

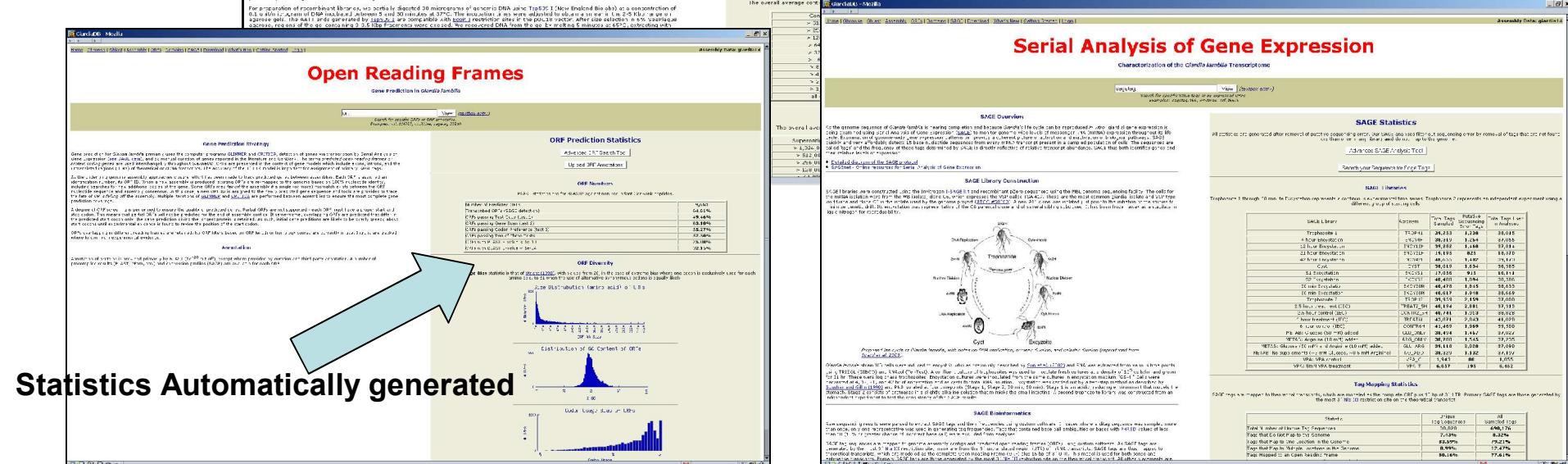
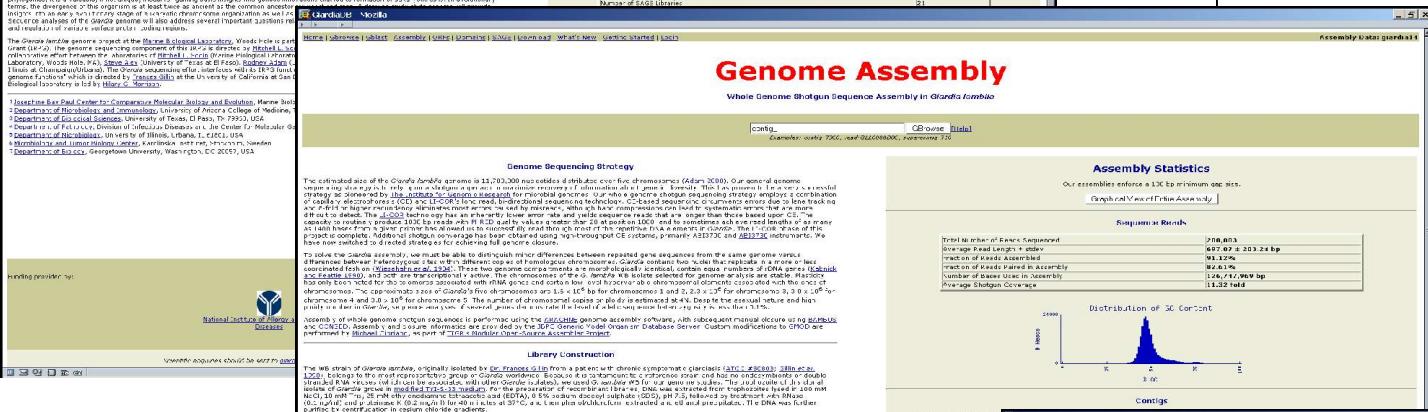
# **GiardiaDB**

(<http://www.mbl.edu/giardia>)

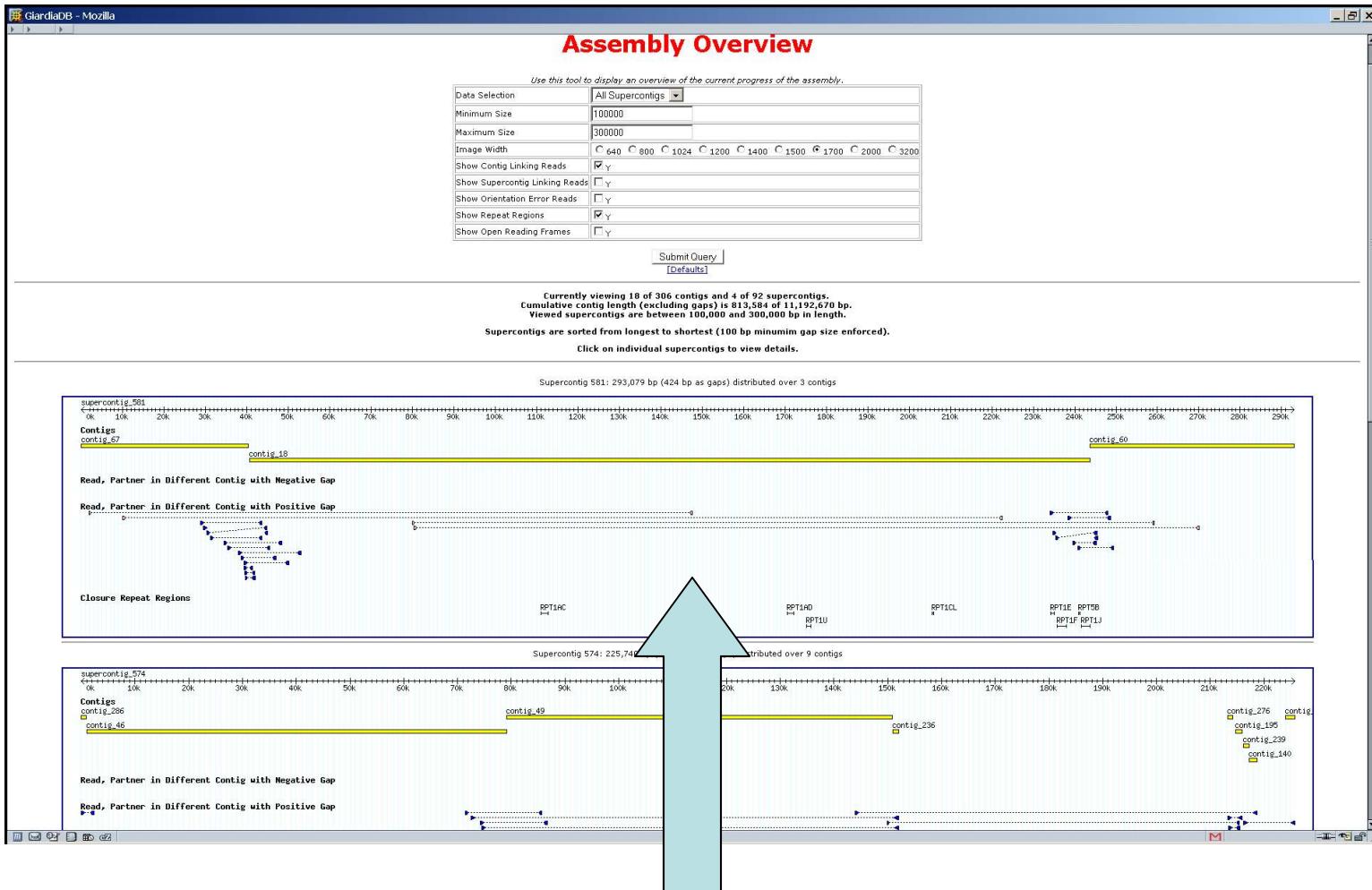
**Overview of features**

## Overview Pages

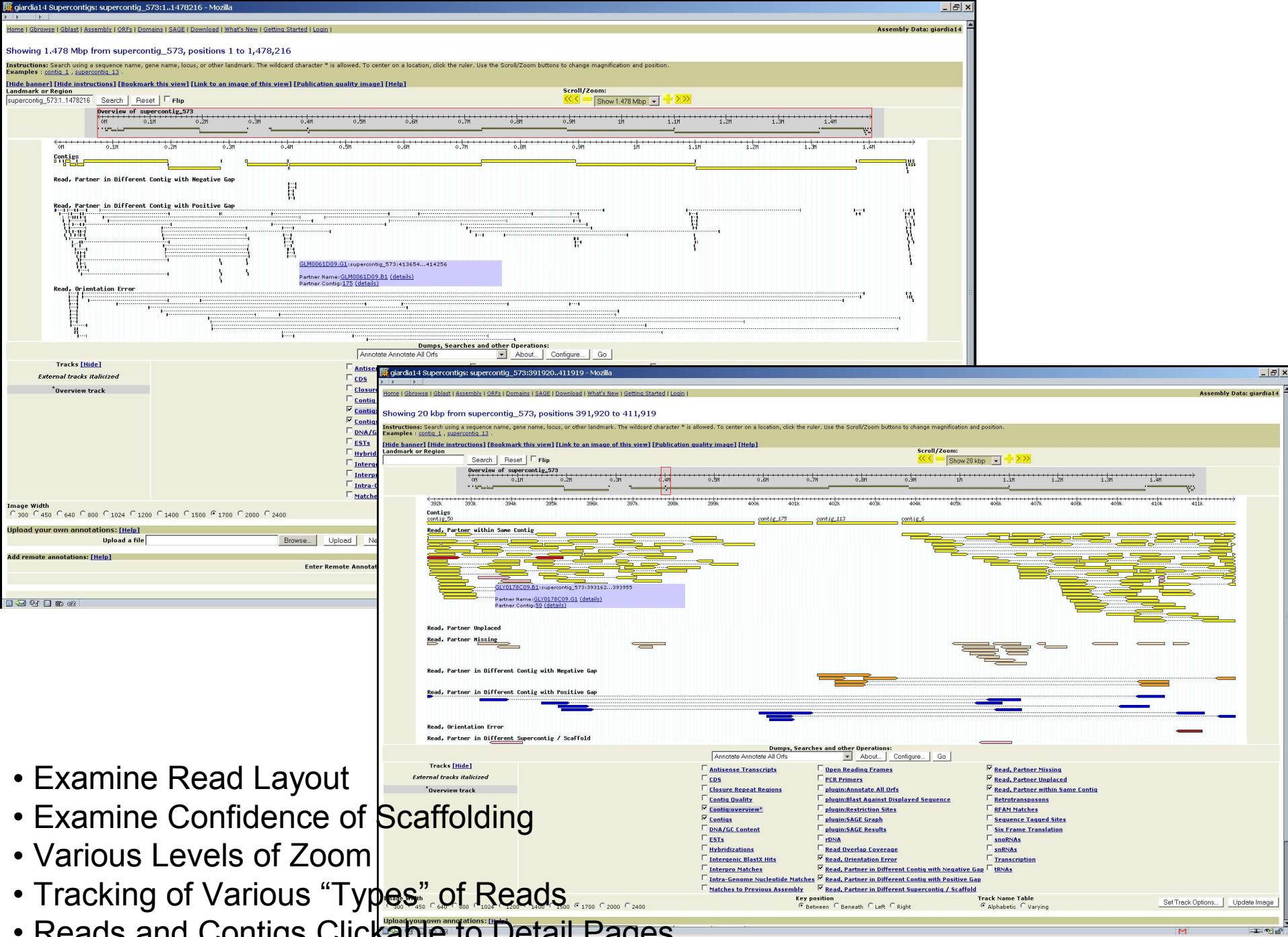
**Text blocks updateable via web interface.**



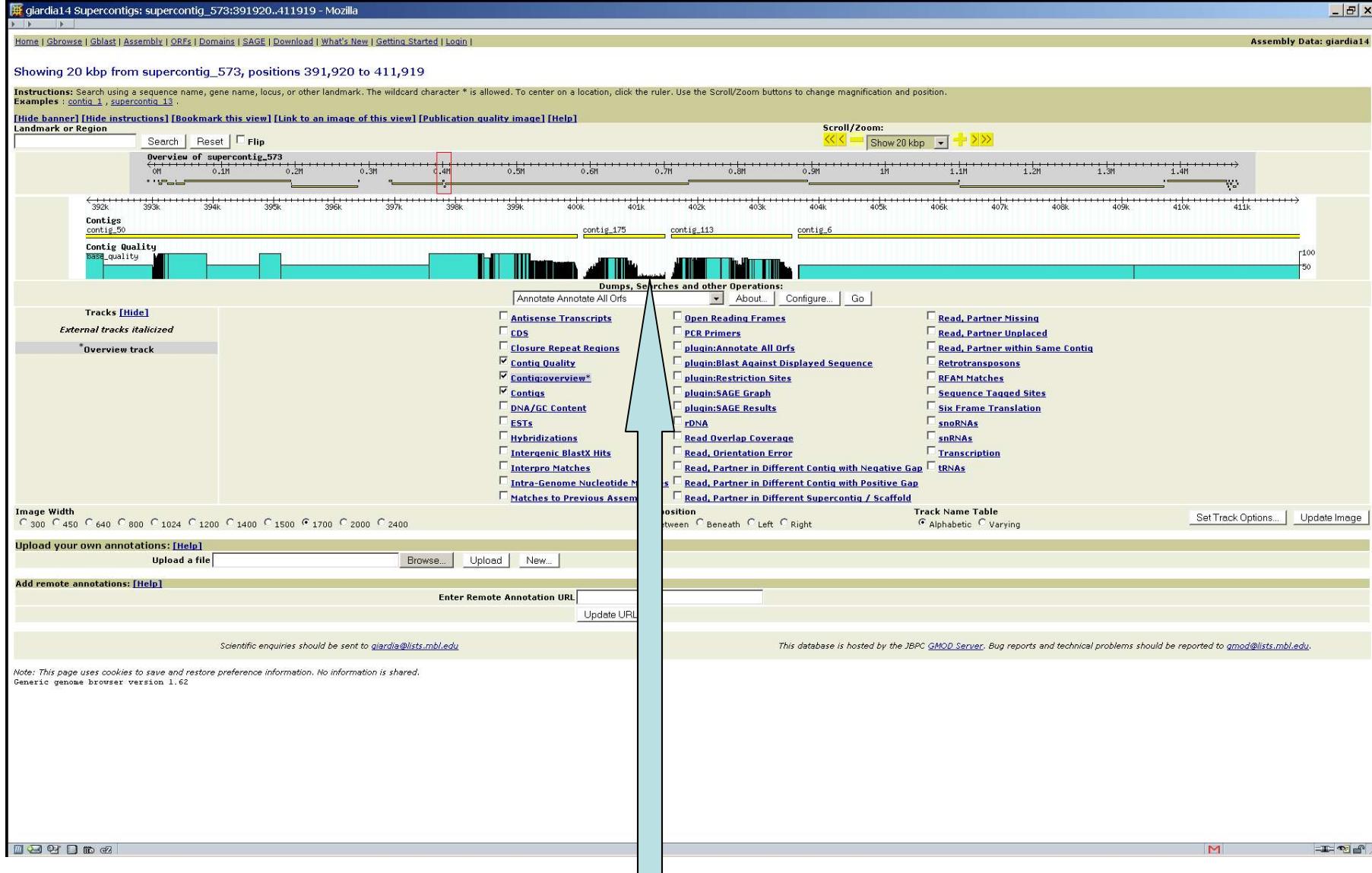
# Web based overview of Assembly Quality



Overview clickable to Gbrowse detail page.



- Examine Read Layout
- Examine Confidence of Scaffolding
- Various Levels of Zoom
- Tracking of Various “Types” of Reads
- Reads and Contigs Clickable to Detail Pages



## Examine Areas of Low Quality

GiardiaDB - Mozilla

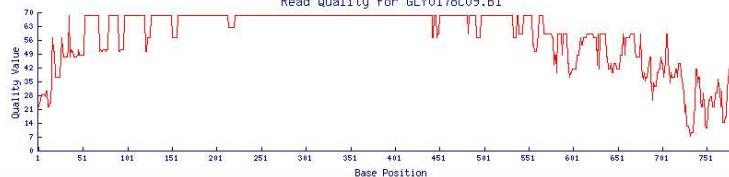
Home | Gbrowse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login | Assembly Data: giardia14

# Read Information

**Read Information**

Read: GLY0178C09.B1

<b>Requested Read</b> Download: Fasta Sequence BLAST Against own Genome: BLAST NT Sequence Against Contigs Read Name: GLY0178C09.B1 Sequencing Center: MBL Template Name: 133103 Plate: Well: Library: Y Trace End: F Trace Direction: Status: Read Status: Untrimmed Length: 793 Trimmed Length: 792 First Base of Trim: 1 Contig: 50 Supercontig: 573 Orientation: + Start Location in Contig: 62014 Stop Location in Contig: 62807 Read Mate Name: GLY0178C09.G1 Read Mate Contig: 50 Read Mate Supercontig: 573 Observed Insert Size: 1303 Given Insert Size: 3000 Given Insert StdDev: 300 Observed Insert StdDev:	<b>Partner Read</b> Download: Fasta Sequence Blast Vs own Genome: BLAST NT Sequence Against Contigs Read Name: GLY0178C09.G1 Sequencing Center: MBL Template Name: 133103 Plate: Well: Library: Trace End: R Trace Direction: Status: Read Status: Untrimmed Length: 757 Trimmed Length: 741 First Base of Trim: 16 Contig: 50 Supercontig: 573 Orientation: - Start Location in Contig: 62575 Stop Location in Contig: 63317 Read Mate Name: GLY0178C09.B1 Read Mate Contig: 50 Read Mate Supercontig: 573 Observed Insert Size: 1303 Given Insert Size: 3000 Given Insert StdDev: 300 Observed Insert StdDev:
--	--


  
 Read Quality for GLY0178C09.B1

Quality Value: 0, 7, 14, 21, 28, 35, 42, 49, 56, 63, 70  
 Base Position: 1, 51, 101, 151, 201, 251, 301, 351, 401, 451, 501, 551, 601, 651, 701, 751, 79

Scientific enquiries should be sent to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu).

This database is hosted by the JBCG GMOD Server. Bug reports and technical problems should be reported to [gmod@lists.mbl.edu](mailto:gmod@lists.mbl.edu).

- Detailed Read Information
- View Quality of Read
- Download Sequence or Chromatogram
- View Partners.

GiardiaDB - Mozilla

Home | Gbrowse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login | Assembly Data: giardia14

# Contig Information

Contig: 50

Download:Fasta Sequence

Contig	50
Super Contig	573
Contig Length	68869
Ordinal Number	14
Supercontig Length	1336650
Supercontig Length (with 100bp minimum gap)	1478216
Contigs in This Supercontig	32
Gap Before Contig	0
Gap After Contig	-1444
Start Position in Supercontig (with 100bp minimum gap)	331148

- View Details on Contigs
- Examine Contig Linking Reads
- Examine Possible Links to other Contigs or Supercontigs

Supercontig	Contig	Paired Links
589	2	2
578	3	2
572	4	1
573	6	8
578	10	1
589	12	1
573	16	2
589	21	2
579	22	1
579	23	1
577	27	2
573	28	6
571	29	2
589	32	1
611	38	1
577	55	1
622	61	1
573	98	2
618	215	1

Scientific enquiries should be sent to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu)

This database is hosted by the JGPC GMOD Server. Bug reports and technical problems should be reported to [gmod@lists.mbl.edu](mailto:gmod@lists.mbl.edu).

GiardiaDB - Mozilla

Home | Gbrowse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login | Assembly Data: giardia14

# Contig Linking Reads

Between Contig 50 and 98

Read Name	Orientation	Pair Name
EJ4198RA	-	EJ4198RA
KJ5625RA	-	KJ5625RA

GiardiaDB - Mozilla

Home | Gbrowse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login | Assembly Data: giardia14

Scientific enquiries should be sent to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu)

This database is hosted by the JGPC GMOD Server. Bug reports and technical problems should be reported to [gmod@lists.mbl.edu](mailto:gmod@lists.mbl.edu).

# Automated and Manual Annotation of Open Reading Frames

GiardiaDB - Mozilla

Home | Gbrowse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login | Assembly Data: giardia14

## Open Reading Frames

Gene Prediction in *Giardia lamblia*

Search for specific ORFs or ORF annotation.  
Examples: orf:114787, gmlm1hne, sagetag:32220

orf: View (textbox entry)

**Gene Prediction Strategy**

Gene prediction for *Giardia lamblia* primarily uses the computer programs GLIMMER and CRITICA detection of genes via transcription by Serial Analysis of Gene Expression (see SAGE data), and by manual curation of genes reported in the literature and GenBank. The terms *predicted open reading frames* or *protein coding genes* are used interchangeably throughout GiardiaDB. ORFs are presented in the context of gene models which include exons, introns, and the untranslated regions (UTRs) of theoretical or cDNA transcripts. The accuracy of the 3'-UTR model is important for assignment of primary SAGE tags.

As the underlying genome assembly approaches closure, effort has been made to track predicted genes between assemblies. Each ORF is assigned an identification number, its ORF ID. When a new assembly is produced, existing ORFs are re-mapped to the genome based on 100% nucleotide identity, including searching for new additional copies of the gene. Some ORFs may fall off the assembly if a single (or more) mismatch exists between the ORF nucleotide sequence and assembly consensus. In this case, a new ORF ID is assigned to the newly predicted gene sequence and tools are provided to trace the fate of ORFs falling off the assembly. Multiple iterations of GLIMMER and CRITICA are performed between assemblies to ensure the most complete gene prediction coverage.

A degree of ORF screening is performed to ensure the quality of predicted genes. Partial ORFs are not supported - each ORF must have a proper start and stop codon. This means that partial ORFs will not be predicted for the end of assembly contigs. If same-frame, overlapping ORFs are predicted that differ in the predicted start codon only, the gene prediction giving the longest protein is retained. As such, initial gene predictions are likely to be overly greedy about start codons until experimental evidence is found to revise the position of the start codon.

ORFs overlapping in different reading frames are retained. No ORF filters based on ORF length or homology scores are currently in use. Introns are tracked where known from experimental evidence.

**Annotation**

Annotation of proteins is provided primarily by BLAST ( $10^{-20}$  cut-off), except where provided by curation and third party annotation. A number of precompiled results (BLAST, PFam, etc.) and expression profiles (SAGE) are available for each ORF.

**ORF Prediction Statistics**

Advanced ORF Search Tool | Upload ORF Annotations

**ORF Numbers**

BLAST statistics are for BLASTP against non-redundant GenBank peptides.

Number of Predicted ORFs	9,663
Transcribed ORFs (SAGE detection)	64.61%
ORFs passing Test Code (test 1)	49.46%
ORFs passing Gen Scan (test 2)	63.10%
ORFs passing Codon Preference (test 3)	55.27%
ORFs passing Two of Three Tests	57.30%
ORFs with BLAST E-value < 1e-10	25.08%
ORFs with BLAST E-value < 1e-04	32.15%

**ORF Diversity**

The Codon Usage Bias statistic is that of Wright (1990), with values from 20, in the case of extreme bias where one codon is exclusively used for each amino acid, to 61 when the use of alternative synonymous codons is equally likely.

**Size Distribution (amino acid) of ORFs**

**Distribution of GC Content of ORFs**

**Codon Usage Bias of ORFs**

# Detailed information on Open Reading Frames

GiardiaDB - Mozilla

File Edit View Bookmarks Tools Window Help

http://bluebook.mbl.edu/perl/site/giardia14?page=showorf&orf=112681

Home Giardia GBlast Assembly ORFs Domains SAGE Download What's New Getting Started Login Admin Help Pages Update Statistics Import Data User Feature Annotation Sage GBrowse User

Assembly Data: giardia14

## ORF 112681

Vesicular-fusion protein SEC18 homolog

\*\*\* You can now click on orfs within this window to goto that orf detail page. To goto Gbrowse, click the link below\*\*\*\*

Goto Genome Browser

contig\_77

Open Reading Frames

orf:188438 G416426-PN [Drosophila pseudoboscana]

orf:188436 Hypothetical protein

orf:137725 Potential phospholipid-transporting ATPase IIB (HUSY-20)

orf:112681 Vesicular-fusion protein SEC18 homolog

orf:197501 Hypothetical protein

orf:112682 alpha-9 giardin, partial ORF

orf:18058 Giardin alpha-1 chain

Click here for publication quality image (SVG)

View in Genomic Context →

Various Statistics ↓

Annotation(s)

Gene Prediction Method: GLIMMER2

Annotation: Vesicular-fusion protein SEC18 homolog ([view details](#))

Last Updated: 2005-03-23 10:20:51 Annotator: Admin Deleted: N Blessed: Y

(add)

Protfun Annotation

Category: Gene Category: Enzyme category:Growth\_factor

Odds: 2.425

Enzyme: Enzyme

Odds: 2.055

Enzyme: Lyase (EC 4.-.-.-)

Odds: 2.783

Function: Purines\_and\_pyrimidines

Odds: 2.308

SAGE tags assigned to ORF 112681 ([view all](#))

Primary Sense: 1 Alternate Sense: 1 Primary Anti-Sense: 1 Alternate Anti-Sense: 1

Levels of Primary Sense Tag 2118

Expression Graph

Statistics ↑

Test	Score	Result
Test Code	Pass	10.90
Gene Scan	Pass	5.66
Codon Preference	Pass	0.858
Codon Usage*	55.952	n/a
Amino Acid Sequence Length	831	n/a
Contig Coordinates	contig_77:4257..6749	n/a

\*Uses EMBOSS default as reference: highly expressed S. pombe genes.

Features within Protein ↑

Cut-off for Pfam is an e-value of 0.0 or better

Type	Feature	Score / E-Value	Position
Pfam_ls	AAA_2: ATPase family associated with various	0.09	300..441
Pfam_ls	AAA_4: ATPase family associated with various	5.9e-69	304..503
Pfam_ls	AAA_5: ATPase family associated with various	0.034	304..542
Pfam_ls	AAA_3: ATPase family associated with various	0.24	304..440
Pfam_ls	DUF370: Domain of unknown function (DUF370)	0.87	607..681
interpro	hmmPfam PF00004.12 ATPase family associated with various cellular processes	4.4e-69	304..503
interpro	hmmsmart SM00382 no description	6.8e-18	301..448
interpro	scanregexp PS00674 AAA	8e-05	416..434

Community Driven Annotation Capabilities

Protein Features Automatically Annotated

Enzyme class:Lyase (EC 4.-.-.-)  
Functional category:Purines and pyrimidines

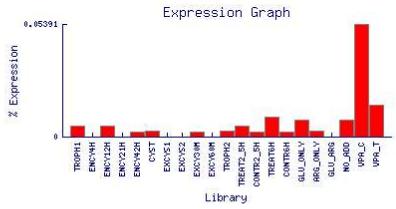
## Serial Analysis of Gene Expression (SAGE) Tags

**SAGE tags** are mapped to theoretical transcripts, which are modeled as the complete Open Reading Frame (ORF) plus 5' or 3' UTR. This model is used for both sense and anti-sense transcripts. Primary SAGE tags are annotated as **Sense** or **Anti-sense**. AS (alternate sense tag), AA (alternative anti-sense tag), UK (unknown - not resolved to an ORF), PA (primary anti-sense tag), AS (alternate sense tag), AA (alternative anti-sense tag), UK (unknown - not resolved to an ORF).

### Number of SAGE tags assigned to ORF 112681 [view all]

*Primary Sense*                    *Alternate Sense*                    *Primary Anti-Sense*                    *Alternate Anti-Sense*

Expression Levels of Primary Sense Tag 2118



# Expression Analysis Results

Automatically  
Generated  
Precompiled Results

### Features within Proteins

Cut-off for Pfam is an e-value of 0.0 or better

Type	Feature	Score / E-Value	Position
Pfam_ls	AAA_2: ATPase family associated with various	0.094	300..441
Pfam_ls	AAA: ATPase family associated with various	5.9e-69	304..503
Pfam_ls	AAA_5: ATPase family associated with various	0.034	304..542
Pfam_ls	AAA_3: ATPase family associated with various	0.24	304..440
Pfam_ls	DUF370: Domain of unknown function (DUF370)	0.89	607..681
interpro	hmmfpam PF00004.12 ATPase family associated with various	4.4e-69	304..503
interpro	hmmsmart SM00382 no description	6.8e-18	301..448
interpro	scanregexp PS00674 AAA	8e-05	416..434

## Precompiled Results

- Blast Reports
    - BLASTP Search Against SwissProt
    - BLASTP Search Against Non-Redundant GenBank
    - HHM Search Against Pfam\_1s
    - BLASTP Against MITOP Database
    - BLASTP Against RefSeqs Database
  - ProFun Results

Alignments and Trees

  - BLASTP alignment against RefSeqs database
  - Consensus tree against RefSeqs database

Tools

- [Create Multiple Sequence Alignment](#)
  - [BLAST NT Sequence Against Contigs](#)
  - [BLAST AA Sequence Against ORFs](#)
  - [BLAST AA Sequence Against Translated Contigs](#)
  - [BLAST AA Sequence Against GenBank](#)

## **Administration Tools**

- Edit Orf Coordinate  
Delete/Reinstate

## On the fly analysis →

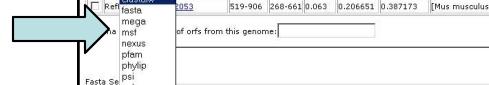
## Administration

**MFHSTPLSNTFNLIELKKQGCQMSGCGFSYKACKYPKPGKYPSTSLLTNCAYLCPSDYAAVLGNEKSDNVYLERNGHVV**

# On the Fly multiple alignments



# Various outputs



Format: **clustalw** | Create Multiple Alignment | Download Sequences | [Help](#) | [About](#) | [Contact](#) | [Logout](#)

## Multiple Sequence Alignment

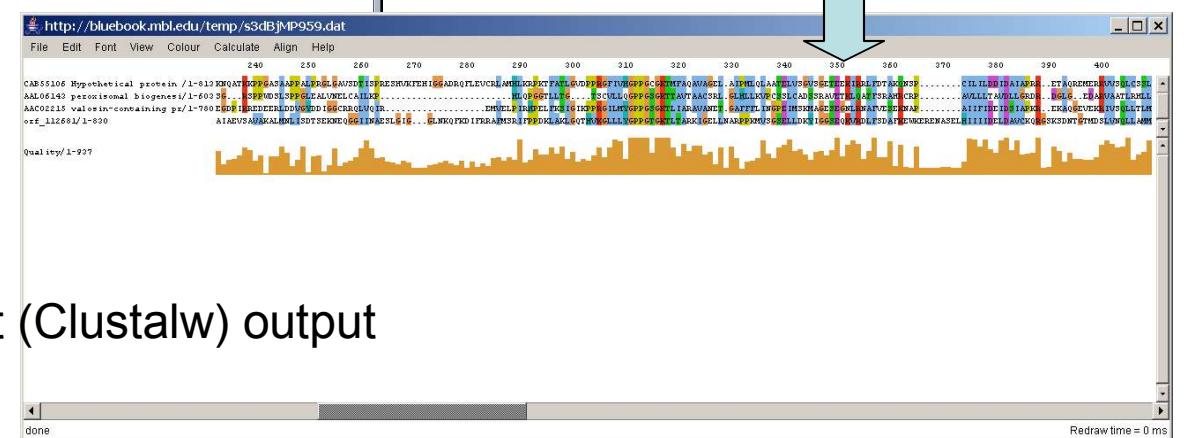
```
Create "file1.txt" var  
Multiple alignment 11/26/01.  
  
Personal Tax ID #: 26-2530756-147822  
Length: 997  
Number Sequences: 4  
Number Residues: 4626  
-----Cut below here for alignment file-----  
CUSTAL 9.1.61) multiple sequence alignment
```

**CAL2105** Hypothetical protein /I-510 1PCKAASDPAAPGQCLDPKPTTIDPDRSHHFFPHICCAADEQPLIVLDRBMLHLFDTCTPATI  
**XAD015** Hypothetical protein /I-503 1PCKAASDPAAPGQCLDPKPTTIDPDRSHHFFPHICCAADEQPLIVLDRBMLHLFDTCTPATI

**QUELLOS MÉTODOS SON LOS MEJORES?** C-210. ¿TIPOS DE ECONOMÍAS Y SUS CARACTERÍSTICAS? DIFERENCIAS ENTRE LAS ECONOMÍAS.

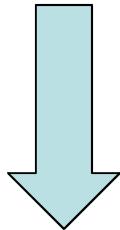
Digitized by srujanika@gmail.com

## Text (Clustalw) output

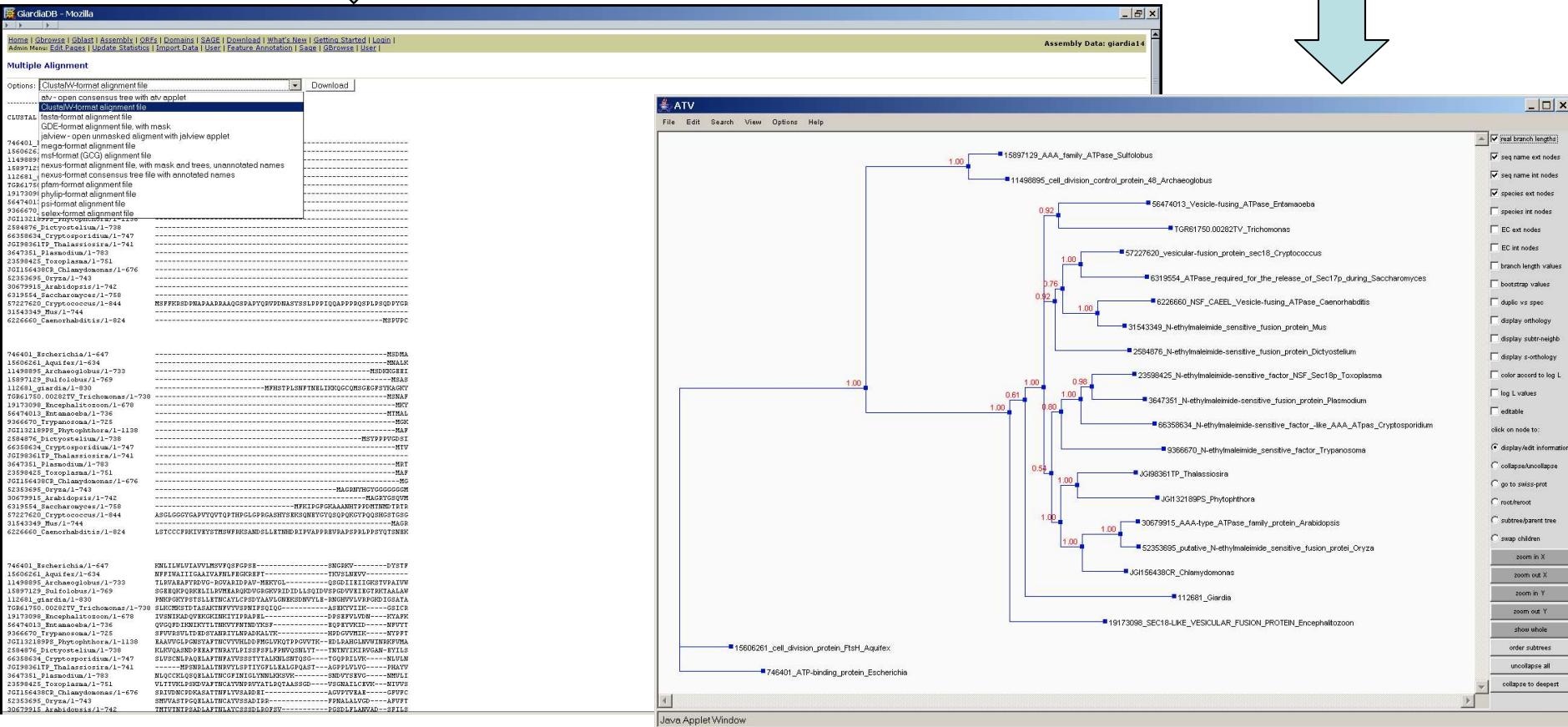
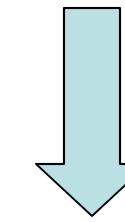


## Alignment/Tree Generation Pipeline on All Open Reading Frames

# Precompiled MUSCLE Alignments



# Precompiled MrBayes Trees



# Advanced Analysis of Precompiled Results

GiardiaDB - Mozilla

Home | Browse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login | Admin Menu: EOR Pages | Update Statistics | Import Data | User | Feature Annotation | SAGE | GBrowse | User |

## Assembly Data: giardia14

### ORF Analysis

**Search Criteria**

Select Precompiled BLAST Reports to Search

Swissprot  
 Non-Redundant GenBank

**Keyword(s) for searching BLAST reports.** Comma separated items are combined with OR automatically. The default setting does not apply a keyword filter.

**BLAST results must include hits to the following specific organisms** (use NCBI Taxonomy Names or ID Numbers). Comma separated items are combined with OR automatically. The default setting does not apply a taxonomy filter. You must include an E-Value if you want to search using this section.

Keyword:  Taxonomy:  Alpha proteobacteria E-Value:  1e-7

**Excluded keyword(s) for searching BLAST reports.** Comma separated items are combined with OR automatically. The default setting does not apply a keyword filter.

**BLAST results cannot include hits to the following specific organisms** (use NCBI Taxonomy Names or ID Numbers). Comma separated items are combined with OR automatically. The default setting does not apply a taxonomy filter. You must include an E-Value if you want to search using this section.

Keyword:  Taxonomy:  Eukaryota E-Value:  1e-10

Note on comma separated items: the use of OR means that both separate and joint instances of search terms will be accepted. For example, a search for *Entamoeba*, *Trichomonas* will find all hits with *Entamoeba* or *Trichomonas* or both.

**BLAST hit to MITOP**

If this option is selected, peptides matching the search criteria must additionally have a significant BLASTP hit to the Database for Mitochondria-Related Proteins.

Not Relevant  1e-7

**Interpro hit**

If a domain is listed below, peptides with that domain as defined by Interpro will be listed in the search. Use any string within the domain name or description. Comma separated items are combined with OR automatically. You may also place an AND between search terms to find multi-domain proteins (e.g. *Zinc finger* AND *kinase*, *lipid transferase* will find all proteins with either a zinc finger domain and a kinase domain (multi domain protein) OR proteins with a lipid transferase domain).

E-value:  1e-10

**Interacting Domains**

If this option is selected, peptides containing an interacting domain as defined by Pfam with the Pfam domain entered in will be returned. Comma separated items are combined with OR automatically. Search terms must be the short Pfam name (ex. *Itrb*, *trypsin*)

E-value:  1e-10

**Transcription**

If this option is selected, peptides matching the search criteria must additionally be from genes with transcription detected by Serial Analysis of Gene Expression, such that only genes with at least one assigned SAGE tag will be accepted. All available SAGE libraries are used in this search, even if the raw data has not yet been released to the public.

Evidence of Transcription Must Exist

**Regulation**

If this option is selected, peptides matching the search criteria must additionally be from genes with the regulation pattern chosen as obtained from Serial Analysis of Gene Expression data. A sum of the percent expression is taken for all sense tags for a particular ORF. (Note: A ORF might be shown as unregulated in one library and regulated in another. Total sampling between the libraries differ and the raw tag frequencies are the same. To be inclusive, the other library must have no sense sage tags set. Sage tags were assigned to ORFs based first on EST data if available, and then via an electronic assignment by averaging the 3 prime UTR length.)

Is:  Not Relevant

In:  Trophozoite 1

Vs:  Trophozoite 1

**Signal Peptide/Anchor**

GiardiaDB - Mozilla

Home | Browse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login |

## Assembly Data: giardia14

### Open Reading Frame Search Results

Scientific enquiries should be sent to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu)

orf ID	Annotations	Transmembrane Domains	Signal sequence
4622	Putative reductase	0	Signal peptide
4622	Peptidyl prolyl isomerase	0	Non-secretory protein
4622	Protein involved in glycosidic bond formation (EC 2.7.0.5)	0	Non-secretory protein
4648	PDR-type permease	0	Non-secretory protein
9311	Phosphotransfatty acid lyase (EC 3.1.1.6)	0	Non-secretory protein
10145	Nucleotide-diphosphate-sugar epimerases	0	Signed peptide
12630	Phosphotransfatty acid lyase	0	Non-secretory protein
14242	putative transport protein	0	Non-secretory protein
14241	Phosphoglycolate phosphatase (EC 3.1.1.10)	0	Non-secretory protein
14241	Phosphoglycolate phosphatase (EC 3.1.1.19)	0	Non-secretory protein
14552	Autocatalytic related protein	0	Non-secretory protein
16652	Acetyl CoA synthetase (NDF forming)	0	Non-secretory protein
16810	Pulldown resistance protein B	0	Non-secretory protein
16810	Pulldown resistance protein A	0	Non-secretory protein
21274	Long-chain-fatty-acid-CoA ligase 3 (EC 6.1.1.9)	0	Non-secretory protein
21298	F50, F5R repeat	0	Non-secretory protein
27208	Ribose 5'-phosphate isomerase (EC 5.3.1.6)	0	Non-secretory protein
28235	Acetyl-CoA acetyltransferase	0	Non-secretory protein
34134	Ribonuclease H (EC 3.1.26.4)	0	Non-secretory protein
66511	Acetyl-CoA acetyltransferase (NDF forming)	0	Non-secretory protein

Scientific enquiries should be sent to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu)

This database is hosted by the JEPG GMOD Server. Bug reports and technical problems should be reported to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu).

# Advanced Analysis of Precompiled Results

GiardiaDB - Mozilla

File Edit View Go Bookmarks Tools Window Help

Home | Gbrowse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login | Admin Menu: Edit Pages | Update Statistics | Import Data | User | Feature Annotation | Sage | GBrowse | User

Assembly Data: giardia14

## ORF Analysis

**Search Criteria**

Select Precompiled BLAST Reports to Search

Swissprot  
 Non Redundant GenBank

**Keyword(s) for searching BLAST reports.** Commas separated items are combined with OR automatically. The default setting does not apply a keyword filter.

**BLAST results must include hits to the following specific organisms (use NCBI Taxonomy Names or ID Numbers).** Commas separated items are combined with OR automatically. The default setting does not apply a taxonomy filter. You must include an E Value if you want to search using this section.

Keywords: TaxonID: E-Value: Not Relevant

**Excluded Keyword(s) for searching BLAST reports.** Commas separated items are combined with OR automatically. The default setting does not apply a keyword filter.

**BLAST results cannot include hits to the following specific organisms (use NCBI Taxonomy Names or ID Numbers).** Commas separated items are combined with OR automatically. The default setting does not apply a taxonomy filter. You must include an E Value if you want to search using this section.

Keywords: TaxonID: E-Value: Not Relevant

Note on comma separated items: the use of OR means that both separate and joint instances of search terms will be accepted. For example, a search for *Entamoeba*, *Trophozoites* will find hits with *Entamoeba* or *Trophozoite* or both.

**BLAST hit to NTOP**

If this option is selected, peptides matching the search criteria must additionally have a significant BLAST hit to the Database for Mitochondria-related Proteins.

(Not Relevant) (1e-10)

**View Data As:** HTML

**Submit Query**

Scientific enquiries should be sent to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu).

If a domain is listed below, peptides containing that domain as defined by Interpro will be listed in the search. Use any string, including domain name or description, to search for items that are combined with OR automatically. You may also click on this field to enter standard terms for two main domains: proteins (e.g. Zinc AND kinase, topoisomerase will find all proteins with either a zinc finger domain and a kinase domain (multi domain protein) OR proteins with a topoisomerase domain).

Zinc Finger 1e-10

**Interpro hit**

If this option is selected, peptides containing an interacting domain as defined by Pfam with the Pfam domain entered in will be returned. Commas separated items are combined with OR automatically. Search terms must be the start Pfam name (e.g. JDN, tyrosine).

JDN\_Finger 1e-10

**Interacting Domains**

If this option is selected, peptides containing an interacting domain as defined by Pfam with the Pfam domain entered in will be returned. Commas separated items are combined with OR automatically. Search terms must be the start Pfam name (e.g. JDN, tyrosine).

**Transcription**

If this option is selected, peptides matching the search criteria must additionally be from genes with transcription detected by Serial Analysis of Gene Expression (SAGE) analysis. Verses will be used. One assigned SAGE tag will be accepted. All available SAGE libraries are used in this search, even if the raw data has not yet been released to the public.

Evidence of "transcript on Must Exist"

**Regulation**

If this option is selected, peptides matching the search criteria must additionally be from genes with the regulation pattern chosen as obtained from Serial Analysis of Gene Expression data. A sum of the percent expression is taken for all series tags for a particular cell. (Note: A cell might be shown as unregulated, but it might still have some expression. If this is the case, then the cell is considered to be regulated.) (Note: Regulation patterns must have no zero tags for each cell. SAGE tags were assigned to ORFs based first on EST data if available, and then via an electronic assignment by averaging the 1 prime UTR length.)

Is upregulated

In 4 hour Encystation  
Vs Trophozoite

**Signal Peptide/Anchor**

If this option is selected, you may restrict the analysis to peptides containing or not containing a Signal Peptide or Signal Anchor as defined by SignalP. A protein will be predicted as having either a signal peptide, a signal anchor, or as being a non-secretory protein.

(Not Relevant) (Not Relevant)

**Transmembrane Domain**

If this option is selected, you may restrict the analysis to peptides containing or not containing a Transmembrane Domain as defined by Tmmem.

(Not Relevant)

Find all genes that are upregulated in the 4 Hour Encystation library verses the Trophozoite library and have a zinc finger domain.

GiardiaDB - Mozilla

File Edit View Go Bookmarks Tools Window Help

Home | Gbrowse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login | Admin Menu: Edit Pages | Update Statistics | Import Data | User | Feature Annotation | Sage | GBrowse | User

Assembly Data: giardia14

## Open Reading Frame Search Results

Included in Interpro/Pfam description (Zinc finger)  
Is upregulated in 4 hour Encystation verses Trophozoite 1

Orf ID	Annotation	Transmembrane Domains	Signal Sequence
6733	DHHC-type zinc finger domain-containing protein 0	0	Signal anchor
8619	DHHC-type zinc finger domain-containing protein 0	0	Non-secretory protein
9376	Protein transport protein SEC23 0	0	Non-secretory protein
16928	Zinc finger DHHC domain containing protein 2 0	0	Non-secretory protein

GBrowse [Help]

Scientific enquiries should be sent to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu).

This database is hosted by the JBCP GMOD Server. Bug reports and technical problems should be reported to [gmod@lists.mbl.edu](mailto:gmod@lists.mbl.edu).

# Advanced Analysis of Precompiled Results

GiardiaDB - Mozilla

Home | Gbrowse | Gblast | Assembly | ORES | Domains | SAGE | Download | What's New | Getting Started | Login | Admin Menu: Edit Pages | Update Statistics | Import Data | User | Feature Annotation | Sane | GBrowse | User |

Assembly Data: giardia14

## ORF Analysis

**Search Criteria**

Select Precompiled BLAST Reports to Search

Swissprot  
 Non-Redundant GenBank

**Keyword(s) for searching BLAST reports.** Comma separated items are combined with OR automatically. The default setting does not apply a keyword filter.

**BLAST results must include hits to the following specific organisms** (use NCBI Taxonomy Names or ID Numbers). Comma separated items are combined with OR automatically. The default setting does not apply a taxonomy filter. You must include an E-Value if you want to search using this section.

Keywords:  Taxonomy:  E-Value:

**Excluded keyword(s) for searching BLAST reports.** Comma separated items are combined with OR automatically. The default setting does not apply a keyword filter.

**BLAST results cannot include hits to the following specific organisms** (use NCBI Taxonomy Names or ID Numbers). Comma separated items are combined with OR automatically. The default setting does not apply a taxonomy filter. You must include an E-Value if you want to search using this section.

Keywords:  Taxonomy:  E-Value:

Note on comma separated items: the use of OR means that both separate and joint instances of search terms will be accepted. For example, a search for *Entamoeba*, *Trichomonas* will find all hits with *Entamoeba* or *Trichomonas* or both.

**BLAST hit to MITOP**

If this option is selected, peptides matching the search criteria must additionally have a significant BLASTP hit to the Database for Mitochondria-Related Proteins.

Not Relevant

**Interpro hit**

If a domain is listed below, peptides with that domain as defined by Interpro will be listed in the search. Use any string within the domain name or description. Comma separated items are combined with OR automatically. You may also place an AND between search terms to find multi-domain proteins.(ex. *zinc AND kinase, topoisomerase* will find all proteins with either a zinc finger domain and a kinase domain (multi domain protein) OR proteins with a topoisomerase domain)

**Interacting Domains**

If this option is selected, peptides containing an interacting domain as defined by iPfam with the Pfam domain entered in will be returned. Comma separated items are combined with OR automatically. Search terms must be the short Pfam name (ex. *Ithr, trypsin*)

**Transcription**

If this option is selected, peptides matching the search criteria must additionally be from genes with transcription detected by Serial Analysis of Gene Expression, such that only genes with at least one assigned SAGE tag will be accepted. All available SAGE libraries are used in this search, even if the raw data has not yet been released to the public.

Evidence of Transcription Must Exist

**Regulation**

If this option is selected, peptides matching the search criteria must additionally be from genes with the regulation pattern chosen as obtained from Serial Analysis of Gene Expression data. A sum of the percent expression is taken for all sense tags for a particular orf. (Note: A orf might be shown as upregulated or downregulated if total sampling between the libraries differ and the raw tag frequencies are the same. To be exclusive, the other library must have no sense sage tags seen. Sage tags were assigned to orfs based first on EST data if available, and then via an electronic assignment by averaging the 3 prime UTR length.)

Is  In  vs

**Signal Peptide/Anchor**

If this option is selected, you may restrict the analysis to peptides containing or not containing a Signal Peptide or Signal Anchor as defined by SignalP. (A protein will be predicted as having either a Signal peptide, a Signal anchor, or as being a non-secretory protein)

Not Relevant

**Transmembrane Domain**

If this option is selected, you may restrict the analysis to peptides containing or not containing a Transmembrane Domain as defined by tmhmm.

Not Relevant

**View Data As**

Scientific enquiries should be sent to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu)

This database is hosted by the JBCP GMOD Server. Bug reports and technical problems should be reported to [gmod@lists.mbl.edu](mailto:gmod@lists.mbl.edu).

Domains

Interacting Domains

Transcription based on Expression Analysis

Signal Sequence

Transmembrane Domains

Search Precompiled Blast reports  
by keyword or based on taxonomic  
classification of hits

# Gene Expression Analysis

**GiardiaDB - Mozilla**

## Serial Analysis of Gene Expression

Characterization of the *Giardia lamblia* Transcriptome

**SAGE Overview**

As the genome sequence of *Giardia lamblia* is nearing completion and because *Giardia*'s life cycle can be reproduced *in vitro*, giardial gene expression is being examined using Serial Analysis of Gene Expression (SAGE) to monitor genome-wide levels of messenger RNA (mRNA) expression throughout the life cycle. Examination of genome-wide gene expression patterns will provide a coherent picture of activation and inactivation of biological pathways. SAGE quickly and very affordably detects 15 base nucleotide sequences from every mRNA transcript present in a sampled population of cells. The sequences are called 'tags' and the frequency of these tags determined by SAGE is directly reflective of relative transcript abundance. SAGE thus both identifies genes and their relative levels of expression.

- [Detailed diagram of the SAGE protocol](#)
- [SAGEnet - Online resources for Serial Analysis of Gene Expression](#)

**SAGE Library Construction**

SAGE libraries were constructed using the Invitrogen I-SAGE Kit and recombinant pZero sequenced using the MBL genome sequencing facility. The cells for the mRNA isolation were from the WB isolate, clone C6, which expresses the VSP called TSA 417. These are the most common giardial isolate and VSP type worldwide and clone C6 is the isolate used by the genome project (ATCC #50803). A new A11 clone was isolated just prior to the initiation of the studies to minimize genetic drift. Its encystation was representative of the C6 parental clone and of several sibling subclones. It has been frozen away as a stabilite in liquid nitrogen for reproducibility.

Proposed life cycle of *Giardia lamblia*, with notes on DNA replication, nuclear division, and cellular division (reproduced from Svard et al. 2003).

*Giardia lamblia* strain WB cells were induced to encyst *in vitro* as previously described by Sun et al. (2002) and RNA was extracted from various time points using TRIzol (GIBCO) and RNeasy (Tel-Test). A confluent culture of trophozoites was used to inoculate fresh cultures at a density of  $10^5$  cells/ml and grown for 21 hr. These were log phase trophozoites. Encystation cultures were inoculated from the same cultures in encystation medium ("GS-4"). Cells were harvested at 4, 12, 21, and 42 hr of encystation and as cysts for total RNA isolation. Encystation was carried out by a two-step method as described by Boucher and Gillin (1990) and RNA isolated at four timepoints (Stage 1, Stage 2, 30 min, 60 min). Stage 1 is an acidic, reducing environment that models the stomach. Stage 2 consists of proteases in a slightly alkaline solution that mimicks the small intestine. A second trophozoite library was constructed from an independent experiment to test the consistency of the SAGE results.

**SAGE Bioinformatics**

Raw sequencing results were parsed to extract SAGE tags and their frequencies using custom software. In cases where a ditag sequence was sampled more than once, only one representative was used in generating tag frequencies. Tags that contained base call ambiguities or bases with PHRED values of less than 10% (or greater chance of incorrect base call) were excluded from analyses.

SAGE tag sequences are mapped to genome assembly contigs and predicted open reading frames (ORFs) using custom software. As SAGE tags are generated by the most 3' *Nla III* restriction site, many are from the 3' untranslated region (UTR) of mRNA transcripts. SAGE tags are thus mapped to theoretical transcripts, which are modeled as the complete Open Reading Frame (ORF) plus 15 bp of 3' UTR. This model is used for both sense and anti-sense transcripts. Primary SAGE tags are those generated by the most 3' *Nla III* restriction site on the theoretical transcript. All other placements are annotated as Alternate tags. Possible tag types are PS (primary sense tag), PA (primary anti-sense tag), AS (alternate sense tag), AA (alternate anti-sense tag), UK (unmapped - not resolved to an ORF). Like any model, assignment of SAGE tags to ORFs includes an inherent error rate. SAGE tags mapping to more than one location in the genome deserve close attention.

**SAGE Statistics**

All statistics are generated after removal of putative sequencing error. Our SAGE analyses filter out sequencing error by removal of tags that are not found more than once in any library and do not map to the genome.

[Advanced SAGE Analysis Tool](#)

[Search your Sequence for Sage Tags](#)

**SAGE Libraries**

Trophozoite 1 through 60 minute Excystation represents a continuous experimental time series. Trophozoite 2 represents an independent experiment using a different group of starting cells.

SAGE Library	Acronym	Total Tags Sampled	Putative Sequencing Error Tags	Total Tags Used in Analyses
Trophozoite 1	TROPH1	39,253	1,238	38,015
4 hour Encystation	ENCY4H	38,319	1,264	37,055
12 hour Encystation	ENCY12H	39,282	1,468	37,814
21 hour Encystation	ENCY21H	19,195	825	18,370
42 hour Encystation	ENCY42H	40,655	1,482	39,173
Cyst	CYST	38,019	1,634	36,385
S1 Excystation	EXCYS1	17,056	915	16,141
S2 Excystation	EXCYS2	40,480	1,894	38,586
30 min Excystation	EXCY30M	40,478	1,845	38,633
60 min Excystation	EXCY60M	40,617	1,948	38,669
Trophozoite 2	TROPH2	39,959	2,159	37,800
2.5 hour treatment (IEC)	TREAT2_5H	40,194	2,881	37,313
2.5 hour control (IEC)	CONTR2_5H	40,741	1,913	38,828
6 hour treatment (IEC)	TREAT6H	43,871	2,843	41,028
6 hour control (IEC)	CONTR6H	41,469	1,969	39,500
METAB: Glucose (50 mM) added	GLU_ONLY	38,494	1,467	37,027
METAB: Arginine (10 mM) added	ARG_ONLY	38,780	1,545	37,235
METAB: Glucose (50 mM) and Arginine (10 mM) added	GLU_ARG	39,119	2,028	37,090
METAB: No supplements (~3 mM Glucose, ~0.5 mM Arginine)	NO_ADD	38,329	1,132	37,197
VPA: VPA control	VPA_C	1,943	88	1,855
VPA: 5mM VPA treatment	VPA_T	6,657	195	6,462

**Tag Mapping Statistics**

SAGE tags are mapped to theoretical transcripts, which are modeled as the complete ORF plus 15 bp of 3' UTR. Primary SAGE tags are those generated by the most 3' *Nla III* restriction site on the theoretical transcript.

Statistic	Unique Tag Sequences	All Sampled Tags
Total Number of Unique Tag Sequences	30,828	690,176
Tags that Do Not Map to the Genome	7.43%	8.32%
Tags that Map to One Location in the Genome	83.59%	79.21%
Tags that Map to Multiple Locations in the Genome	8.99%	12.47%
Tags Mapped to an Open Reading Frame	80.16%	77.61%
Tags Not Mapped to an Open Reading Frame (UK)	19.84%	22.39%
Tags Mapped as Primary Sense Tags (PS)	15.16%	62.00%
Tags Mapped as Primary Anti-Sense Tags (PA)	5.82%	3.00%
Tags Mapped as Alternate Sense Tags (AS)	34.67%	9.00%

# Analyze Sage Data

GiardiaDB - Mozilla

Home | Browse | Blast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login |  
Admin Menu: Edit Pages | Update Statistics | Import Data | User | Feature Annotation | Sage | GBrowse | User | Assembly Data: giardia14

## SAGE Analysis

**Data Access**  
[Login](#) to access all SAGE libraries (SAGE Consortium members only). Data from all libraries will be publicly released upon publication.

**Select Libraries**

Trophozoite 1  
 4 hour Encystation  
 12 hour Encystation  
 21 hour Encystation  
 42 hour Encystation  
 Cyst  
 S1 Encystation  
 S2 Encystation  
 30 min Encystation  
 60 min Encystation  
 Trophozoite 2  
 2.5 hour treatment (IEC)  
 2.5 hour control (IEC)  
 6 hour treatment (IEC)  
 6 hour control (IEC)  
 METAB: Glucose (50 mM) added  
 METAB: Arginine (10 mM) added  
 METAB: Glucose (50 mM) and Arginine (10 mM) added  
 METAB: No supplements (>3 mM Glucose, ~0.5 mM Arginine)  
 VPA: VPA control  
 VPA: 5mM VPA treatment

**Statistical Confidence** →

**R Value**: 4 greater than  True  
The R-Value is the log-likelihood ratio statistic of Stekel *et al.* (2000), which scores tags by their deviation from the null hypothesis of equal frequencies. Higher scores represent a greater deviation from the null hypothesis, while scores close to zero represent near constitutive expression.

**Regulation** →

**Regulation**: Is upregulated  True  
In: Cyst  
Vs: Trophozoite 1  
This allows you to determine tags that are up or downregulated between libraries

**Cluster Data** →

**Perform Clustering**:  True [5] Clusters  
Analysis of SAGE data using this tool automatically includes clustering of gene expression profiles using log transformation, median centering, Pearson's correlation coefficient, and kmeans/median clustering.

**Shade Expression Levels Based on**:  True  
Median Centering  
Median centering forces shading of expression levels to ignore magnitude of expression, thus highlighting correlated patterns of expression among lowly and highly expressed genes in the search results.

**Sort By**: Cluster  descending

**View Data As**: HTML  
For more complex analyses, output options: HTML, Tab Delimited, Cluster 3.0, CLUTO, TableView, IDEG6.

This database is hosted by the JBCP GMOD Server. Bug reports and technical problems should be reported to [gmmod@lists.mbl.edu](mailto:gmmod@lists.mbl.edu).

Choose libraries ↑

Export in various formats for offline analysis ↑

Find me all tags that are unregulated in the Cyst library vs. the Trophozoite library with differential expression with a R-Value confidence of greater then 4, and cluster them using k-means clustering into 5 groups.

# Sage Analysis Results

GiardiaDB - Mozilla

Home | Gbrowse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login | Admin Menu: Edit Pages | Update Statistics | Import Data | User | Feature Annotation | Sage | GBrowse | User | Assembly Data: giardia14

Statistical Confidence of differential expression

**SAGE Analysis Results**

Expression between libraries. Colored based on up/down regulation

SAGE tags matching your search criteria are listed below. Click on the SAGE tag or ORF identification numbers for more details. Tag frequencies are presented as percentages of the total sample of tags, after removal of sequencing error. Possible tag types are PS (primary sense tag), PA (primary anti-sense tag), AA (alternate anti-sense tag), UK (not known - not resolved to an ORF). The R-Value (if shown) is the log-likelihood ratio statistic of [Stekel et al. \(2000\)](#), which scores tags by their deviation from the null hypothesis of equal frequencies. Higher scores represent a greater deviation from the null hypothesis. S (alternate while scores close to zero represent near constitutive expression)

Tag ID	Sequence	Type	Cluster	R-Value	OrfID	Description	TROPH1	ENCY4H	ENCY12H	ENCY21H	ENCY42H	CYST
29250	CATCTAAATTCAAATC	PA	4	5.6079	17465	Structural maintenance of chromosome 4	0.00000	0.00000	0.00793	0.03266	0.02298	0.01099
11788	CATGTTAGCTAAACGG	PA	4	4.3354	39587	gi 49657813 emb CA90649.1  unnamed protein product [Debaromyces hansenii CBS767] gi 50427107 ref XP_462161.1  unnamed protein product [Debaromyces hansenii]4e-17	0.00000	0.00000	0.00000	0.01639	0.01532	0.00825
29200	CATGGGGATGGCGAG	PS	4	39.0041	7982	UDP-glucose 4-epimerase	0.00789	0.00540	0.07669	0.22319	0.21188	0.04947
33314	CATGTTAAATAGTCTA	UK	4	5.0520			0.00000	0.00000	0.00529	0.01089	0.02298	0.01924
8983	CATGCTTTCACTTGCC	PS	4	4.8742	27652		0.00000	0.00000	0.01058	0.01633	0.02996	0.00275
11775	CATGTTAACCTCCGCT	UK	4	5.3734			0.00000	0.00000	0.00000	0.02177	0.01787	0.01099
31826	CATGGGGAGATTCTT	AA	4	6.1152	8850	Protein 21.1	0.01052	0.00540	0.01587	0.05444	0.04850	0.01099
8842	CATGGAGGAAGGAGG	AS	4	21.9859	33230		0.00000	0.00000	0.04231	0.10867	0.07914	0.00550
31345	CATCTGTATAGGCC	PS	4	17.3106	33513		0.00789	0.01349	0.02908	0.08710	0.13274	0.04672
2199	CATGAAGCAGAACAC	UK	4	72.5754			0.00526	0.00540	0.08991	0.35928	0.29868	0.01374
32849	CATGTTAAAAAAAUK	4	11.0677				0.00789	0.01349	0.01058	0.04899	0.08935	0.02748
31671	CATGAGCAGCCACTT	PS	4	8.2686	135269	GLP_608_2922_2404 [Giardia lamblia ATCC 50803] gi 29247911 gb EAA39460.1  GLP_762_52128_52646 [Giardia lamblia ATCC 50803]	0.00000	0.00000	0.00529	0.01633	0.03574	0.03023
11922	CATGTTAAGTGTCT	AS	4	7.1039	11692	Possible surface antigen	0.00000	0.00000	0.00000	0.01089	0.03633	0.00825
11768	CATGACAGACTCGAT	AS	4	4.9339	8377		0.00000	0.00000	0.00264	0.02722	0.01787	0.00550
3852	CATGCTAACCGCTCAC	UK	4	6.8678		Multiple Map Locations	0.00526	0.00270	0.01815	0.04899	0.04084	0.00550
11776	CATGTCCTTTGCTC	PS	4	4.5992	28183	gi 55983926 emb CAE64016.1  Hypothetical protein CBG08611 [Caenorhabditis briggsae] 5e-05	0.00000	0.00000	0.00264	0.02177	0.01787	0.00275
11769	CATGCCCTCATATTCT	PA	4	4.3045	23433	matches Pfam TM2 domain E=6.7e-15	0.00000	0.00000	0.00529	0.02722	0.01532	0.00550
31527	CATGCTCCGACTCCCT	UK	4	6.1935			0.00000	0.00000	0.00793	0.03266	0.02553	0.00550
8853	CATGGAATGCTAAC	UK	4	10.7329			0.00000	0.00270	0.02380	0.06532	0.04340	0.00550
11766	CATGGGGAGGTCTT	AS	4	6.7889	39607		0.00000	0.00270	0.00000	0.03811	0.02298	0.00825
8846	CATGCAGTTCCCTG	AA	4	17.2759	14259	Glucose 6-phosphate N-acetyltransferase (EC 2.3.1.4)	0.00000	0.00000	0.01851	0.08710	0.05871	0.00275
13891	CATGGGGAGACATA	AA	4	7.2083	103606		0.00000	0.00000	0.00000	0.02068	0.02068	0.00275
32320	CATCTAAATTCTATA	AS	4	5.5959	97437	Multiple Map Locations	0.00000	0.00270	0.01324	0.03811	0.02553	0.00550
11793	CATGGGATTTGATA&AA	AS	4	8.8050	112875	gi 5594431 ref XP_525675.1  PREDICTED: similar to membrane-type 1 matrix metalloproteinase cytoplasmic tail binding protein-1; submergence induced protein 2 [Pan troglodytes] 6e-19	0.00000	0.00000	0.00000	0.01633	0.03063	0.02748
9765	CATCTAAATTCTAC	AS	4	9.0956	119598	Multiple Map Locations	0.00000	0.00000	0.00264	0.01633	0.04340	0.00825
9070	CATCTAAATTAGCTC	AA	4	7.4877	16143	Transcription regulatory protein SNF2	0.00000	0.00000	0.00793	0.02177	0.03574	0.02474
30670	CATGCCAAAGACACT	PS	4	8.0095	7374	gi 12247892 gb AAG50081.1  SPEC3-like protein [Sterniella histriomuscorum] 3e-05	0.00263	0.00000	0.00529	0.02722	0.03299	0.03296
32081	CATGGGAAACACTCTA	PS	4	14.3781	10939	gi 7489916 pir S72442 actin-fragmin kinase - slime mold (Physarum polycephalum) gi 1553133 gb AAB08728.1  actin-fragmin kinase [Physarum polycephalum] 2e-05	0.00789	0.00540	0.02624	0.03266	0.07194	0.00646
23946	CATGCTGTTAAATTCT	AA	4	6.9001	16050	gi 55240261 gb EAA09724.2  ENSANGP00000013163 [Anopheles gambiae str. PEST] gi 58385916 ref XP_314309.2  ENSANGP00000013163 [Anopheles gambiae str. PEST] 1e-06	0.00000	0.00000	0.00793	0.02177	0.03574	0.01374
31101	CATGCTCCCTACTAA	AA	4	6.0806	22543		0.00263	0.00276	0.02909	0.03811	0.03029	0.03023
9071	CATGCTGAATGCTGATA	AA	4	5.6169	16143	Transcription regulatory protein SNF2	0.00000	0.00000	0.00529	0.01089	0.02808	0.01649
11804	CATGCTGACTCTG	AA	4	4.4335	16728	Phosphatase and tensin homolog	0.00000	0.00000	0.00000	0.01089	0.01787	0.00275
33485	CATGCCATTACTTCT	PS	4	92.9380	14259	Glucose 6-phosphate N-acetyltransferase (EC 2.3.1.4)	0.01052	0.00270	0.15074	0.46271	0.43397	0.05497
2894	CATGCCATTACCACT	UK	4	10.4535			0.00789	0.00540	0.02909	0.07621	0.07658	0.02748
31401	CATGCTTAATAGCTC	PS	3	20.9127	10313	gi 50938053 ref XP_478554.1  putative serine/threonine-specific protein kinase(gi 7488195)  [Oryza sativa (japonica cultivar-group)] gi 51963464 ref XP_506393.1  PREDICTED P0696F12.4 gene product [Oryza sativa (japonica cultivar-group)] gi 34394939 db BAC84489.1  putative serine/thr 2e-10	0.00789	0.01349	0.00793	0.00544	0.03319	0.13467
32809	CATGGCTCTAGACCC	PS	3	4.5108	88888		0.00263	0.00000	0.00264	0.00000	0.01021	0.02474
33813	CATGGGAAACACTCTA	AS	3	4.0193	19062		0.00000	0.00000	0.00000	0.00511	0.01649	
30794	CATGGCTCTAGCTGTC	PS	3	112.8458	4984		0.00000	0.00000	0.01587	0.13609	0.27315	0.50570
5990	CATGGGGGGAAAAAA	UK	3	5.2922			0.00263	0.01079	0.00793	0.01089	0.02298	0.04672
32560	CATGATGATTCTTCTG	UK	3	6.8211			0.00000	0.00000	0.00264	0.02722	0.01276	0.03023
32532	CATGAATGATTCTCTG	AA	3	8.6250	6624	gi 48870355 ref XP_0032079.1  hypothetical protein PpenA01001008 [Pedioecoccus pentosaceus ATCC 25745] 2e-05	0.00000	0.00000	0.00000	0.00000	0.00766	0.03573
31749	CATGCTGTTAAATTA	UK	3	12.8527			0.00263	0.00000	0.00264	0.03266	0.02808	0.06596
32409	CATGAGATGAAACACT	AS	3	6.8094	16363	Putative alanine aminotransferase	0.00000	0.00000	0.00000	0.00544	0.00511	0.03023
29782	CATGAGACGACCTT	AS	3	6.6644	3475	Protein 21.1	0.00000	0.00000	0.00000	0.00000	0.00766	0.02748
30147	CATGCCCAACAACTG	UK	3	4.9019			0.00000	0.00000	0.00000	0.01633	0.00766	0.01924
2377	CATGGCCCCCTGGTAT	AA	3	5.5161	14758		0.00789	0.01079	0.01851	0.05444	0.02042	0.05497
13921	CATGGGGTCACGGG	AS	3	4.4954	12885		0.00000	0.00000	0.00000	0.01021	0.01649	
33729	CATGGACGACCTT	PS	3	4.8743	11885	Phosphatase 1 regulatory subunit, putative	0.00789	0.01349	0.01058	0.01089	0.02042	0.05497

Position of tag on transcript

Tag/ORF annotation

# SAGE Tag Detail

GiardiaDB - Mozilla

Assembly Data: giardia14

Home | Gbrowse | Gblast | Assembly | ORFs | Domains | SAGE | Download | What's New | Getting Started | Login | Admin Menu: Edit Pages | Update Statistics | Import Data | User | Feature Annotation | Sage | GBrowse | User |

## Raw Results

## SAGE Tag 1691

CATGGCCGACAATAT, Primary Sense Tag for ORF 678  
putative 5'(3')-deoxyribonucleotidase

Library	Raw Count	Percentage
Trophozoite 1	4 of 38015	0.01052%
4 hour Encystation	2 of 37055	0.00540%
12 hour Encystation	5 of 37814	0.01322%
21 hour Encystation	1 of 18370	0.00544%
42 hour Encystation	6 of 39173	0.01532%
Cyst	415 of 36385	1.14058%
S1 Excystation	134 of 16141	0.83018%
S2 Excystation	96 of 38586	0.24879%
30 min Excystation	129 of 38633	0.33391%
60 min Excystation	98 of 38669	0.25343%
Trophozoite 2	5 of 37800	0.01323%
2.5 hour treatment (IEC)	3 of 37313	0.00804%
2.5 hour control (IEC)	8 of 38822	0.02060%
6 hour treatment (IEC)	2 of 41028	0.00487%
6 hour control (IEC)	5 of 39500	0.01266%
METAB: Glucose (50 mM) added	6 of 37027	0.01620%
METAB: Arginine (10 mM) added	3 of 37235	0.00806%
METAB: Glucose (50 mM) and Arginine (10 mM) added	2 of 37094	0.00539%
METAB: No supplements (~3 mM Glucose, ~0.5 mM Arginine)	3 of 37197	0.00807%
VPA: VPA control	0 of 1855	0.00000%
VPA: 5mM VPA treatment	1 of 6462	0.01548%

[Details about SAGE libraries and statistics]

## Graph of Expression

Expression Graph

% Expression

Library

Administration

Assign a new location for this tag

## Tag to ORF Mapping

SAGE tags are mapped to theoretical transcripts, which are modeled as the complete Open Reading Frame (ORF) plus 15 bp of 3' UTR. This model is used for both sense and anti-sense transcripts. Primary SAGE tags are those generated by the most 3' Nla III restriction site on the theoretical transcript. All other placements are annotated as Alternate tags. Possible tag types are PS (primary sense tag), PA (primary anti-sense tag), AS (alternate sense tag), AA (alternate anti-sense tag), UK (unknown - not resolved to an ORF).

contig\_28  
95.7k 95.8k 95.9k 96k 96.1k 96.2k 96.3k 96.4k 96.5k 96.6k 96.7k 96.8k 96.9k 97k 97.1k 97.2k  
Contig  
contig\_28

Open Reading Frames  
orf:102963 tRNA 2-methylthioadenosine synthase  
orf:678 putative 5'(3')-deoxyribonucleotidase  
orf:96818 hypothetical protein 008

View in Genome context

Administration

Remove above assignment.  
Click on image to view details

## All Matches to Assembly Contigs

Tag sequences can match multiple locations within the genome, even if mapped to a single theoretical transcript. SAGE tags mapping to more than one location in the genome deserve close attention. All hits to the genome assembly are presented below. Click on individual images to view details

# Administrative Annotation/Assignment of Sage Tags

GiardiaDB - Mozilla

METAB: Glucose (50 mM) added	47 of 37027 [0.07292%]
METAB: Arginine (10 mM) added	28 of 37235 [0.07520%]
METAB: Glucose (50 mM) and Arginine (10 mM) added	62 of 37090 [0.16716%]
METAB: No supplements (~3 mM Glucose, ~0.5 mM Arginine)	42 of 37197 [0.11291%]
VPA: VPA control	2 of 1855 [0.10782%]
VPA: 5mM VPA treatment	2 of 6462 [0.03095%]

[Details about SAGE libraries and statistics]

**Administration**

Assign a new location for this tag

**All Matches to Assembly Contigs**

Tag sequences can match multiple locations within the genome, even if mapped to a single theoretical transcript. SAGE tags mapping to more than one location in the genome deserve close attention. All hits to the genome assembly are presented below. Click on individual images to view details.

contig\_30 ← 42k 43k 44k 45k 46k →  
Contigs contig\_30  
Open Reading Frames  
orf:10880 hypothetical protein  
orf:10881 hypothetical protein  
orf:10882 hypothetical protein  
orf:10883 hypothetical protein  
orf:10885 hypothetical protein  
sagetag 29943  
enzyme CATG

**Administration**

Assign to ORF 10882 at this location as a Primary Sense Tag  
Assign to other ORF at this location

contig\_35 ← 10k 11k 12k 13k 14k →  
Contigs contig\_35  
Open Reading Frames  
orf:17368 Fused1 protein (Fragment)  
orf:24146 hypothetical protein  
orf:13262 hypothetical protein  
orf:17367 hypothetical protein  
orf:39250 hypothetical protein  
orf:1951 hypothetical protein  
orf:17365 hypothetical protein  
sagetag 29943  
enzyme CATG

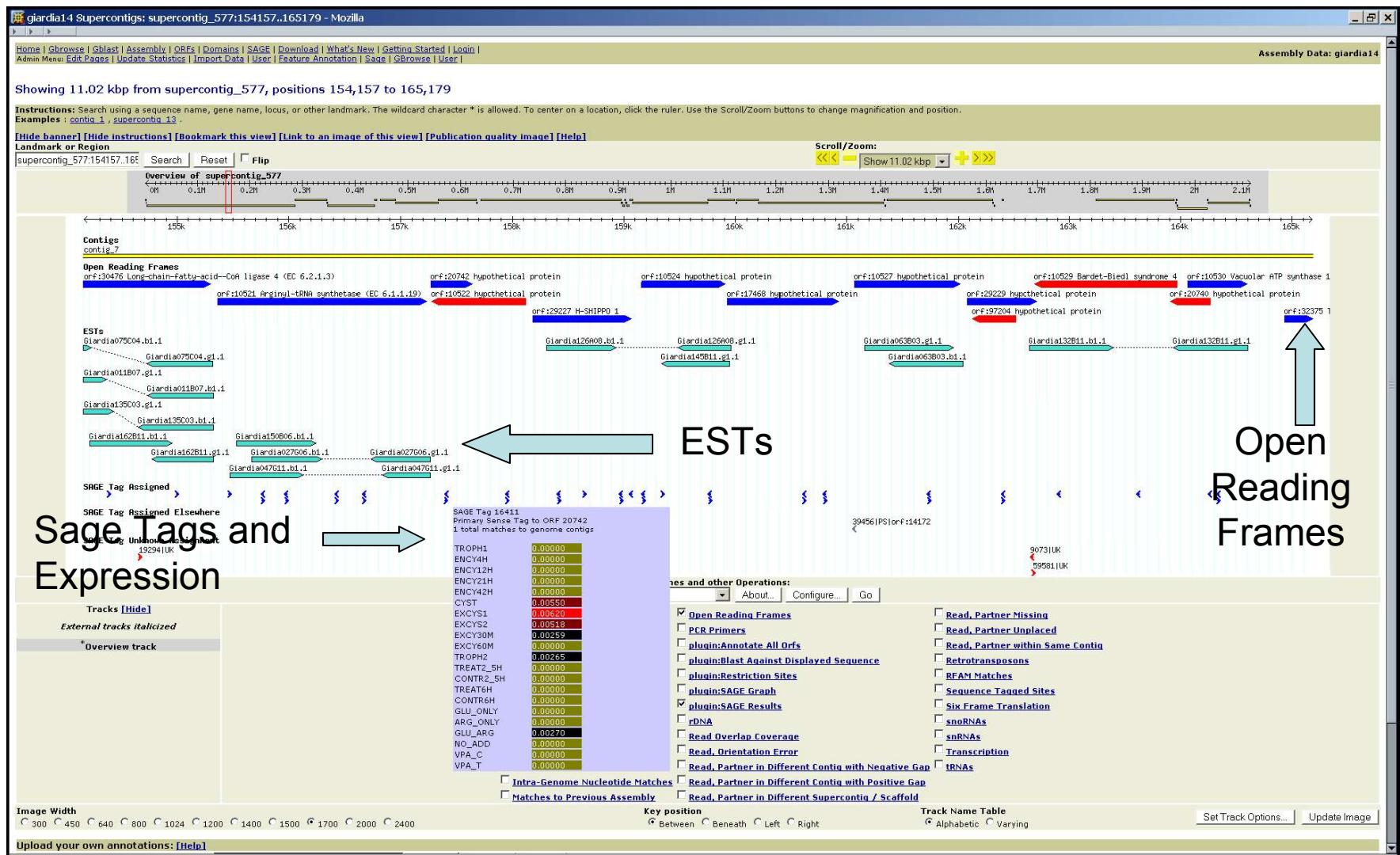
**Administration**

Assign to ORF 17368 at this location as a Alternate Sense Tag  
Assign to other ORF at this location

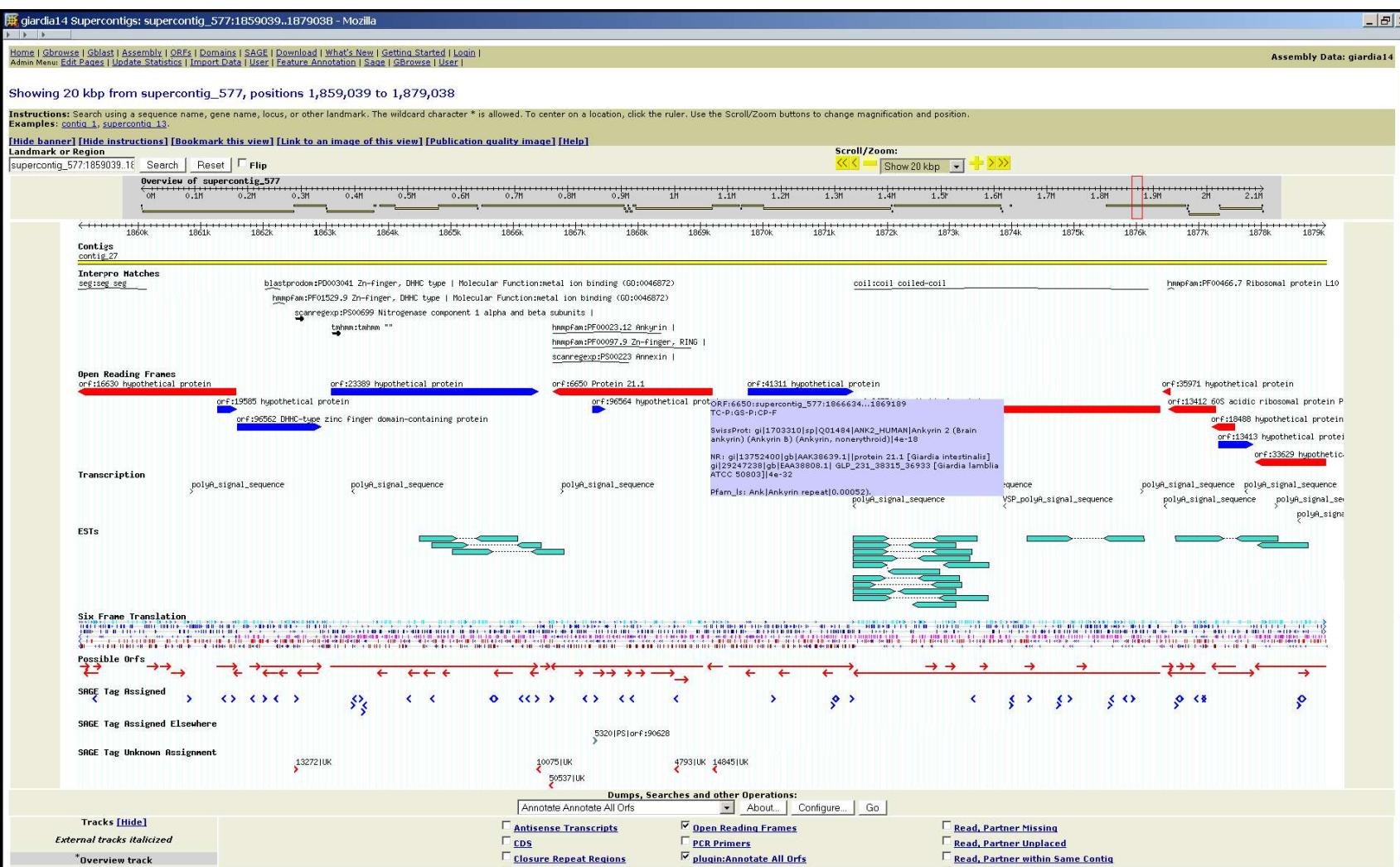
Manually assign Sage Tags to Genes →

Attempt to resolve tags based on position of cutting enzyme site ←

# Data Visualized in GBrowse



# Data Visualized in GBrowse



Many different tracks of data

# Data Visualized in GBrowse

- |  |  |   |
|--|--|---|
| <input type="checkbox"/> <a href="#">Antisense Transcripts</a>           | <input type="checkbox"/> <a href="#">Open Reading Frames</a>                                 | <input type="checkbox"/> <a href="#">Read, Partner Missing</a>            |
| <input type="checkbox"/> <a href="#">CDS</a>                             | <input type="checkbox"/> <a href="#">PCR Primers</a>   | <input type="checkbox"/> <a href="#">Read, Partner Unplaced</a>           |
| <input type="checkbox"/> <a href="#">Closure Repeat Regions</a>          | <input type="checkbox"/> <a href="#">plugin:Annotate All Orfs</a>                            | <input type="checkbox"/> <a href="#">Read, Partner within Same Contig</a> |
| <input type="checkbox"/> <a href="#">Contig Quality</a>                  | <input type="checkbox"/> <a href="#">plugin:Blast Against Displayed Sequence</a>             | <input type="checkbox"/> <a href="#">Retrotransposons</a>                 |
| <input type="checkbox"/> <a href="#">Contig:overview*</a>                | <input type="checkbox"/> <a href="#">plugin:Restriction Sites</a>                            | <input type="checkbox"/> <a href="#">RFAM Matches</a>                     |
| <input type="checkbox"/> <a href="#">Contigs</a>                         | <input type="checkbox"/> <a href="#">plugin:SAGE Graph</a>                                   | <input type="checkbox"/> <a href="#">Sequence Tagged Sites</a>            |
| <input type="checkbox"/> <a href="#">DNA/GC Content</a>                  | <input type="checkbox"/> <a href="#">plugin:SAGE Results</a>                                 | <input type="checkbox"/> <a href="#">Six Frame Translation</a>            |
| <input type="checkbox"/> <a href="#">ESTs</a>                            | <input type="checkbox"/> <a href="#">rDNA</a>  | <input type="checkbox"/> <a href="#">snoRNAs</a>                          |
| <input type="checkbox"/> <a href="#">Hybridizations</a>                  | <input type="checkbox"/> <a href="#">Read Overlap Coverage</a>                               | <input type="checkbox"/> <a href="#">snRNAs</a>                           |
| <input type="checkbox"/> <a href="#">Intergenic BlastX Hits</a>          | <input type="checkbox"/> <a href="#">Read, Orientation Error</a>                             | <input type="checkbox"/> <a href="#">Transcription</a>                    |
| <input type="checkbox"/> <a href="#">Interpro Matches</a>                | <input type="checkbox"/> <a href="#">Read, Partner in Different Contig with Negative Gap</a> | <input type="checkbox"/> <a href="#">tRNAs</a>                            |
| <input type="checkbox"/> <a href="#">Intra-Genome Nucleotide Matches</a> | <input type="checkbox"/> <a href="#">Read, Partner in Different Contig with Positive Gap</a> |   |
| <input type="checkbox"/> <a href="#">Matches to Previous Assembly</a>    | <input type="checkbox"/> <a href="#">Read, Partner in Different Supercontig / Scaffold</a>   |   |

Many different tracks of data

# Administrative Interface

**GiardiaDB - Mozilla**

**GiardiaDB**

Informatics Support for the *Giardia lamblia* Genome and Gene Expression Projects  
The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution at the Marine Biological Laboratory, Woods Hole, MA  
This current release has been constructed using the Generic Model Organism Database paradigm.  
Citation: The *Giardia lamblia* Genome Database (www.mbl.edu/Giardia)

**Sequence Data Import Administration Tool**

Import Data

Import Sequence Data  
Choose the type of data you would like to import  
Fasta Genbank Ace ARACHNE AMOS  
Fasta File:  Browse...  
Delete/Append Options Append to existing Assembly  
Submit Query

Scientific enquiries should be sent to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu). This database is hosted by the problems should be reported to [gmod@lists.mbl.edu](mailto:gmod@lists.mbl.edu).

**Import Data**

**GiardiaDB - Mozilla**

**GiardiaDB**

Informatics Support for the *Giardia lamblia* Genome and Gene Expression Projects  
The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution at the Marine Biological Laboratory, Woods Hole, MA  
This current release has been constructed using the Generic Model Organism Database paradigm.  
Citation: The *Giardia lamblia* Genome Database (www.mbl.edu/Giardia)

**Feature Administration Tool**

Call ORFs Searches

Annotate Open Reading Frames

Upload Training File:  Browse... Extract long orfs from file   
or Extract Training File from Assembly:   
Minimum Length: 100  
Maximum Length: 25000  
Remove Alternate/False Starts:   
Remove No Stop Codon:   
Bacterial Organism:   
Remove Existing Orfs:   
Extract Orfs

Warning, this will remove all orfs, annotations, search results, sage to orf mapping, tree's and alignments

**ORF Calling**

**GiardiaDB - Mozilla**

**GiardiaDB**

Informatics Support for the *Giardia lamblia* Genome and Gene Expression Projects  
The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution at the Marine Biological Laboratory, Woods Hole, MA  
This current release has been constructed using the Generic Model Organism Database paradigm.  
Citation: The *Giardia lamblia* Genome Database (www.mbl.edu/Giardia)

**Feature Administration Tool**

Call ORFs Searches

ORFs Intergenic Space Contigs Contig Slice Sage Tags

blastp	<input type="checkbox"/> nr <input type="checkbox"/> swissprot <input type="checkbox"/> RefEukls <input type="checkbox"/> mitop <input type="checkbox"/> ecoli_aa	blastx	<input type="checkbox"/> nr <input type="checkbox"/> swissprot <input type="checkbox"/> RefEukls <input type="checkbox"/> mitop <input type="checkbox"/> ecoli_aa	blastx - not yet	<input type="checkbox"/> nr <input type="checkbox"/> swissprot <input type="checkbox"/> RefEukls <input type="checkbox"/> mitop <input type="checkbox"/> ecoli_aa
blastn	<input type="checkbox"/> nt	blastn	<input type="checkbox"/> nt	blastn - not yet	<input type="checkbox"/> nt
other - not yet	<input type="checkbox"/> interpro <input type="checkbox"/> hmmpfam <input type="checkbox"/> signalp <input type="checkbox"/> targets				
	Min: 10 Max: 20000		Min: 100 Max: 6000		

Delete Search Queue Before Insert  
not yet -  Start on Cluster [20] Jobs  
Start Searches

Scientific enquiries should be sent to [giardia@lists.mbl.edu](mailto:giardia@lists.mbl.edu). This database is hosted by the JBCP GMOD Server. Bug reports and technical problems should be reported to [gmod@lists.mbl.edu](mailto:gmod@lists.mbl.edu).

**Automated Searches**

**GiardiaDB - Mozilla**

**GiardiaDB**

Informatics Support for the *Giardia lamblia* Genome and Gene Expression Projects  
The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution at the Marine Biological Laboratory, Woods Hole, MA  
This current release has been constructed using the Generic Model Organism Database paradigm.  
Citation: The *Giardia lamblia* Genome Database (www.mbl.edu/Giardia)

**SAGE Administration Tool**

Upload SAGE Library Names User Tools

- E 1. New library
- C 2. Edit library
- C 3. View library
- C 4. New experiment
- C 5. View experiment
- C 6. New sample
- C 7. View sample
- C 8. New replicates
- C 9. View replicates
- C 10. New array
- C 11. View array
- C 12. New analysis
- C 13. View analysis
- C 14. New report
- C 15. View report
- C 16. New transcript
- C 17. View transcript
- C 18. New treatment
- C 19. View treatment
- C 20. New replicates
- C 21. View replicates
- C 22. New array
- C 23. View array
- C 24. New analysis
- C 25. View analysis
- C 26. New report
- C 27. View report
- C 28. New transcript
- C 29. View transcript
- C 30. New treatment
- C 31. View treatment
- C 32. New replicates
- C 33. View replicates
- C 34. New array
- C 35. View array
- C 36. New analysis
- C 37. View analysis
- C 38. New report
- C 39. View report
- C 40. New transcript
- C 41. View transcript
- C 42. New treatment
- C 43. View treatment
- C 44. New replicates
- C 45. View replicates
- C 46. New array
- C 47. View array
- C 48. New analysis
- C 49. View analysis
- C 50. New report
- C 51. View report
- C 52. New transcript
- C 53. View transcript
- C 54. New treatment
- C 55. View treatment
- C 56. New replicates
- C 57. View replicates
- C 58. New array
- C 59. View array
- C 60. New analysis
- C 61. View analysis
- C 62. New report
- C 63. View report
- C 64. New transcript
- C 65. View transcript
- C 66. New treatment
- C 67. View treatment
- C 68. New replicates
- C 69. View replicates
- C 70. New array
- C 71. View array
- C 72. New analysis
- C 73. View analysis
- C 74. New report
- C 75. View report
- C 76. New transcript
- C 77. View transcript
- C 78. New treatment
- C 79. View treatment
- C 80. New replicates
- C 81. View replicates
- C 82. New array
- C 83. View array
- C 84. New analysis
- C 85. View analysis
- C 86. New report
- C 87. View report
- C 88. New transcript
- C 89. View transcript
- C 90. New treatment
- C 91. View treatment
- C 92. New replicates
- C 93. View replicates
- C 94. New array
- C 95. View array
- C 96. New analysis
- C 97. View analysis
- C 98. New report
- C 99. View report
- C 100. New transcript
- C 101. View transcript
- C 102. New treatment
- C 103. View treatment
- C 104. New replicates
- C 105. View replicates
- C 106. New array
- C 107. View array
- C 108. New analysis
- C 109. View analysis
- C 110. New report
- C 111. View report
- C 112. New transcript
- C 113. View transcript
- C 114. New treatment
- C 115. View treatment
- C 116. New replicates
- C 117. View replicates
- C 118. New array
- C 119. View array
- C 120. New analysis
- C 121. View analysis
- C 122. New report
- C 123. View report
- C 124. New transcript
- C 125. View transcript
- C 126. New treatment
- C 127. View treatment
- C 128. New replicates
- C 129. View replicates
- C 130. New array
- C 131. View array
- C 132. New analysis
- C 133. View analysis
- C 134. New report
- C 135. View report
- C 136. New transcript
- C 137. View transcript
- C 138. New treatment
- C 139. View treatment
- C 140. New replicates
- C 141. View replicates
- C 142. New array
- C 143. View array
- C 144. New analysis
- C 145. View analysis
- C 146. New report
- C 147. View report
- C 148. New transcript
- C 149. View transcript
- C 150. New treatment
- C 151. View treatment
- C 152. New replicates
- C 153. View replicates
- C 154. New array
- C 155. View array
- C 156. New analysis
- C 157. View analysis
- C 158. New report
- C 159. View report
- C 160. New transcript
- C 161. View transcript
- C 162. New treatment
- C 163. View treatment
- C 164. New replicates
- C 165. View replicates
- C 166. New array
- C 167. View array
- C 168. New analysis
- C 169. View analysis
- C 170. New report
- C 171. View report
- C 172. New transcript
- C 173. View transcript
- C 174. New treatment
- C 175. View treatment
- C 176. New replicates
- C 177. View replicates
- C 178. New array
- C 179. View array
- C 180. New analysis
- C 181. View analysis
- C 182. New report
- C 183. View report
- C 184. New transcript
- C 185. View transcript
- C 186. New treatment
- C 187. View treatment
- C 188. New replicates
- C 189. View replicates
- C 190. New array
- C 191. View array
- C 192. New analysis
- C 193. View analysis
- C 194. New report
- C 195. View report
- C 196. New transcript
- C 197. View transcript
- C 198. New treatment
- C 199. View treatment
- C 200. New replicates
- C 201. View replicates
- C 202. New array
- C 203. View array
- C 204. New analysis
- C 205. View analysis
- C 206. New report
- C 207. View report
- C 208. New transcript
- C 209. View transcript
- C 210. New treatment
- C 211. View treatment
- C 212. New replicates
- C 213. View replicates
- C 214. New array
- C 215. View array
- C 216. New analysis
- C 217. View analysis
- C 218. New report
- C 219. View report
- C 220. New transcript
- C 221. View transcript
- C 222. New treatment
- C 223. View treatment
- C 224. New replicates
- C 225. View replicates
- C 226. New array
- C 227. View array
- C 228. New analysis
- C 229. View analysis
- C 230. New report
- C 231. View report
- C 232. New transcript
- C 233. View transcript
- C 234. New treatment
- C 235. View treatment
- C 236. New replicates
- C 237. View replicates
- C 238. New array
- C 239. View array
- C 240. New analysis
- C 241. View analysis
- C 242. New report
- C 243. View report
- C 244. New transcript
- C 245. View transcript
- C 246. New treatment
- C 247. View treatment
- C 248. New replicates
- C 249. View replicates
- C 250. New array
- C 251. View array
- C 252. New analysis
- C 253. View analysis
- C 254. New report
- C 255. View report
- C 256. New transcript
- C 257. View transcript
- C 258. New treatment
- C 259. View treatment
- C 260. New replicates
- C 261. View replicates
- C 262. New array
- C 263. View array
- C 264. New analysis
- C 265. View analysis
- C 266. New report
- C 267. View report
- C 268. New transcript
- C 269. View transcript
- C 270. New treatment
- C 271. View treatment
- C 272. New replicates
- C 273. View replicates
- C 274. New array
- C 275. View array
- C 276. New analysis
- C 277. View analysis
- C 278. New report
- C 279. View report
- C 280. New transcript
- C 281. View transcript
- C 282. New treatment
- C 283. View treatment
- C 284. New replicates
- C 285. View replicates
- C 286. New array
- C 287. View array
- C 288. New analysis
- C 289. View analysis
- C 290. New report
- C 291. View report
- C 292. New transcript
- C 293. View transcript
- C 294. New treatment
- C 295. View treatment
- C 296. New replicates
- C 297. View replicates
- C 298. New array
- C 299. View array
- C 300. New analysis
- C 301. View analysis
- C 302. New report
- C 303. View report
- C 304. New transcript
- C 305. View transcript
- C 306. New treatment
- C 307. View treatment
- C 308. New replicates
- C 309. View replicates
- C 310. New array
- C 311. View array
- C 312. New analysis
- C 313. View analysis
- C 314. New report
- C 315. View report
- C 316. New transcript
- C 317. View transcript
- C 318. New treatment
- C 319. View treatment
- C 320. New replicates
- C 321. View replicates
- C 322. New array
- C 323. View array
- C 324. New analysis
- C 325. View analysis
- C 326. New report
- C 327. View report
- C 328. New transcript
- C 329. View transcript
- C 330. New treatment
- C 331. View treatment
- C 332. New replicates
- C 333. View replicates
- C 334. New array
- C 335. View array
- C 336. New analysis
- C 337. View analysis
- C 338. New report
- C 339. View report
- C 340. New transcript
- C 341. View transcript
- C 342. New treatment
- C 343. View treatment
- C 344. New replicates
- C 345. View replicates
- C 346. New array
- C 347. View array
- C 348. New analysis
- C 349. View analysis
- C 350. New report
- C 351. View report
- C 352. New transcript
- C 353. View transcript
- C 354. New treatment
- C 355. View treatment
- C 356. New replicates
- C 357. View replicates
- C 358. New array
- C 359. View array
- C 360. New analysis
- C 361. View analysis
- C 362. New report
- C 363. View report
- C 364. New transcript
- C 365. View transcript
- C 366. New treatment
- C 367. View treatment
- C 368. New replicates
- C 369. View replicates
- C 370. New array
- C 371. View array
- C 372. New analysis
- C 373. View analysis
- C 374. New report
- C 375. View report
- C 376. New transcript
- C 377. View transcript
- C 378. New treatment
- C 379. View treatment
- C 380. New replicates
- C 381. View replicates
- C 382. New array
- C 383. View array
- C 384. New analysis
- C 385. View analysis
- C 386. New report
- C 387. View report
- C 388. New transcript
- C 389. View transcript
- C 390. New treatment
- C 391. View treatment
- C 392. New replicates
- C 393. View replicates
- C 394. New array
- C 395. View array
- C 396. New analysis
- C 397. View analysis
- C 398. New report
- C 399. View report
- C 400. New transcript
- C 401. View transcript
- C 402. New treatment
- C 403. View treatment
- C 404. New replicates
- C 405. View replicates
- C 406. New array
- C 407. View array
- C 408. New analysis
- C 409. View analysis
- C 410. New report
- C 411. View report
- C 412. New transcript
- C 413. View transcript
- C 414. New treatment
- C 415. View treatment
- C 416. New replicates
- C 417. View replicates
- C 418. New array
- C 419. View array
- C 420. New analysis
- C 421. View analysis
- C 422. New report
- C 423. View report
- C 424. New transcript
- C 425. View transcript
- C 426. New treatment
- C 427. View treatment
- C 428. New replicates
- C 429. View replicates
- C 430. New array
- C 431. View array
- C 432. New analysis
- C 433. View analysis
- C 434. New report
- C 435. View report
- C 436. New transcript
- C 437. View transcript
- C 438. New treatment
- C 439. View treatment
- C 440. New replicates
- C 441. View replicates
- C 442. New array
- C 443. View array
- C 444. New analysis
- C 445. View analysis
- C 446. New report
- C 447. View report
- C 448. New transcript
- C 449. View transcript
- C 450. New treatment
- C 451. View treatment
- C 452. New replicates
- C 453. View replicates
- C 454. New array
- C 455. View array
- C 456. New analysis
- C 457. View analysis
- C 458. New report
- C 459. View report
- C 460. New transcript
- C 461. View transcript
- C 462. New treatment
- C 463. View treatment
- C 464. New replicates
- C 465. View replicates
- C 466. New array
- C 467. View array
- C 468. New analysis
- C 469. View analysis
- C 470. New report
- C 471. View report
- C 472. New transcript
- C 473. View transcript
- C 474. New treatment
- C 475. View treatment
- C 476. New replicates
- C 477. View replicates
- C 478. New array
- C 479. View array
- C 480. New analysis
- C 481. View analysis
- C 482. New report
- C 483. View report
- C 484. New transcript
- C 485. View transcript
- C 486. New treatment
- C 487. View treatment
- C 488. New replicates
- C 489. View replicates
- C 490. New array
- C 491. View array
- C 492. New analysis
- C 493. View analysis
- C 494. New report
- C 495. View report
- C 496. New transcript
- C 497. View transcript
- C 498. New treatment
- C 499. View treatment
- C 500. New replicates
- C 501. View replicates
- C 502. New array
- C 503. View array
- C 504. New analysis
- C 505. View analysis
- C 506. New report
- C 507. View report
- C 508. New transcript
- C 509. View transcript
- C 510. New treatment
- C 511. View treatment
- C 512. New replicates
- C 513. View replicates
- C 514. New array
- C 515. View array
- C 516. New analysis
- C 517. View analysis
- C 518. New report
- C 519. View report
- C 520. New transcript
- C 521. View transcript
- C 522. New treatment
- C 523. View treatment
- C 524. New replicates
- C 525. View replicates
- C 526. New array
- C 527. View array
- C 528. New analysis
- C 529. View analysis
- C 530. New report
- C 531. View report
- C 532. New transcript
- C 533. View transcript
- C 534. New treatment
- C 535. View treatment
- C 536. New replicates
- C 537. View replicates
- C 538. New array
- C 539. View array
- C 540. New analysis
- C 541. View analysis
- C 542. New report
- C 543. View report
- C 544. New transcript
- C 545. View transcript
- C 546. New treatment
- C 547. View treatment
- C 548. New replicates
- C 549. View replicates
- C 550. New array
- C 551. View array
- C 552. New analysis
- C 553. View analysis
- C 554. New report
- C 555. View report
- C 556. New transcript
- C 557. View transcript
- C 558. New treatment
- C 559. View treatment
- C 560. New replicates
- C 561. View replicates
- C 562. New array
- C 563. View array
- C 564. New analysis
- C 565. View analysis
- C 566. New report
- C 567. View report
- C 568. New transcript
- C 569. View transcript
- C 570. New treatment
- C 571. View treatment
- C 572. New replicates
- C 573. View replicates
- C 574. New array
- C 575. View array
- C 576. New analysis
- C 577. View analysis
- C 578. New report
- C 579. View report
- C 580. New transcript
- C 581. View transcript
- C 582. New treatment
- C 583. View treatment
- C 584. New replicates
- C 585. View replicates
- C 586. New array
- C 587. View array
- C 588. New analysis
- C 589. View analysis
- C 590. New report
- C 591. View report
- C 592. New transcript
- C 593. View transcript
- C 594. New treatment
- C 595. View treatment
- C 596. New replicates
- C 597. View replicates
- C 598. New array
- C 599. View array
- C 600. New analysis
- C 601. View analysis
- C 602. New report
- C 603. View report
- C 604. New transcript
- C 605. View transcript
- C 606. New treatment
- C 607. View treatment
- C 608. New replicates
- C 609. View replicates
- C 610. New array
- C 611. View array
- C 612. New analysis
- C 613. View analysis
- C 614. New report
- C 615. View report
- C 616. New transcript
- C 617. View transcript
- C 618. New treatment
- C 619. View treatment
- C 620. New replicates
- C 621. View replicates
- C 622. New array
- C 623. View array
- C 624. New analysis
- C 625. View analysis
- C 626. New report
- C 627. View report
- C 628. New transcript
- C 629. View transcript
- C 630. New treatment
- C 631. View treatment
- C 632. New replicates
- C 633. View replicates
- C 634. New array
- C 635. View array
- C 636. New analysis
- C 637. View analysis
- C 638. New report
- C 639. View report
- C 640. New transcript
- C 641. View transcript
- C 642. New treatment
- C 643. View treatment
- C 644. New replicates
- C 645. View replicates
- C 646. New array
- C 647. View array
- C 648. New analysis
- C 649. View analysis
- C 650. New report
- C 651. View report
- C 652. New transcript
- C 653. View transcript
- C 654. New treatment
- C 655. View treatment
- C 656. New replicates
- C 657. View replicates
- C 658. New array
- C 659. View array
- C 660. New analysis
- C 661. View analysis
- C 662. New report
- C 663. View report
- C 664. New transcript
- C 665. View transcript
- C 666. New treatment
- C 667. View treatment
- C 668. New replicates
- C 669. View replicates
- C 670. New array
- C 671. View array
- C 672. New analysis
- C 673. View analysis
- C 674. New report
- C 675. View report
- C 676. New transcript
- C 677. View transcript
- C 678. New treatment
- C 679. View treatment
- C 680. New replicates
- C 681. View replicates
- C 682. New array
- C 683. View array
- C 684. New analysis
- C 685. View analysis
- C 686. New report
- C 687. View report
- C 688. New transcript
- C 689. View transcript
- C 690. New treatment
- C 691. View treatment
- C 692. New replicates
- C 693. View replicates
- C 694. New array
- C 695. View array
- C 696. New analysis
- C 697. View analysis
- C 698. New report
- C 699. View report
- C 700. New transcript
- C 701. View transcript
- C 702. New treatment
- C 703. View treatment
- C 704. New replicates
- C 705. View replicates
- C 706. New array
- C 707. View array
- C 708. New analysis
- C 709. View analysis
- C 710. New report
- C 711. View report
- C 712. New transcript
- C 713. View transcript
- C 714. New treatment
- C 715. View treatment
- C 716. New replicates
- C 717. View replicates
- C 718. New array
- C 719. View array
- C 720. New analysis
- C 721. View analysis
- C 722. New report
- C 723. View report
- C 724. New transcript
- C 725. View transcript
- C 726. New treatment
- C 727. View treatment
- C 728. New replicates
- C 729. View replicates
- C 730. New array
- C 731. View array
- C 732. New analysis
- C 733. View analysis
- C 734. New report
- C 735. View report
- C 736. New transcript
- C 737. View transcript
- C 738. New treatment
- C 739. View treatment
- C 740. New replicates
- C 741. View replicates
- C 742. New array
- C 743. View array
- C 744. New analysis
- C 745. View analysis
- C 746. New report
- C 747. View report
- C 748. New transcript
- C 749. View transcript
- C 750. New treatment
- C 751. View treatment
- C 752. New replicates
- C 753. View replicates
- C 754. New array
- C 755. View array
- C 756. New analysis
- C 757. View analysis
- C 758. New report
- C 759. View report
- C 760. New transcript
- C 761. View transcript
- C 762. New treatment
- C 763. View treatment
- C 764. New replicates
- C 765. View replicates
- C 766. New array
- C 767. View array
- C 768. New analysis
- C 769. View analysis
- C 770. New report
- C 771. View report
- C 772. New transcript
- C 773. View transcript
- C 774. New treatment
- C 775. View treatment
- C 776. New replicates
- C 777. View replicates
- C 778. New array
- C 779. View array
- C 780. New analysis
- C 781. View analysis
- C 782. New report
- C 783. View report
- C 784. New transcript
- C 785. View transcript
- C 786. New treatment
- C 787. View treatment
- C 788. New replicates
- C 789. View replicates
- C 790. New array
- C 791. View array
- C 792. New analysis
- C 793. View analysis
- C 794. New report
- C 795. View report
- C 796. New transcript
- C 797. View transcript
- C 798. New treatment
- C 799. View treatment
- C 800. New replicates
- C 801. View replicates
- C 802. New array
- C 803. View array
- C 804. New analysis
- C 805. View analysis
- C 806. New report
- C 807. View report
- C 808. New transcript
- C 809. View transcript
- C 810. New treatment
- C 811. View treatment
- C 812. New replicates
- C 813. View replicates
- C 814. New array
- C 815. View array
- C 816. New analysis
- C 817. View analysis
- C 818. New report
- C 819. View report
- C 820. New transcript
- C 821. View transcript
- C 822. New treatment
- C 823. View treatment
- C 824. New replicates
- C 825. View replicates
- C 826. New array
- C 827. View array
- C 828. New analysis
- C 829. View analysis
- C 830. New report
- C 831. View report
- C 832. New transcript
- C 833. View transcript
- C 834. New treatment
- C 835. View treatment
- C 836. New replicates
- C 837. View replicates
- C 838. New array
- C 839. View array
- C 840. New analysis
- C 841. View analysis
- C 842. New report
- C 843. View report
- C 844. New transcript
- C 845. View transcript
- C 846. New treatment
- C 847. View treatment
- C 848. New replicates
- C 849. View replicates
- C 850. New array
- C 851. View array
- C 852. New analysis
- C 853. View analysis
- C 854. New report
- C 855. View report
- C 856. New transcript
- C 857. View transcript
- C 858. New treatment
- C 859. View treatment
- C 860. New replicates
- C 861. View replicates
- C 862. New array
- C 863. View array
- C 864. New analysis
- C 865. View analysis
- C 866. New report
- C 867. View report
- C 868. New transcript
- C 869. View transcript
- C 870. New treatment
- C 871. View treatment
- C 872. New replicates
- C 873. View replicates
- C 874. New array
- C 875. View array
- C 876. New analysis
- C 877. View analysis
- C 878. New report
- C 879. View report
- C 880. New transcript
- C 881. View transcript
- C 882. New treatment
- C 883. View treatment
- C 884. New replicates
- C 885. View replicates
- C 886. New array
- C 887. View array
- C 888. New analysis
- C 889. View analysis
- C 890

# Many Datasets

**All Other Projects**

Project	Description	Access	Investigator
<a href="#">AntonosporaDB</a>	Website for the <i>Antonospora locustae</i> genome project.	Public	Hilary Morrison
<a href="#">Aureococcus</a>	The <i>Aureococcus</i> Serial Analysis of Gene Expression (SAGE) project. Genome and EST data acquired from GenBank and other external sources.	Private	Elizabeth Orchard
<a href="#">Blochmannia pennsylvanicus</a>	The <i>Blochmannia pennsylvanicus</i> genome sequencing assembly and annotation.	Private	Jennifer Wernegreen
<a href="#">Emiliania huxleyi</a>	The <i>Emiliania huxleyi</i> Serial Analysis of Gene Expression (SAGE) project. Genome and EST data acquired from GenBank and other external sources.	Public	Sonya Dyrman
<a href="#">Entamoeba histolytica</a>	Annotation of the <i>Entamoeba histolytica</i> genome, using genome data acquired from the Sanger and TIGR genome centers.	Coming Soon	<a href="mailto:gmod@lists.mbl.edu">gmod@lists.mbl.edu</a>
<a href="#">Gallus gallus</a>	The <i>Gallus gallus</i> Serial Analysis of Gene Expression (SAGE) project. EST data acquired from TIGR.	Private	Sean Kennedy
<a href="#">GiardiaDB</a>	Public informatics support for the <i>Giardia lamblia</i> genome and gene expression projects. Currently, GiardiaDB is based on the <i>giardia22</i> genome assembly.	Public	<a href="mailto:giardia@lists.mbl.edu">giardia@lists.mbl.edu</a>
<a href="#">giardia14</a>	Analysis, SAGE mapping, and annotation of the <i>giardia14</i> genome assembly. To be released to the public once completed.	Private	<a href="mailto:giardia@lists.mbl.edu">giardia@lists.mbl.edu</a>
<a href="#">Leishmania major</a>	Annotation of the <i>Leishmania major</i> genome, using genome data acquired from the Sanger and TIGR genome centers.	Coming Soon	<a href="mailto:gmod@lists.mbl.edu">gmod@lists.mbl.edu</a>
<a href="#">Leishmania tarentolae RNA editing database</a>	A database of <i>Leishmania tarentolae</i> maxicircle, minicircle, and guide RNA sequences.	Private	Larry Simpson
<a href="#">Marinobacter aquaeolei FOSMID 38/33, (manual annotation)</a>	Sequence assembly and manual annotation of <i>Marinobacter aquaeolei</i> FOSMID 38/33.	Private	<a href="mailto:Ashita.Dhillon@jgi.doe.gov">Ashita Dhillon</a>
<a href="#">Marinobacter aquaeolei FOSMID 38/33, (automated annotation)</a>	Sequence assembly and automated annotation of <i>Marinobacter aquaeolei</i> FOSMID 38/33.	Private	<a href="mailto:Ashita.Dhillon@jgi.doe.gov">Ashita Dhillon</a>
<a href="#">Marinobacter aquaeolei (complete genome)</a>	Annotation of the complete <i>Marinobacter aquaeolei</i> genome, using genome data acquired from the DOE Joint Genome Institute.	Private	<a href="mailto:Susan.Huse@jgi.doe.gov">Susan Huse</a>
<a href="#">Pholidina roseola</a>	Sequence assembly and automated annotation of cosmid and other sequences of the rotifer <i>Pholidina roseola</i> .	Private	<a href="mailto:David.Welch@jgi.doe.gov">David Mark Welch</a>
<a href="#">Schistosoma mansoni</a>	Annotation of the <i>Schistosoma mansoni</i> genome, including mapping of results of Serial Analysis of Gene Expression (SAGE) experiments. Genome data acquired from the Sanger and TIGR genome centers.	Private	<a href="mailto:gmod@lists.mbl.edu">gmod@lists.mbl.edu</a>
<a href="#">Schistosoma mansoni EST</a>	Annotation of <i>Schistosoma mansoni</i> EST data, including mapping of results of Serial Analysis of Gene Expression (SAGE) experiments. EST sequence data acquired from the TIGR and ONSA genome centers.	Private	<a href="mailto:gmod@lists.mbl.edu">gmod@lists.mbl.edu</a>
<a href="#">Trypanosoma brucei brucei</a>	Annotation of the <i>Trypanosoma brucei brucei</i> genome, including mapping of results of Serial Analysis of Gene Expression (SAGE) experiments in <i>T. b. brucei</i> and <i>T. b. rhodesiense</i> . Genome sequence data acquired from the Sanger and TIGR genome centers.	Private	<a href="mailto:gmod@lists.mbl.edu">gmod@lists.mbl.edu</a>
<a href="#">Trypanosoma brucei brucei GenBank Entries</a>	Annotation of the <i>Trypanosoma brucei brucei</i> GenBank sequences, including mapping of results of Serial Analysis of Gene Expression (SAGE) experiments in <i>T. b. brucei</i> and <i>T. b. rhodesiense</i> . Genome survey sequences (GSS) and genome sequence data acquired from the Sanger and TIGR genome centers not included (see above database).	Private	<a href="mailto:gmod@lists.mbl.edu">gmod@lists.mbl.edu</a>
<a href="#">Trypanosoma brucei brucei RNA</a>	A database of <i>Trypanosoma brucei brucei</i> maxicircle, minicircle, guide RNA, and sampled mitochondrial mRNA sequences with various	Private	Torsten O�senreiter

The AMOS consortium is committed to the development of open-source whole genome assembly software. The project acronym (AMOS) represents its primary goal - to produce A Modular, Open-Source whole genome assembler. Open-source so that everyone is welcome to contribute code and documentation. Assembly tools, and modular in nature, so that new contributions can be easily inserted into an existing assembly pipeline. This modular design will foster the development of new assembly algorithms and allow the AMOS project to continually grow and improve in hopes of eventually becoming a widely accepted and deployed assembly infrastructure. The [McArthur Laboratory](#) is currently developing AMOS modules to incorporate genome assembly information within GMOD/GBrowse genome databases, such as is currently available for [GiardiaDB](#).



Development of the JBPC GMOD Servers and development of the JBPC computing infrastructure have been supported by the National Institutes of Health, the National Science Foundation, NASA, the Josephine Bay Paul and C. Michael Paul Foundation, the W.M. Keck Foundation, the G. Unger Vetlesen Foundation, and the Ellison Medical Foundation.

# PHRAP Assembly

# Ability to import many different datasets

# Sanger/TIGR data

The screenshot shows a Mozilla Firefox browser window with the following details:

- Title Bar:** BlochmanniaDB - Mozilla
- Menu Bar:** File Edit View Go Bookmark Tools Window Help
- Address Bar:** http://bluebook.mbl.edu/perl/site/blochmannia06
- Toolbar:** Back, Forward, Stop, Home, Bookmarks, Gmail, Bugzilla, Taxonomy browser, GiardiaDB-14, GiardiaDB, GiardiaDB Admin, T. cruzi, T brucei brucei, T brucei gamblei, Rna Editing.
- Content Area:**
  - Header:** BlochmanniaDB
  - Text:** Informatics Support for the *Blochmannia pennsylvanicus* Genome and Gene Expression Projects
  - Text (below):** This current release is a development version constructed using the Generic Model Organism Database paradigm. The assembly data, database structure, and analytical tools are evolving on a daily basis. Plan your experiments accordingly.
- Bottom Navigation:** search available annotation, GBrowse, Help
- Bottom Footer:** Evidence of Host-Level Selection on the Genome Evolution of an Insect

The figure shows a screenshot of a Firefox browser window. The title bar reads "Trypanosoma brucei brucei GMOD - Mozilla Firefox". The address bar shows the URL "http://bluebook.mbl.edu/perl/site/t\_brucei\_brucel02". The main content area features a large red header "T. b. brucei Genome Browser". Below it, a sub-header states "This current release is a development version constructed using the [Generic Model Organism Database](#) paradigm." A search bar contains the placeholder "search available annotation" and a link "GBrowse [Help]". On the left, there's a sidebar with sections like "Introduction" and "Summary Statistics". The "Summary Statistics" table includes rows for Total Number of Contigs (113), Total Number of Supercontigs (20), Average Shotgun Coverage (0.00), Estimated Closure (of 26 Mbp) (99.06%), Predicted Open Reading Frames (9,737), Transcribed ORFs (SAGE detection) (5,905), Number of SAGE Libraries (11), Number of Unique SAGE Tags (filtered) (28,200), and SAGE Tags mapped to ORFs (11,779). A large green arrow points from the top right towards the "Summary Statistics" table.

The screenshot shows a Mozilla Firefox browser window with the following details:

- Title Bar:** OrganismDB - Mozilla
- Menu Bar:** File Edit View Go Bookmarks Tools Window Help
- Address Bar:** http://bluebook.mbl.edu/perl/site/\_smansoniest
- Toolbar:** Home Bookmarks Gmail Bugzilla Taxonomy brows... GiardiaDB-14 GiardiaDB Admin T. cruzi T brucei brucei
- Page Content:**
  - Header:** Schistosoma mansoni EST DataBase
  - Search Bar:** search available annotation
  - Links:** GBrowse [Help]
  - Section:** My Organism
  - Text:** Place information here
  - Summary Statistics:** A table showing database statistics:

Total Number of Contigs	33,704
Total Number of Supercontigs	33,704
Average Shotgun Coverage	1.00
Estimated Closure (of 270 Mbp)	6.48%
Predicted Open Reading Frames	8,375
Transcribed ORFs (SAGE detection)	3,280
Number of SAGE Libraries	9

## EST Dataset

ARACHNE Assembly