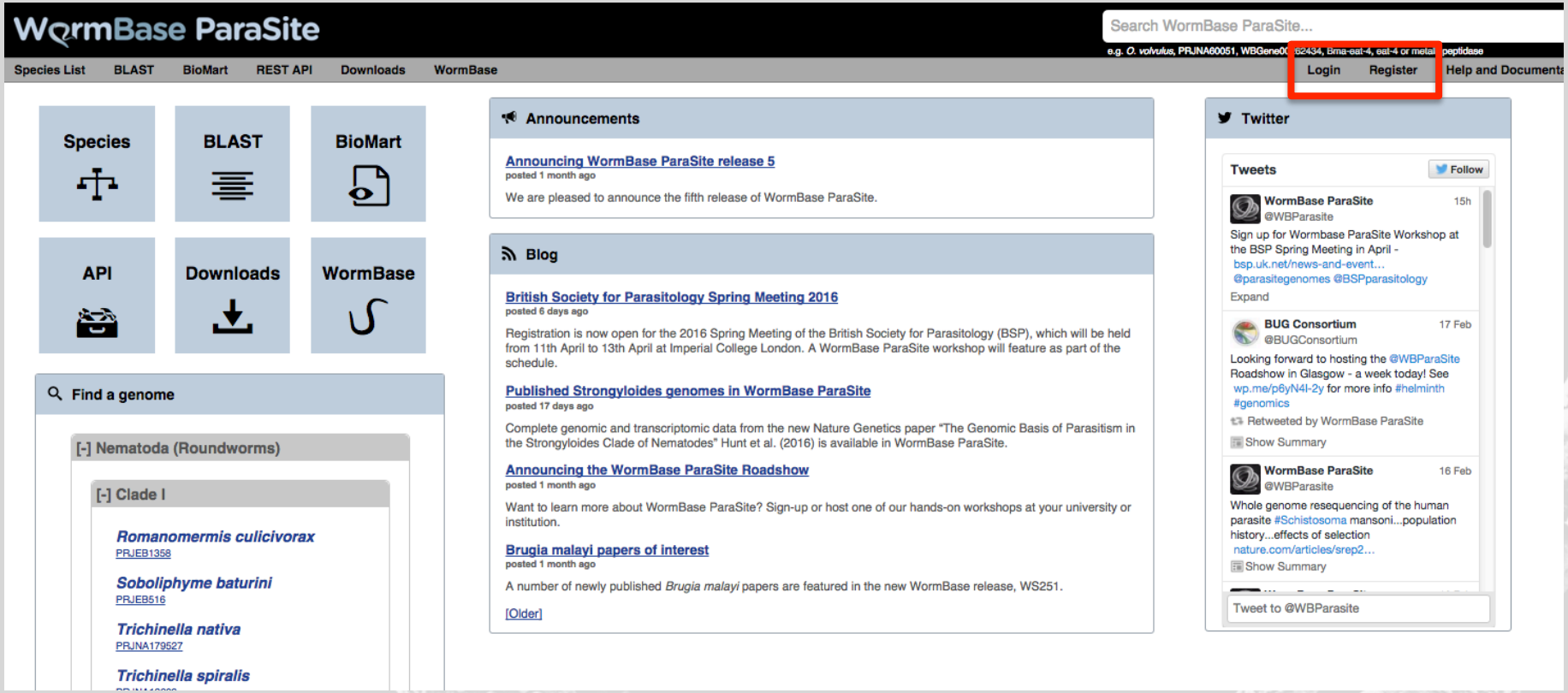


7. User Accounts

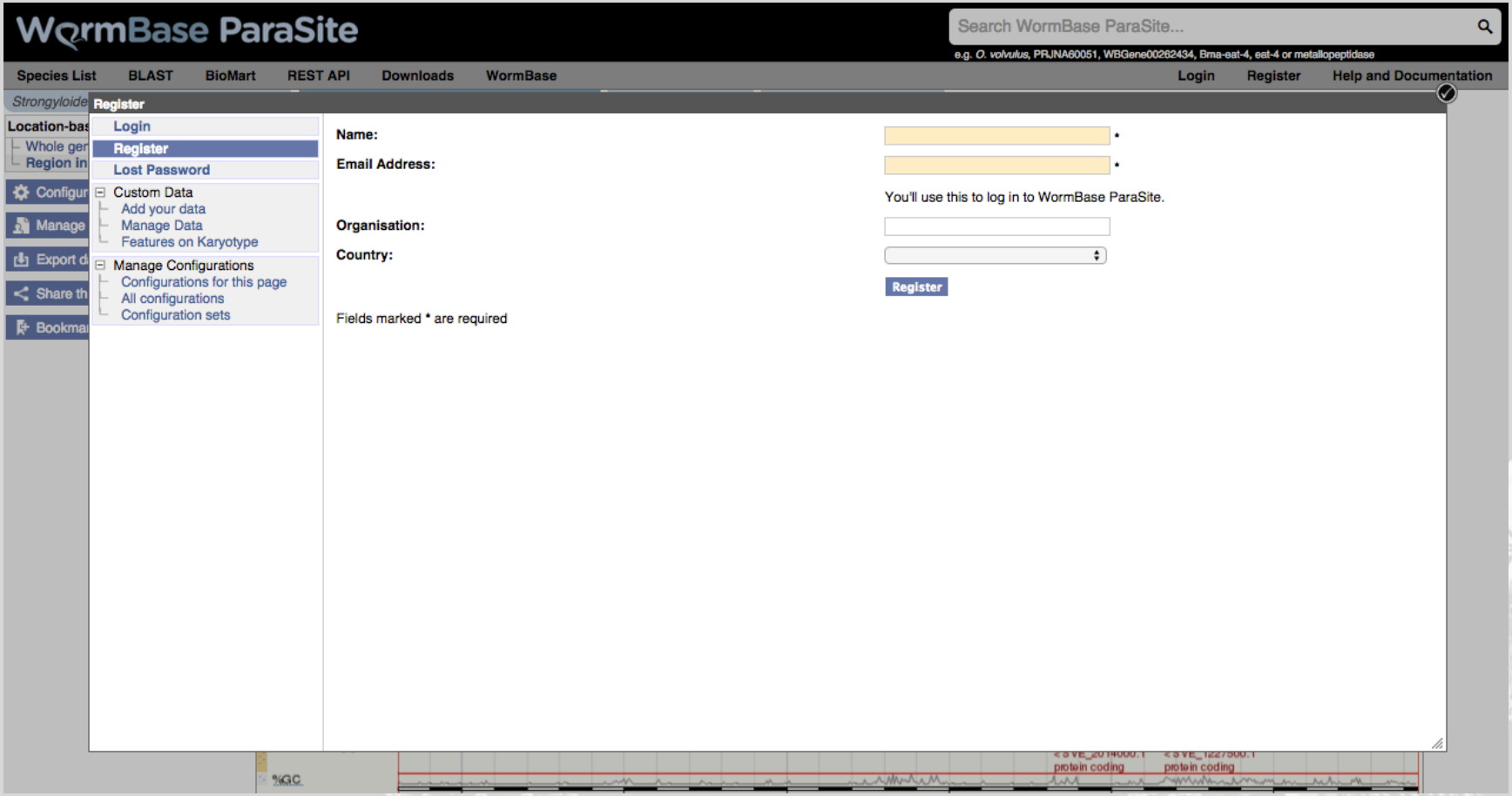
User accounts

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

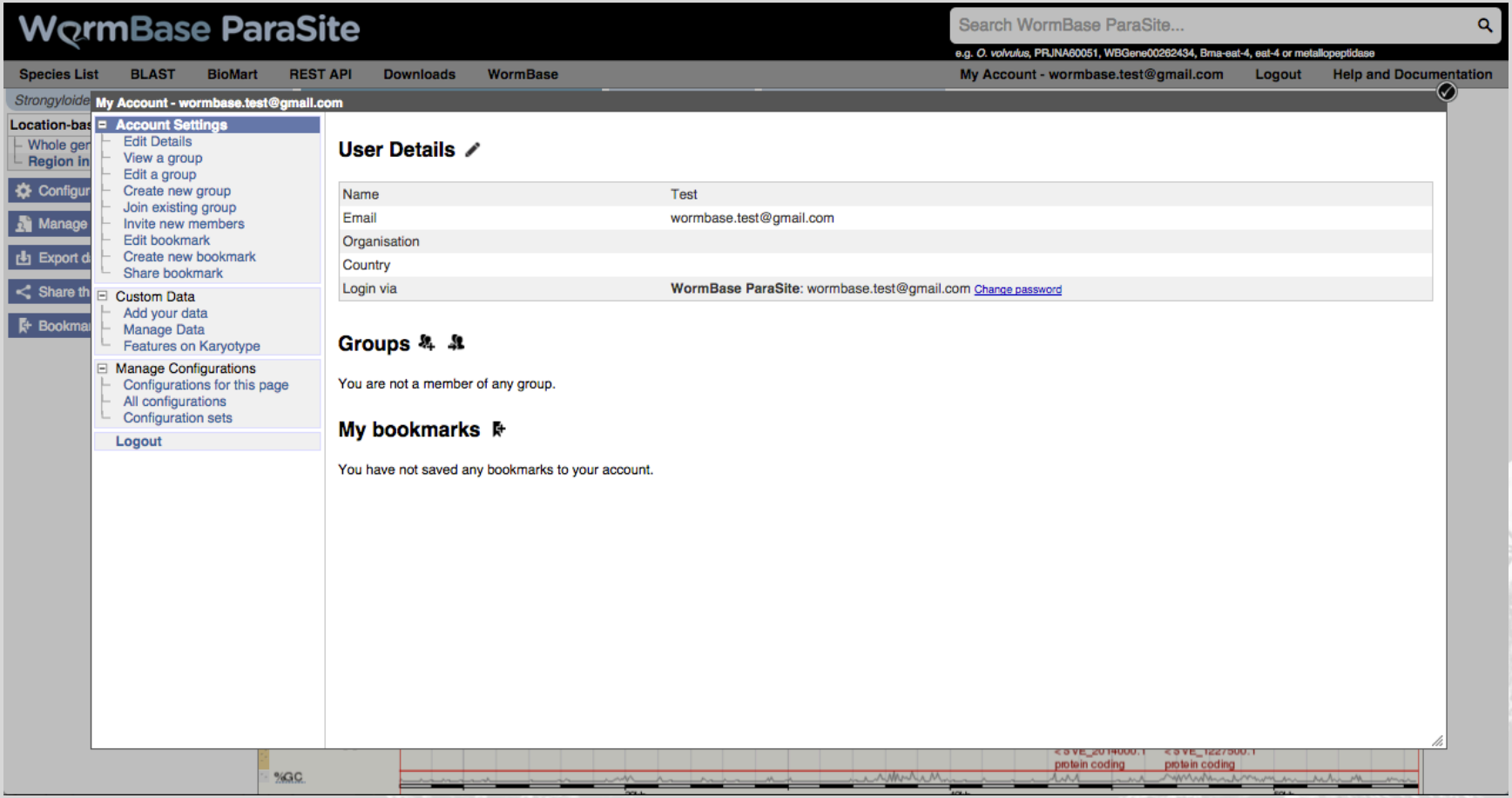
User accounts



User accounts: registering



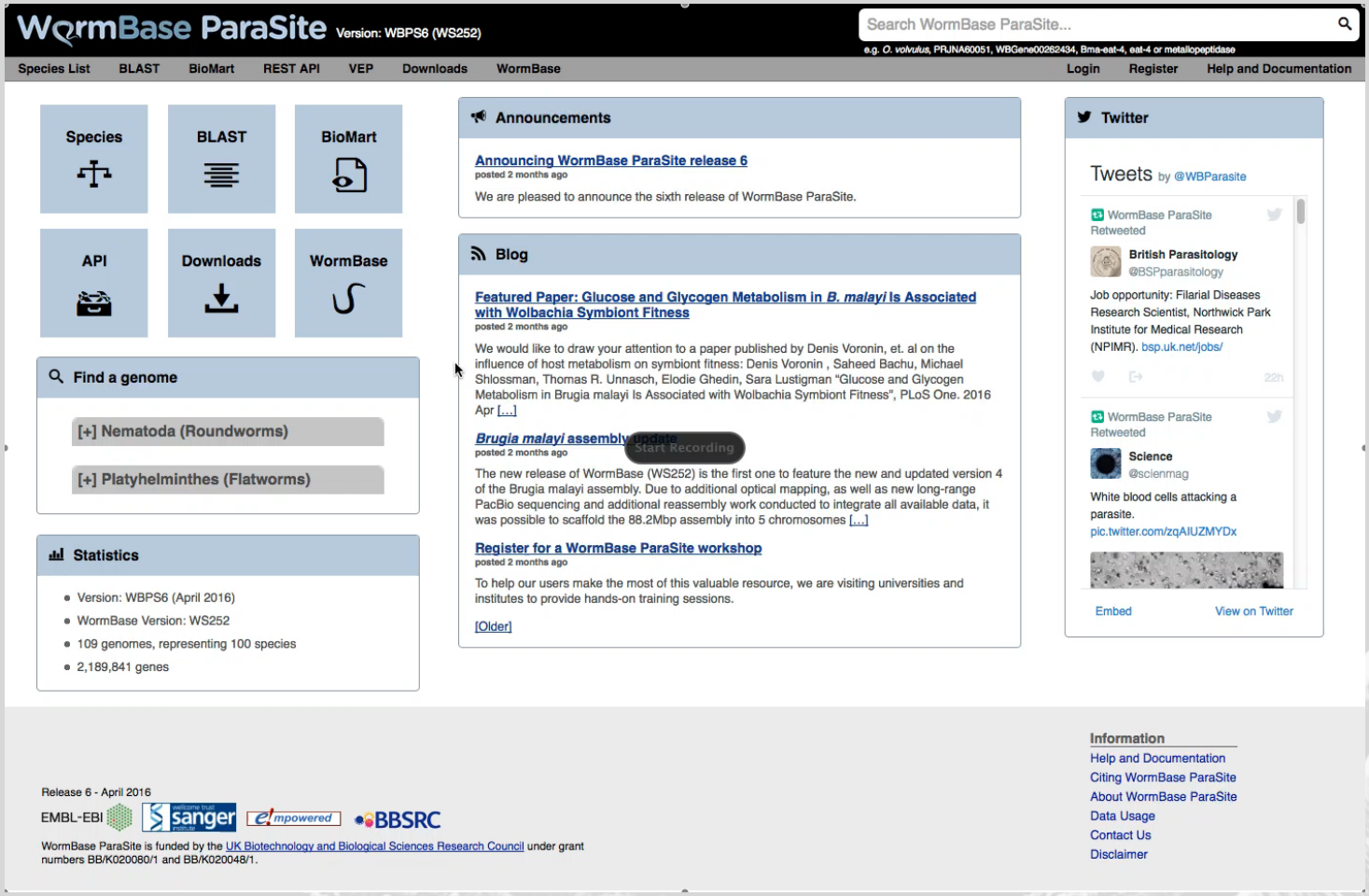
User accounts



Exercises

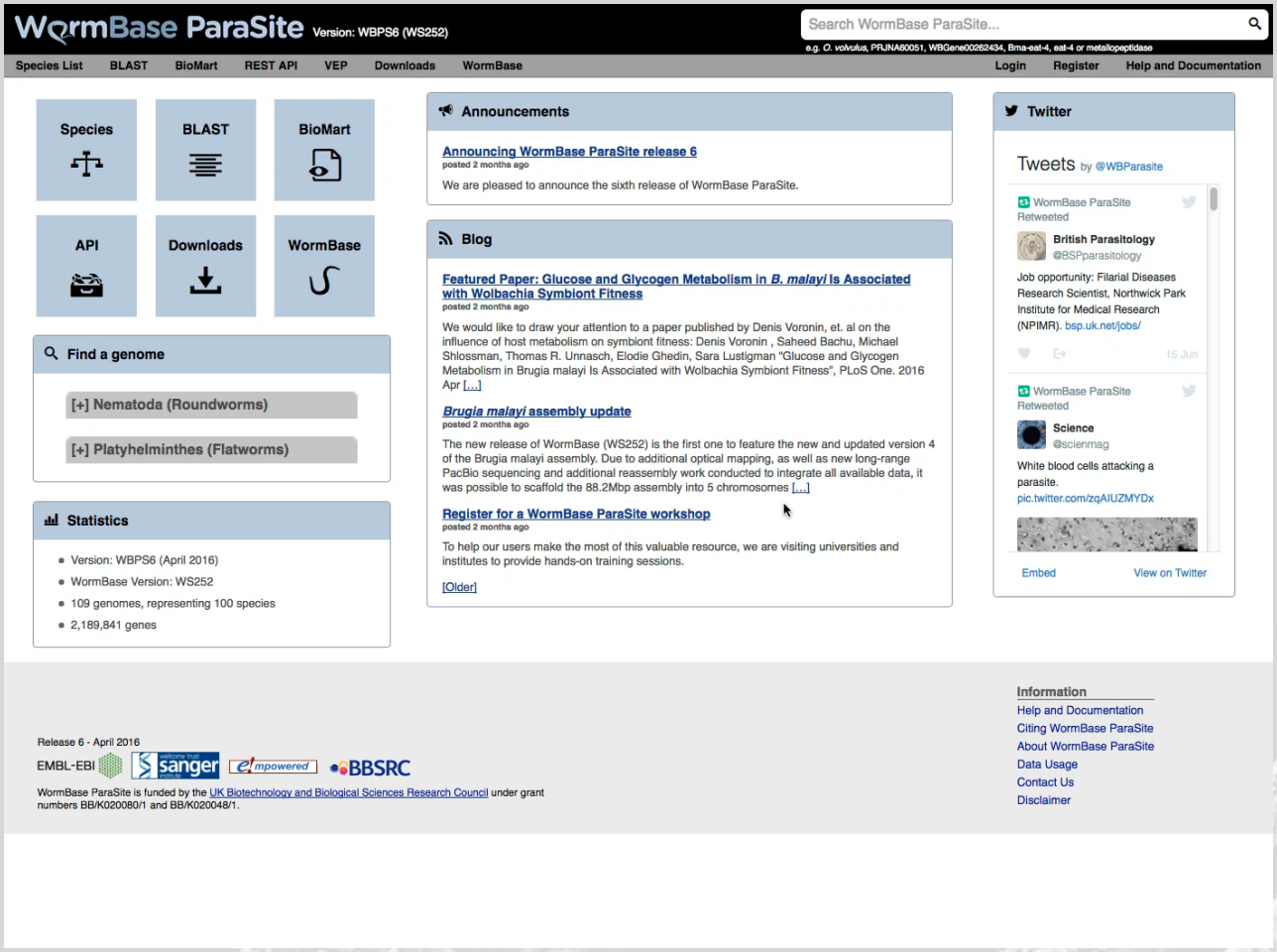
Answers to exercises

1. 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)
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 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
Start Recording
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/
22h
WormBase ParaSite Retweeted
Science
@scienmag
White blood cells attacking a parasite.
pic.twitter.com/zqAIUZMYDx
Embed
View on Twitter
Information
Help and Documentation
Citing WormBase ParaSite
About WormBase ParaSite
Data Usage
Contact Us
Disclaimer
Release 6 - April 2016
EMBL-EBI
sanger
empowered
BBSRC
WormBase ParaSite is funded by the UK Biotechnology and Biological Sciences Research Council under grant numbers BB/K020080/1 and BB/K020048/1.

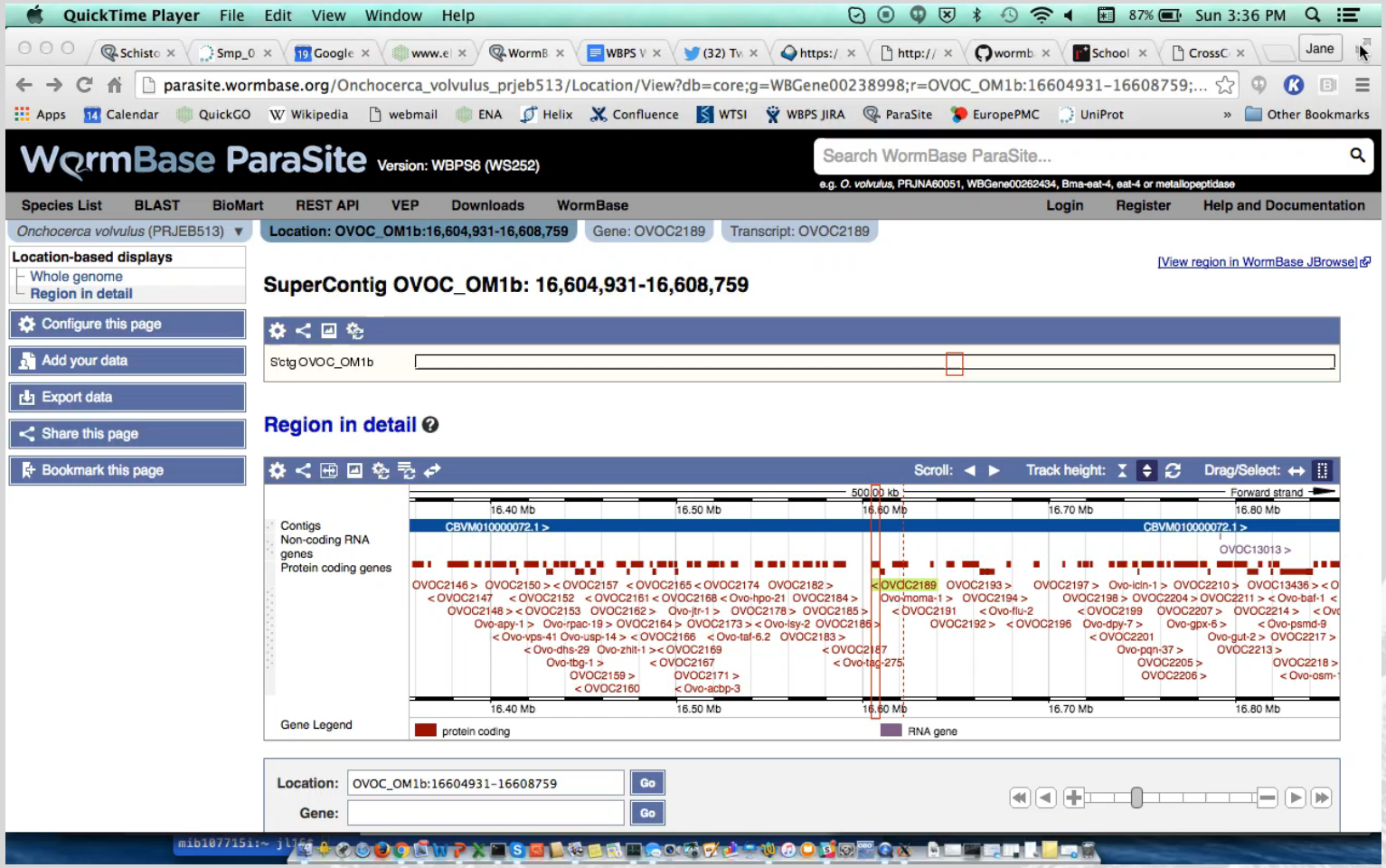
Answers to exercises

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_OM1b: 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)
Search WormBase ParaSite...
e.g. O. volvulus, PRJNA80051, 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 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
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
EMBL-EBI
sanger
empowered
BBSRC
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.

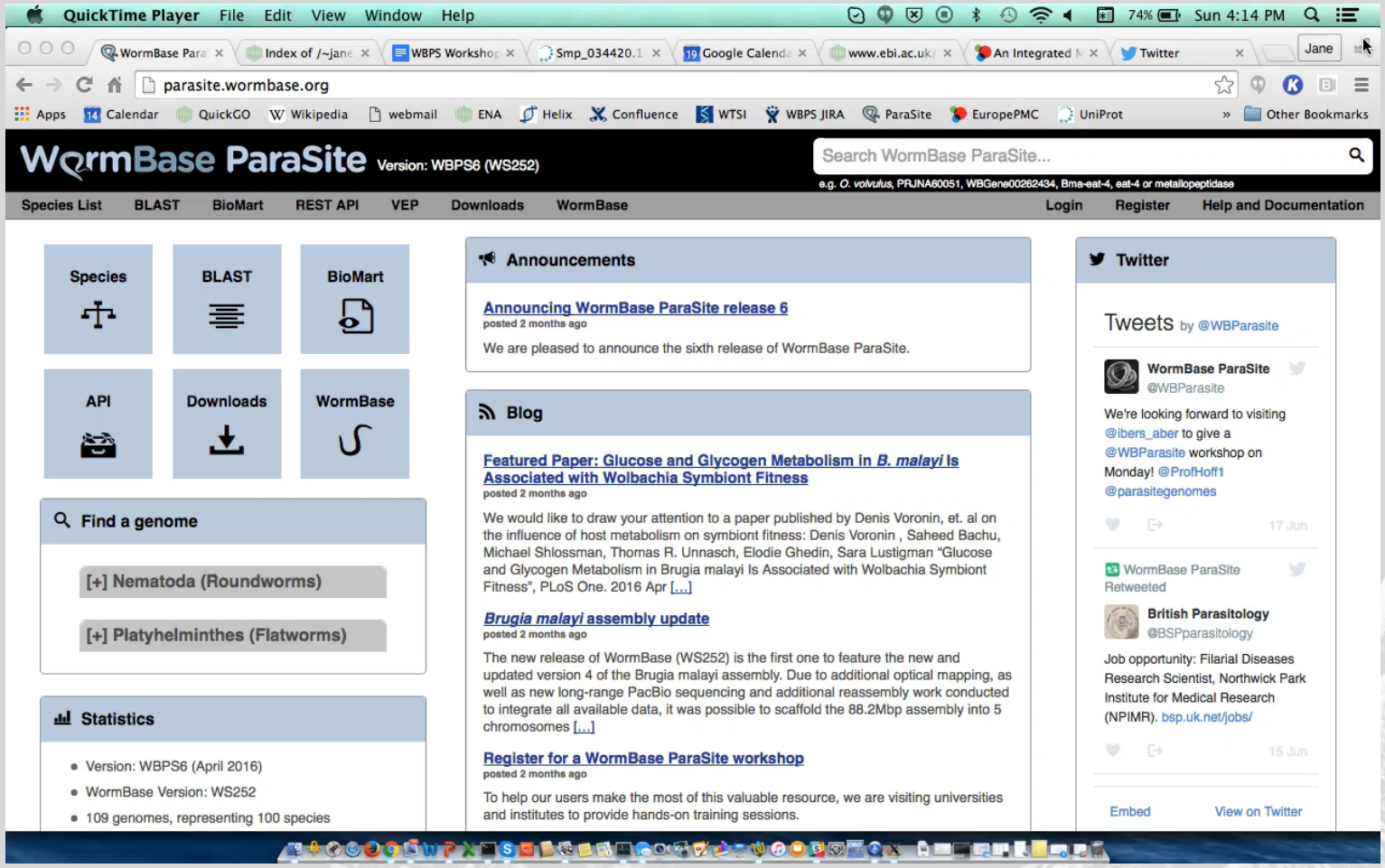

QuickTime Player File Edit View Window Help
Sun 3:36 PM
parasite.wormbase.org/Onchocerca_volvulus_prjeb513/Location/View?db=core;g=WBGene00238998;r=OVOC_OM1b:16604931-16608759;...
WormBase ParaSite Version: WBPS6 (WS252)
Search WormBase ParaSite...
e.g. O. volvulus, PRJNA60051, WBGene00262434, Bma-eat-4, 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
Location-based displays
Whole genome
Region in detail
Configure this page
Add your data
Export data
Share this page
Bookmark this page
[View region in WormBase JBrowse]
SuperContig OVOC_OM1b: 16,604,931-16,608,759
S'ctg OVOC_OM1b
Region in detail
Scroll:
Track height:
Drag/Select:
500.00 kb
Forward strand
16.40 Mb
16.50 Mb
16.60 Mb
16.70 Mb
16.80 Mb
Contigs
Non-coding RNA genes
Protein coding genes
CBVM010000072.1 >
OVOC13013 >
OVOC2189
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:

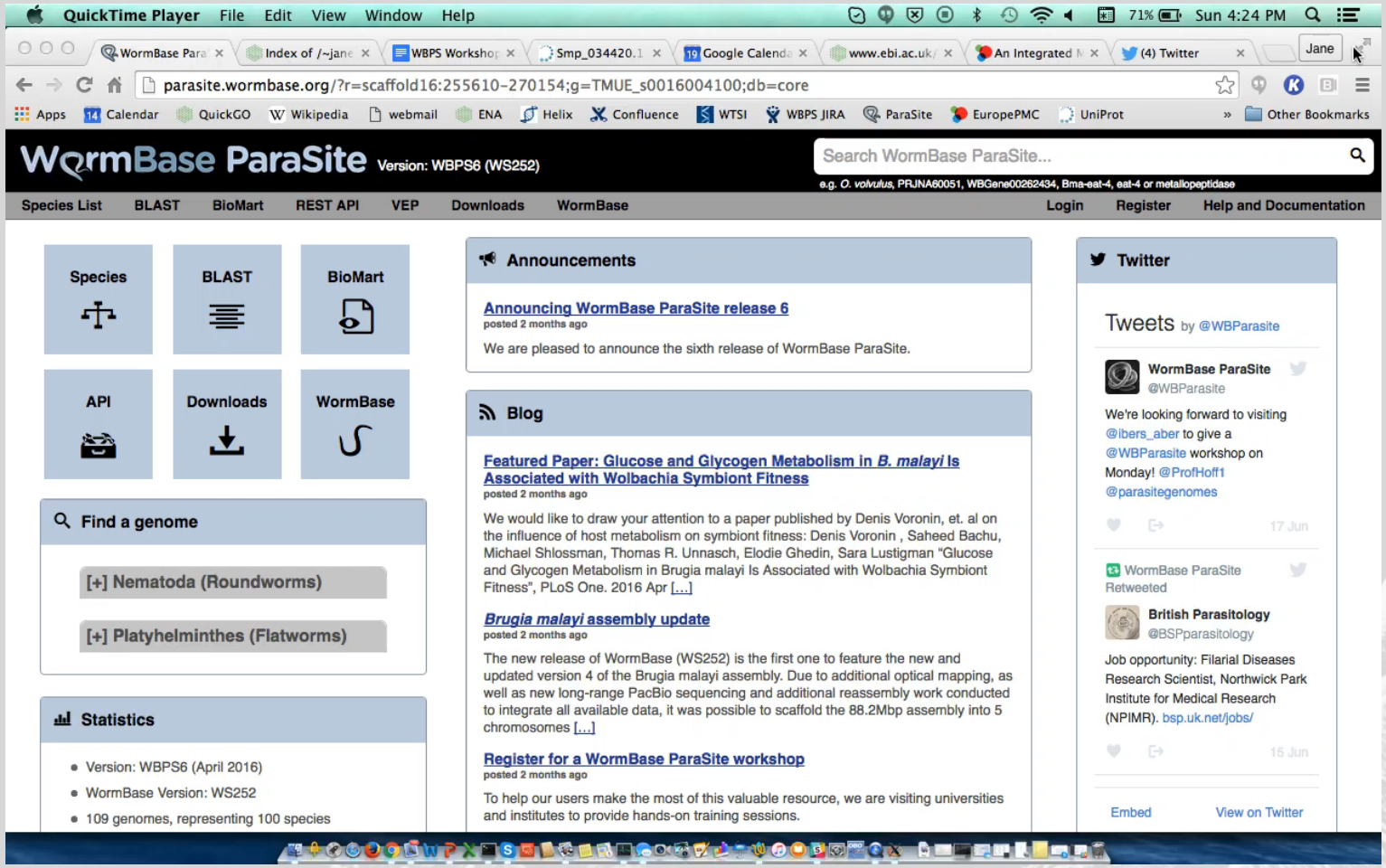
     `http://www.ebi.ac.uk/~jane/testdata/`

     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
Sun 4:14 PM
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
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
17 Jun
WormBase ParaSite Retweeted
British Parasitology
@BSParasitology
Job opportunity: Filarial Diseases Research Scientist, Northwick Park Institute for Medical Research (NPIMR). bsp.uk.net/jobs/
15 Jun
Embed
View on Twitter

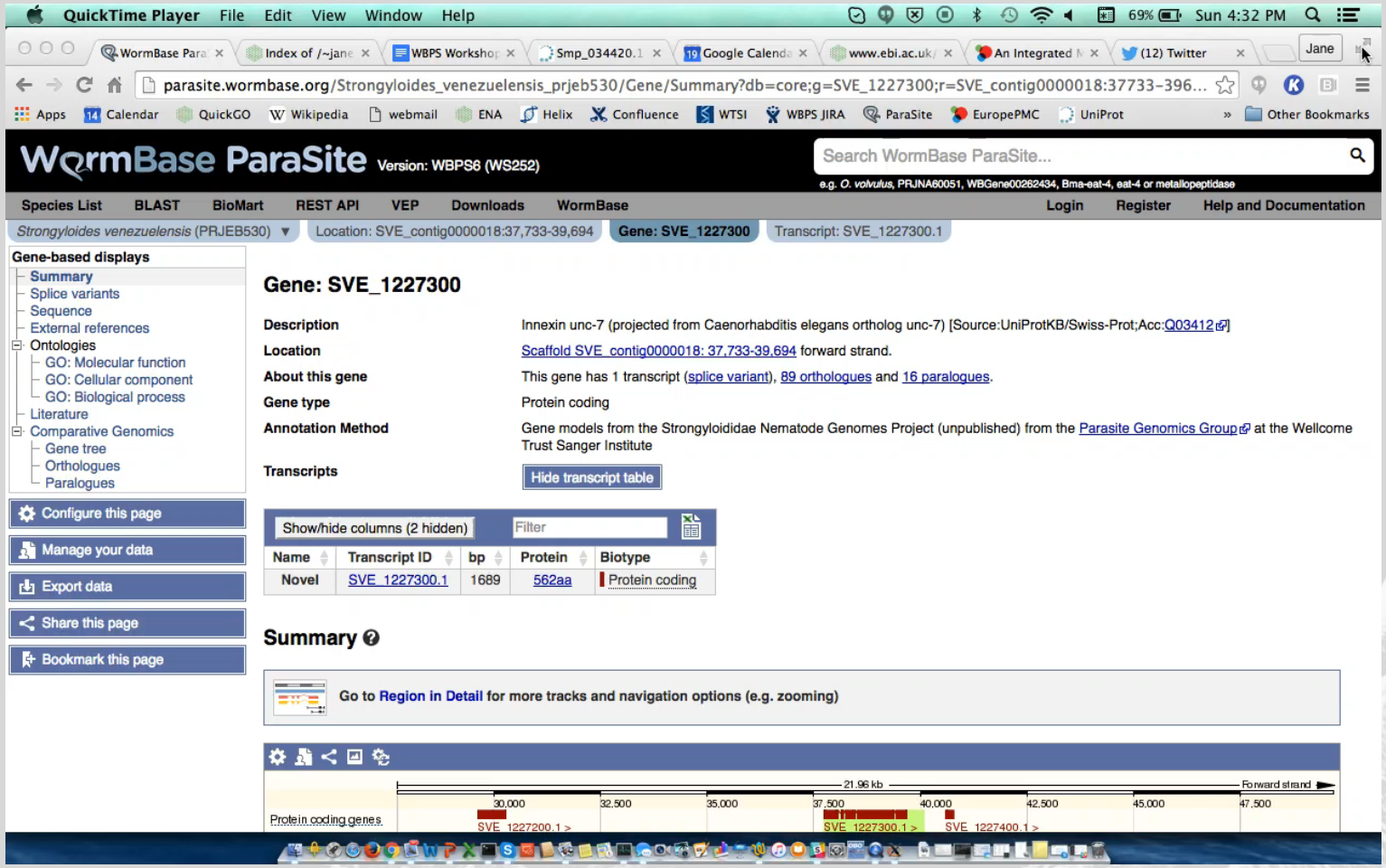
Answers to exercises

5. Locate the gene SVE_1227300
   - 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)

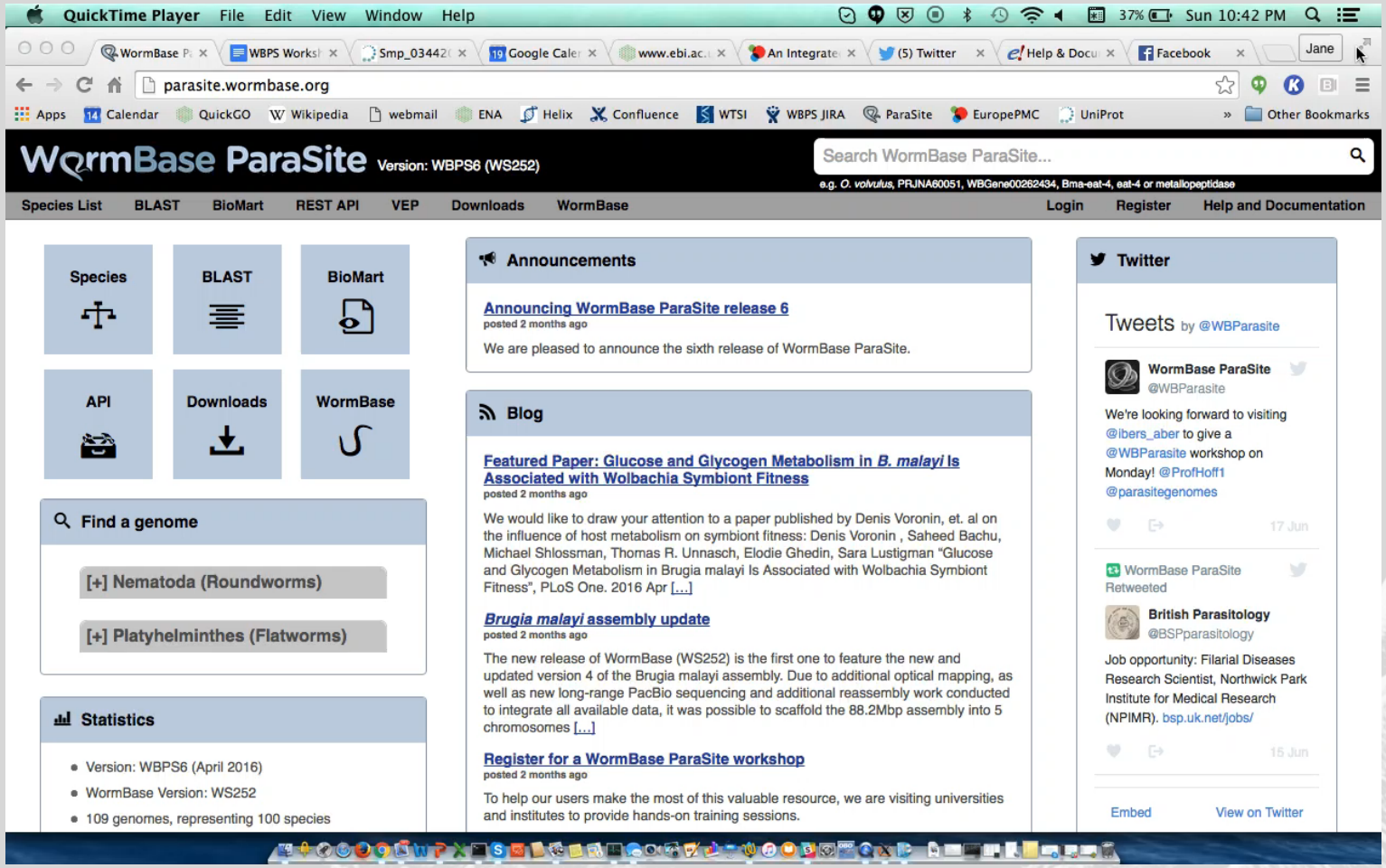

QuickTime Player File Edit View Window Help
Sun 4:24 PM
parasite.wormbase.org/?r=scaffold16:255610-270154;g=TMUE_s0016004100;db=core
WormBase ParaSite Version: WBPS6 (WS252)
Search WormBase ParaSite...
e.g. O. volvulus, PRJNA60051, WBGene00262434, Bma-eat-4, eat-4 or metallopeptidase
Species List BLAST 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
17 Jun
WormBase ParaSite Retweeted
British Parasitology
@BSParasitology
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

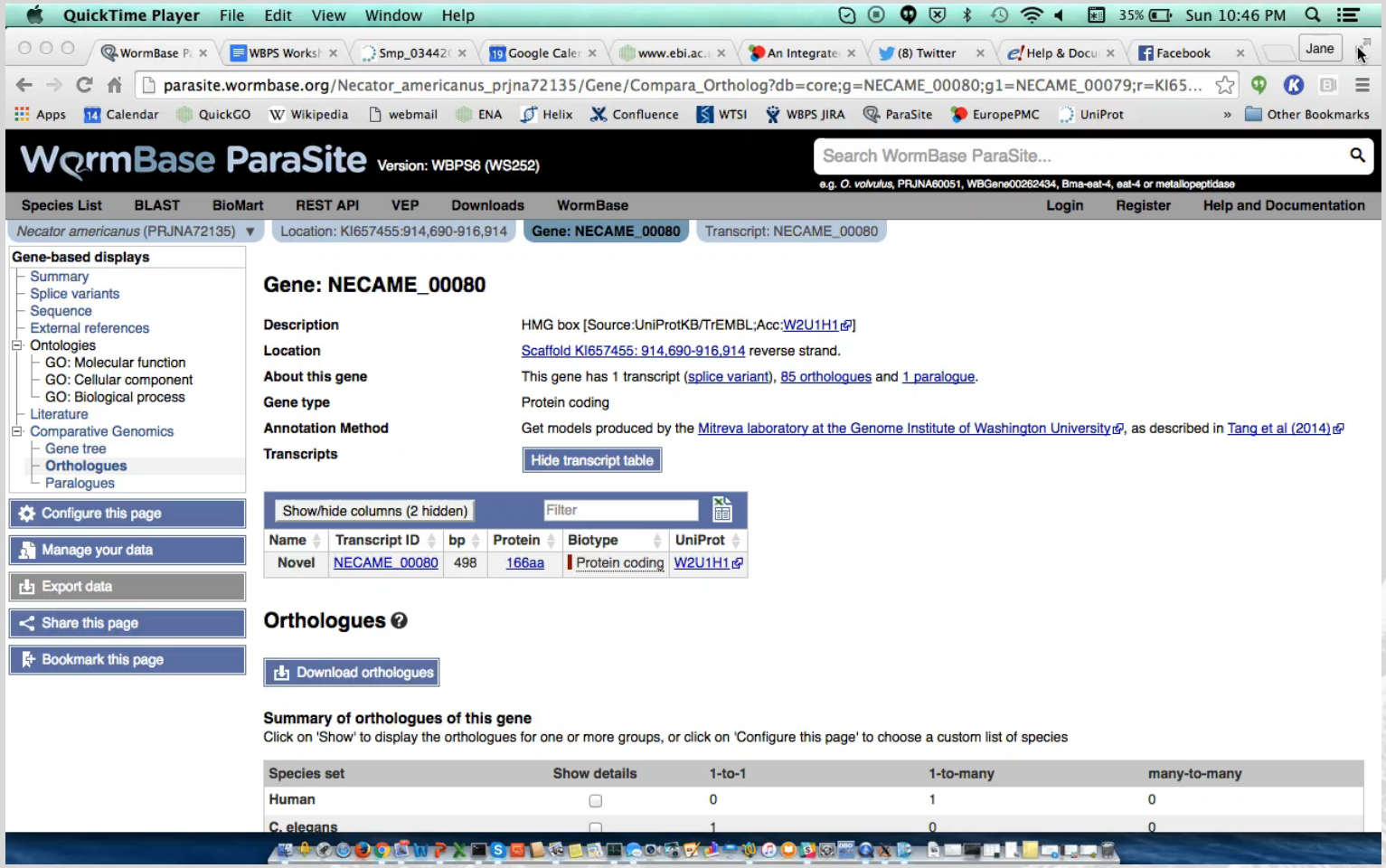
6. Move onto the 'transcript' tab for SVE_1227300
   - How many exons does the single transcript of this gene have? 6
   - Which Pfam domain has been assigned to the protein product of this gene? Innexin (PF00876)


QuickTime Player File Edit View Window Help
Sun 4:32 PM
WormBase Para
Index of /~jane
WBPS Worksho
Smp_034420.1
Google Calenda
www.ebi.ac.uk/
An Integrated M
(12) Twitter
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
Configure this page
Manage your data
Export data
Share this page
Bookmark this page
Gene: SVE_1227300
Description
Innexin 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)
21.96 kb
Forward strand
30,000
32,500
35,000
37,500
40,000
42,500
45,000
47,500
Protein coding genes
SVE_1227200.1 >
SVE_1227300.1 >
SVE_1227400.1 >

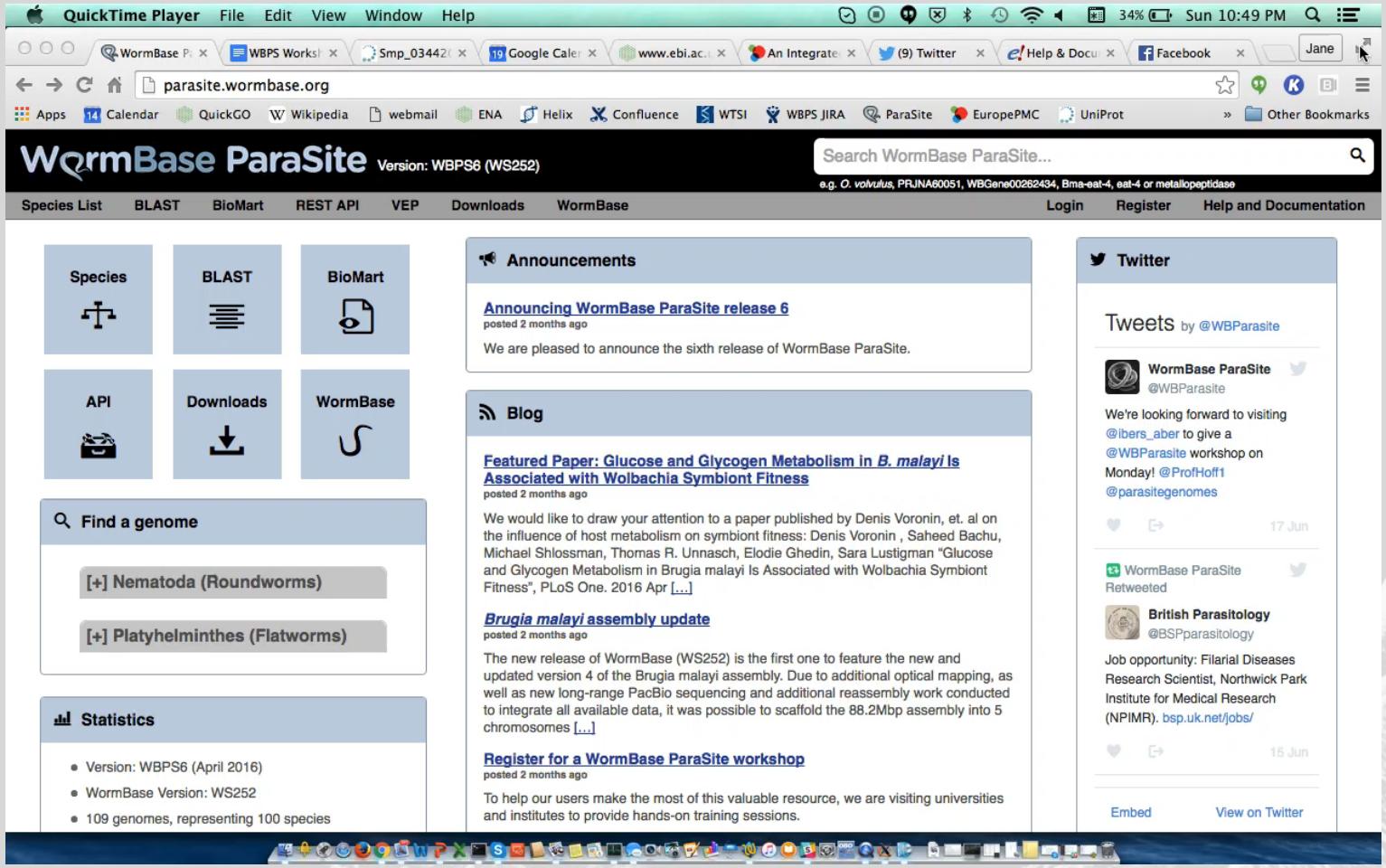
7. Navigate to the gene page for the *Necator americanus* gene NECAME_00080
   - How many orthologues have been predicted in flatworm genomes? 13
   - Which species has an orthologue with the highest percentage identity? *Brugia timori* (PRJEB4663)
   - View the alignment between this gene's protein product and the protein of the *Ancylostoma duodenale* orthologue.


QuickTime Player File Edit View Window Help
Sun 10:42 PM
parasite.wormbase.org
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
17 Jun
WormBase ParaSite Retweeted
British Parasitology
@BSParasitology
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 paralogue?


QuickTime Player File Edit View Window Help
Sun 10:46 PM
parasite.wormbase.org/Necator_americanus_prjna72135/Gene/Compara_Ortholog?db=core;g=NECAME_00080;g1=NECAME_00079;r=KI65...
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
Configure this page
Manage your data
Export data
Share this page
Bookmark this page
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 paralogue.
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
Show/hide columns (2 hidden)
Filter
Name Transcript ID bp Protein Biotype UniProt
Novel NECAME_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 0 1 0
C. elegans 1 0 0

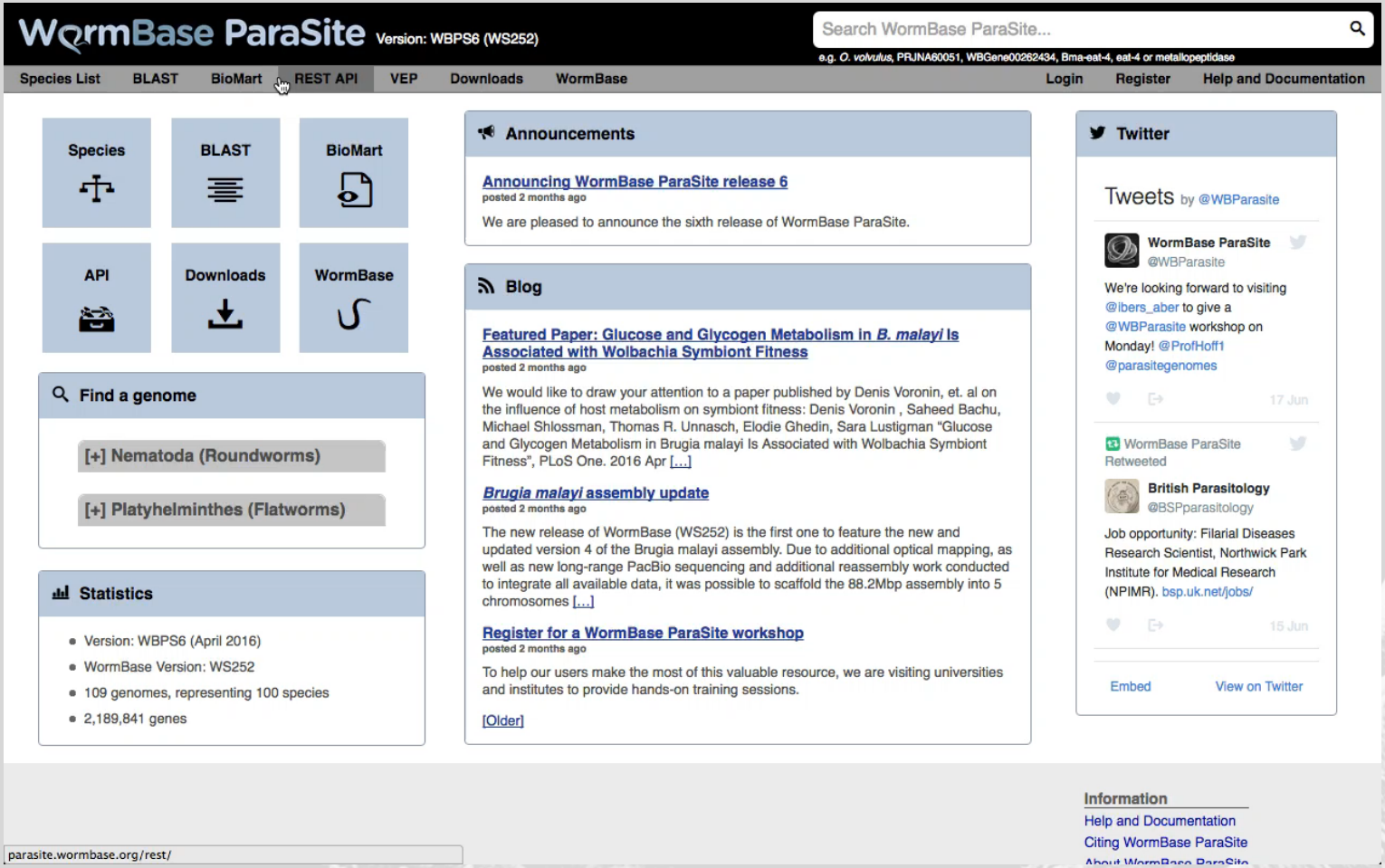
9. Locate the *Fasciola hepatica* (PRJNA179522) orthologue of the human gene BRCA2. Using the gene trees:
   - How close in evolutionary history is this gene located to its orthologue?
   - Are there any duplication events in the evolution of this gene and its homologues? Yes, 2


QuickTime Player File Edit View Window Help
Sun 10:49 PM
parasite.wormbase.org
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
17 Jun
WormBase ParaSite Retweeted
British Parasitology
@BSParasitology
Job opportunity: Filarial Diseases Research Scientist, Northwick Park Institute for Medical Research (NPIMR). bsp.uk.net/jobs/
15 Jun
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?

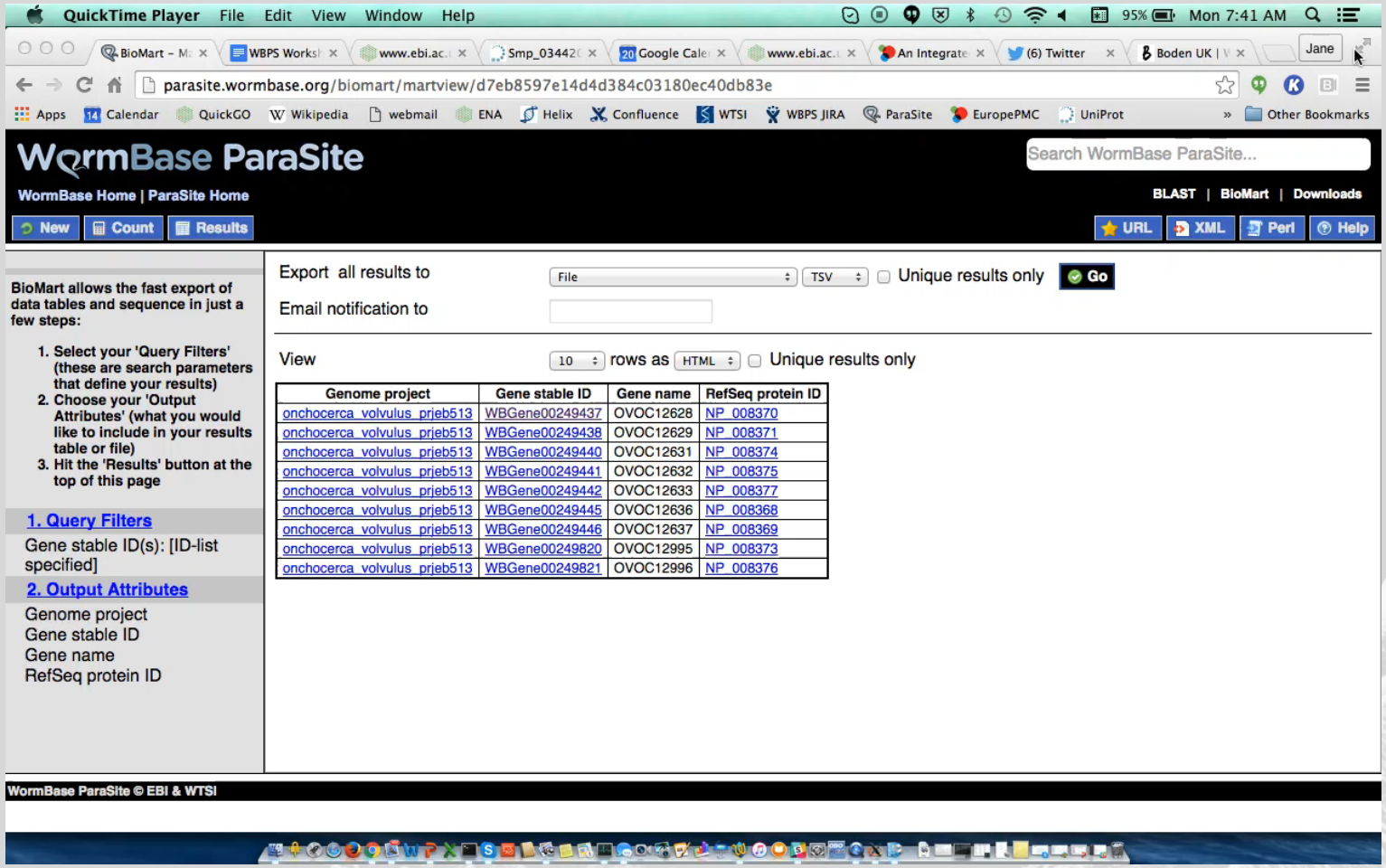
BioMart part II answers

1. 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.


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 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
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
@BSParasitology
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
parasite.wormbase.org/rest/

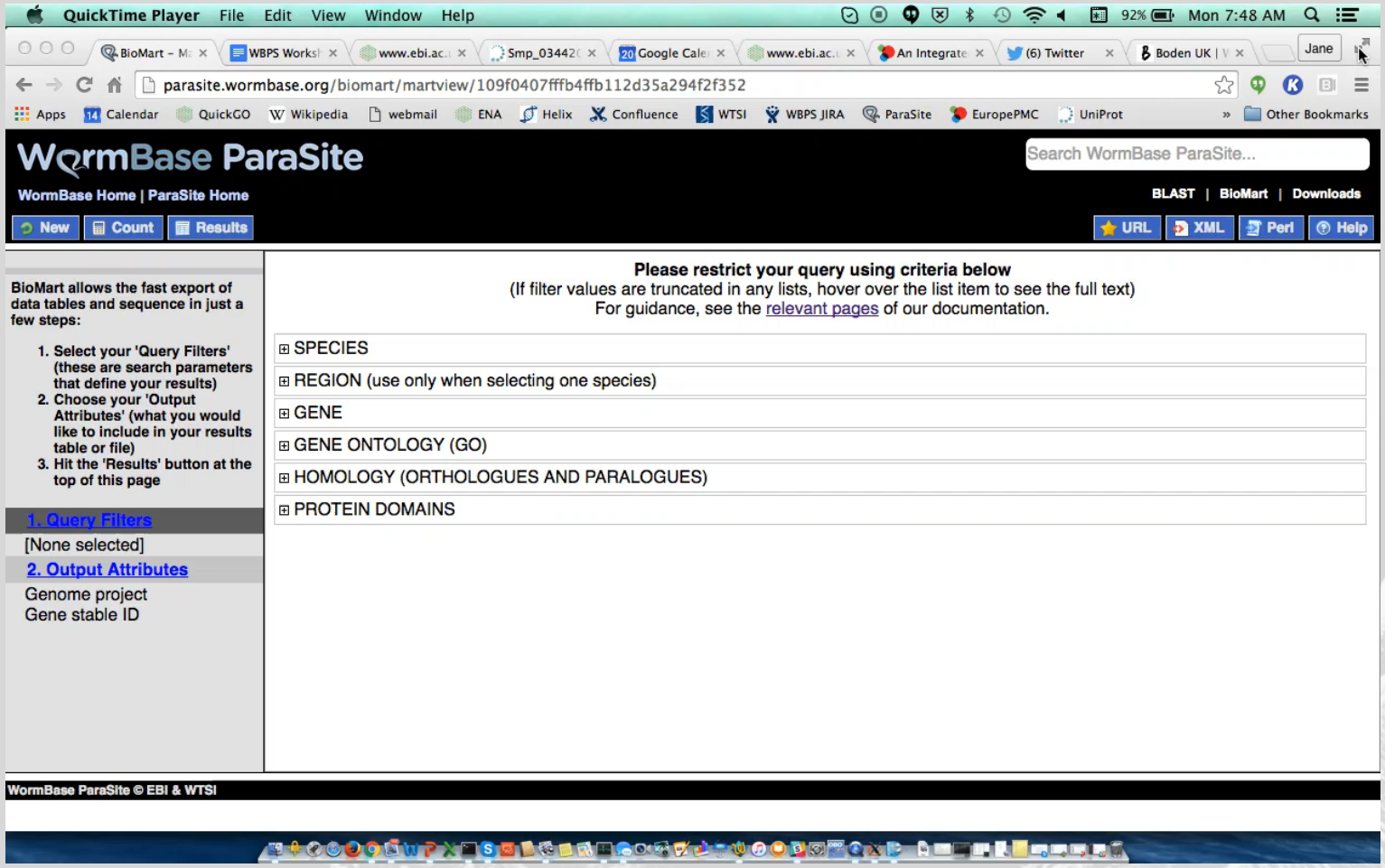
BioMart part II answers

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
Mon 7:41 AM
parasite.wormbase.org/biomart/martview/d7eb8597e14d4d384c03180ec40db83e
WormBase ParaSite
Search WormBase ParaSite...
WormBase Home | ParaSite Home
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
Gene stable ID(s): [ID-list specified]
2. Output Attributes
Genome project
Gene stable ID
Gene name
RefSeq protein ID
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_prjeb513 | WBGene00249437 | OVOC12628 | NP_008370
onchocerca_volvulus_prjeb513 | WBGene00249438 | OVOC12629 | NP_008371
onchocerca_volvulus_prjeb513 | WBGene00249440 | OVOC12631 | NP_008374
onchocerca_volvulus_prjeb513 | WBGene00249441 | OVOC12632 | NP_008375
onchocerca_volvulus_prjeb513 | WBGene00249442 | OVOC12633 | NP_008377
onchocerca_volvulus_prjeb513 | WBGene00249445 | OVOC12636 | NP_008368
onchocerca_volvulus_prjeb513 | WBGene00249446 | OVOC12637 | NP_008369
onchocerca_volvulus_prjeb513 | WBGene00249820 | OVOC12995 | NP_008373
onchocerca_volvulus_prjeb513 | WBGene00249821 | OVOC12996 | NP_008376
WormBase ParaSite © EBI & WTSI

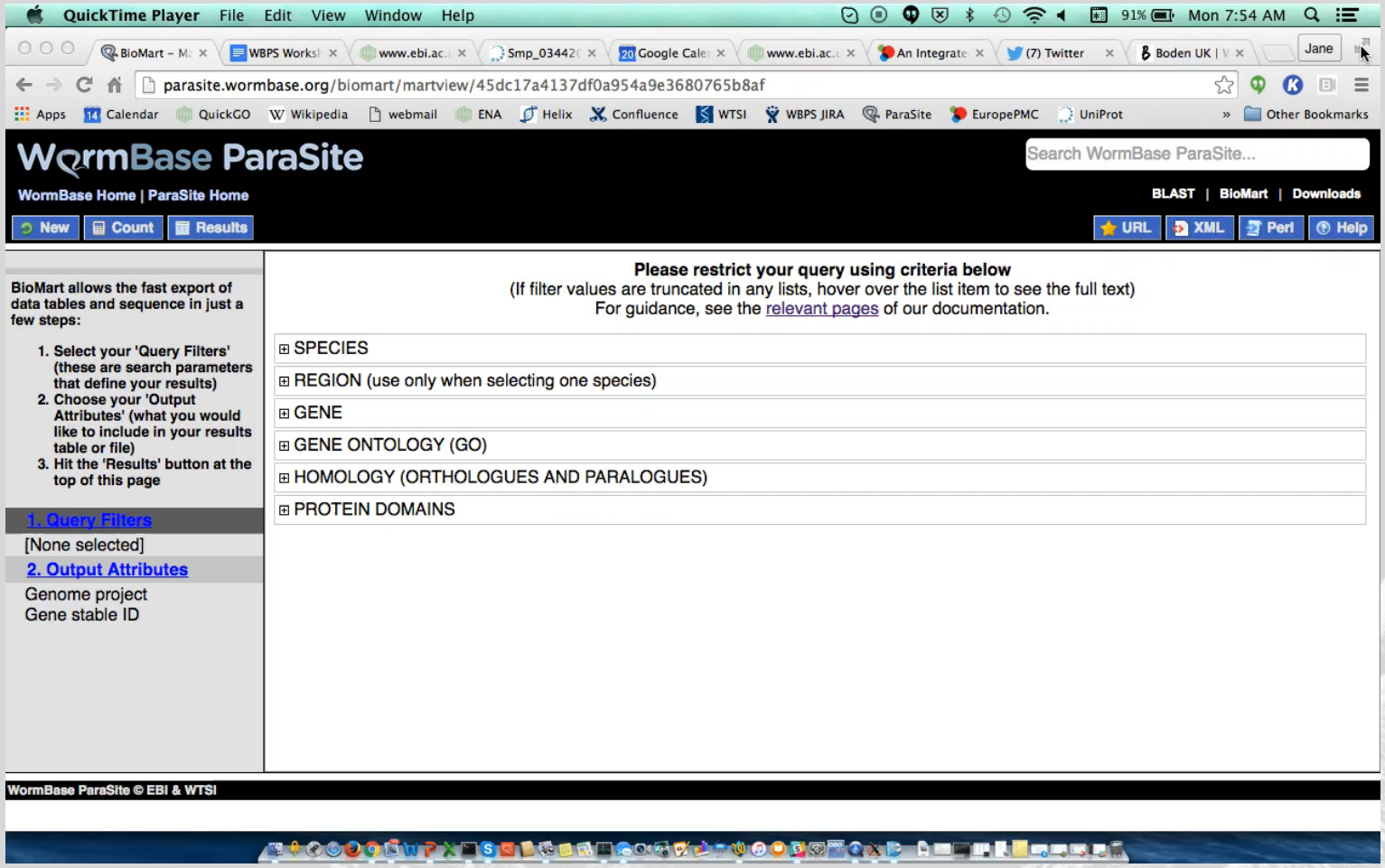
BioMart part II answers

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
92% Mon 7:48 AM
BioMart – M
WBPS Works
www.ebi.ac.
Smp_03442
Google Cale
www.ebi.ac.
An Integrate
(6) Twitter
Boden UK | V
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
Search WormBase ParaSite...
WormBase Home | ParaSite Home
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
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
Mon 7:54 AM
parasite.wormbase.org/biomart/martview/45dc17a4137df0a954a9e3680765b8af
WormBase ParaSite
Search WormBase ParaSite...
WormBase Home | ParaSite Home
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
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	*String*	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

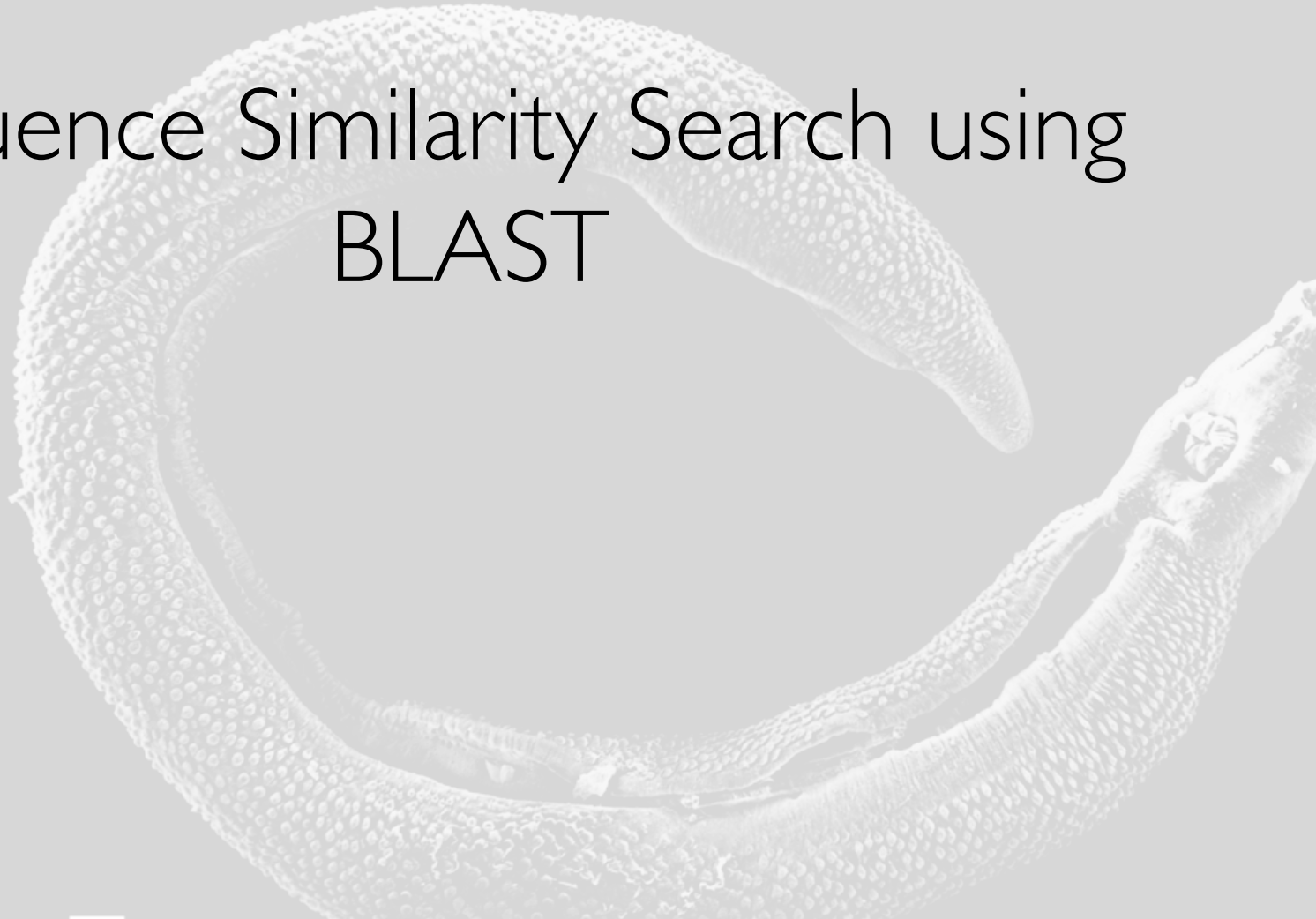
Example Requests

/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

```perl
use strict;
use warnings;

use HTTP::Tiny;

my $http = HTTP::Tiny->new();

my $server = 'http://parasite.wormbase.org';
my $ext = '/genetree/member/symbol/brugia_malayi_prjna10729/Bma-unc-1?';
my $response = $http->get($server.$ext, {
  headers => { 'Content-type' => 'text/x-phyloxml+xml' }
});

die "Failed!\n" unless $response->{success};


print "$response->{status} $response->{reason}\n";

```

Sequence Similarity Search using BLAST



What is BLAST?

- BLAST = **B**asic **L**ocal **A**lignment **S**earch **T**ool
- 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?

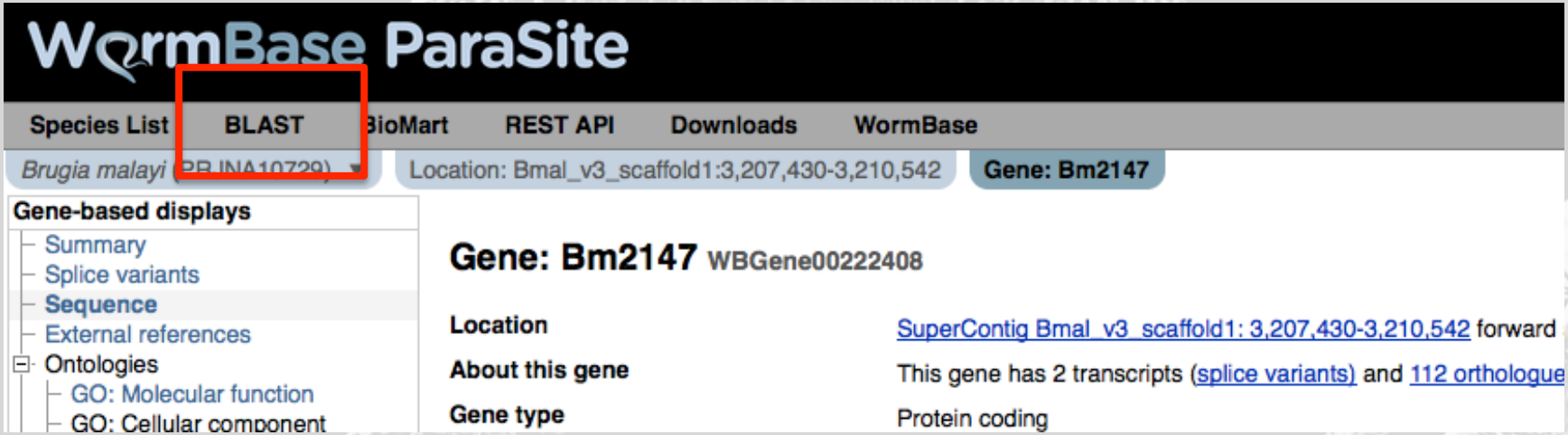
- Input:
  Nucleotide or protein sequence
  Search Parameters

- Output:
  List of all hits ranked in order of statistical significance

Types of BLAST

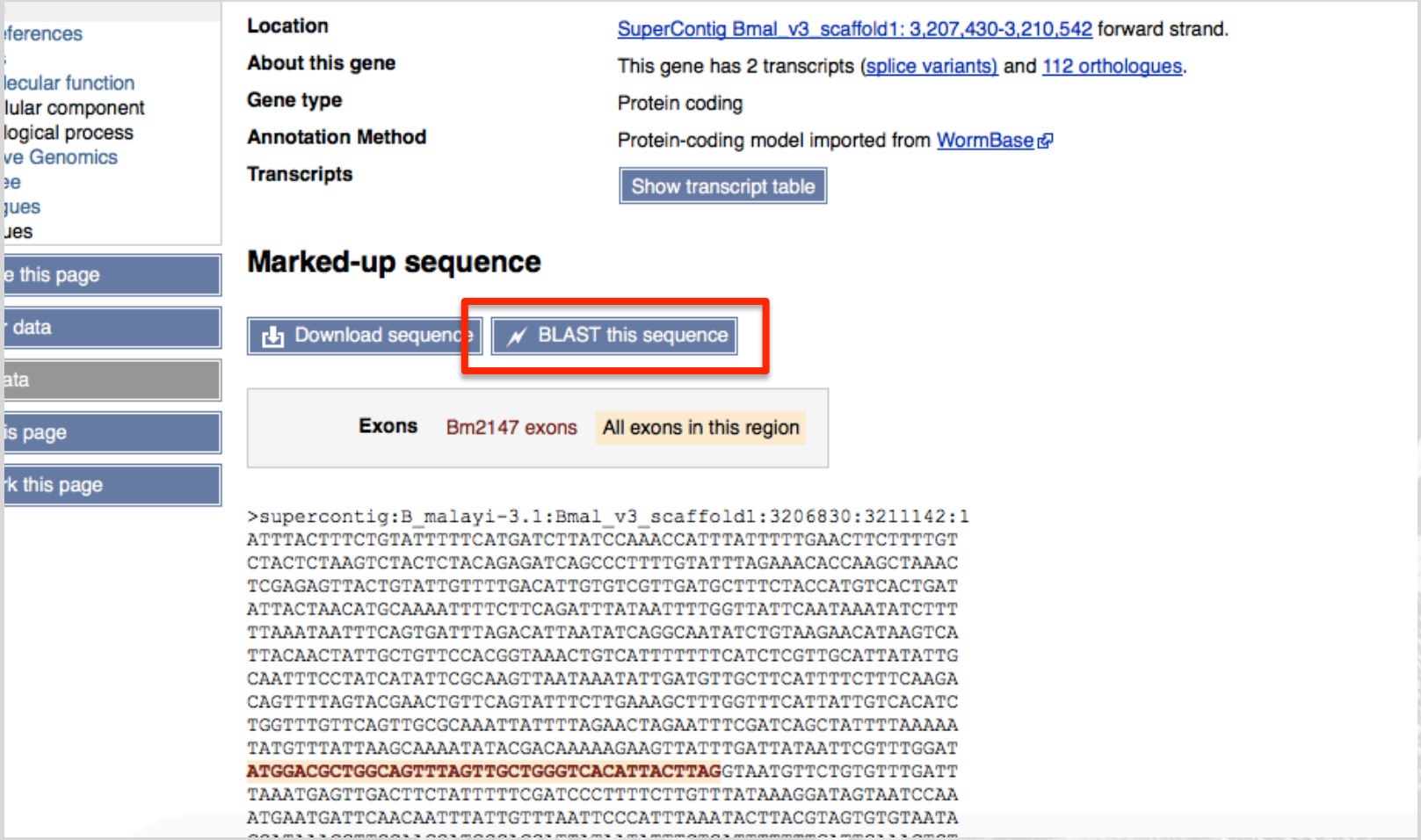
BLAST Type	Query Sequence	Target Database
BLASTN	Nucleotide	Genome (nucleotide)
BLASTP	Peptide	Proteome (peptide)
BLASTX	Six frame translation of a nucleotide sequence	Proteome (peptide)
TBLASTX (slowest)	Six frame translation of a nucleotide sequence	Six frame translation of genome
TBLASTN	Peptide	Six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

Using the ParaSite BLAST

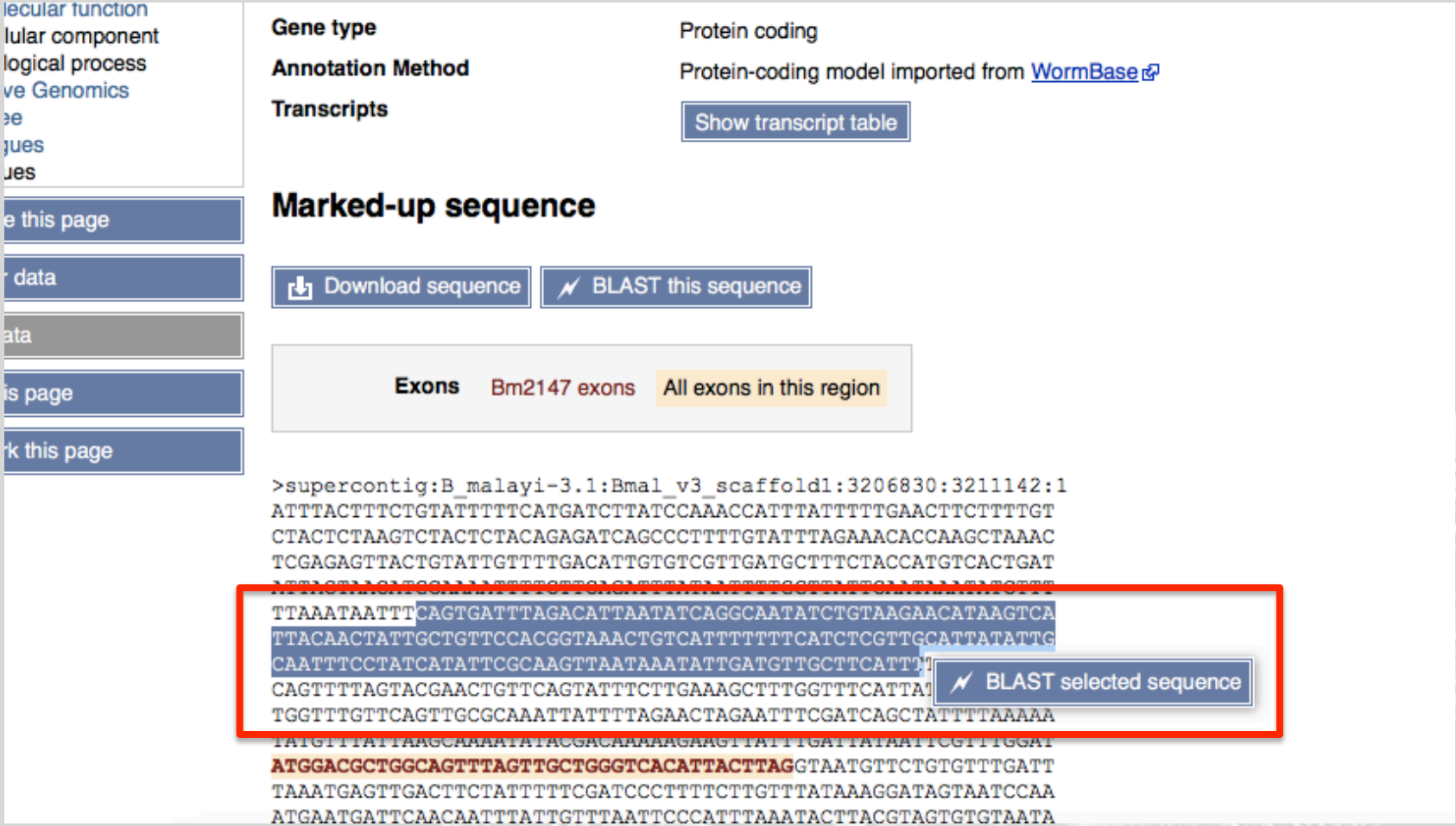


Defaults to the species you are currently browsing

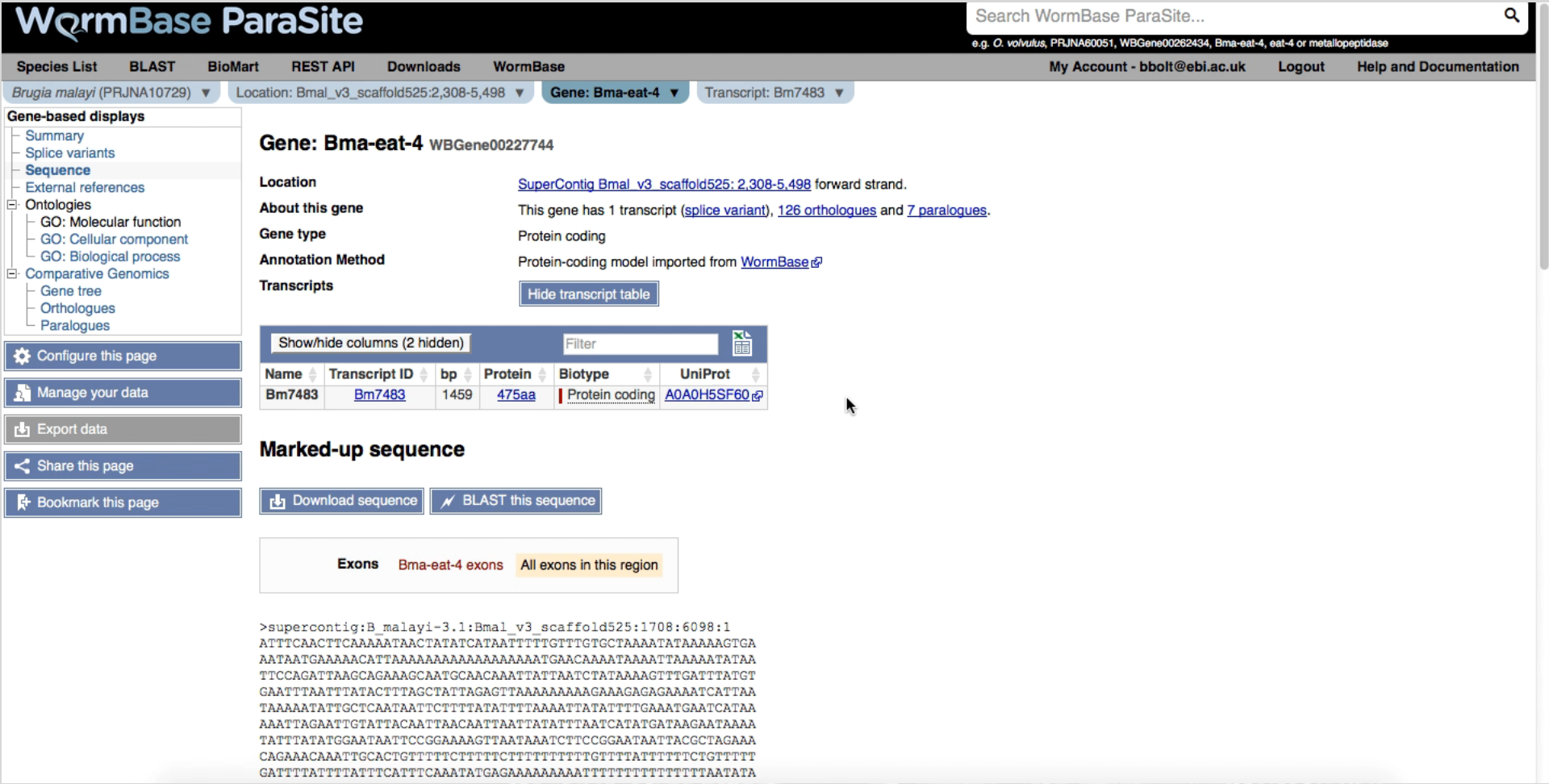
Using the ParaSite BLAST



Using the ParaSite BLAST



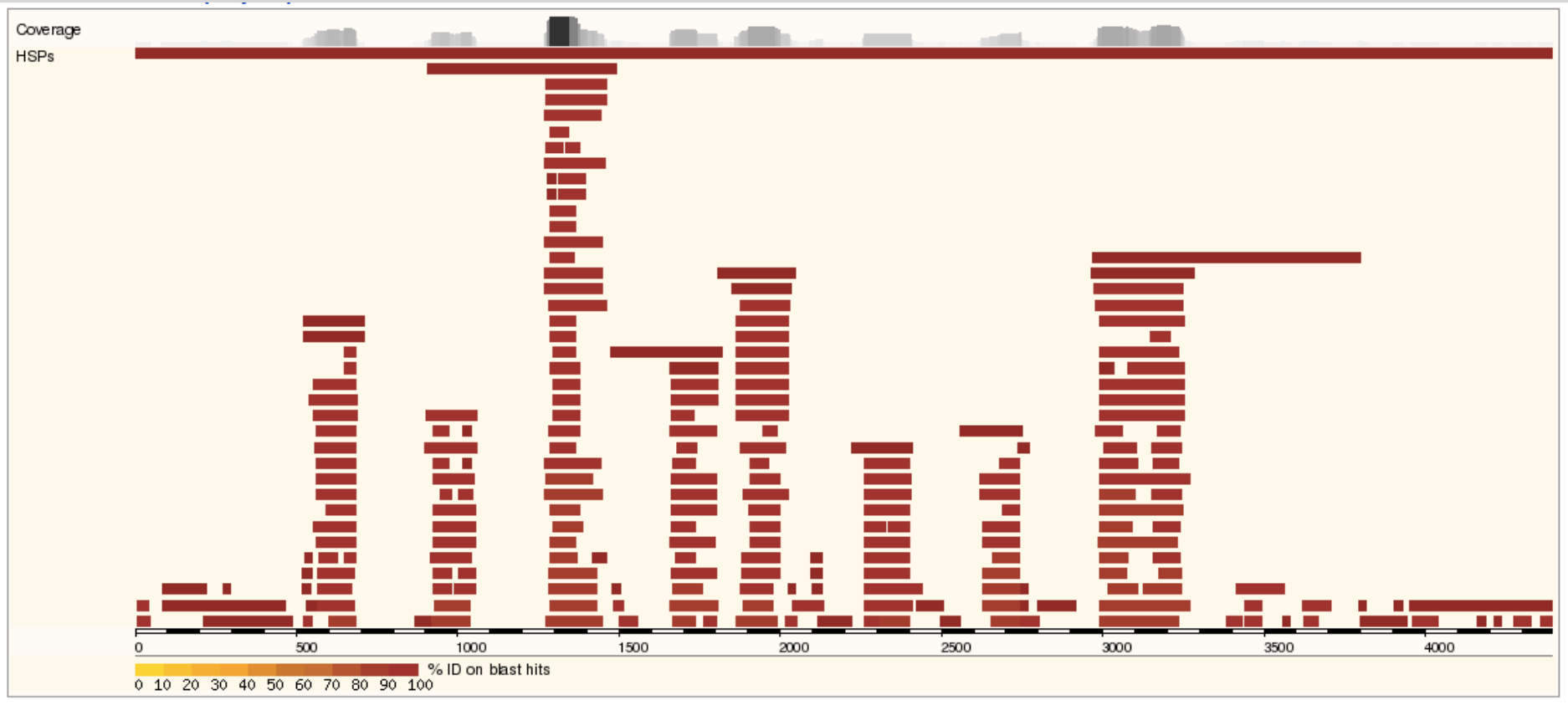
Using the ParaSite BLAST



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 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?

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

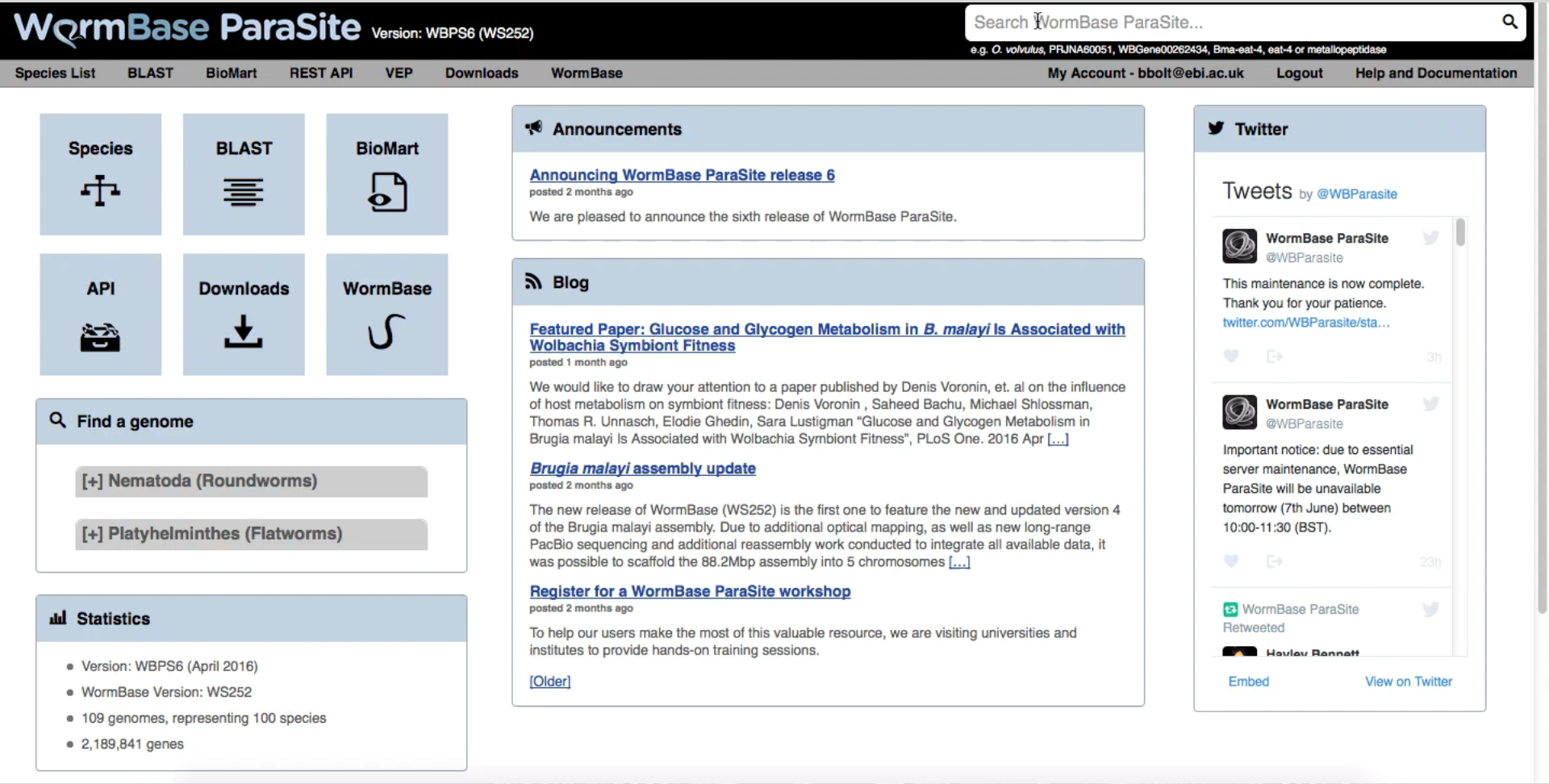
1. Locate the peptide sequence for the *Brugia malayi* gene Bma-eat-4. Using BLAST, find which other nematode(s) this peptide sequence occurs in at 100% similarity?

2. Locate the cDNA sequence for the *Clonorchis sinensis* gene 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?

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?

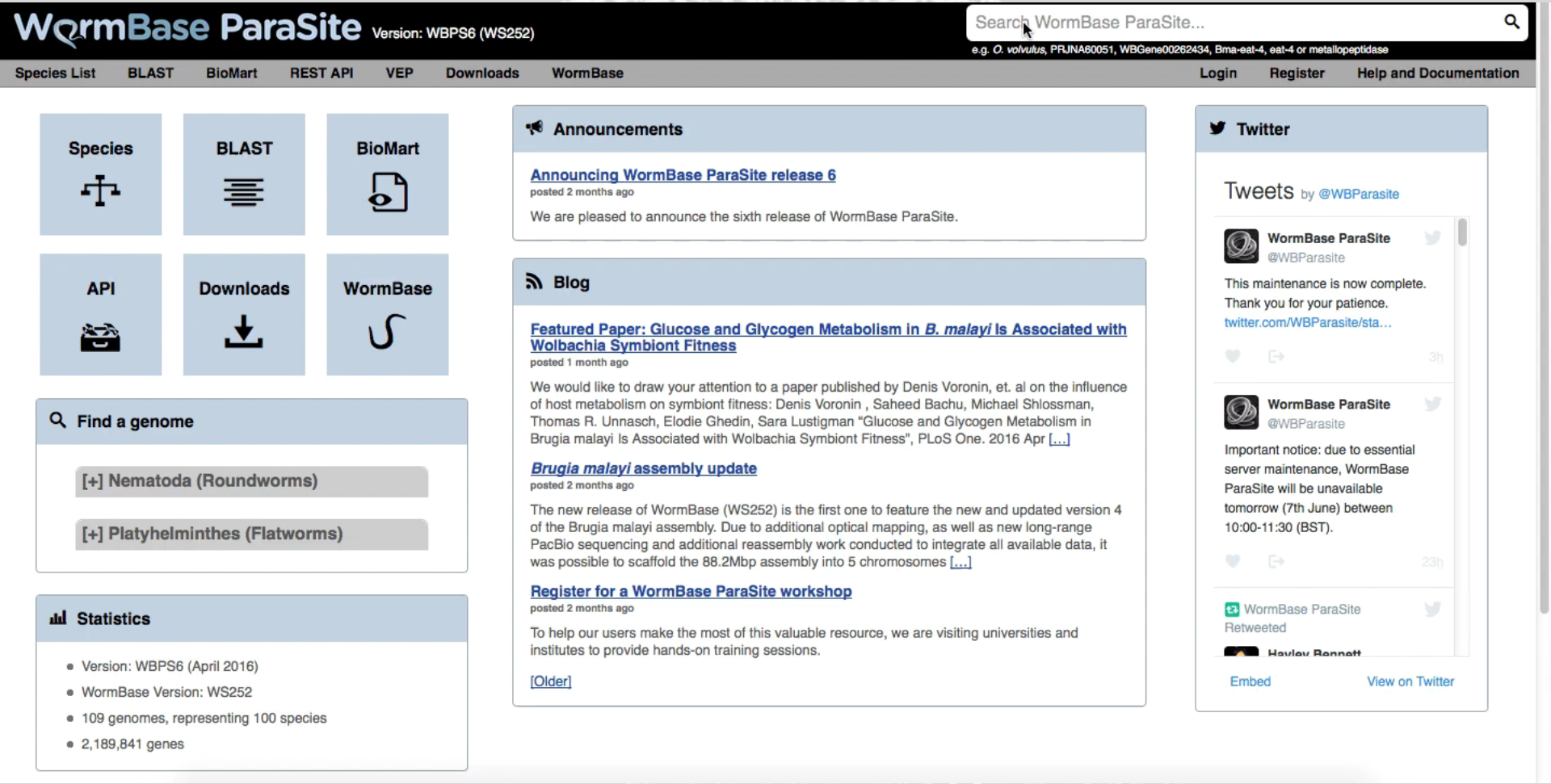
4. Find a gene of choice from your favourite species. BLAST the sequence of the first exon against the database of all helminth transcriptomes.

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 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



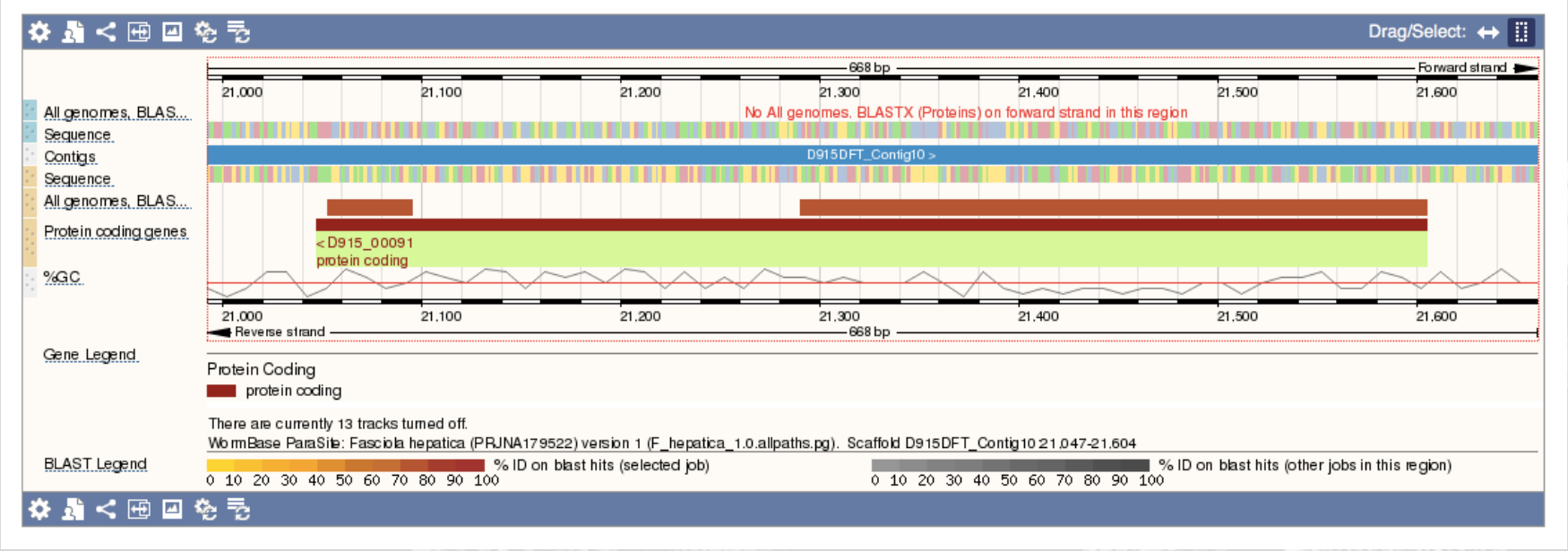
3. Use the 'Edit' button to populate the input form with the sequence from (2). There are many advanced options that can be set. In most cases, these can be left at the default values. Decrease the maximum number of alignments displayed to 5. Why are there more than five results in the table?

Show/hide columns (8 hidden) Filter

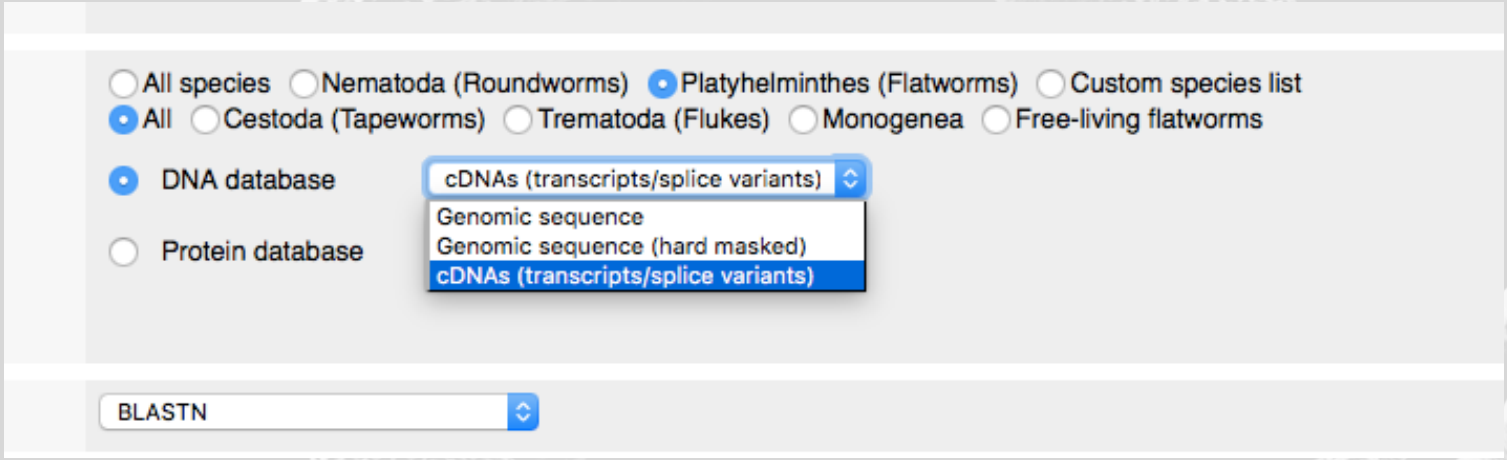
Genome	Subject name	Subject description	Gene hit	Query start	Length	Score	E-val	%ID
Clonorchis sinensis (PRJDA72781)	csin111107	DNA-binding protein A	csin111107	427	128 [Sequence]	677	7.1E-144	100.0 [Alignment]
Clonorchis sinensis (PRJDA72781)	csin111107	DNA-binding protein A	csin111107	1	108 [Sequence]	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

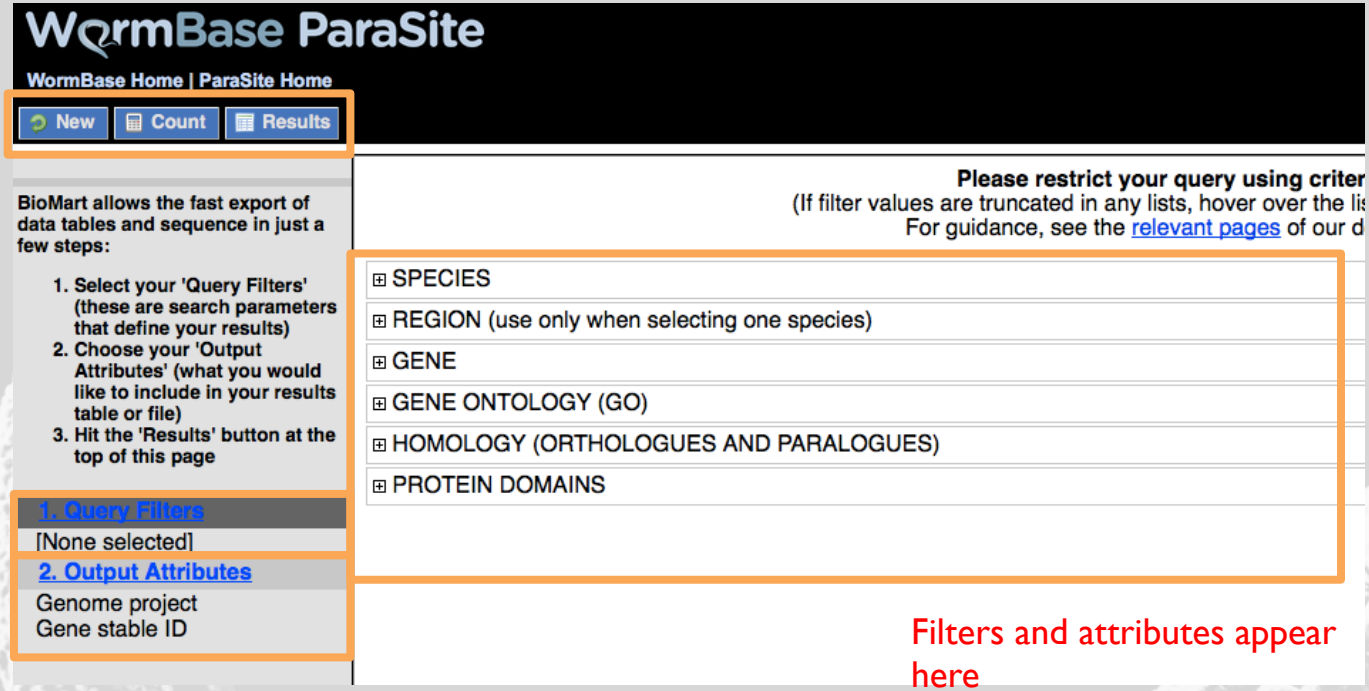
- Sequences (genomic, cDNA, UTR, flanking, cDNA, peptide)
- Gene IDs, names and descriptions
- Identifiers for data from external databases (e.g. UniProt)
- Gene structure (e.g. exons)
- Protein domains and function (e.g. InterPro, Gene3D, PANTHER etc.)
- Gene ontology terms
- Orthologues and paralogues (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)



Filters and attributes appear here

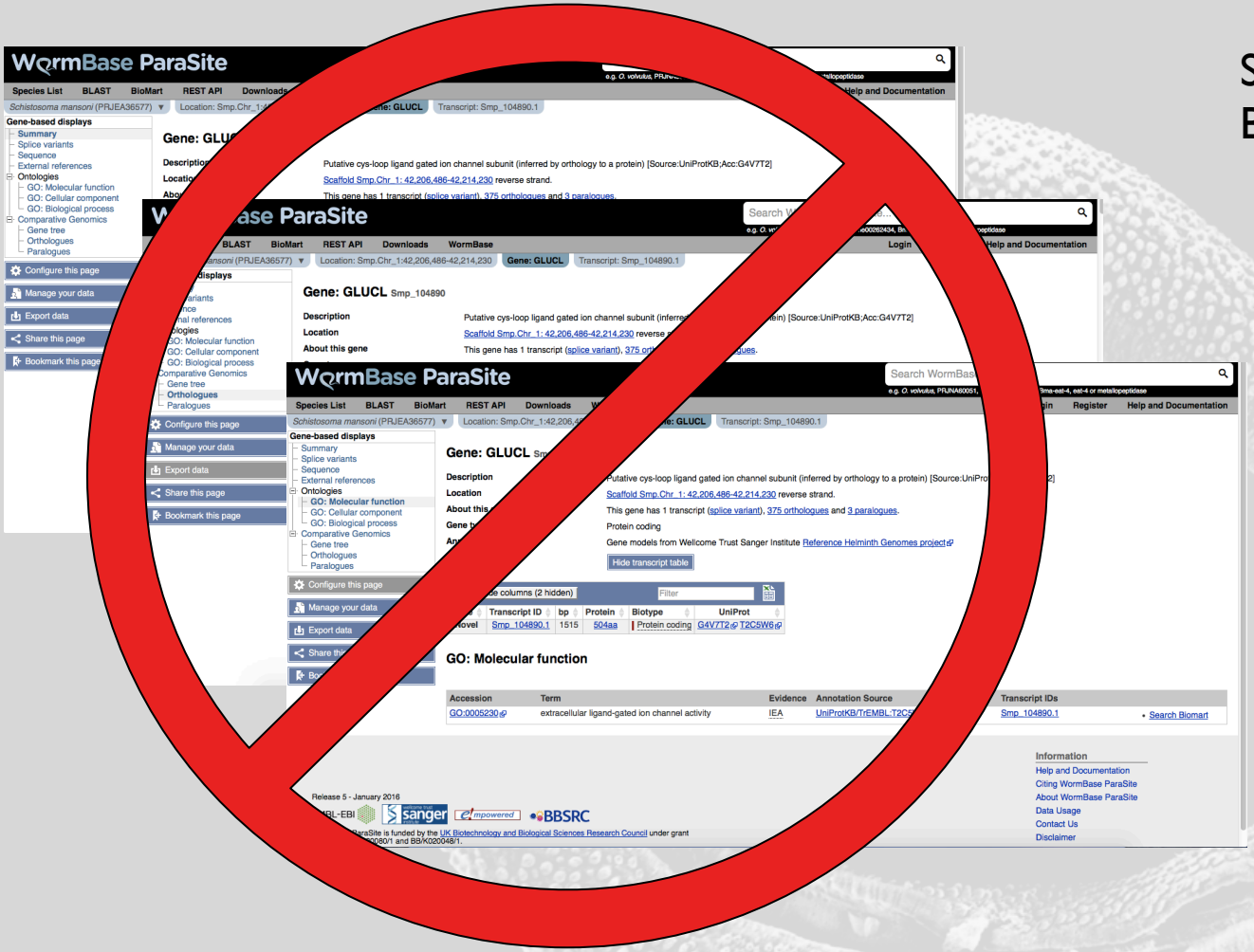
BioMart Example 1 – Working with a list

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

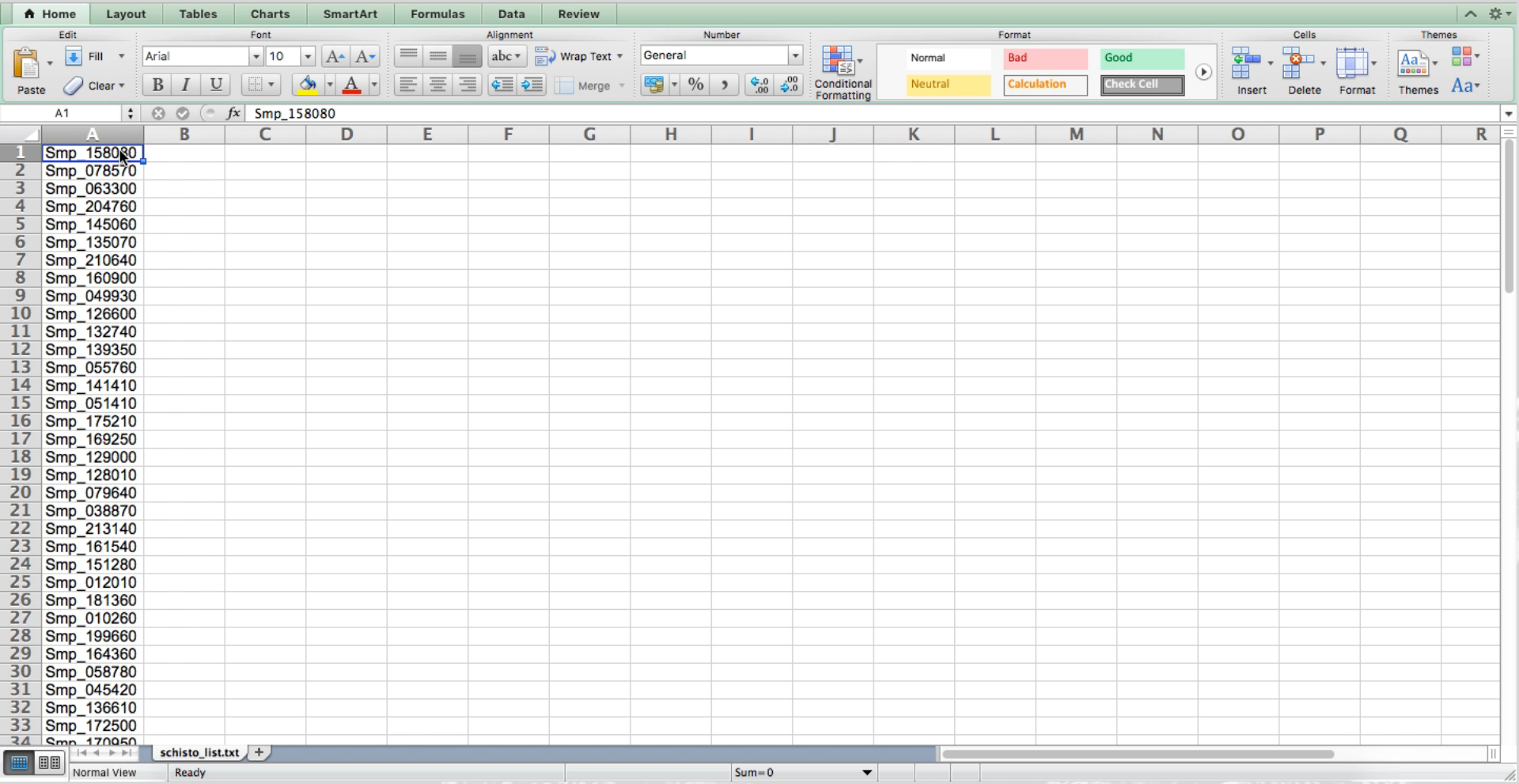
BioMart Example 1 – Working with a list



Save time by using BioMart!

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

1. The gene name and gene description



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
schistosoma_mansoni_prjea36577	Smp_002160		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4VLL1]
schistosoma_mansoni_prjea36577	Smp_007180		Sr-related ctd associated factor (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VB15]
schistosoma_mansoni_prjea36577	Smp_010260		
schistosoma_mansoni_prjea36577	Smp_012010	FOXO	Forkhead transcription factor (inferred by orthology to a protein) [Source:UniProtKB;Acc:A6XDL3]
schistosoma_mansoni_prjea36577	Smp_013610		
schistosoma_mansoni_prjea36577	Smp_018150		Putative mannosyl-oligosaccharide alpha-1,2-mannosidase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VP10]
schistosoma_mansoni_prjea36577	Smp_029520		Villin, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LW89]
schistosoma_mansoni_prjea36577	Smp_036550		
schistosoma_mansoni_prjea36577	Smp_038870		NADH-ubiquinone oxidoreductase, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LUK7]
schistosoma_mansoni_prjea36577	Smp_044010		Tropomyosin-1 [Source:UniProtKB/Swiss-Prot;Acc:P42637]
schistosoma_mansoni_prjea36577	Smp_045420		Adiponectin receptor, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LXQ3]
schistosoma_mansoni_prjea36577	Smp_046890		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4VCL8]
schistosoma_mansoni_prjea36577	Smp_048430	TGR	Thioredoxin glutathione reductase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V8J4]
schistosoma_mansoni_prjea36577	Smp_049930		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4LWN1]
schistosoma_mansoni_prjea36577	Smp_051410		Septate junction protein (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LY05]
schistosoma_mansoni_prjea36577	Smp_055760		ATP-dependent zinc metalloprotease YME1 homolog [Source:UniProtKB/Swiss-Prot;Acc:P54813] (projected from Caenorhabditis elegans ortholog ymel-1)
schistosoma_mansoni_prjea36577	Smp_058780		
schistosoma_mansoni_prjea36577	Smp_060480		Putative copine (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V6J0]
schistosoma_mansoni_prjea36577	Smp_063300		60S ribosomal protein L36-like protein; Putative 60s ribosomal protein L36e [Source:UniProtKB/TrEMBL;Acc:Q15ET2]
schistosoma_mansoni_prjea36577	Smp_069170		Putative cation efflux protein/ zinc transporter (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VIP1]
schistosoma_mansoni_prjea36577	Smp_074990		Regulator of chromosome condensation-related (inferred by orthology to a protein) [Source:UniProtKB;Acc:C4QHH6]
schistosoma_mansoni_prjea36577	Smp_075110		Putative rad1 DNA damage checkpoint protein (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VH02]
schistosoma_mansoni_prjea36577	Smp_077790		Putative histone H2B (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VL73]
schistosoma_mansoni_prjea36577	Smp_078570		Putative zinc finger protein (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VM27]
schistosoma_mansoni_prjea36577	Smp_079640		Putative monocarboxylate transporter (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V7W3]
schistosoma_mansoni_prjea36577	Smp_085080		
schistosoma_mansoni_prjea36577	Smp_093760		Strain Puerto Rico chromosome W, complete genome [Source:UniProtKB/TrEMBL;Acc:G4VTD3]
schistosoma_mansoni_prjea36577	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...

WormBase Home | ParaSite Home

BLAST | BioMart | Downloads

New | Count | Results | URL | XML | Perl | Help

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 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	Human gene stable ID
schistosoma_mansoni_prjea36577	Smp_002160		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4VLL1]	ENSG00000178028
schistosoma_mansoni_prjea36577	Smp_007180		Sr-related ctd associated factor (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VB15]	ENSG00000085872
schistosoma_mansoni_prjea36577	Smp_010260			ENSG00000136936
schistosoma_mansoni_prjea36577	Smp_013610			ENSG00000170264
schistosoma_mansoni_prjea36577	Smp_013610			ENSG00000156050
schistosoma_mansoni_prjea36577	Smp_018150		Putative mannosyl-oligosaccharide alpha-1,2-mannosidase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VP10]	ENSG00000111885
schistosoma_mansoni_prjea36577	Smp_018150		Putative mannosyl-oligosaccharide alpha-1,2-mannosidase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VP10]	ENSG00000198162
schistosoma_mansoni_prjea36577	Smp_018150		Putative mannosyl-oligosaccharide alpha-1,2-mannosidase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4VP10]	ENSG00000117643
schistosoma_mansoni_prjea36577	Smp_038870		NADH-ubiquinone oxidoreductase, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LUK7]	ENSG00000023228
schistosoma_mansoni_prjea36577	Smp_044010		Tropomyosin-1 [Source:UniProtKB/Swiss-Prot;Acc:P42637]	ENSG00000167460
schistosoma_mansoni_prjea36577	Smp_044010		Tropomyosin-1 [Source:UniProtKB/Swiss-Prot;Acc:P42637]	ENSG00000140416
schistosoma_mansoni_prjea36577	Smp_044010		Tropomyosin-1 [Source:UniProtKB/Swiss-Prot;Acc:P42637]	ENSG00000198467
schistosoma_mansoni_prjea36577	Smp_044010		Tropomyosin-1 [Source:UniProtKB/Swiss-Prot;Acc:P42637]	ENSG00000143549
schistosoma_mansoni_prjea36577	Smp_045420		Adiponectin receptor, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LXQ3]	ENSG00000006831
schistosoma_mansoni_prjea36577	Smp_045420		Adiponectin receptor, putative (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4LXQ3]	ENSG00000159346
schistosoma_mansoni_prjea36577	Smp_046890		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4VCL8]	ENSG00000196850
schistosoma_mansoni_prjea36577	Smp_048430	TGR	Thioredoxin glutathione reductase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V8J4]	ENSG00000197763
schistosoma_mansoni_prjea36577	Smp_048430	TGR	Thioredoxin glutathione reductase (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V8J4]	ENSG00000198431
schistosoma_mansoni_prjea36577	Smp_049930		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4LWN1]	ENSG00000148343
schistosoma_mansoni_prjea36577	Smp_049930		Putative uncharacterized protein [Source:UniProtKB/TrEMBL;Acc:G4LWN1]	ENSG00000180488
schistosoma_mansoni_prjea36577	Smp_058780			ENSG00000105617
schistosoma_mansoni_prjea36577	Smp_060480		Putative copine (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V6J0]	ENSG00000139117
schistosoma_mansoni_prjea36577	Smp_060480		Putative copine (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V6J0]	ENSG00000144550
schistosoma_mansoni_prjea36577	Smp_060480		Putative copine (inferred by orthology to a protein) [Source:UniProtKB;Acc:G4V6J0]	ENSG00000124772
schistosoma_mansoni_prjea36577	Smp_063300		60S ribosomal protein L36-like protein; Putative 60s ribosomal protein L36e [Source:UniProtKB/TrEMBL;Acc:Q15ET2]	ENSG00000130255

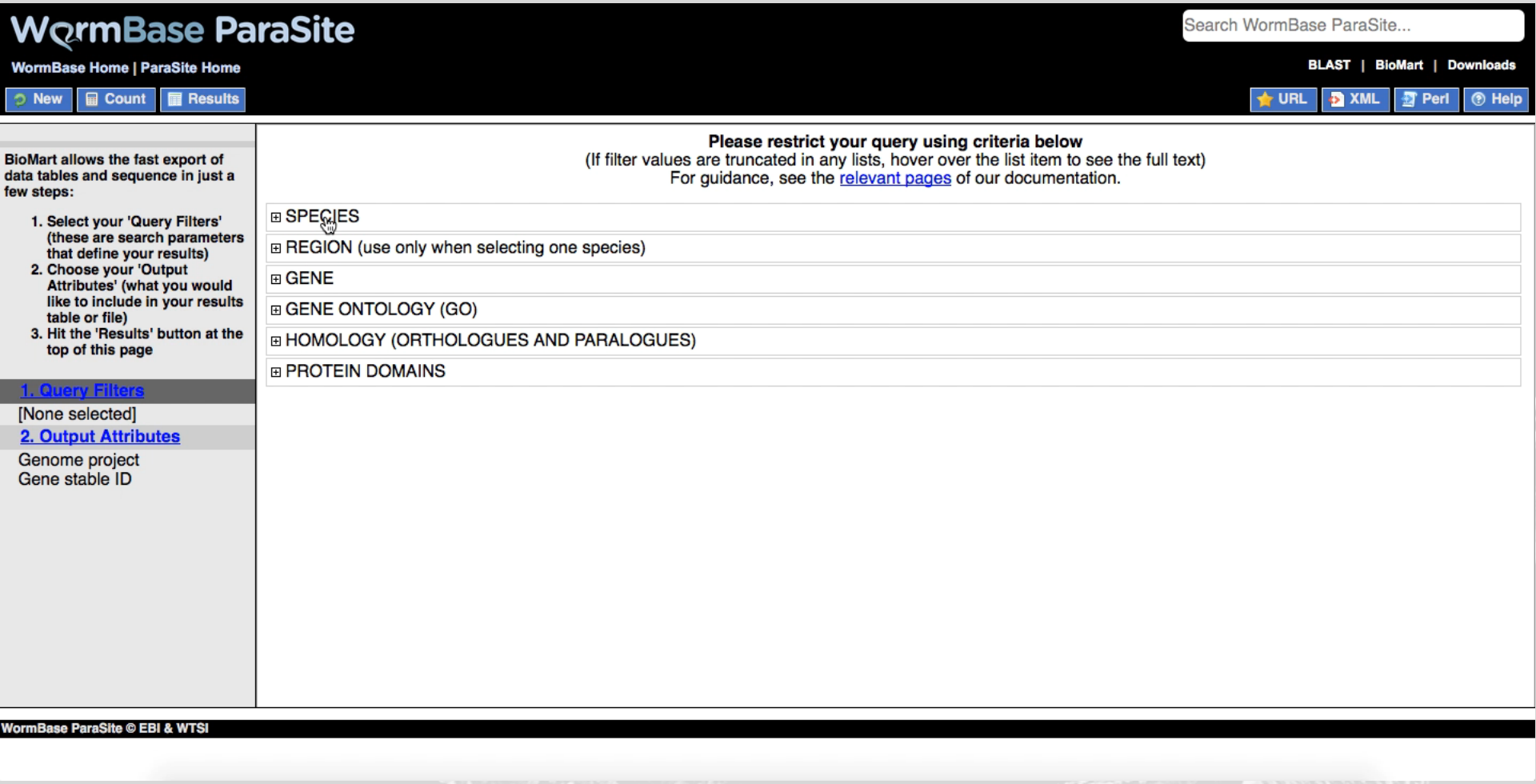
BioMart Example 2 – Generating a list

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:

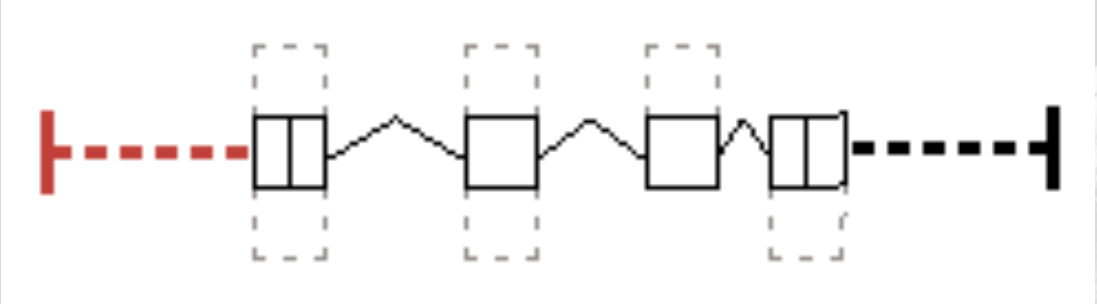
- Do not have an orthologue in humans
- Are associated with transmembrane signalling receptor activity
- Annotated with the *C. elegans* orthologue, if available

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

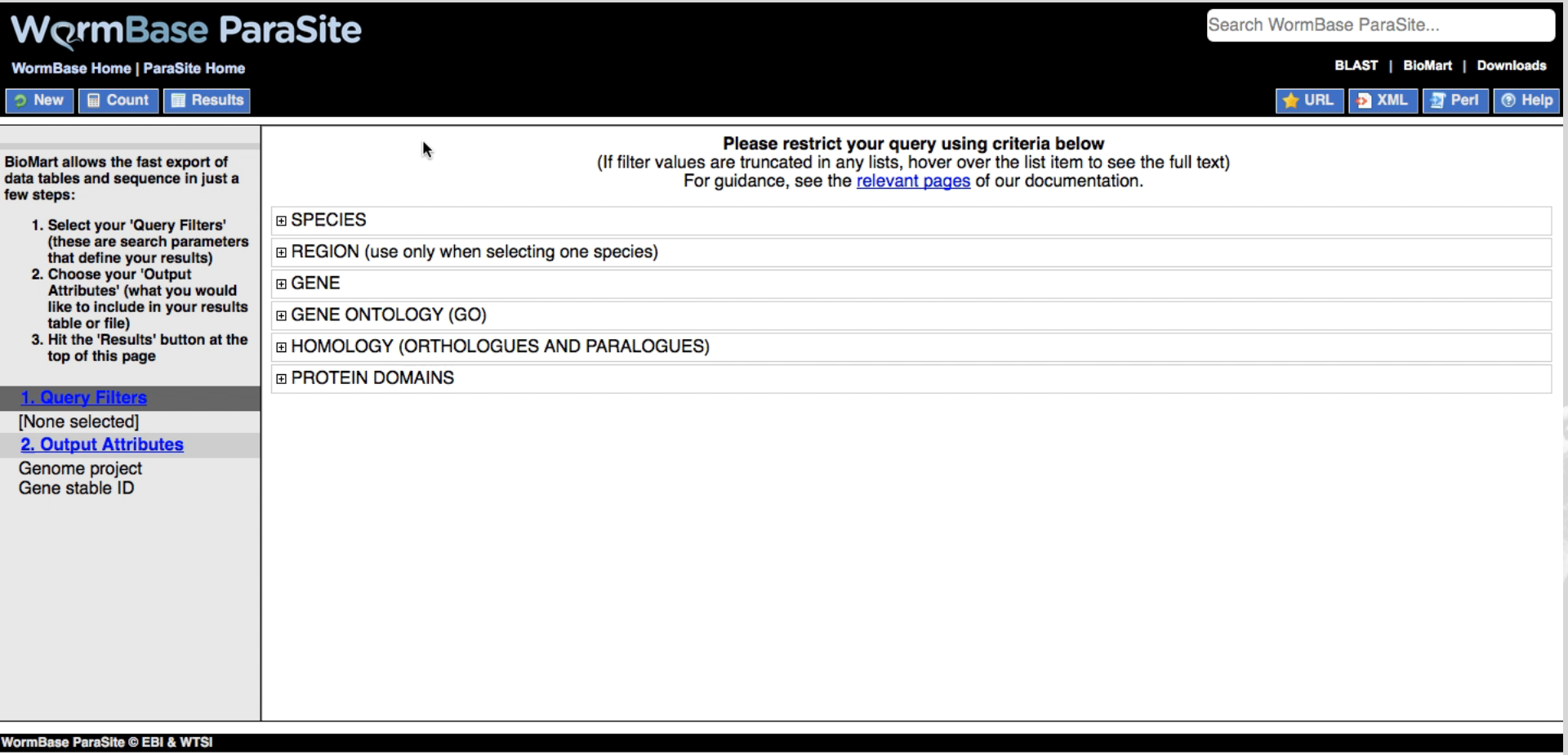


BioMart Example 3 – Retrieving sequences

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



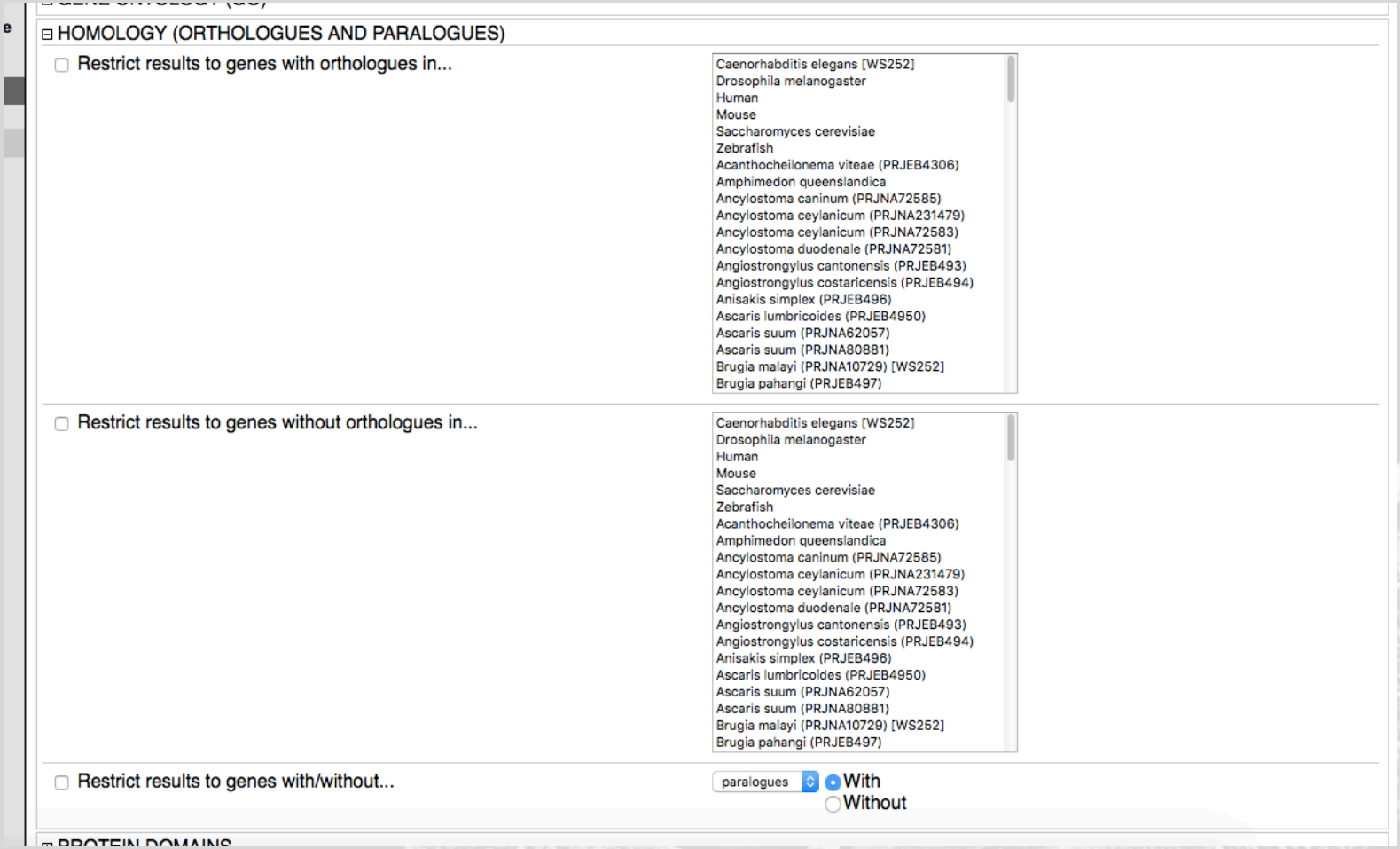
I would like to retrieve the sequence located 500bp upstream of each *Brugia malayi* gene



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)

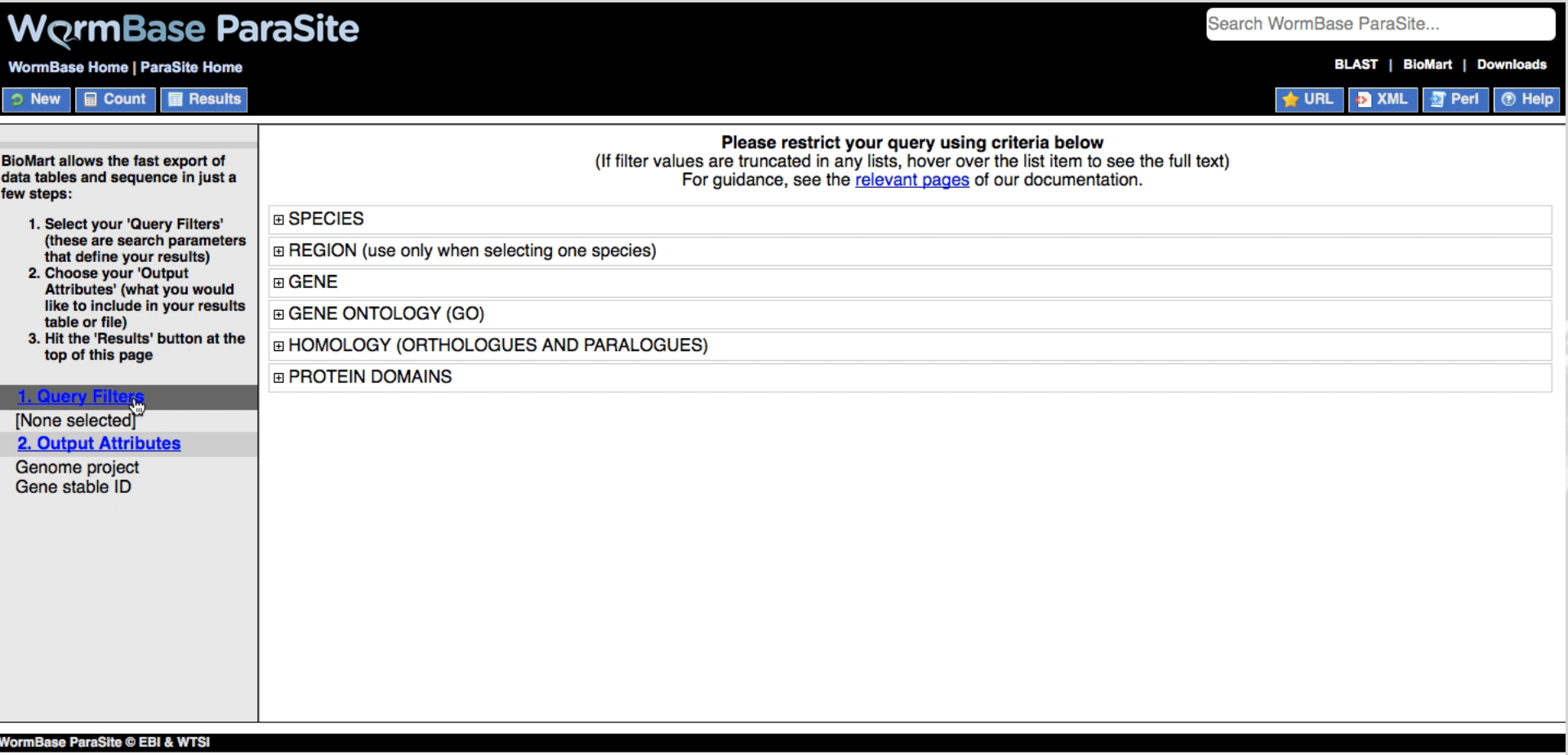
Coming soon… (release 7 in August)



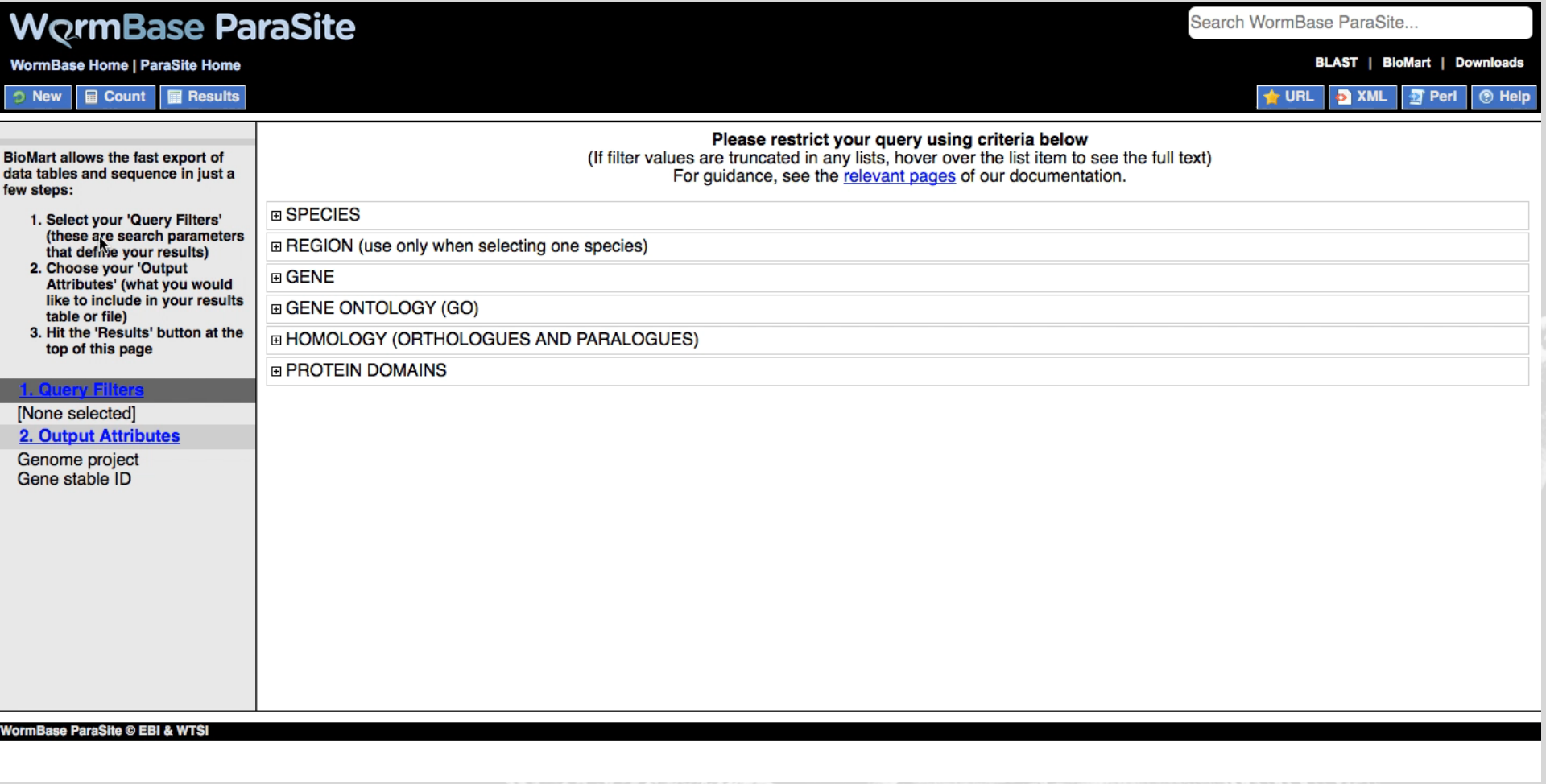
Exercises – Part 1

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

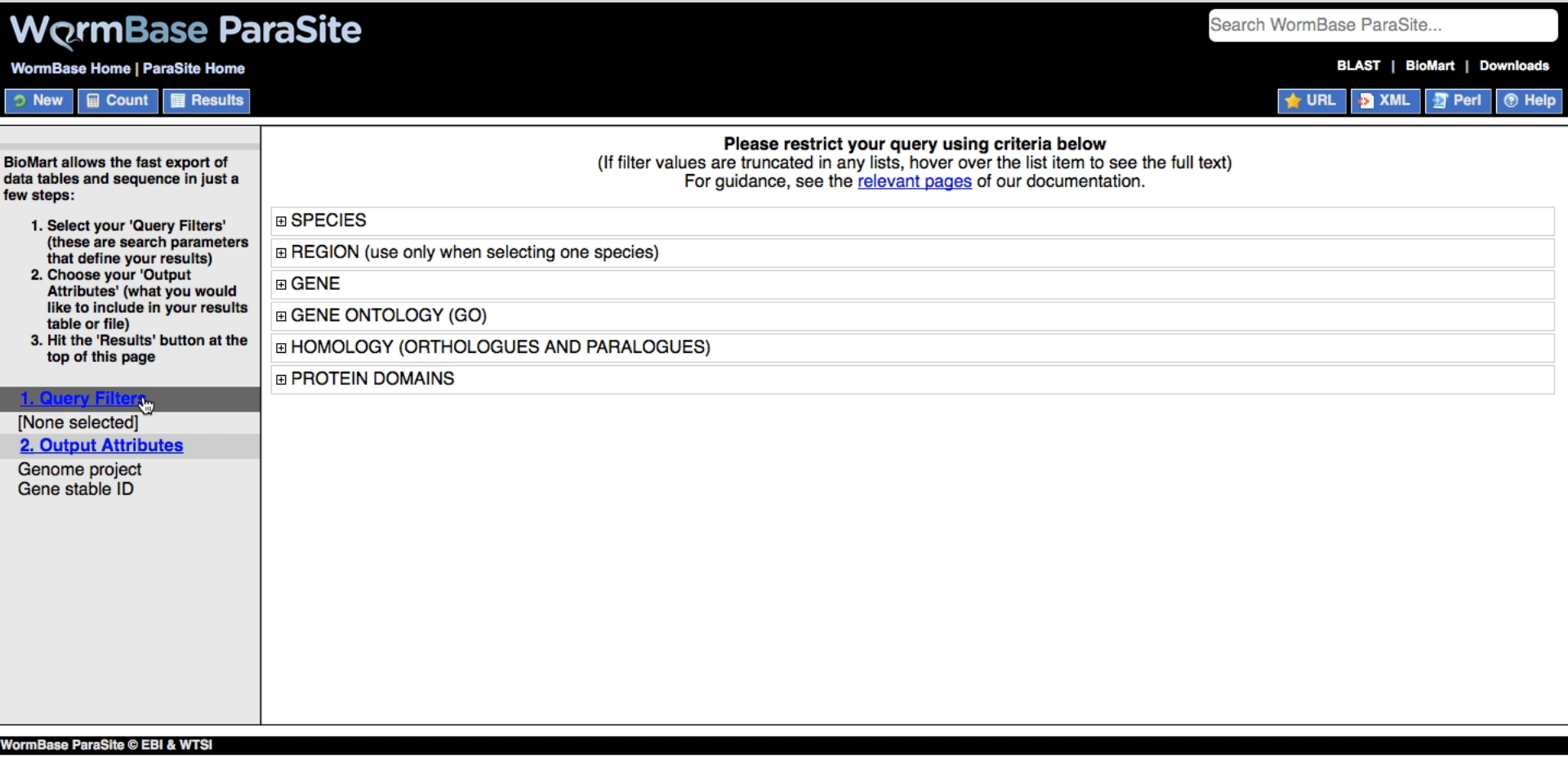
Exercise 1


WormBase ParaSite
Search WormBase ParaSite...
WormBase Home | ParaSite Home
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
WormBase ParaSite © EBI & WTSI

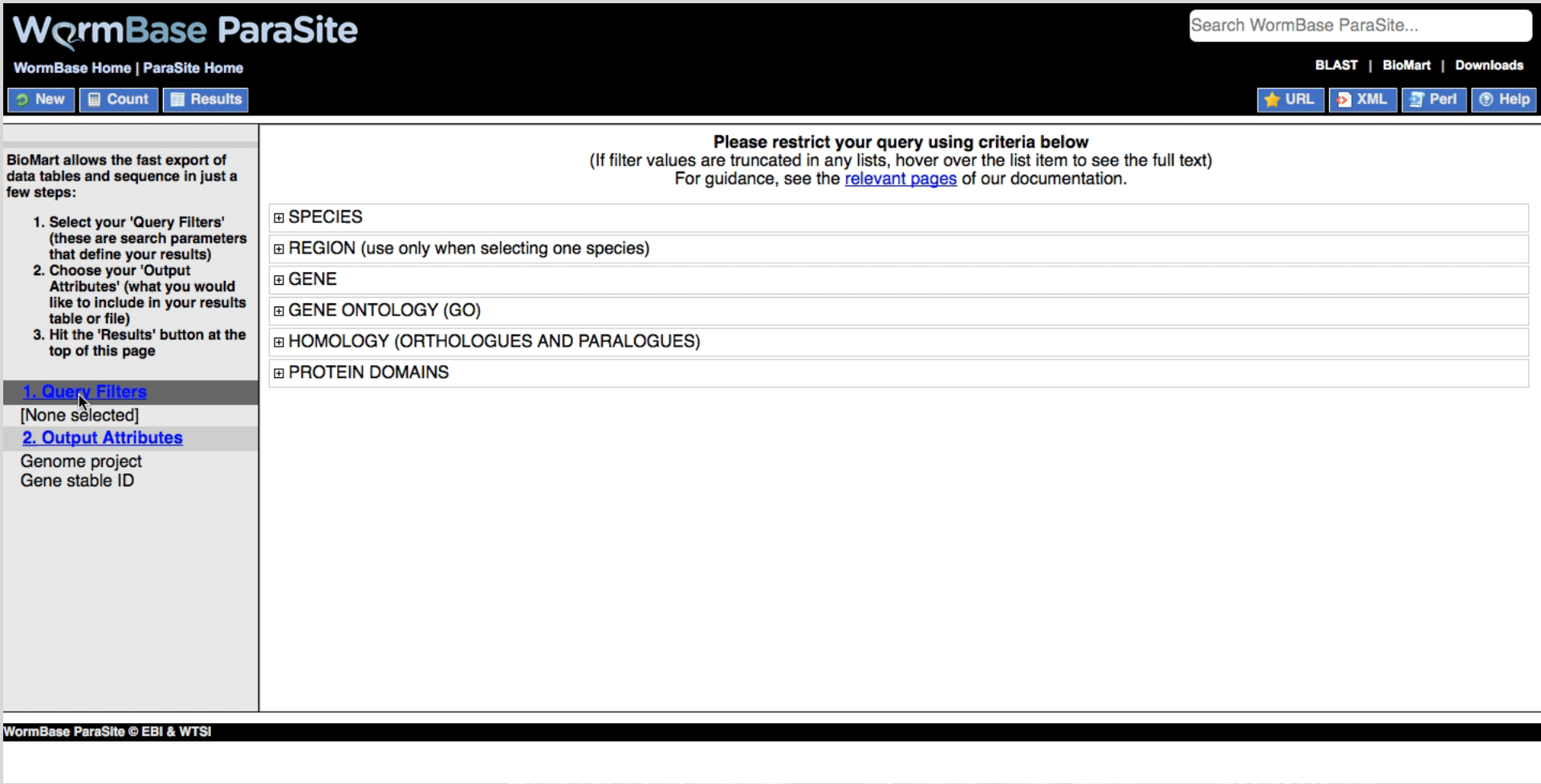
Exercise 2


WormBase ParaSite
Search WormBase ParaSite...
WormBase Home | ParaSite Home
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
WormBase ParaSite © EBI & WTSI

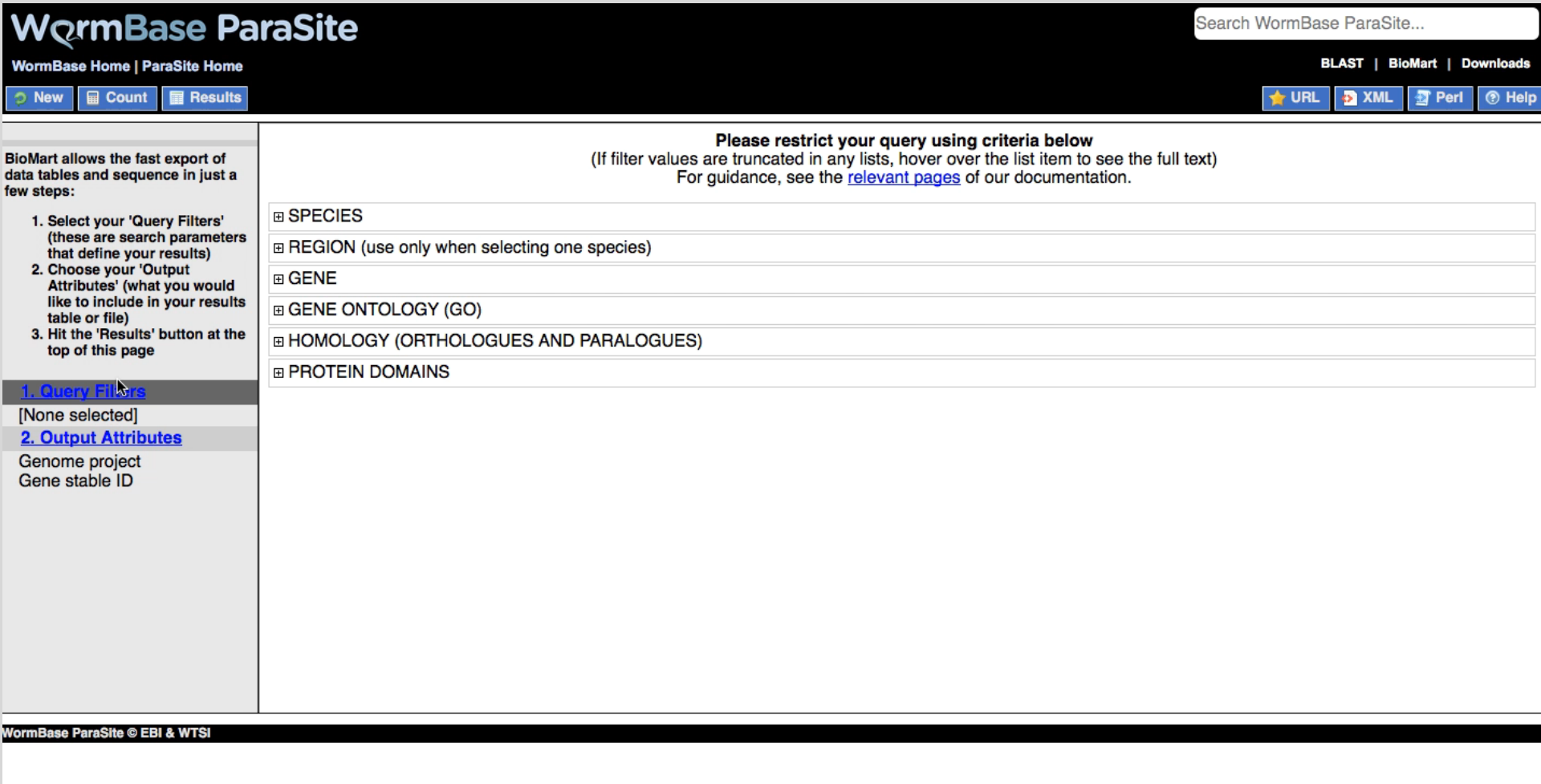
Exercise 3


WormBase ParaSite
Search WormBase ParaSite...
WormBase Home | ParaSite Home
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
WormBase ParaSite © EBI & WTSI

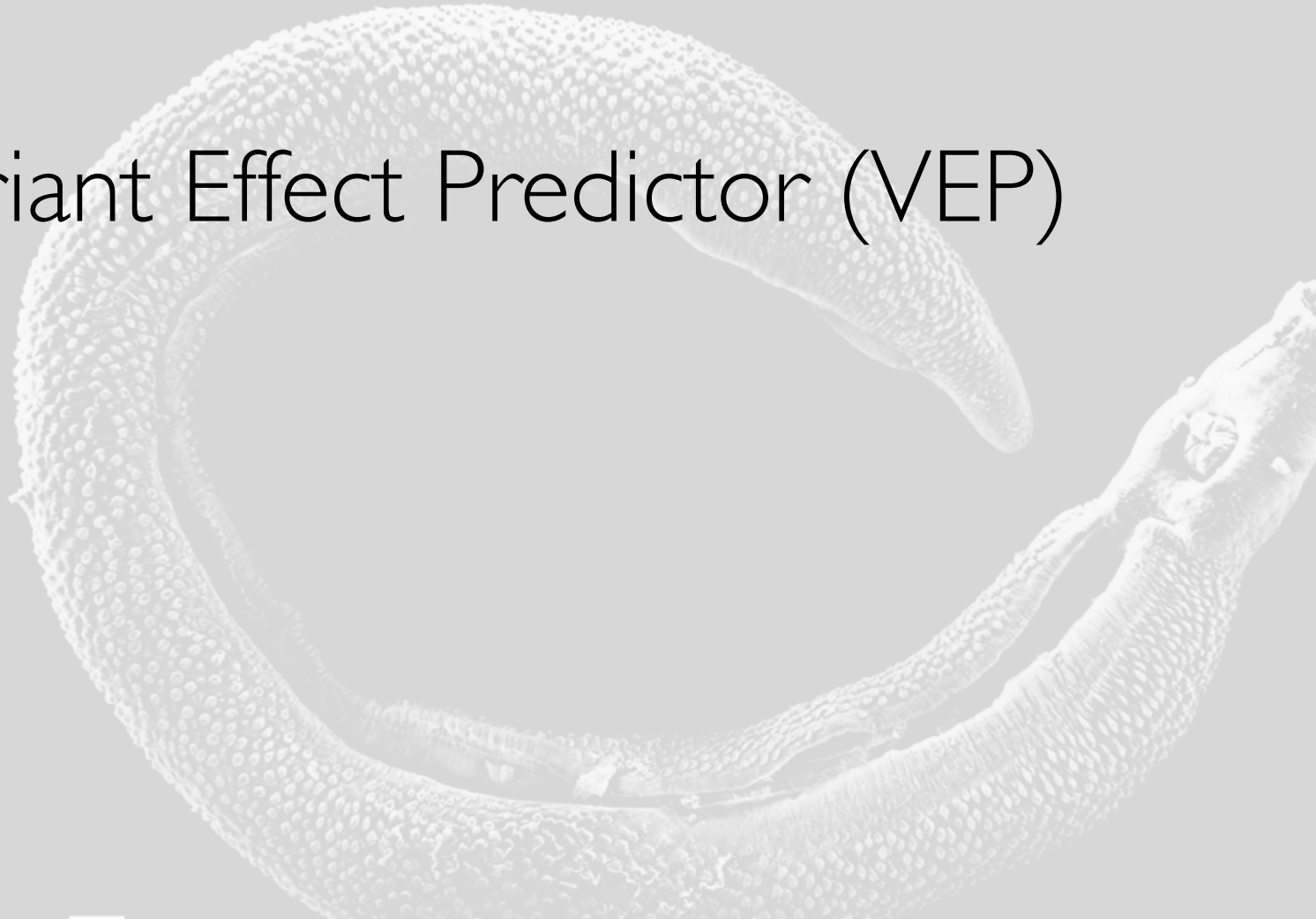
Exercise 4


WormBase ParaSite
Search WormBase ParaSite...
WormBase Home | ParaSite Home
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
WormBase ParaSite © EBI & WTSI

Exercise 5


WormBase ParaSite
Search WormBase ParaSite...
WormBase Home | ParaSite Home
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
WormBase ParaSite © EBI & WTSI

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)

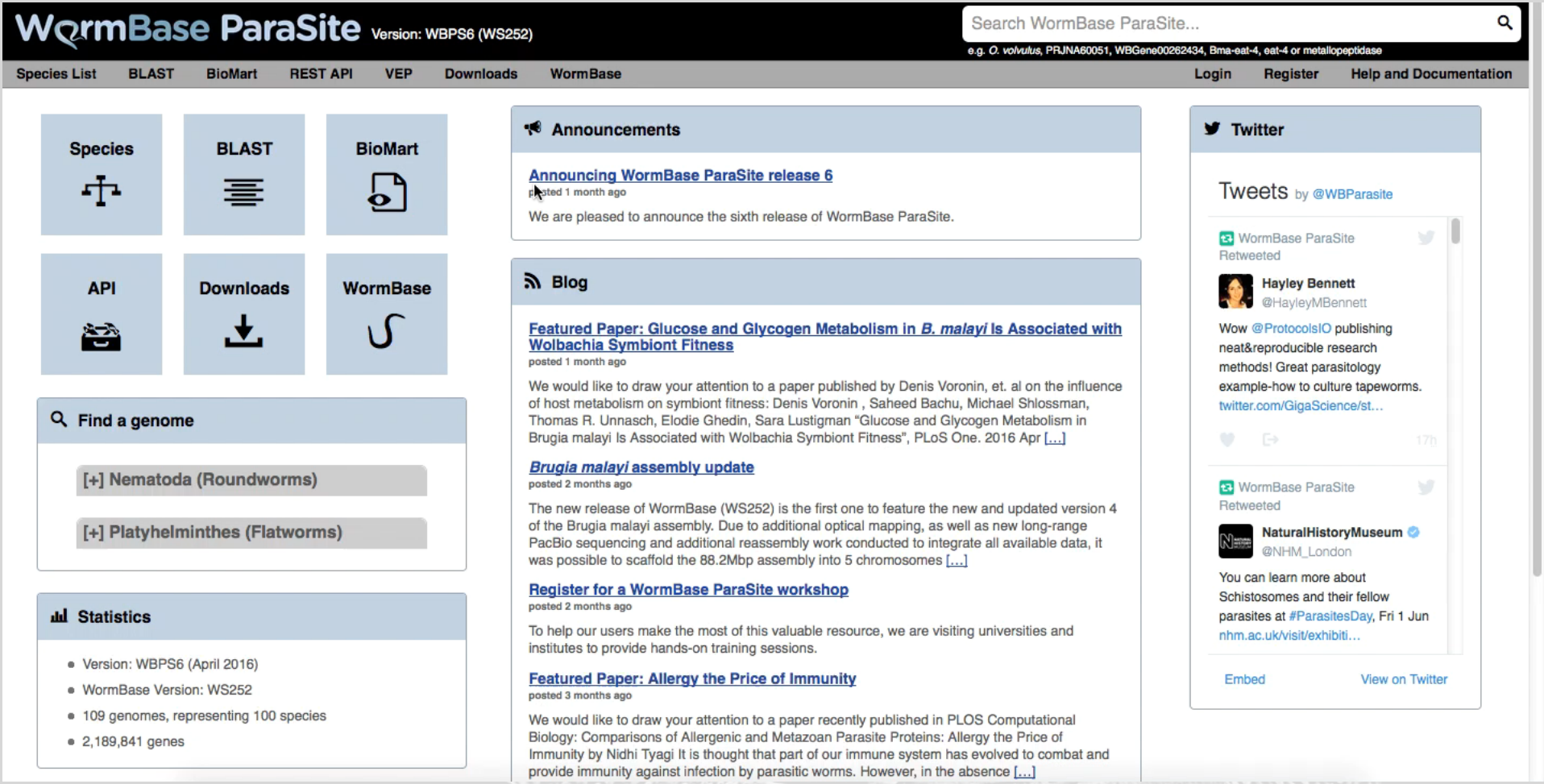
Introduction to the VEP

- 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

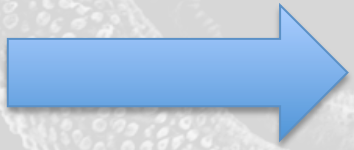
- Runs online using our servers (ideal for small number of variants)
- Possible to download and run locally for large amounts of data
  - Pre-computed "VEP caches" available for use offline

Using the VEP



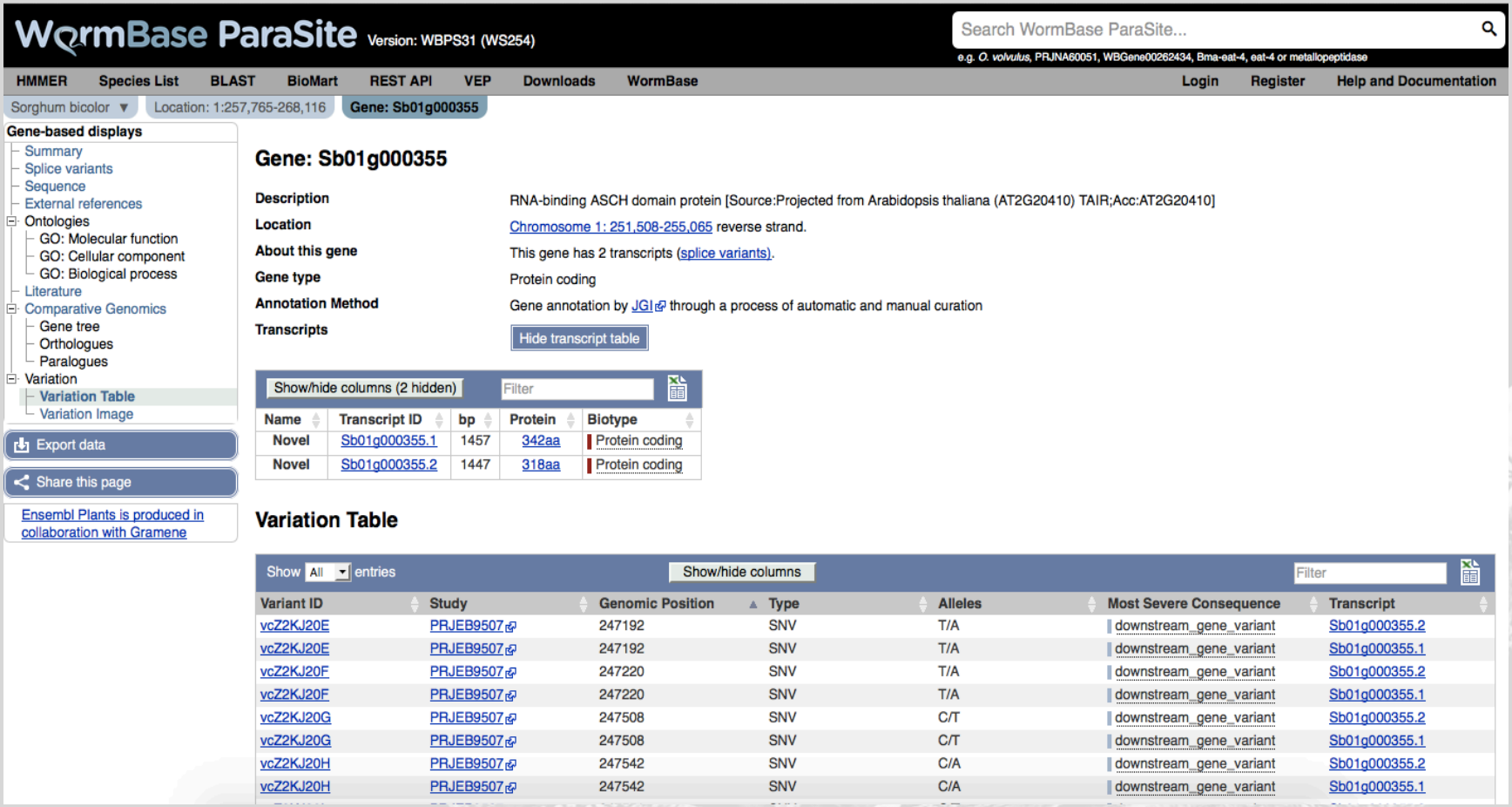
Other variation data

- 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

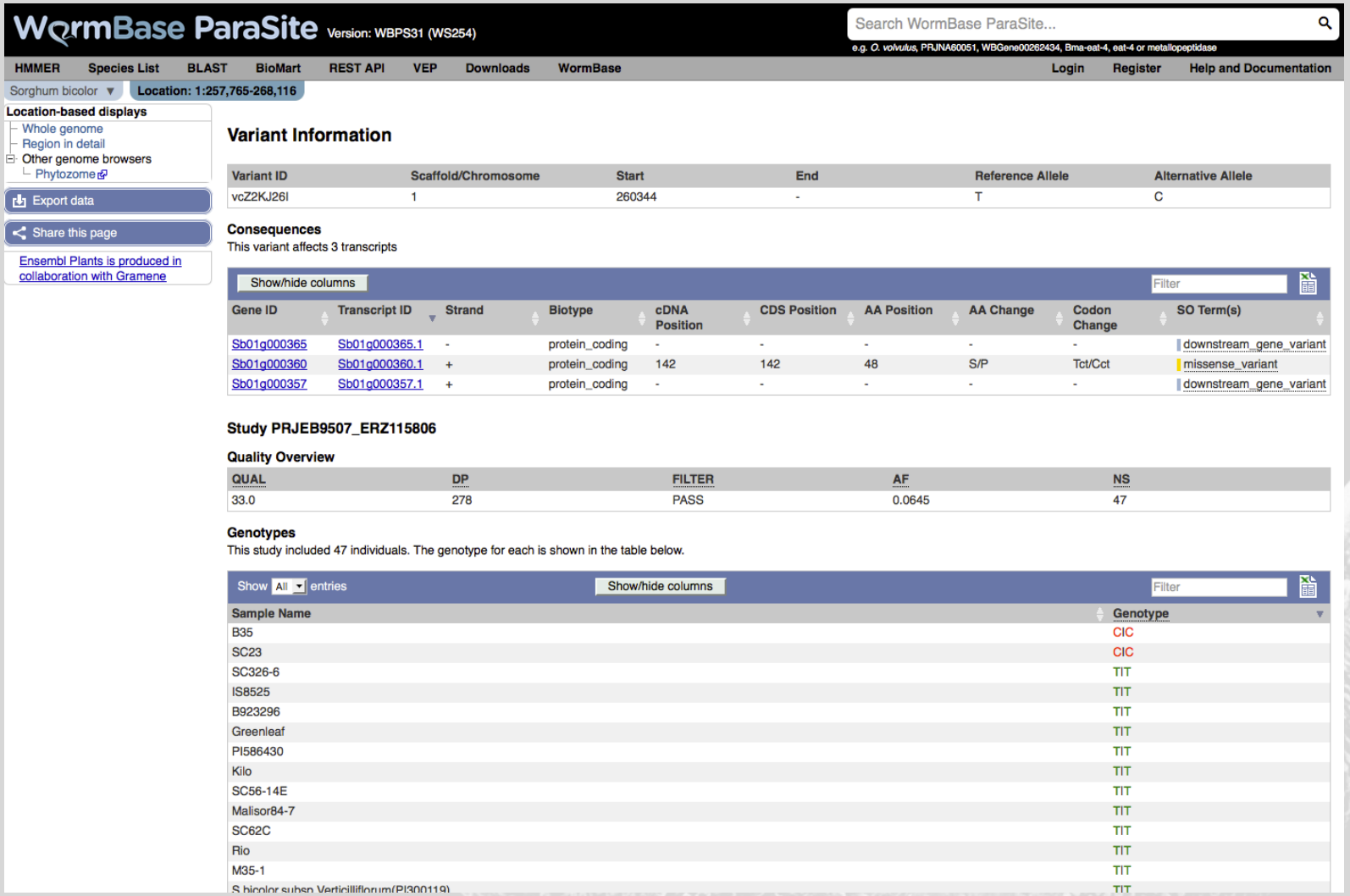


WormBase ParaSite

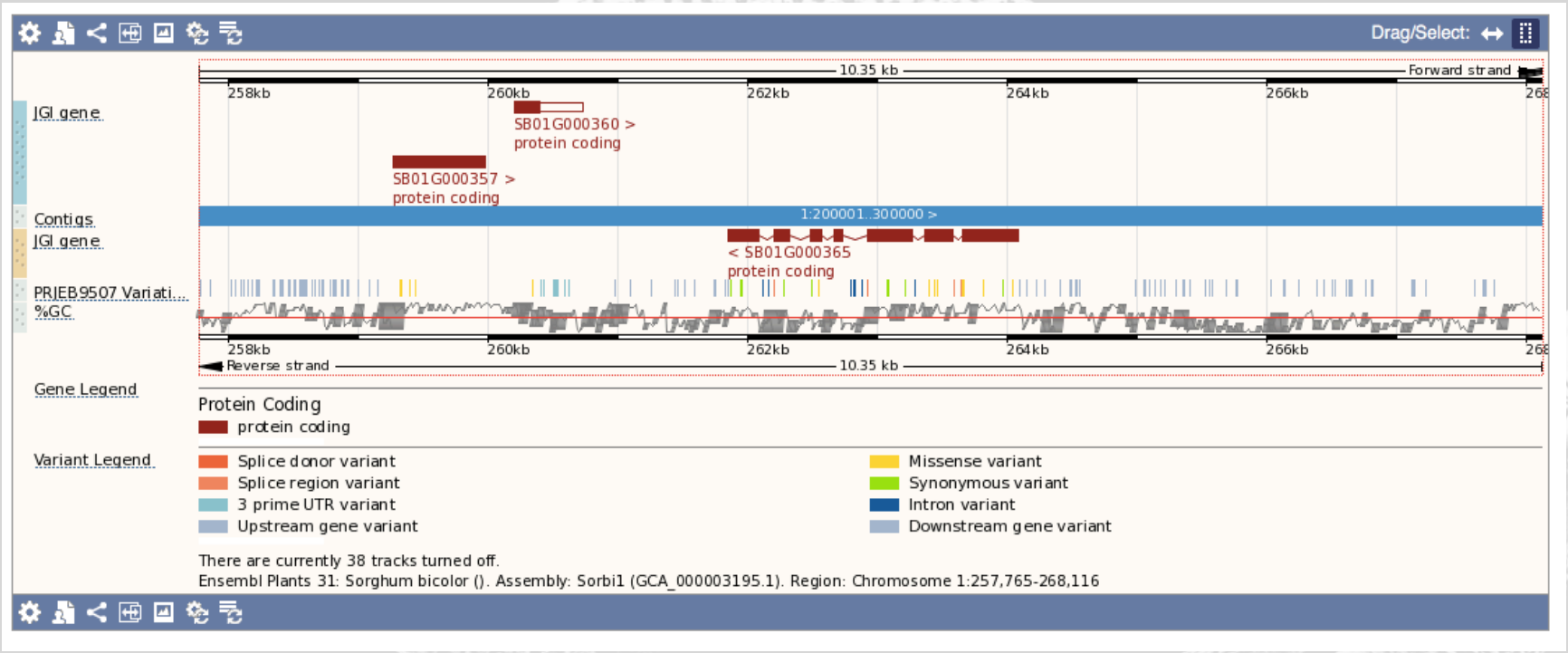
Gene Variation Table



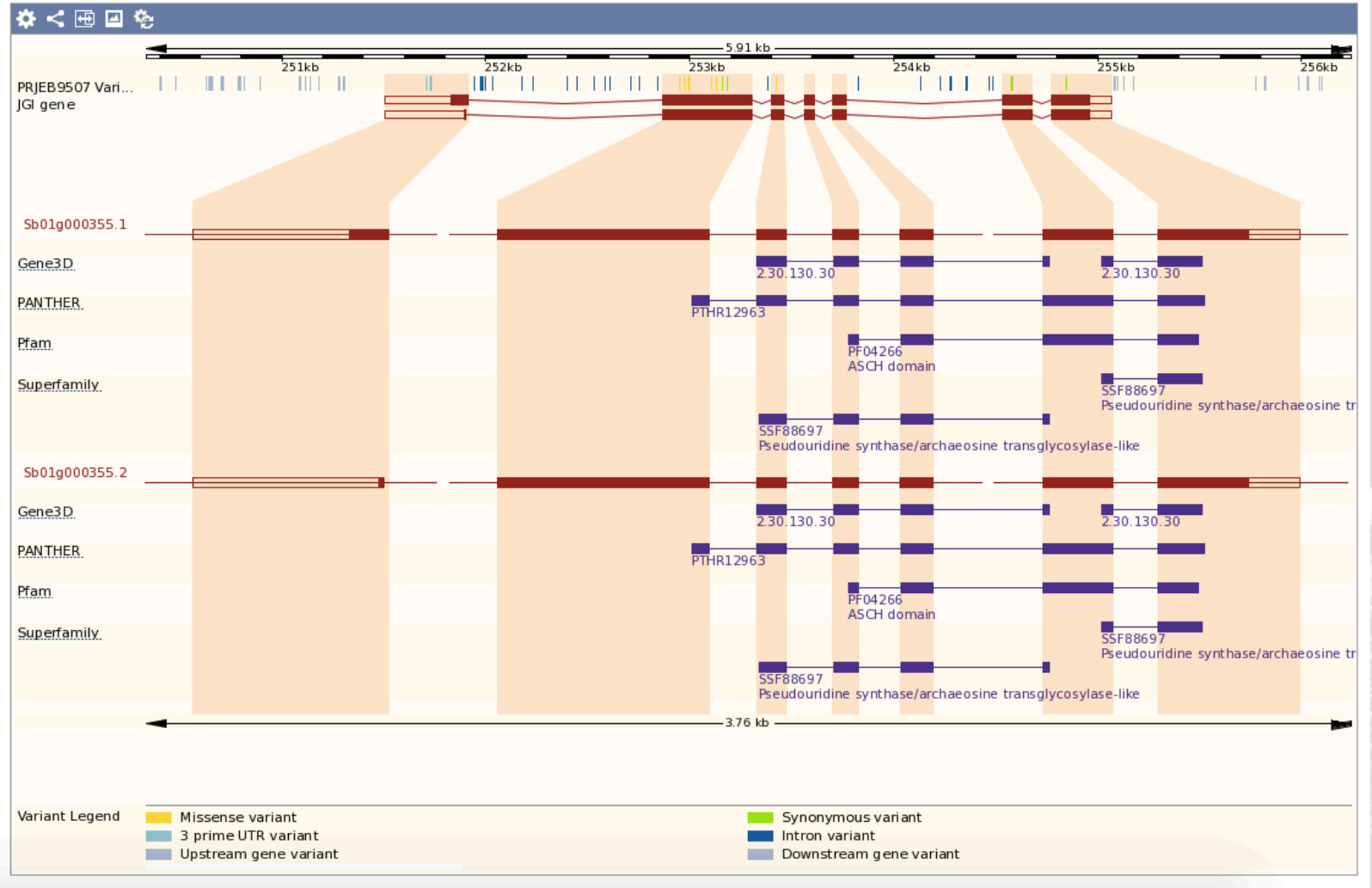
Variant Information Page



Genome Browser Tracks



Variant/Transcript Image

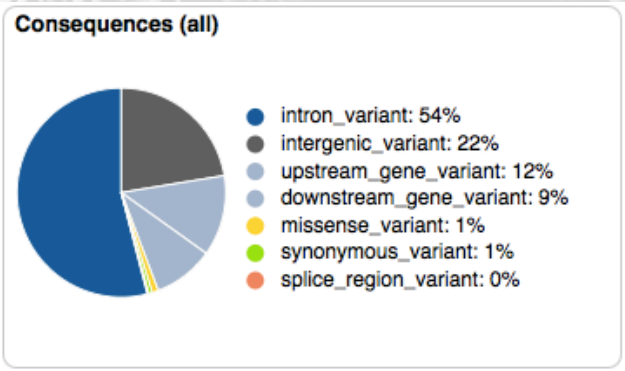


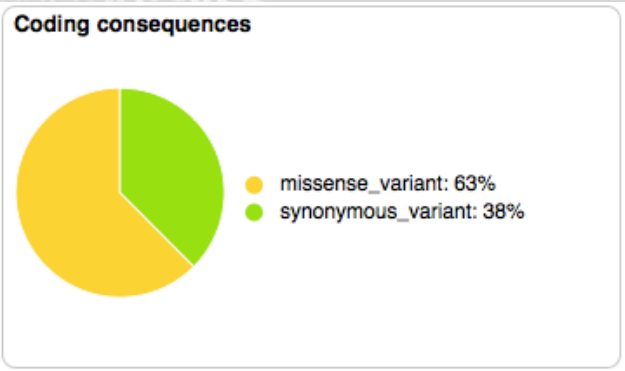
Exercises

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

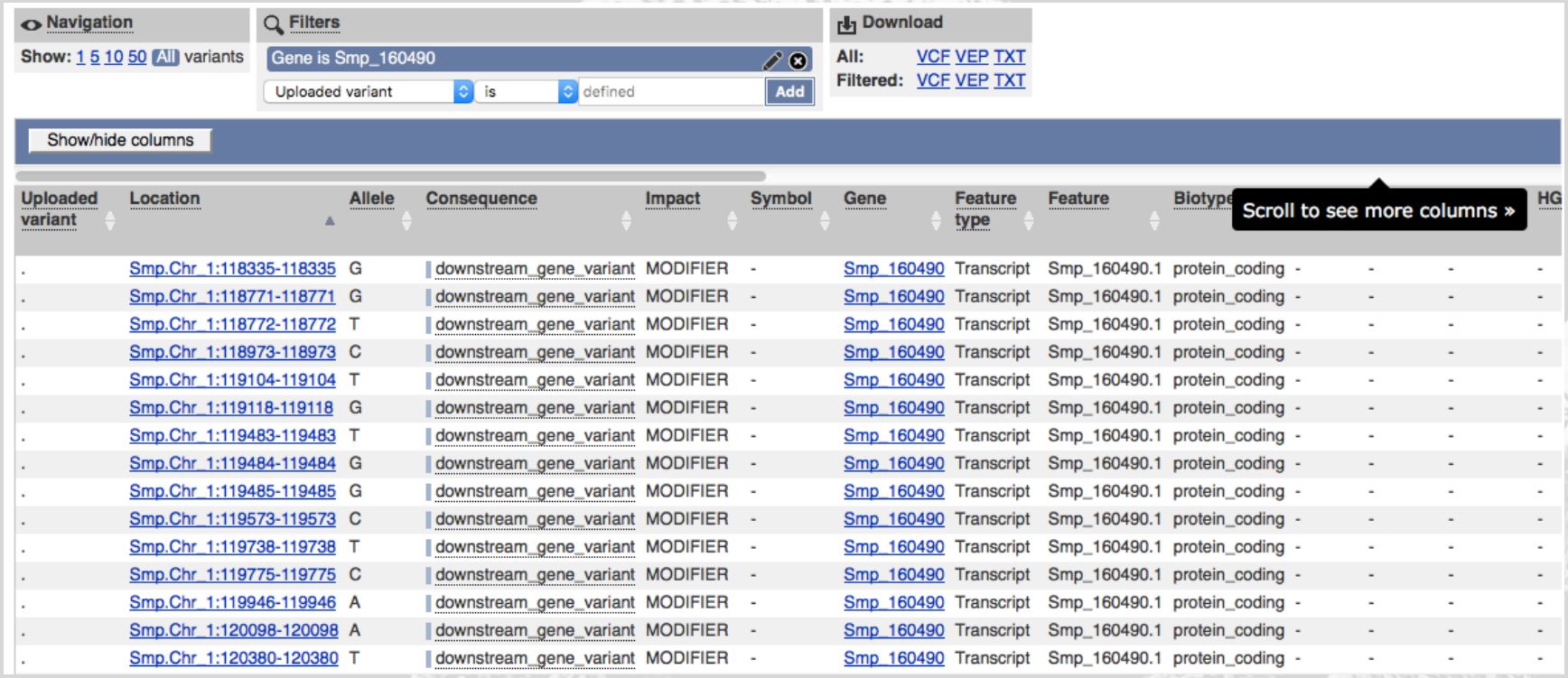
Exercise Solutions

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





Exercise Solutions



Navigation

Show: 1 5 10 50 All variants

Filters

Gene is Smp_160490

Uploaded variant | is | defined | Add

Download

All: VCF VEP TXT

Filtered: VCF VEP TXT

Show/hide columns

Scroll to see more columns »

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	-	-	-	-