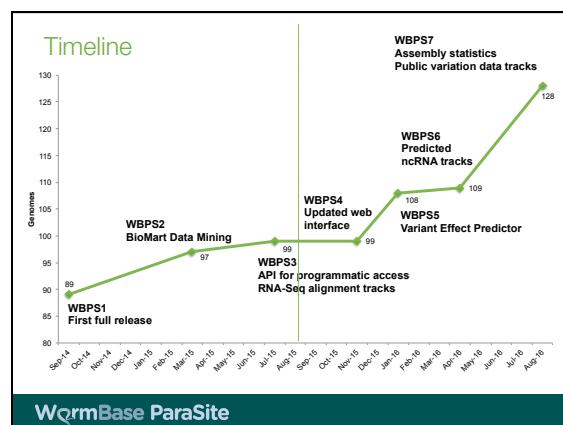


## Overview

- Introduction to WormBase ParaSite
- Website interface update
- New and updated data
- Features and tools
- Q&A session

- Portal to collect and display genomes of helminths (both nematodes and platyhelminthes)
- Integrated with other genomic databases
- Data annotated (mainly) using computational pipelines
- Updated approximately three times per year
- Community driven resource  
(new data and feature requests are always welcome!)
- Free and open-access at [parasite.wormbase.org](http://parasite.wormbase.org)



# Help & Documentation

**WormBase ParaSite** Version: WPS7 (WS08)

Sessions List    BLAST    Search    REST API    VSP    Downloads    Microarray    Search Microarray Publications

Log in    Register    Help and Documentation

**Help and Documentation**

- Getting Started
  - Setting Up Your Environment
  - Session List
  - REST API
  - VSP
  - Downloads
  - Microarray
- Browsing
  - Admin Custom Tracks
  - Annotations
  - ChIP-seq
  - Comparative Genomics
  - Gene Pages
  - Genes
  - Proteins
  - Regions (detail view)
  - Transcripts
  - Transgenic Lines
  - Transgenic Pages
  - Transcripts
  - Whole Genome
  - Tables
  - Tools
  - REST API
  - BLAST
  - Microarray
  - Downloads
  - Help and Documentation
- Frequently Asked Questions
- Announcements
- Data Integrity Policy
- Downloads

**Help and Documentation**

- Getting Started
  - Front Page
  - Search
  - Annotations
  - Comparative Genomics
  - Gene Pages
  - Genes
  - Regions (detail view)
  - Transcripts
  - Transgenic Lines
  - Transgenic Pages
  - Whole Genome
- Browsing
  - Admin Custom Tracks
  - Annotations
  - BLAST File Format
  - ChIP-seq
  - Comparative Genomics
  - Gene Pages
  - Genes
  - Regions (detail view)
  - Transcripts
  - Transgenic Lines

# Help & Documentation

**WormBase ParaSite** WormBase ParaSite

Help & Documentation | Home | Search | Log In | Help & Documentation

**Browsing**

Help topics

- [About WormBase ParaSite](#)
- [Using WormBase ParaSite](#)
- [Data Sources](#)
- [Data Types](#)
- [Browsing](#)
- [Adding Tools](#)
- [About WormBase ParaSite](#)
- [Contact Us](#)

**INDEXES** [TRINITY](#)

For some species, where data is available, WormBase ParaSite displays policy available transcriptomic data as tracks aligned to the genome. These can be seen in the [Location view](#).

[!\[\]\(6af8fb3374762cb7dc918a112e102b36\_img.jpg\)](#)

You can visualize your own RNA, [Add the Custom Tracks](#)

Having one or the short track name to the left of the track brings up a window providing additional metadata about the track, and options to configure the track.

[!\[\]\(31dc830bf8206b94b5a585ce61ce9013\_img.jpg\)](#)

Metadata included under the  icon are:

- DNA sequence ID
- DNA strand ID
- RNA sequence ID (where available)

## Features and tools

- Comparative Genomics
  - Quality Statistics
  - BLAST
  - Variant Effect Predictor (VEP)
  - BioMart
  - Genome Browser Tracks (RNA-Seq, ncRNA, Variation)
  - Programmatic Access (REST API)  
  - Bed = not covered in this session, speak to us at poster

WormBase ParaSite

## New and updated genomes

- 21 new genomes:
    - 16 new *Trichinella* genomes
    - Alternate assemblies for: *Echinococcus granulosus*, *Onchocerca ochengi*, *Toxocara canis* and *Wuchereria bancrofti*
    - Free-living flatworm *Macrostomum lignano*
  - Updated genomes:
    - New *Brugia malayi* assembly and annotation
    - Updated annotations: *Echinococcus multilocularis*, *Hymenolepis microstoma* and *Onchocerca ochengi*

WormBase ParaSite

- All genes are aligned against all others
- Gene family trees generated
- Orthologues/paralogues predicted
- Use this data to produce gene coverage and quality statistics
- Re-calculated at each release to incorporate new or updated genomes

WormBase ParaSite

## Frequently asked question

"How do I know whether a genome is high or low quality?"

We now approach this in two ways:

- Statistics based on comparative genomics data
- BUSCO and CEGMA scores for gene set coverage

WormBase ParaSite

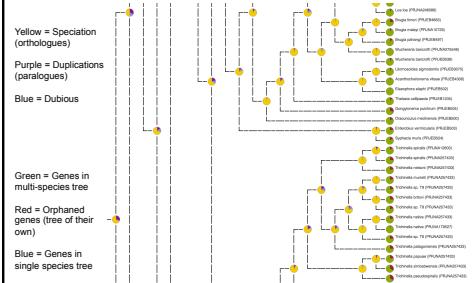
## Comparative Genomics – Summary Statistics

Species	Genes		Sequences		Show hide columns		Coverage	# species in multi-species tree	# genes in multi-species tree
	# genes in tree	# genes in tree	# sequences in tree	# sequences in tree	Genes in tree	Genes in tree			
Ascaris suum virus	10297	10297	9753	644	46	9708	46	41	46
Ascaris suum	30188	30188	26750	3448	312	25408	4933	1230	4933
Ascaris suum pseudocount	36687	66583	26879	10812	2915	22980	5592	273	5592
Ascaris suum pseudocount (PRJEB100)	36687	10882	14801	961	18	14913	510	377	510
Ascaris suum pseudocount (PRJEB100)	36687	27485	27485	3346	231	23928	518	3706	518
Ascaris suum pseudocount (PRJEB100)	14030	14030	12665	1635	106	12779	842	88	842
Ascaris suum pseudocount (PRJEB100)	13417	13417	11955	1462	90	11885	621	127	621
Ascaris suum pseudocount (PRJEB100)	20971	20971	16469	4475	1056	15490	2998	26	2998
Ascaris lumbricoides	23804	23804	17797	587	153	17544	1253	51	1253
Ascaris suum (PRJNA0037)	15260	15260	13574	1696	4	13570	158	117	158
Ascaris suum (PRJNA0031)	18542	18542	15523	3019	27	15496	305	457	305
Bugia malayi (PRJNA10729)	11021	11021	10553	968	16	10337	349	40	349

Green = Genes in multi-species tree  
Red = Orphaned genes (genes not in a tree)  
Blue = Genes in single species tree

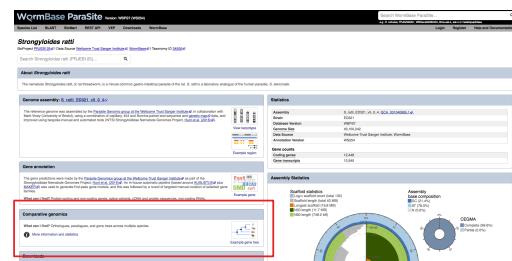
WormBase ParaSite

## Comparative Genomics – Summary Statistics



WormBase ParaSite

## Comparative Genomics – Summary Statistics



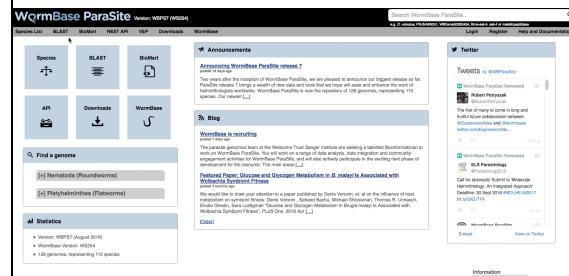
WormBase ParaSite

## Assembly quality statistics

- We additionally calculate statistics which show the quality of the assembly and gene set prediction
- CEGMA and BUSCO scores calculated for each genome:
  - CEGMA looks for a set of highly conserved genes found in most eukaryotes
  - BUSCO looks for single-copy orthologues found in more than 90% of animal species
  - In both cases, higher score generally represents a higher quality genome assembly and gene set annotation

WormBase ParaSite

## Quality metrics



WormBase ParaSite

### Quality metrics

Species		Assembly		File	
Species Name	Provider	Assembly ID	CSGMA	BSUCCO	Size (MB)
<i>Steinermettia carpocapsae</i>	California Institute of Technology	RWbase3	5, contig v1 submitted	PRJNA202124	239,569
<i>Caenorhabditis elegans</i>	WormBase	RWbase35		PRJNA213738	17,493,823
<i>Caenorhabditis briggsae</i>	WormBase	C54		PRJNA213721	17,485,433
<i>Caenorhabditis brenneri</i>	WormBase	C_brenneri4.0.1b		PRJNA203336	377,350
<i>Steinermettia asperiferae</i>	California Institute of Technology	R, contig v1 submitted		PRJNA202142	80,783
<i>Panagrolimus redivivus</i>	WormBase	Pred		PRJNA194777	292,414
<i>Steinermettia fitchii</i>	California Institute of Technology	S, left_v1_submitted		PRJNA202011	47,473
<i>Caenorhabditis remanei</i>	WormBase	C_remanei10.0.1		PRJNA202027	436,512
<i>Alloelophontes sp. KR9201</i>	Wellcome Trust Sanger Institute	Phasmidopines_sp_KR9201_v2_5_4		PRJNA321209	137,193
<i>Steinermettia glaseri</i>	California Institute of Technology	S, right_v1_submitted		PRJNA202034	37,382
<i>Caenorhabditis iricolorata</i>	WormBase	Caenorhabditis_iricolorata_1.0.1		PRJNA202027	20,021,886
<i>Strongyloides ratti</i>	Wellcome Trust Sanger Institute	S, ratti_2020_v5_0_4		PRJNA321209	11,033,554
<i>Strongyloides papillaeus</i>	Wellcome Trust Sanger Institute	S_papillaeus_LN_v2_1_4		PRJNA321209	84,047
<i>Bursaphelenchus xylophilus</i>	Wellcome Trust Sanger Institute	AR093115v1_submitted		PRJNA44027	849,851

WormBase ParaSite

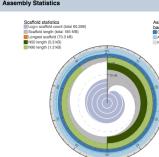
### Assembly quality widget

**Assembly Statistics**



Scaffold statistics  
Scaffolds > 100 kb: 1,076  
Scaffolds > 10 kb: 60,465  
Scaffolds > 1 kb: 1,076,000  
Scaffolds > 100 bp: 1,076,000  
Total length: 14.7 MB

**Assembly Statistics**



Assembly  
Contig distribution  
Scaffolds > 100 kb: 1,076  
Scaffolds > 10 kb: 60,465  
Scaffolds > 1 kb: 1,076,000  
Scaffolds > 100 bp: 1,076,000  
Total length: 14.7 MB

Learn more about this widget in our help section

This widget has been derived from the [assembly-stats.js](#) developed by the LepBase project at the University of Edinburgh.

Learn more about this widget in our help section

This widget has been derived from the [assembly-stats.js](#) developed by the LepBase project at the University of Edinburgh.

Adapted from code developed by Richard Challis for use in LepBase

WormBase ParaSite

### Variant Effect Predictor (VEP)

- Determine the effect of your variants:
- SNPs
- Insertions
- Deletions
- CNVs
- Structural variants

WormBase ParaSite

### Variant Effect Predictor (VEP)

- Input:
  - Coordinates of the variant and nucleotide change (or VCF file)
- Output:
  - Genes and transcripts affected by the variants
  - Location of the variants (e.g. upstream of a transcript, in coding sequence, 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

WormBase ParaSite

### Using the VEP

**WormBase ParaSite** Version: WMP86 (WS20)

[Search WormBase ParaSite](#) | [Log in](#) | [Register](#) | [Help and Documentation](#)

**Species List**: BLAST, BLAST+, REST API, VEP, Downloads, WormBase

**Announcements**: Announcing WormBase ParaSite release 5

**Blog**: Featured Paper: Glucose and Oxygen Metabolism in *B. malayi* Is Associated with Wormbase SymbolFitness

**Twitter**: [View Twitter feed](#)

**Blog**: [View Blog feed](#)

**Statistics**: WormBase ParaSite statistics

**Find a gene**: [\(1\) Nematoda \(Roundworms\)](#), [\(1\) Platyhelminthes \(Flatworms\)](#)

**Statistics**: WormBase ParaSite statistics

**Using the VEP**: The new release of WormBase (WS20) is the first one to feature the new and updated version 4 of the Variant Effect Predictor (VEP). This update includes many improvements, including: better PEG sequencing and additional assembly work conducted to integrate all available data, it also provides improved support for structural variants and more detailed gene annotation. [Read more](#)

**Request for a WormBase ParaSite workshop**: To help our users make the most of this valuable resource, we are raising awareness and interest in the use of WormBase ParaSite through a series of regional workshops. [Read more](#)

**Featured Paper: Allergy to the Price of Immunity**: We would like to draw your attention to a paper recently published in PLoS Computational Biology: Comparison of Allergen and Malaria Parasite Protein Allergy: The Price of Immunity. [Read more](#)

**European Variation Archive**: European Variation Archive

**WormBase ParaSite**

WormBase ParaSite

### Display of public variation data

- From release 7, we have begun importing data from the European Variation Archive (EVA)
- Make your data public by depositing in EVA
- Variation track will become available in WormBase ParaSite
- Speak to us for advice before depositing in EVA


→


→


WormBase ParaSite

## 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

WormBase ParaSite

Wormbase Home | Parasite Home

New Count Results

Blockout shows the last export of your query. Click here to clear it or edit steps.

Select your "Query Filter" (These are search parameters that will filter your results).  
 1. Choose your "Output" (Additional filters you would like to include in your results)  
 2. Click "Run Query" (After the top of this page)

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)

Please restrict your query using criter (If filter values are truncated in any lists, hover over the list for guidance, see the truncated values on the right)

eSPECIES

- eREGION (use only when selecting one species)
- eGENE
- eGENE ONTOLOGY (GO)
- eHOMOLOGY (ORTHOLOGUES AND PARALOGUES)
- ePROTEIN DOMAINS

Filters and attributes appear here

A
1 Smp_158080
2 Smp_158070
3 Smp_063300
4 Smp_204760
5 Smp_145060
6 Smp_145070
7 Smp_210640
8 Smp_169900
9 Smp_049930
10 Smp_132700
11 Smp_132700
12 Smp_139350
13 Smp_055760
14 Smp_139350
15 Smp_261410
16 Smp_175210
17 Smp_169250
18 Smp_128000
19 Smp_128000
20 Smp_079640
21 Smp_038870
22 Smp_213140
23 Smp_151540
24 Smp_151280
25 Smp_012010
26 Smp_181360
27 Smp_010260
28 Smp_199990

## 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*.

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

- Have an orthologue in *C. elegans*
- Do not have an orthologue in humans
- Are associated with transmembrane signalling receptor activity

- 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)

### Contacting us

- Please get in touch with us to report:
  - Feature and data requests
  - Bug reports
  - Questions and comments
- E-mail: link at bottom of website
- Twitter: @WBParaSite

WormBase ParaSite

### Acknowledgements

- WormBase ParaSite
  - Kevin Howe
  - Myriam Shafie
  - Jane Lomax
  - Michael Paulini
  - Matt Berriman
  - Paul Kersey
- Ensembl and Ensembl Genomes
- Data Providers
- WormBase Funding: BBSRC, MRC, NHGRI

WormBase ParaSite