Genome Suite | 1.0.2

GLBRC Genome Suite Manual

Documentation for users interacting with the GLBRC Genome Suite

https://gs.glbrc.org/

Additional questions can be directed to helpdesk@glbrc.org

1 Table of Contents

2 Getting Started 2
2.1 Logging into the GLBRC Genome Suite 2
2.2 Logging out of the GLBRC Genome Suite 2
2.3 If you need help at any time 2

3 Profile 3
3.1 Account Management 3
3.2 Favorites 3
3.3 Recent Blasts 3
3.4 Groups 3
3.5 Samples 4

4 Samples 4
4.1 Sample Types 4
4.2 Sample Permissions 5
4.3 Listing Samples 5
4.4 Sample Details 6
4.5 Editing Samples 6

5 Sequence 8
5.1 Sequence Permissions 9
5.2 Listing Genomes and Transcriptomes 9
5.3 Searching Sequence 10

6 Features 11
6.1 Listing Genomic Features 11
6.2 Searching Features 12
6.3 Feature Detail Formats 13
6.4 Editing Feature Properties 18

7 Quantitative Expression Tool 21
7.1 Building a Matrix 21
7.2 Building a ratio 22
7.3 Expression interface 22

8 Genomic Context Tool 26
8.1 Navigation 26
8.2 Data Panels 28
8.3 Display 31

9 Local Blast Tool 33
9.1 Options 34
9.2 Viewing Blast Reports 35

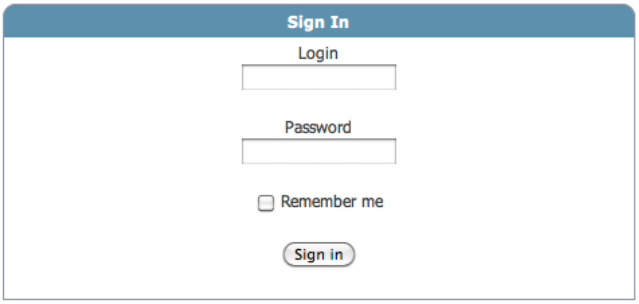
2 Getting Started

2.1 Logging into the GLBRC Genome Suite

Go to https://gs.glbrc.org

Select the Sign In link on the top right

Log in with your Hub credentials. You can make logging in easier next time by checking the "remember me" box. After login, you will be shown your User Profile page.



2.2 Logging out of the GLBRC Genome Suite

Select the Log Out link on the top right

Your session will be removed and you will have to login again to view any authorized data. Logging out does not remove any saved items from your account. They will be waiting for you when you return.

2.3 If you need help at any time

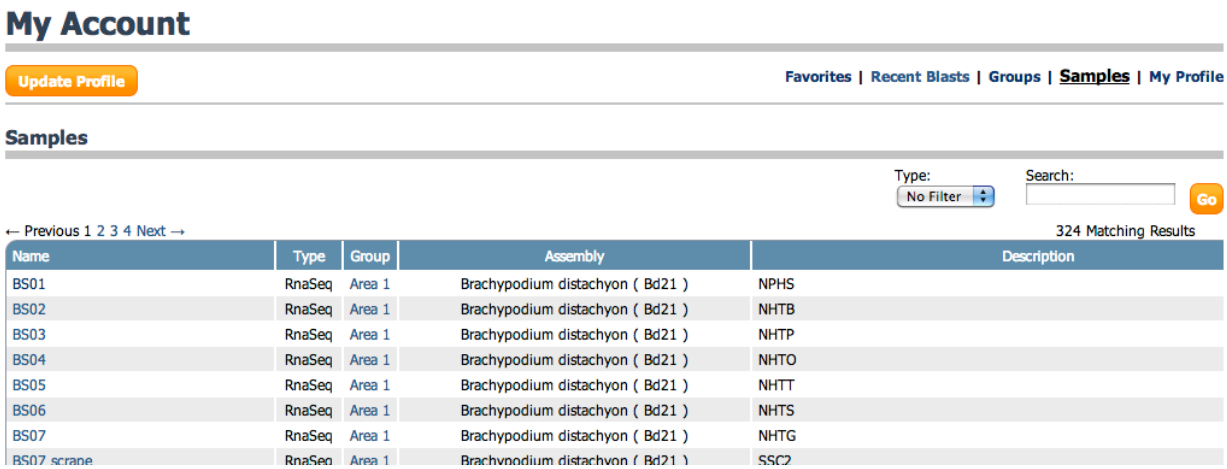
First look for a box and click on it.

If you need more information you can check the FAQ and Tutorial sections of the help tab at the top right of the page.

If you are still struggling and you cannot find the information you are looking for please contact helpdesk@glbrc.org

3 Profile

To access user profile information select the My Account link on the top right after logging in. Your account information page will be displayed. The information available for a user profile is detailed below.



3.1 Account Management

User Profiles have basic account management. Select update profile  on the top left to change the e-mail associated with your account. If you are a member of GLBRC your login and password are connected with the GLBRC center credentials and cannot be changed. If you are not using center wide credentials you can also change your login and password here.

3.2 Favorites

Throughout the site there are star links to add a feature to your favorites list  . These links are toggled with a simple click. Yellow stars mark items that will be saved in your favorites. If your account has favorites, this list is loaded by default when you login. The listing includes the feature locus, definition and a link to more details on the feature. Favorites can also be used to filter listings throughout the site.

3.3 Recent Blasts

The blast tool described in Ch. 9 tracks recent blast runs by user account. The history of recent blasts includes information about your query and the database used along with the BLAST result. This information is displayed in a listing with a link to the results.

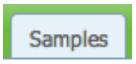
3.4 Groups

Each account can be assigned to multiple groups in the site. By default all accounts will be members of the **public** group and all GLBRC members belong to the **Glbrc** group. The listing displays the group name, number of members, and number samples in each group.

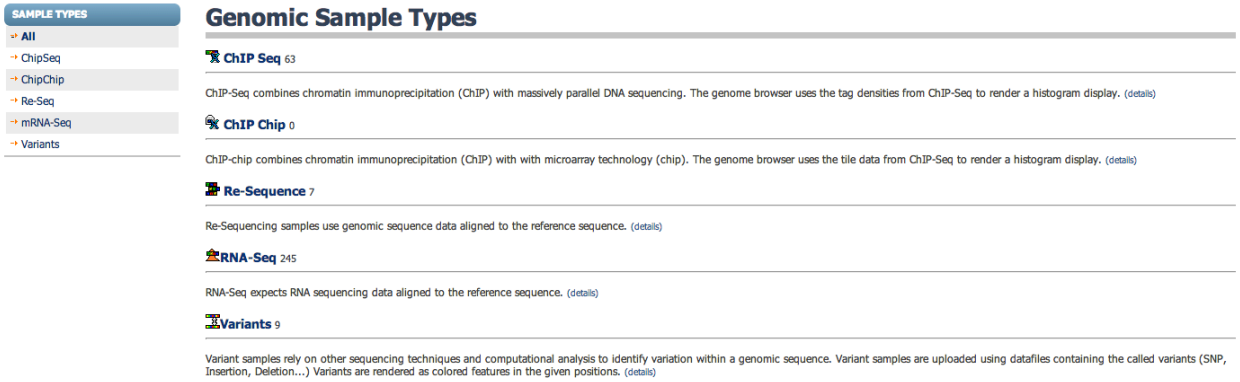
3.5 Samples

Every accessible sample is presented in this listing. You can filter by sample type and search on name and description. Two links are provided on this table, a link to View more sample details and a link to Edit sample data. Viewing and editing samples are described in more detail in sections 3.3 and 3.4.

4 Samples

Selecting the samples tab 

on the top right will display a listing and description of each sample type, along with the number of each type accessible by your account.



4.1 Sample Types

There are 5 sample types available in the Genome Suite to help organize different types of genomic experimental data. Each sample type is described below.

4.1.1 ChIP-Seq

Chip-Seq samples combine chromatin immunoprecipitation with massively parallel DNA sequencing. Tag densities from sequence data are used to render histograms and display peak data locations.

4.1.2 ChIP-Chip

ChIP-Chip samples combine chromatin immunoprecipitation with microarray technology. Signal measures from tile data are used to render histograms and display peak data locations.

4.1.3 Re-Sequence

Re–sequencing samples use massively parallel DNA sequencing to collect new genomic data for a sequenced genome. Sequence alignments are used to render histograms and read mapping details.

4.1.4 RNA-Seq

RNA-Seq samples use massively parallel RNA sequencing to collect sequence data from RNA. Sequence alignments are used to render histograms and read mapping details. Expression quantified from sequence alignments is used to render tabular listings.

4.1.5 Variant

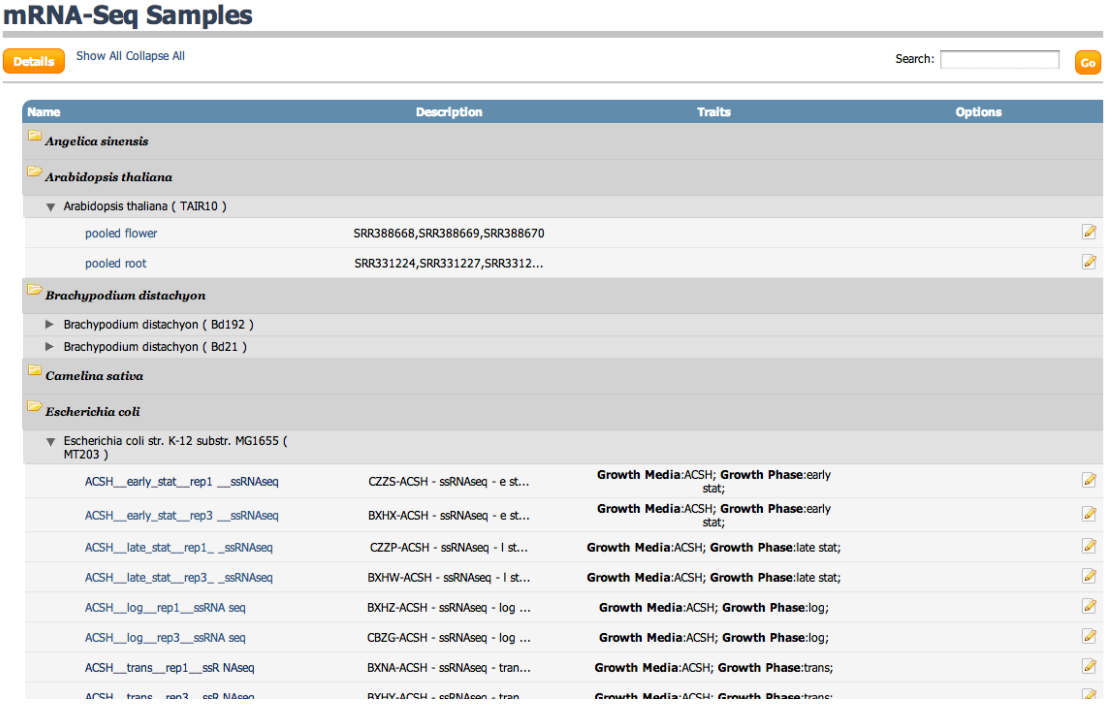
Variant samples use massively parallel DNA sequencing and computational analysis to identify variation within a known genomic sequence. Variant samples use the identified differences to render information at the given positions and generate altered sequence for features.

4.2 Sample Permissions

Samples are made accessible to a user account based on group membership. Each sample belongs to a group. If your account is also a member of the same group, you will have access to that sample. A site administrator manages group membership for user accounts, and the creator of each sample defines its group.

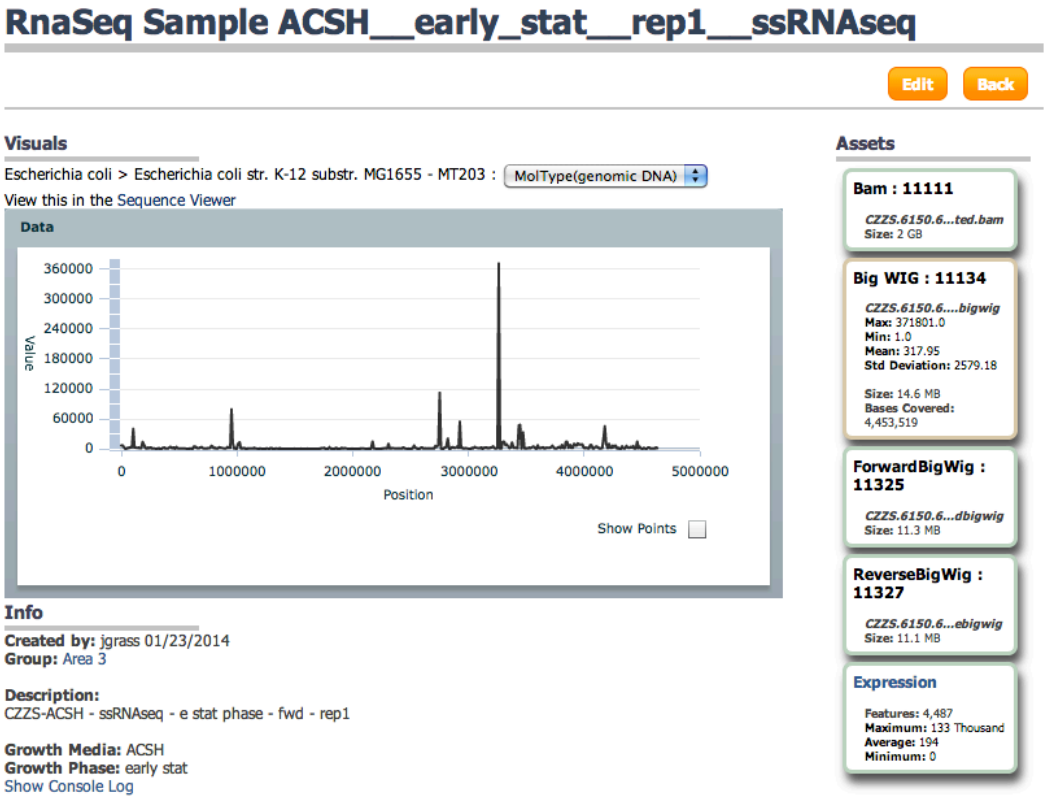
4.3 Listing Samples

Select a sample type from the description or the navigational pane on the top left to view a listing of those samples. A hierarchical tree organized by taxonomy will be displayed. The first level is species and the second level is strain or assembly version. On the third level each sample is listed with its name, description, and user defined traits. Samples can be searched by name using the keyword box in the upper right. Sample Names are linked to more details about the sample. A quick link to edit  is displayed on the far right if the user account has update permission.



4.4 Sample Details

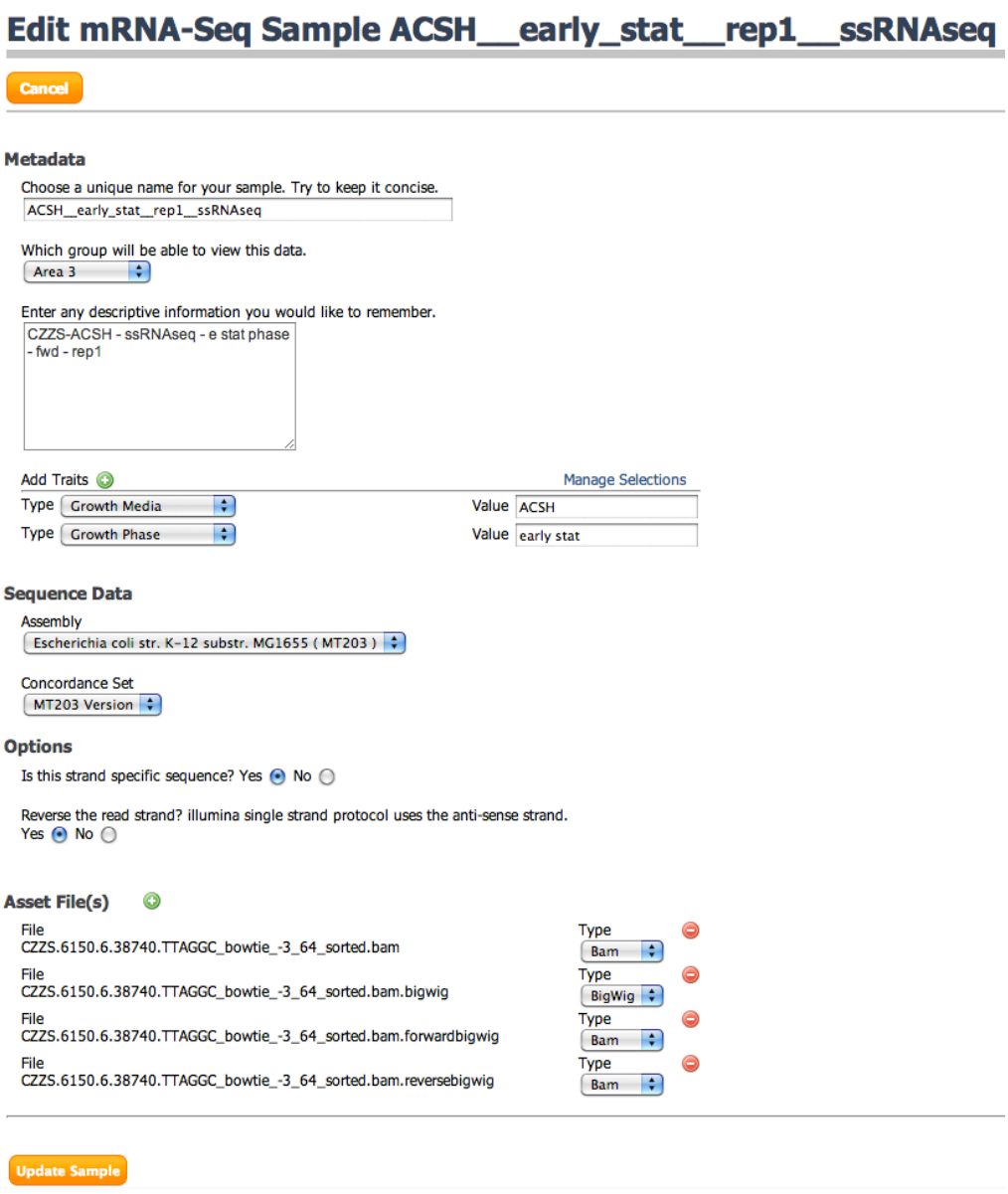
Each Sample has a details page displaying information stored by the system. Sample data are stored in the system as Assets. Each Asset is displayed with an information card showing basic details. Selecting an Asset card will display more details about that data including processing output and header details. Samples are connected to an assembly and a group. This information is displayed under the graphical rendering along with the sample description and named traits.



4.5 Editing Samples

If the current user account has edit permissions on a sample, an edit button 

will be displayed on the top right of the sample details page. Member accounts can edit samples they have access to. Selecting the edit button will open the sample form with options to manage all of the sample details. Each item in the sample form is described below.



4.5.1 Name

Each sample must have a name unique to the assembly. This name is an identifier used on many tables and columns and should be concise.

4.5.2 Description

The description is used for free form text allowing any information pertaining to the sample to be recorded. This could include processing steps, or metadata about the sample.

4.5.3 Group

As the owner of a sample you have the ability to control access to it. Changing the group for a sample will immediately give access to all users in the new group.

4.5.4 Managing Sample Traits

Sample traits can be used for key value descriptive information about sample metadata. Traits are more powerful than free text desc

4.5.5 Assets Types

4.5.5.1 Bam

Compressed binary file containing reads aligned to reference sequence. Bam files can be used by RNA-Seq and Re-Sequencing samples.

4.5.5.2 Wig

Density data along a reference sequence. Wig files can be used by Chip-Chip, Chip-Seq and RNA-Seq samples.

4.5.5.3 BigWig

The indexed binary format of Wig data files. BigWig files will automatically be generated from Bam and Wig files. BigWig files can be used on Chip-Chip, Chip-Seq, and RNA-Seq samples.

4.5.5.4 VCF

Variants called from sequence alignments. SNP's, Insertions and Deletions from genomic coordinates. VCF files can only be used with Variant samples.

4.5.5.5 TabixVcf

An internal file type not likely to be provided by users. This is VCF file compressed by bgzip in preparation for indexing with Tabix. TabixVcf files will automatically be created from VCF files.

4.5.6 Managing Assets

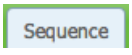
As the owner of a sample you can manage the sample asset data.

The green plus  icon can be used to add a new asset row. New assets have a file field to choose the file source from your local system. Existing assets display the original filename. You cannot change the file for an existing asset.

Asset rows also include a type dropdown. Selecting the correct type is critical to properly displaying data from the asset.

The red minus  icon next to each asset can be used to remove the asset from a sample. Clicking the icon will remove the asset row from view, but it will not immediately remove the asset. Submitting the form will save the changes and permanently remove the asset. **Removing assets cannot be undone!**

5 Sequence

Selecting the sequence tab 

on the top right will display a listing of sequence information. Sequence is categorized into Genomes and Transcriptomes. Both sequence

groups are referred to as Assemblies. The left navigation panel has links to view Genomes, view Transcriptomes, and Search sequence.

5.1 Sequence Permissions

Assemblies and the sequence attached to them are accessible through group permissions and sample settings. Accounts are granted access to an assembly if they have access to a sample connected with that assembly. Assemblies can also be tied directly to a group. In this case if an account is a member of the same group the have access to the assembly. As a GLBRC member, having access to an assembly will also grant access to manage annotation on that assembly. Guest accounts and public access are read only and cannot manipulate annotations.

5.2 Listing Genomes and Transcriptomes

The listings for Genomes and Transcriptomes are setup the same. The listing is grouped into folders for each species. Within each species folder the assemblies are listed by taxon and version along with links to more information about each assembly. The option links are detailed below.

Species	Taxon	Version	Options
Arabidopsis thaliana			
	Arabidopsis thaliana	TAIR10	Explore 7 entries | View First entry
Brachypodium distachyon			
	Brachypodium distachyon	Bd192	Explore 83 entries | View First entry | Expression
	Brachypodium distachyon	Bd21	Explore 45 entries | View First entry | Expression
Escherichia coli			
	Escherichia coli str. K-12 substr. MG1655	MT203	Explore 1 entries | View First entry | Expression
Saccharomyces cerevisiae			
	Saccharomyces cerevisiae	R63 with JAY291	Explore 48 entries | View First entry
	EF4	68.72	Explore 17 entries | View First entry | Expression
	Saccharomyces cerevisiae S288c	R64	Explore 17 entries | View First entry | Expression
Sorghum bicolor			
	Sorghum bicolor	1.4	Explore 3304 entries | View First entry | Expression

5.2.1 Explore entries

The link to Explore X entries is a quick link to the search interface. The search will be filtered for the given assembly. The number X shows the count of sequences connected to this assembly. The search interface is described in more detail in section 4.2.

5.2.2 View First entry

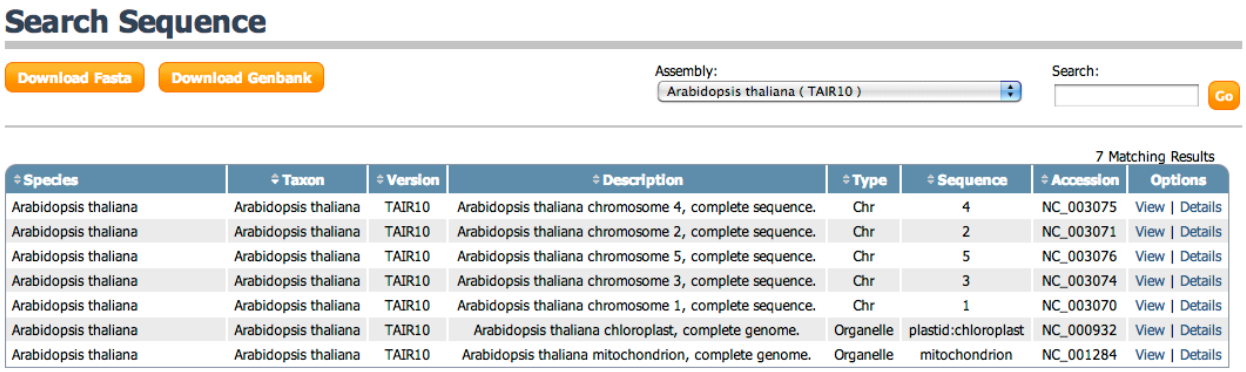
This link lets a user quickly navigate to the Genomic Context for an assembly. It uses the first sequence attached to the assembly to initialize the view.

5.2.3 Expression

This link will direct users to the Quantitative Expression Tool with the given assembly selected. The quick link helps to connect users with other areas within the site.

5.3 Searching Sequence

The Search link in the left navigation pane and the Explore X entries link from the assembly listings will open up the sequence search page. This page can be used to find any accessible sequence within the site. The basic search form on the top right allows filtering on assembly and text based searching of assembly information. The keywords are compared to all columns using a tokenized search.



Species	Taxon	Version	Description	Type	Sequence	Accession	Options
Arabidopsis thaliana	Arabidopsis thaliana	TAIR10	Arabidopsis thaliana chromosome 4, complete sequence.	Chr	4	NC_003075	View | Details
Arabidopsis thaliana	Arabidopsis thaliana	TAIR10	Arabidopsis thaliana chromosome 2, complete sequence.	Chr	2	NC_003071	View | Details
Arabidopsis thaliana	Arabidopsis thaliana	TAIR10	Arabidopsis thaliana chromosome 5, complete sequence.	Chr	5	NC_003076	View | Details
Arabidopsis thaliana	Arabidopsis thaliana	TAIR10	Arabidopsis thaliana chromosome 3, complete sequence.	Chr	3	NC_003074	View | Details
Arabidopsis thaliana	Arabidopsis thaliana	TAIR10	Arabidopsis thaliana chromosome 1, complete sequence.	Chr	1	NC_003070	View | Details
Arabidopsis thaliana	Arabidopsis thaliana	TAIR10	Arabidopsis thaliana chloroplast, complete genome.	Organelle	plastid:chloroplast	NC_000932	View | Details
Arabidopsis thaliana	Arabidopsis thaliana	TAIR10	Arabidopsis thaliana mitochondrion, complete genome.	Organelle	mitochondrion	NC_001284	View | Details

5.3.1 Downloading Sequence

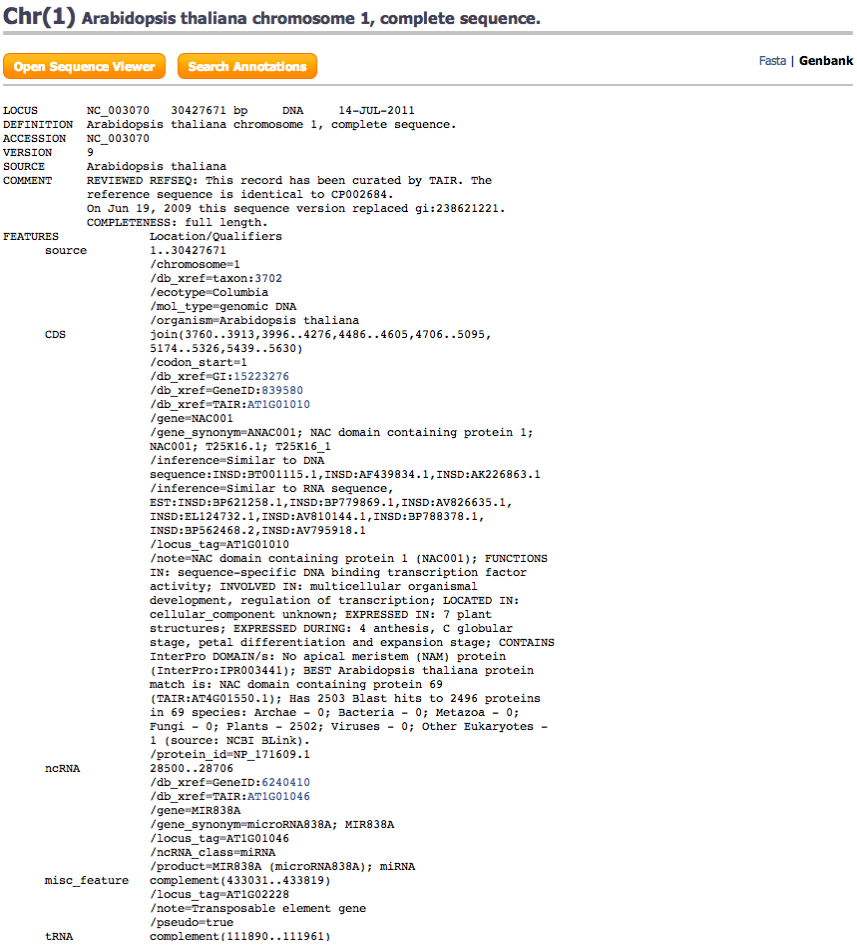
The filtered sequence list can be downloaded for further use outside the Genome Suite. The buttons on the top left allow users to download a Fasta file containing sequence names and bases or a Genbank file containing all annotations as well as sequence base date.

5.3.2 View Option

The View link will open the Genomic Context for this sequence. More information about the Genomic Context is given in Ch. 8 Genomic Context Tool.

5.3.3 Details Option

The Details link will open a detailed page for the specific sequence entry. This page can be rendered in Genbank or Fasta format. An example details page for Arabidopsis thaliana chromosome 1 is displayed below.



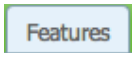
Chr(1) Arabidopsis thaliana chromosome 1, complete sequence.

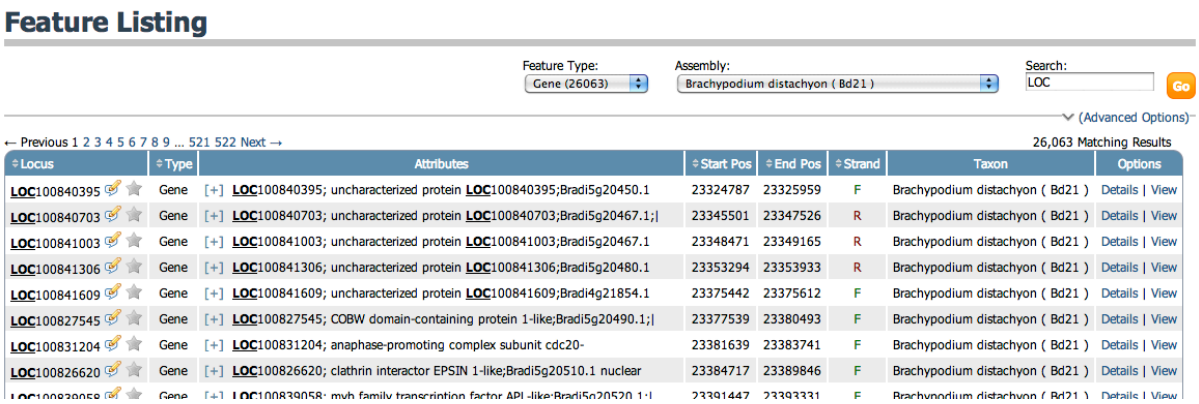
Open Sequence Viewer | Search Annotations | Fasta | **Genbank**

```
LOCUS       NC_003070   30427671 bp     DNA     14-JUL-2011
DEFINITION  Arabidopsis thaliana chromosome 1, complete sequence.
ACCESSION   NC_003070
VERSION     9
SOURCE      Arabidopsis thaliana
COMMENT     REVIEWED REFSEQ: This record has been curated by TAIR. The
            reference sequence is identical to CP002684.
            On Jun 19, 2009 this sequence version replaced gi:238621221.
            COMPLETENESS: full length.
FEATURES             Location/Qualifiers
     source          1..30427671
                     /chromosome=1
                     /db_xref=taxon:3702
                     /ecotype=Columbia
                     /mol_type=genomic DNA
                     /organism=Arabidopsis thaliana
     CDS             join(3760..3913,3996..4276,4486..4605,4706..5095,
                     5174..5326,5439..5630)
                     /codon_start=1
                     /db_xref=GI:15223276
                     /db_xref=GeneID:839580
                     /db_xref=TAIR:AT1G01010
                     /gene=NAC001
                     /gene_synonym=ANAC001; NAC domain containing protein 1;
                     NAC001; T25K16.1; T25K16_1
                     /inference=Similar to DNA
                     sequence:INSD:BT001115.1,INSD:AF439834.1,INSD:AK226863.1
                     /inference=Similar to RNA sequence,
                     EST:INSD:BP621258.1,INSD:BP779869.1,INSD:AV826635.1,
                     INSD:EL124732.1,INSD:AV810144.1,INSD:BP788378.1,
                     INSD:BP562468.2,INSD:AV795918.1
                     /locus_tag=AT1G01010
                     /note=NAC domain containing protein 1 (NAC001); FUNCTIONS
                     IN: sequence-specific DNA binding transcription factor
                     activity; INVOLVED IN: multicellular organismal
                     development, regulation of transcription; LOCATED IN:
                     cellular_component unknown; EXPRESSED IN: 7 plant
                     structures; EXPRESSED DURING: 4 anthesis, C globular
                     stage, petal differentiation and expansion stage; CONTAINS
                     InterPro DOMAIN/s: No apical meristem (NAM) protein
                     (InterPro:IPR003441); BEST Arabidopsis thaliana protein
                     match is: NAC domain containing protein 69
                     (TAIR:AT4G01550.1); Has 2503 Blast hits to 2496 proteins
                     in 69 species: Archae - 0; Bacteria - 0; Metazoa - 0;
                     Fungi - 0; Plants - 2502; Viruses - 0; Other Eukaryotes -
                     1 (source: NCBI BLink).
                     /protein_id=NP_171609.1
     ncRNA           28500..28706
                     /db_xref=GeneID:6240410
                     /db_xref=TAIR:AT1G01046
                     /gene=MIR838A
                     /gene_synonym=microRNA838A; MIR838A
                     /locus_tag=AT1G01046
                     /ncRNA_class=miRNA
                     /product=MIR838A (microRNA838A); miRNA
     misc_feature    complement(433031..433819)
                     /locus_tag=AT1G02228
                     /note=Transposable element gene
                     /pseudo=true
     tRNA            complement(111890..111961)
```

6 Features

6.1 Listing Genomic Features

Selecting the Features tab  on the top right will display a listing of all the accessible features in the site. Each of the table columns are detailed below.



Feature Listing

Feature Type: Gene (26063) | Assembly: Brachypodium distachyon (Bd21) | Search: LOC | Go

(Advanced Options)

← Previous 1 2 3 4 5 6 7 8 9 ... 521 522 Next →

26,063 Matching Results

Locus	Type	Attributes	Start Pos	End Pos	Strand	Taxon	Options
LOC100840395	Gene	[+] LOC100840395; uncharacterized protein LOC100840395;Bradi5g20450.1	23324787	23325959	F	Brachypodium distachyon (Bd21)	Details | View
LOC100840703	Gene	[+] LOC100840703; uncharacterized protein LOC100840703;Bradi5g20467.1;|	23345501	23347526	R	Brachypodium distachyon (Bd21)	Details | View
LOC100841003	Gene	[+] LOC100841003; uncharacterized protein LOC100841003;Bradi5g20467.1	23348471	23349165	R	Brachypodium distachyon (Bd21)	Details | View
LOC100841306	Gene	[+] LOC100841306; uncharacterized protein LOC100841306;Bradi5g20480.1	23353294	23353933	R	Brachypodium distachyon (Bd21)	Details | View
LOC100841609	Gene	[+] LOC100841609; uncharacterized protein LOC100841609;Bradi4g21854.1	23375442	23375612	F	Brachypodium distachyon (Bd21)	Details | View
LOC100827545	Gene	[+] LOC100827545; COBW domain-containing protein 1-like;Bradi5g20490.1;|	23377539	23380493	F	Brachypodium distachyon (Bd21)	Details | View
LOC100831204	Gene	[+] LOC100831204; anaphase-promoting complex subunit cdc20-	23381639	23383741	F	Brachypodium distachyon (Bd21)	Details | View
LOC100826620	Gene	[+] LOC100826620; clathrin interactor EPSIN 1-like;Bradi5g20510.1 nuclear	23384717	23389846	F	Brachypodium distachyon (Bd21)	Details | View
LOC100839058	Gene	[+] LOC100839058; myb family transcription factor APL-like;Bradi5g20520.1;|	23391447	23393331	F	Brachypodium distachyon (Bd21)	Details | View

6.1.1 Locus

The locus column displays the unique locus id given for the feature. A locus is only unique to an assembly and may appear in alternate versions. A locus can also be repeated for different feature types. For example, a Gene and mRNA may share the same locus. Two additional link icons are contained within the locus column. They are detailed below.

6.1.1.1 Quick Edit

The edit icon  can be clicked to open an interactive form. This form will allow updates to the functional annotation for the feature. See section 5.4.1 for more information.

6.1.1.2 Favorites

The favorite icon  /  can be used to add this feature to the favorites list describe in section 2.2. These favorites can be used to filter the expression listing. The filter option is a basic search option described in section 7.3.7.

6.1.2 Type

The type column displays the textual name of the feature type. Examples include Gene, mRNA and CDS.

6.1.3 Attributes

The attributes column lists the description and any blast definitions attached to the feature.

6.1.4 Start Pos

The minimum start position for the feature.

6.1.5 End Pos

The maximum end position for the feature.

6.1.6 Taxon

The assembly taxonomy this feature is associated with.

6.1.1 View Option

The View link will open the Genomic Context for the sequence this feature is attached to. More information about the Genomic Context is given in Ch. 8 Genomic Context Tool.

6.1.2 Details Option

The Details link will open the detailed format pages for this feature.

6.2 Searching Features

The basic search form on the top right of the feature listing allows filtering on assembly, feature type, and text based searching of feature annotations. The keywords are compared to the Attribute and Locus columns. The Advanced search options add filters for start and end location as well as strand orientation.

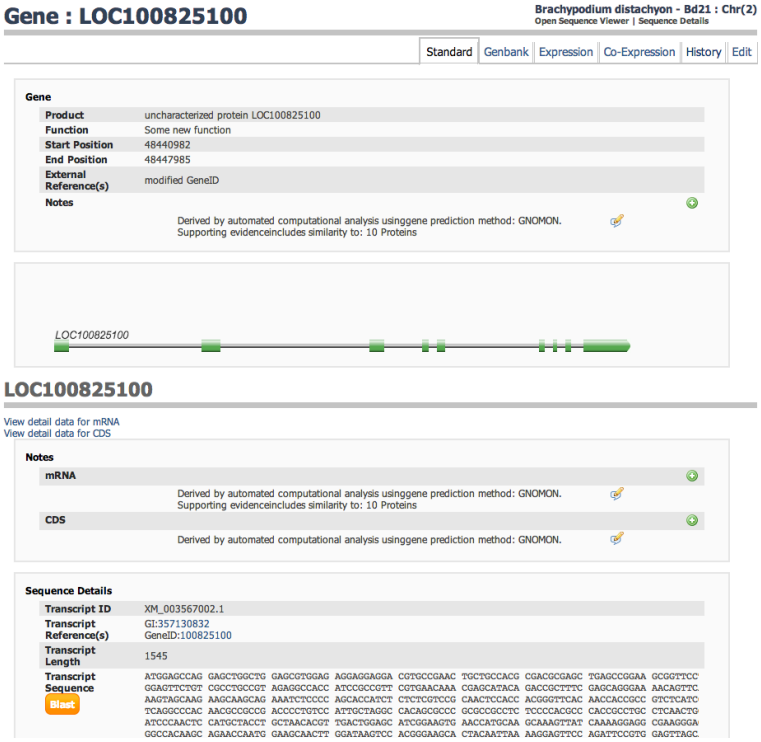
6.3 Feature Detail Formats

Throughout the site there are links to view Details for a feature. These links take you to the detailed feature format pages with the standard format displayed by default.

6.3.1 Standard

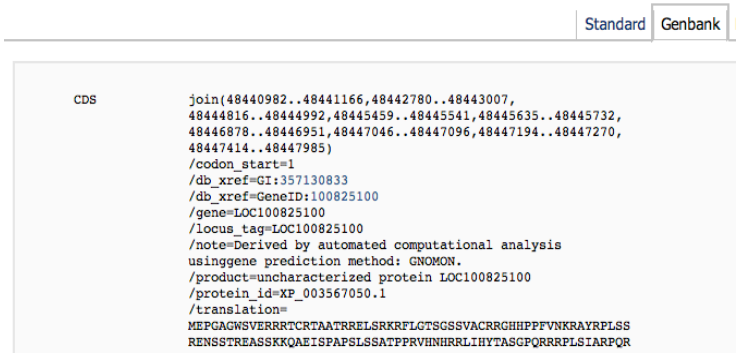
The standard format displays feature attributes, a graphical rendering of the feature, and sequence taken from the features location. This format also renders displays icons to quickly manage notes on the feature. If the feature is a CDS the translated protein sequence will also be displayed.

The standard format for a Gene feature is special. This page will also display Gene Model information including CDS and Mrna and the attributes associated with each.



6.3.2 Genbank

The Genbank format displays all of the features and attributes found for the give locus tag in the NCBI Genbank text format.

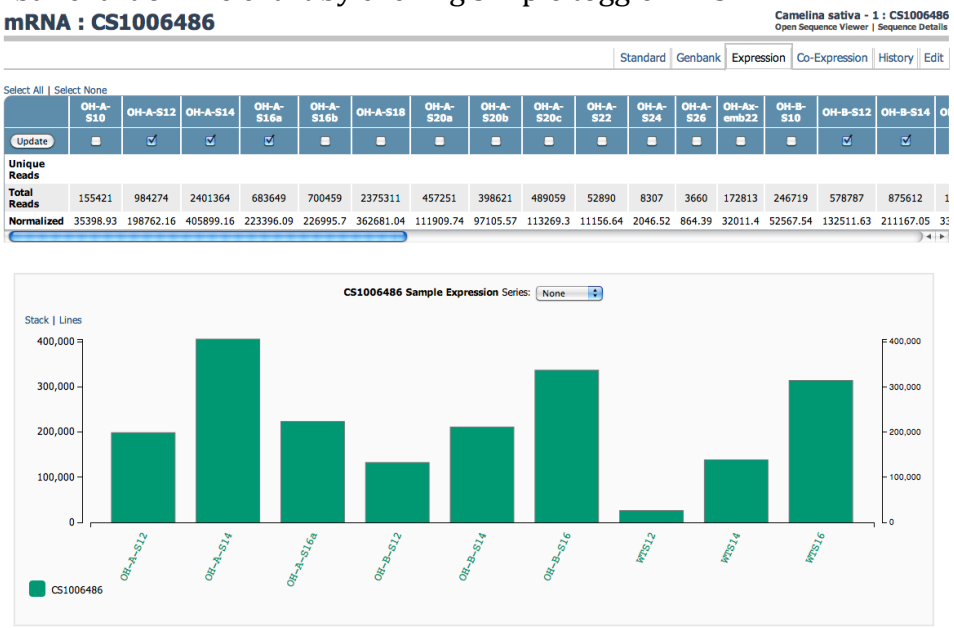


6.3.3 Expression

If the feature has sample expression data the Expression tab will be available. This format shows information from all sample expression data assigned to the feature. A selection table across the top of the page allows users to limit the datasets that are displayed. This table also shows the unique, total, and normalized values loaded for each sample. The selections made on this table are stored in user accounts. The expression tab and co-expression tab described in section 5.3.4 will load the most recent settings from the user profile. The expression page also renders two visualizations, the expression profile across samples, and the read density across the feature.

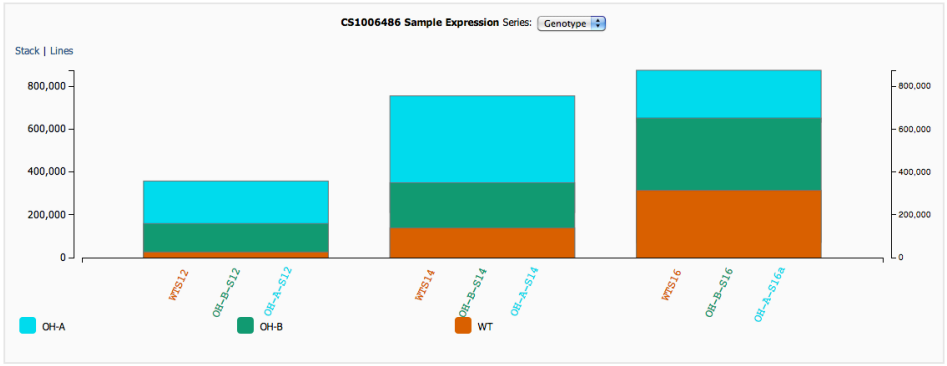
6.3.3.1 Expression Profile

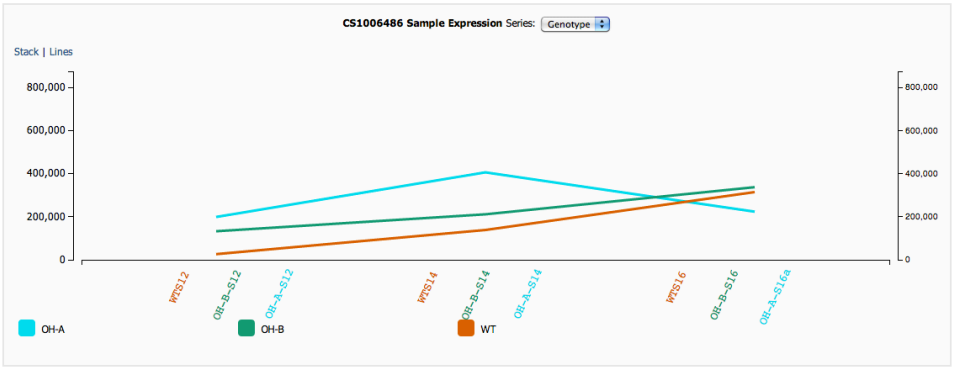
The expression profile renders each selected sample along the x-axis using the samples normalized expression value for the y-axis. This graph can be rendered with a bar chart or line chart by clicking simple toggle links.



6.3.3.2 Grouped Profile

The expression profile chart can be modified based on user defined sample traits. If sample traits are defined they can be used to generate multiple series on the expression profile. These series are then rendered as stacked bars or multiple lines depending on the users choice.

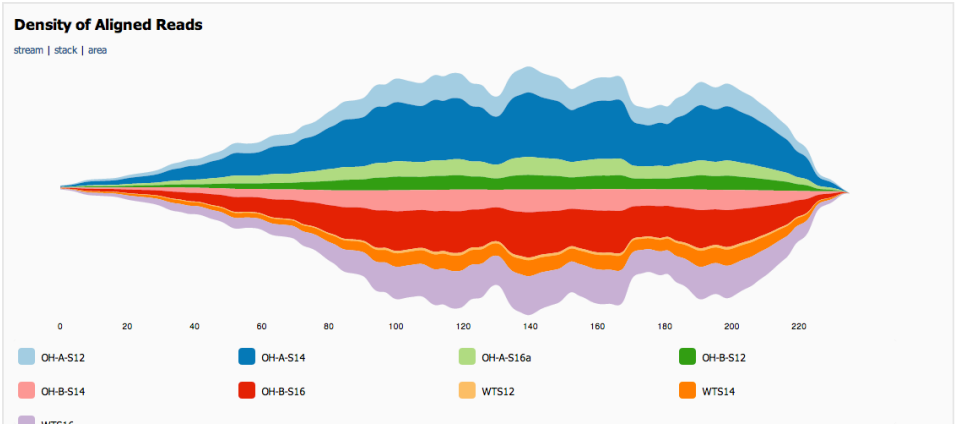




6.3.3.3 Read Density

The read density visualization shows read alignment depth on the feature. The x-axis is the location on the feature, and the y-axis is the number of aligned reads. Each selected sample is added as a new series to the chart. This chart can be rendered in 3 different formats.

Stream – An area chart with a dynamic y-axis allowing the data to adjust its center based on total density.
Stack - An area chart with total contributions adding to the overall height so that Y values from each series stack on top of each other.
Area - A basic area chart with semi-transparent colors. This transparency allows completely occluded samples to be seen.

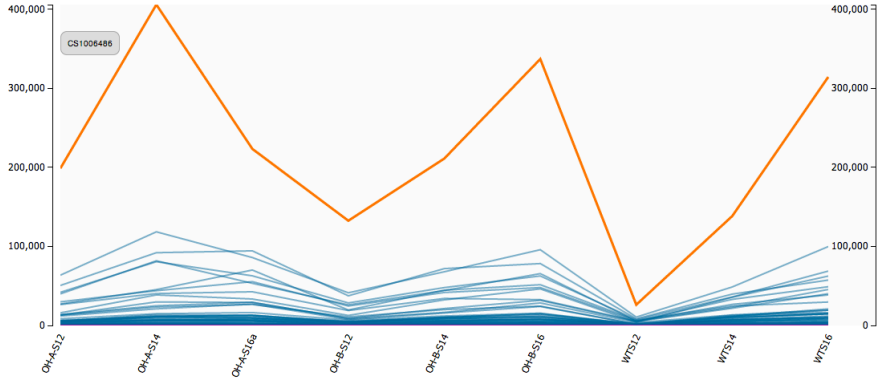


6.3.4 Co-Expression

If the feature has sample expression data the Co-Expression tab will also be available. This format lists information about features with similar expression profiles. A selection table across the top of the page allows users to limit the datasets that are displayed. This table also shows the unique, total, and normalized values loaded for each sample. The page also renders a visualization of expression profiles from similar genes linked to a tabular listing of the same genes. This view displays the top 500 similar genes determined by Pearson correlation coefficient.

6.3.4.1 Co-expressed graph

The co-expression graph displays samples along the x-axis and normalized expression values on the y-axis. Each co-expressed gene is added as a series to the graph and rendered as a line. Positive correlation is blue and negative correlation is pink. Using the mouse to hover over a line will show the Locus associated with that feature. Clicking the line will highlight the feature in the table below.



6.3.4.2 Co-expressed table

The co-expression table lists details for all 500 co-expressed genes. The table columns are Name, Definition, R, R^2, and average normalized value. Each of the columns can be used to sort the table. Clicking on a row highlights the corresponding gene in the graph and clicking on the Feature Name (Locus) will open the co-expression view for that feature. To enable outside processing a link to download the table is provided on the top right. This will send the table in textual CSV (comma separated value) format. This format can be opened directly with excel.

Download Table

Name	Definition	R	R2	AVG NORM
CS2006486		1	1	2.85
CS2076116		0.8005	0.6408	13.65
CS2051551	**| Symbols: | Major facilitator superfamily protein | chr5:3420876-3423166...**	**0.7298**	**0.5326**	**2.46**
CS2036298		0.6904	0.4767	8.42
CS2007039		0.6994	0.4892	1.2
CS2035997		0.235	0.0552	4.91
CS2062260	| Symbols: ATSGT1B, ETA3, RPR1, EDM1, SGT1B | phosphatase-related | chr4:6...	0.5941	0.353	6.57
CS2053549		0.8578	0.7358	8.6
CS2049775		0.2486	0.0618	1.39
CS2048710	| Symbols: | Mitochondrial transcription termination factor family protein | chr1...	0.9123	0.8323	0.83
CS2035516	| Symbols: | Helicase protein with RING/U-box domain | chr1:1471624-1476067...	-0.0279	0.0008	4.95
CS2015695		0.7956	0.633	2.15
CS2066787	| Symbols: | RNI-like superfamily protein | chr5:20872783-20874192 FORWARD...	0.3472	0.1205	8.62
CS2032250	| Symbols: | PLC-like phosphodiesterases superfamily protein | chr1:26885556-...	0.7778	0.605	1.32
CS2082196		0.8887	0.7898	6.03
CS2035132	| Symbols: | DNA-binding storekeeper protein-related transcriptional regulator |...	0.7318	0.5355	79.86
CS2035487		0.5394	0.291	2.74
CS2029289	| Symbols: | Plant stearoyl-acyl-carrier-protein desaturase family protein | chr5...	0.645	0.416	19.48
CS2009624	| Symbols: anac094, NAC094 | NAC domain containing protein 94 | chr5:159393...	0.3569	0.1274	1.34
CS2082152		0.7279	0.5298	1.15
CS2001898		0.716	0.5127	0.5
CS2010522		0.888	0.7885	1.83
CS2083002		0.8964	0.8035	3.39

6.3.5 Blast

Each feature can have multiple BLAST sequence alignment results. If BLAST results are attached to a feature the Blast tab will be shown. This tab displays a drop down list of each database result available for the feature. The first item in the list is selected by default. The Blast Report for the selected database is rendered below with all of the information discussed in section 8.2 Viewing Blast Reports.

6.3.6 Variants

If the feature has variant sample data the Variant tab will be displayed. This tab lists the feature sequence after applying alterations found within each sample. There is a new entry listed for every variant sample. Each entry displays altered nucleotide and protein sequence. The sequence is colored to show variant locations. Single nucleotide changes are colored blue and insertions are colored green.

GLBRCY2A::GLBRCY2A

```
>YAL064W-B Saccharomyces cerevisiae  - R63 with JAY291 : Chr(I) from 12047 to 12427
ATGGCAGGTGAAGCAGTTTCGGAACACACACCAGATTCGCAGGAAGTAACAGTAACTAGCGTAGTTTGTTGCCTCGATTCTGTGGTGGAAATAGGACACCATGTCGTGTAT
TCTGTGGTAACGCCGTTAATAGTAGCAGTGCTTGTAGATACAATGGCTGGAGAAGCAGTCTTAGAGCATACGTCAGATTCACAAGAAGAAATTGTAACTACTGTGGTTTGC
TCTGTTGTCCCCTTGGTTTGCTTAGTTGTCTCCGTAGTTTGCTTTGTTATCTCTGTGGTAGAAATAGGGCACCATGTGGTATACTCTGTTGTGACACCGCTAACAGTAACG
GTGGCCGTGGAAACAATGCCAGAGGAGATGGATTCAGTGCACACATGA

>YAL064W-B Saccharomyces cerevisiae  - R63 with JAY291 : Chr(I) from 12047 to 12427
MAGEAVSEHTPDSQEVTVTSVVCCLDSVVEIGHHVVYSVVTPLIVAVLVDTMAGEAVLEHTSDSQEEIVTTVVCSVVPLVCLVVSVVCFVISVVEIGHHVVYSVVTPLTVT
VAVETIAEEMDSVHT*
```

GLBRCY4A::GLBRCY4A

```
>YAL064W-B Saccharomyces cerevisiae  - R63 with JAY291 : Chr(I) from 12047 to 12427
ATGGCAGGTGAAGCAGTTTCGGAACACACACCAGATTCGCAGGAAGTAACAGTAACTAGCGTAGTTTGTTGCCTCGATTCTGTGGTGGAAATAGGACACCATGTCGTGTAT
TCTGTGGTAACGCCGTTAATAGTAGCAGTGCTTGTAGATACAATGGCTGGAGAAGCAGTCTTAGAGCATACGTCAGATTCACAAGAAGAAATTGTAACTACTGTGGTTTGT
TCTGTTGTCCCCTTGGTTTGCTTTGTTGTCTCCGTAGTTTGCTTTGTTATCTCTGTGGTAGAAATAGGGCACCATGTGGTATACTCTGTTGTGACACCGCTAACAGTAACG
GTGGCCGTGGAAACAATGCCAGAGGAGATGGATTCAGTGCACACATGA

>YAL064W-B Saccharomyces cerevisiae  - R63 with JAY291 : Chr(I) from 12047 to 12427
MAGEAVSEHTPDSQEVTVTSVVCCLDSVVEIGHHVVYSVVTPLIVAVLVDTMAGEAVLEHTSDSQEEIVTTVVCSVVPLVCFVVSVVCFVISVVEIGHHVVYSVVTPLTVT
VAVETIAEEMDSVHT*
```

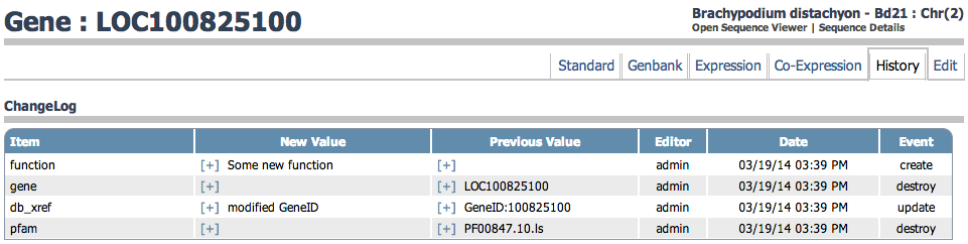
GLBRCY70::GLBRCY70

```
>YAL064W-B Saccharomyces cerevisiae  - R63 with JAY291 : Chr(I) from 12047 to 12427
ATGGCAGGTGAAGCAGTTTCGGAACACACACCAGATTCGCAGGAAGTAACAGTAACTAGCGTAGTTTGTTGCCTCGATTCTGTGGTGGAAATAGGACACCATGTCGTGTAT
TCTGTGGTAACGCCGTTAATAGTAGCAGTGCTTGTAGATACAATGGCTGGAGAAGCAGTCTTAGACCATACGTCAGATTCACAAGAAGAAATTGTAACTACTGTGGTTTGT
TTTGTTGTCCCCTTGGTTTGCTTTGTTGTCTCCGTAGTTTGCTTTGTTATCTCTGTGGTAGAAATAGGGCACCATGTGGTATACTCTGTTGTGACACCGCTAACAGTAACG
GTGGCCGTGGAAACAATGCCAGAGGAGATGGATTCAGTGCACACATGA

>YAL064W-B Saccharomyces cerevisiae  - R63 with JAY291 : Chr(I) from 12047 to 12427
MAGEAVSEHTPDSQEVTVTSVVCCLDSVVEIGHHVVYSVVTPLIVAVLVDTMAGEAVLEHTSDSQEEIVTTVVCPVVPLVCFVVSVVCFVISVVEIGHHVVYSVVTPLTVT
VAVETIAEEMDSVHT*
```

6.3.7 History

Users can manipulate feature information stored in the Genome Suite. All of these changes are tracked and stored in a set of change logs including the item changed, its new and previous value, and the user who made the change. The history format displays a listing of every modification to the feature.



Item	New Value	Previous Value	Editor	Date	Event
function	[+] Some new function	[+]	admin	03/19/14 03:39 PM	create
gene	[+]	[+] LOC100825100	admin	03/19/14 03:39 PM	destroy
db_xref	[+] modified GeneID	[+] GeneID:100825100	admin	03/19/14 03:39 PM	update
pfam	[+]	[+] PF00847.10.ls	admin	03/19/14 03:39 PM	destroy

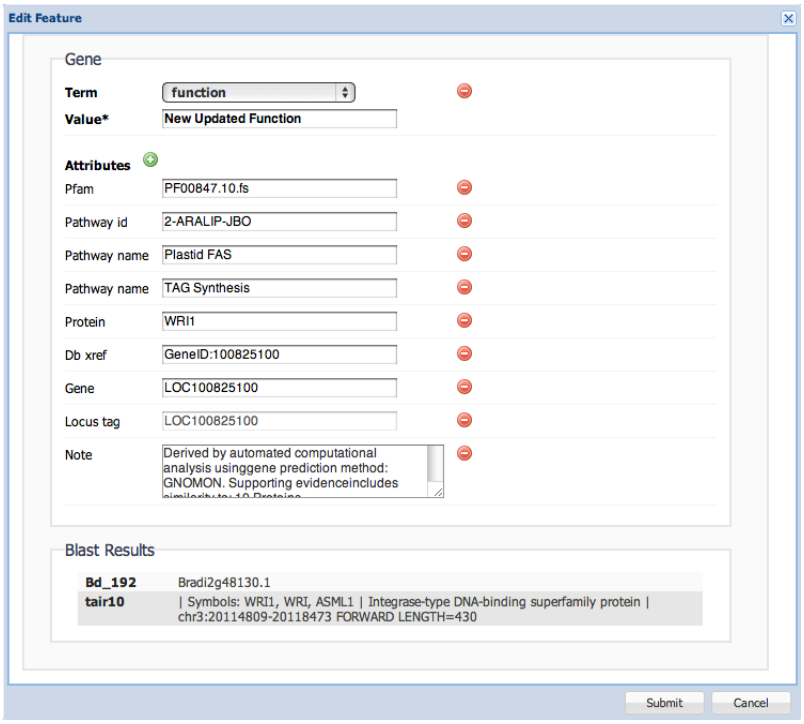
6.3.8 Edit

The Edit view allows users to add, change and remove functional annotations as well as manipulate properties of the feature. A portion of this view is also available from the Expression Tool. More details on this view are provided in section 5.4 Editing Feature Properties.

6.4 Editing Feature Properties

6.4.1 Editing Functional Annotations

There are quick links throughout the site to open a popup feature edit form. This form allows users to manage all of the feature attributes. Each attribute has a Term Name and a value. Users can modify values of existing attributes by changing the text field values. The green plus icon is used to add new attributes to the list. A new attribute needs a Term name selected from a dropdown along with the new value. New and existing attributes can be removed with the red minus icon .



6.4.1.1 ChangeLogs

If any modifications have been made to the feature a link to [+] Show History ... will be displayed. Selecting this link will expand the form with a list of Change Logs documenting the previous updates. This list is the same table described in section 5.3.6.

6.4.2 Editing Feature Annotations

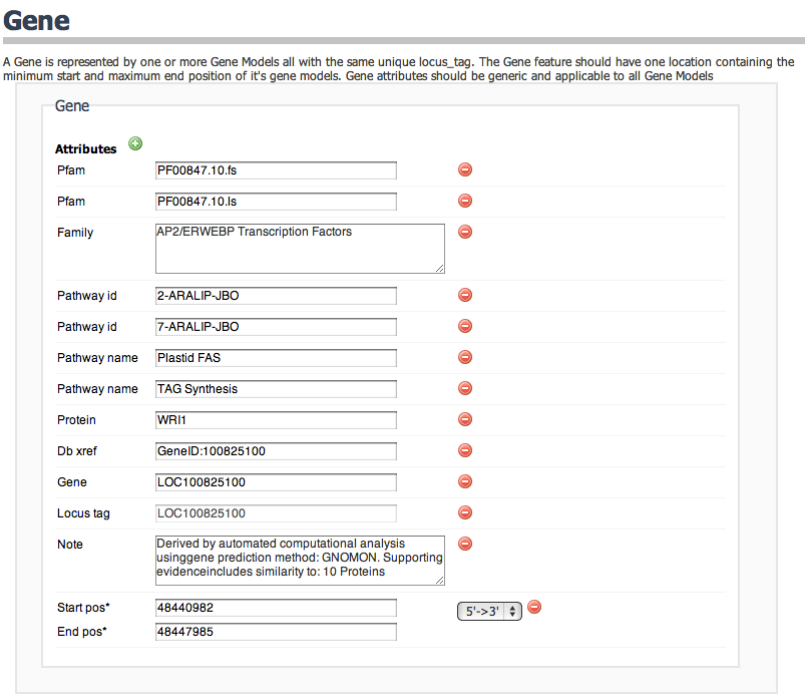
The feature annotation form is available from the Edit format tab. On this form attributes are managed as described in the previous section. The full edit form also adds the option to manage the start and stop locations as well as the strand orientation for the feature.

6.4.3 Editing Gene Annotations

Gene features have special edit forms designed around the concept of gene models. Each gene has one or more gene models that represent alternate isoforms. A gene model is made from a Gene an optional CDS and optional mRNA.

6.4.3.1 Gene feature

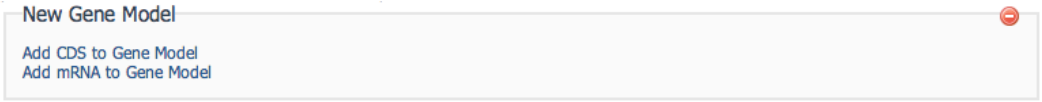
The gene feature is managed similar to feature annotations. Functional annotations can be removed, updated or created. The location can be updated and the strand can be changed. The Locus Tag for a Gene is frozen after it's been created.



6.4.3.2 Building Gene Models

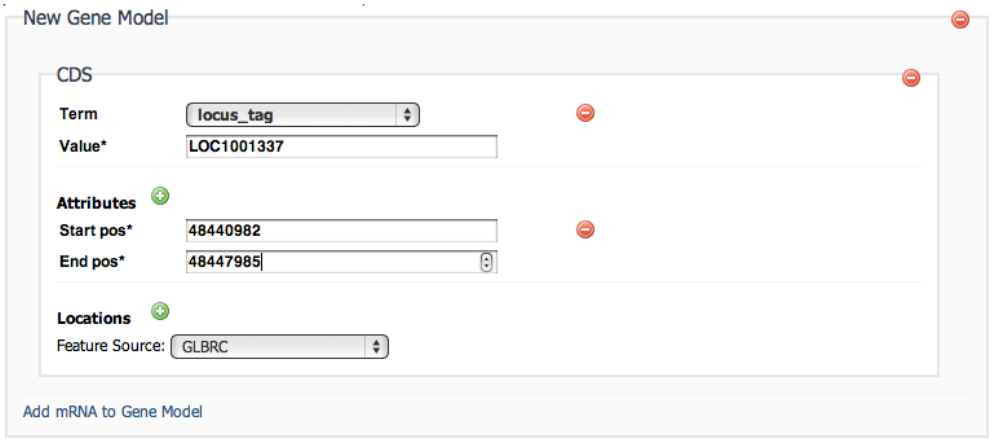
New gene models can be added to existing genes when a new isoform is identified or built while creating a new gene. Gene Models are created with the green plus next to the "Gene Model(s)" text on the gene form.

Clicking this icon will add a new empty gene model to the page with links to add CDS or mRNA.



Clicking the links will add an empty feature ready to be defined with new attributes and locations. This nested form is identical to the feature edit form described in section 5.4.2.

**It is recommended that you assign the Gene's Locus Tag to the new CDS and mRNA to enable administrator tasks to properly find them although this step is not required.*



Submitting the form will save the new gene model and features along with all of the attributes and locations. The new Gene Model will be rendered in the Genomic Context and the information will be displayed in feature listings.

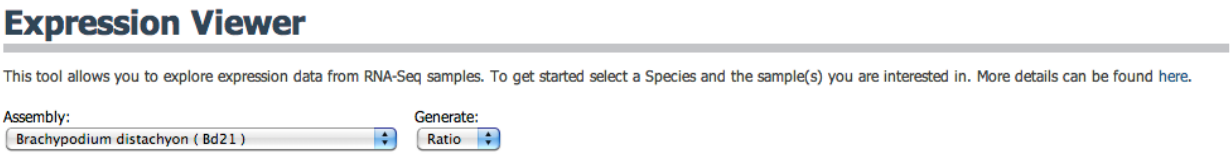
6.4.3.3 Editing Gene Models

Existing gene models can be updated from the Edit page for a gene. Similar to feature annotation described in section 5.4.2 attributes and locations can be added, removed or changed. CDS and mRNA can also be redefined or removed. The changes are not saved until the form has been submitted.

7 Quantitative Expression Tool

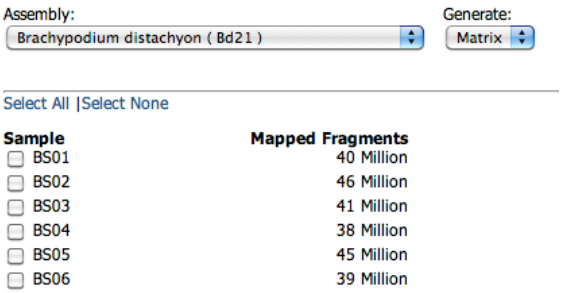
To open the expression tool, first navigate to the tools tab 

. Then select Expression Viewer from the list of tools. The expression tool allows you to build tabular listings of genomic features including expression data and functional annotations. The expression tool begins with a form used to select samples for inspection. The Assembly dropdown is used to choose a Genome or Transcriptome to view. The second dropdown is used to choose between generating a matrix of sample data or a ratio of two sample sets.



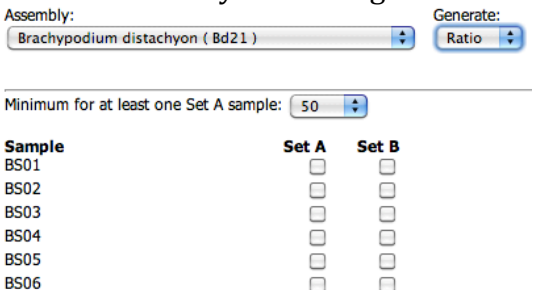
7.1 Building a Matrix

To build a matrix of data make sure 'Matrix' is selected in the dropdown menu. A single checkbox column will be displayed. Select the samples you would like to include by checking the box next to each sample's name. After making your selection click submit to view the expression matrix.



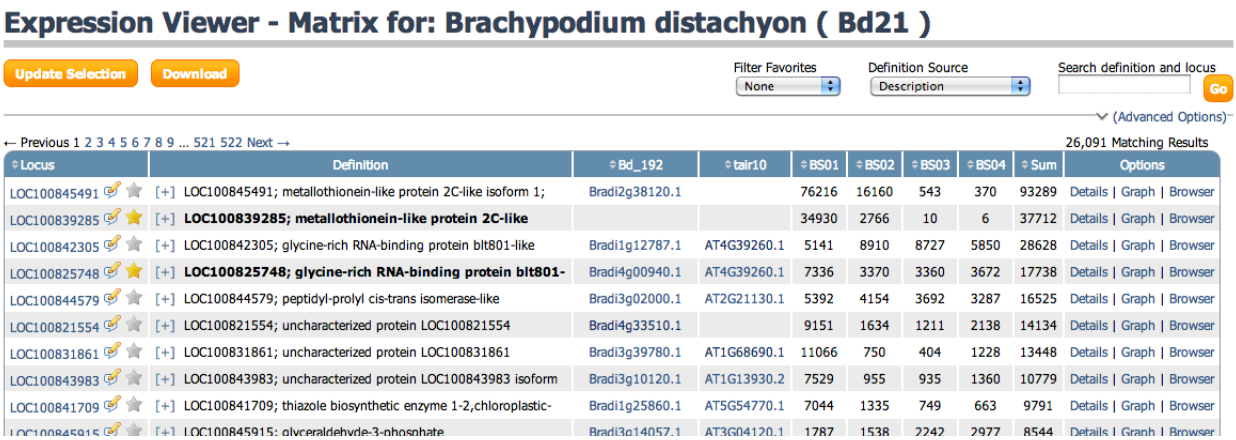
7.2 Building a ratio

To build a ratio of samples first select 'Ratio' from the dropdown menu. Two checkbox columns will be displayed titled Set A and Set B. Samples chosen from the Set A column will be averaged together and divided by the average of Set B.



7.3 Expression interface

Each row in the Matrix or Ratio table represents a genomic feature from the selected assembly. Information about this feature and its expression profile is presented in the columns described below.



7.3.1 Locus

The locus column displays the unique locus id given for the feature. A locus is only unique to an assembly and may appear in alternate versions. A locus can also be repeated for different feature types. For example, a Gene and mRNA may share the same locus. Two additional link icons are contained within the locus column. They are detailed below.

7.3.1.1 Quick Edit

The edit icon  can be clicked to open an interactive form. This form will allow updates to the functional annotation for the feature. See section 5.4.1 for more information.

7.3.1.2 Favorites

The favorite icon  /  can be used to add this feature to the favorites list describe in section 2.2. These favorites can be used to filter the expression listing. The filter option is a basic search option described in section 7.3.7.

7.3.2 Definition

The definition column presents the functional annotation data chosen for display. The Definition Source dropdown changes the information presented here. Definitions can be displayed from any functional annotation or best blast hit. The advanced search pane displays a multiple select box. This multi select allows multiple sources to be displayed, each concatenated with a '|' symbol. The search field only applies to definitions selected for display.

7.3.3 Blast Accession(s)

For each blast report that is loaded onto an assembly, a column of ID's will be displayed representing the best blast result for the feature. This ID is linked to a detailed view of the feature's blast reports including additional hits and alignment information. BLAST reports are described in more detail in Ch. 8.

7.3.4 Sample(s), Sum / Ratio

For a Matrix view, each sample chosen will be presented as a column. This column will display numerical data representing the expression value for the given feature. The count columns are followed by a total sum of all selected samples.

	BS01	BS02	BS03	BS04	Sum
	76216	16160	543	370	93289
	34930	2766	10	6	37712
1	5141	8910	8727	5850	28628
1	7336	3370	3360	3672	17738

For a Ratio view, the two sets of samples are combined and a column of averages for Set A and Set B are displayed. The ratio column is displayed after the two average columns. This column shows the ratio of Set A averages divided by Set B averages.

Set A	Set B	A / B
145	0	2906.90
1835	1	2659.24
18848	8	2290.17
1601	1	2106.76

7.3.5 Details Option

The Details link will open the detailed format pages for this feature and load the Standard format

7.3.6 Graph Option

The Graph link will open the detailed format pages for this feature and load the Expression format

7.3.7 Browser Option

The Browser link will open the Genomic Context for the sequence this feature is attached to. More information about the Genomic Context is given in Ch. 8 Genomic Context Tool.

7.3.8 Basic Search Options

The basic search options are located on the top right of the window. Making changes to these items and clicking Go will update the listing with the search filters applied.

7.3.8.1 Keyword Search

The search box allows for text based keyword searching. It will search against text in the selected definition source and locus tag. Keywords are used in a token-based search matching each search keyword to the beginning of white space separated words in the text. Keyword matches can be separated by non-matching text. All of the keywords must be found for an item to be returned but the keywords can span any of the selected definitions.

7.3.8.2 Definition Source

The Definition column can display data from numerous definition sources. This dropdown allows users to change the source of information displayed. Definitions can come from the genomic annotations or Annotation Tags, results from BLAST alignments, or user provided annotation files.

7.3.8.3 Filter Favorites

The favorites filter has three options:

None - makes no change to the listing.
My Favorites – Only shows items in the current user account favorite list.
All Favorites – Shows items that are in any user account favorite list.

7.3.9 Advanced Search Option

The advanced search options are available by clicking the (Advanced Options) link under the basic search form. This link will open several new filter options as detailed below.

7.3.9.1 Definition Builder

The definition builder expands on the Definition Source selection from the basic search form. This builder is a multiple select listing allowing several different sources to be displayed at the same time. Use ctrl/cmd click or shift to highlight multiple entries. The table will then join each entry with a '|' symbol in the definition column.

7.3.9.2 Match (BLAST) ID

For each BLAST report loaded on the assembly a 'Match {blastname} ID' field is presented. This field allows querying based on the best hit ID from the Blast alignment. It is exact and case-sensitive.

7.3.9.3 Expression Type

There are three types of expression that can be stored.

Normalized - The default and required type. This type can represent any normalization method from the processing steps.

Total Counts - The number of raw reads mapped to a feature regardless of alternate mapping locations for the read. These are non-unique counts.

Unique Counts – The number of raw reads that mapped to a feature uniquely without aligning to any other feature.

7.3.9.4 Show Empty Results

Some of the features in the listing might not have any information from the chosen definition source(s). This option controls what to do with empty definitions.

Yes – Show all results regardless of definition.
No – Show only results with a value for the chosen definition(s)
Only Empty – Shown only results that do not have a value definition value.

7.3.9.5 Show Best Blast E-value

This option allows displaying the e-value from the best blast hit along with the ID. This can be useful for quickly comparing multiple blast results. Yes to enable, No to disable.

7.3.10 Downloading Expression

On the top left is a button to download 

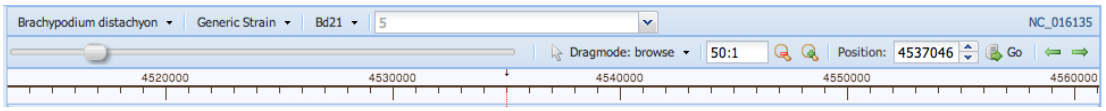
the expression data. This button will send a text file with all of the currently displayed results using the search filters. These results will be in CSV format and can be opened directly in Microsoft Excel.

8 Genomic Context Tool

The Genomic Context is accessed from the View links on tables listing sequence and annotation. It is also accessible on the Feature details page from the Open Sequence Viewer link on the top right. The Genomic Context Tool displays experimental data alongside genomic sequence and annotations. The view is split into three regions. Data management panels are displayed on the left. A navigation toolbar is on the top right, and the main track display is on the bottom right.

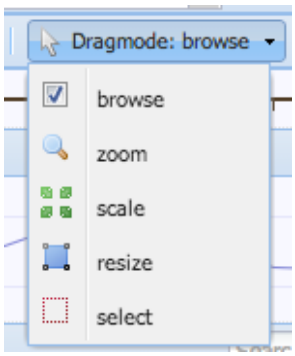
8.1 Navigation

The navigation toolbar is used to manage the view settings for the genomic sequence. When working with the Genomic Context Tool a single sequence is displayed. The viewable region is defined with a zoom level and position. Adjusting zoom will change the size of the region and adjusting position change the region's sequence location.



8.1.1 Dragmode

Interacting with the tool is based on different mouse modes. The Dragmode selector is used to change the current mode. Each mode has a different effect on the tracks.



8.1.1.1 Browse

The default mode allows changing position by clicking and dragging the view. Tracks will shift left or right based on how far they were dragged.

8.1.1.2 Zoom

Show a highlighted area on the tracks. This area will become the new viewable region when dragging ends.

8.1.1.3 Scale

Changes the scale of rendering in a track increasing letter size, height and y-scale.

8.1.1.4 Resize

Changes the height of each track.

8.1.1.5 Select

Highlights an area of track. The Sequence Track will return a FASTA window of the selected region. The Variant track will return a list of alternate tracks matching the selected location.

8.1.2 Sequence Selector

The sequence selector on the top toolbar allows switching between all of the available sequence. It uses a series of dropdowns to present users with a hierarchical selection. The leftmost dropdown contains species, followed by a dropdown for strain and then assembly version. The final field is a dropdown search field to choose a sequence for display. Typing into the field will search the current assembly and display results.

8.1.3 Managing Position

The current position is shown in a text field on the navigation bar. Typing a new position into this field and clicking Go 

will change the current position. A slider and dynamic button give an overall representation of the current region and location on the sequence. Moving this slider will also change the position. Finally, left and right arrow icons are quick links to shift the position one full screen left or right.

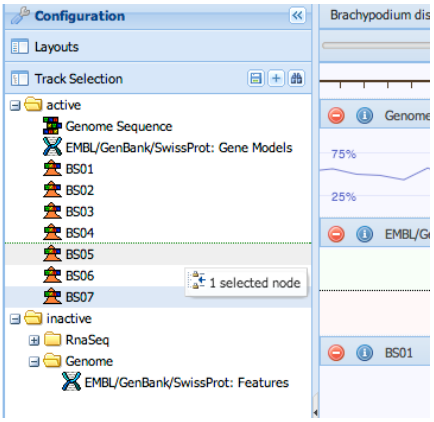
8.1.4 Managing Zoom

The current zoom is displayed as a ratio of bases to pixels. Typing a new ratio in this field and clicking Go  will change the size of the region displayed. Two icons for zoom out and zoom in also allow incremental changes to the zoom level.

8.2 Data Panels

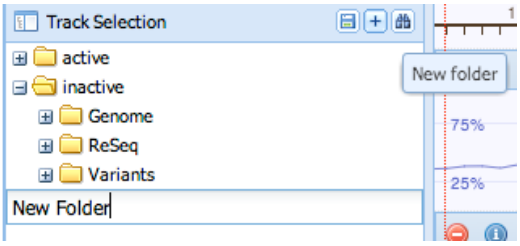
8.2.1 Track Selector

The track selector is used to manage the state and order of tracks. All of the genomic features and sample data are represented by individual items in a tree structure similar to file management views. There are two top-level folders named active and inactive. Items can be dragged and dropped into different folders. Placing an item in the active folder will open it on the main track display. The order of items in this folder also determines the order for display.



8.2.1.1 Folders

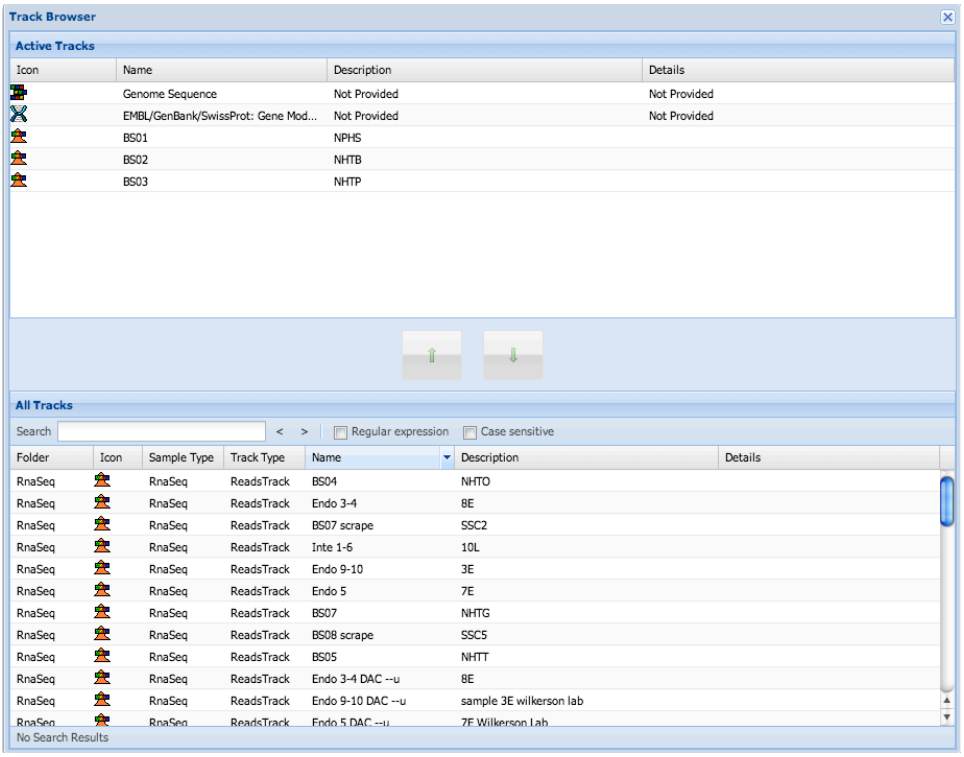
By default each track is organized into folders by type. The folder structure can be changed by each user and is saved to the user account. Dragging a track into a new folder automatically saves the new location. New folders can be created and named with the plus icon.



8.2.1.2 Track Browser

A more detailed track management view is available from the browse icon  . Clicking this will open the Track Browser with two tables. The top table lists

active tracks in order of display. The bottom table lists a sortable and searchable table of all the tracks including descriptions and sample traits. Tracks are moved between tables with drag and drop, double-click or by using the up and down arrow buttons.



8.2.2 Track Layouts

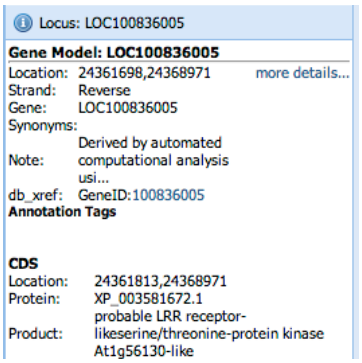
The current set of active tracks along with track configurations can be saved to a named layout. Click the save icon  to create a new layout. You will be required to enter a new unique layout name and all of the configuration will be saved. All saved layouts are displayed by name in the Layout panel. Clicking on a name will refresh the view using the layout to load and configure tracks.

8.2.3 Information Panels

Certain items rendered in the tracks are interactive. Clicking on one of them will load more details about the item. The information panel is used to render these details.

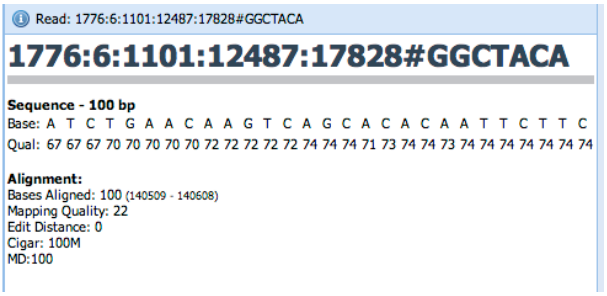
8.2.3.1 Gene Models / Features

The information panel for genomic features will show location, strand, locus and all of the attribute values. This window also includes a quick link in the upper right to view the standard format page for the feature. This page is detailed in section 5.3.



8.2.3.2 Reads

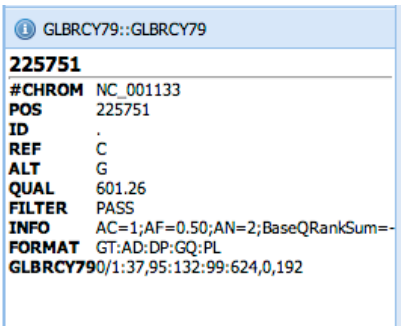
The information panel for reads will show the reads ID, sequence and quality strings. Details of the alignment are also displayed including the number of bases, position matched, overall mapping quality, edit distance, the SAM cigar string, and if present, the MD tag. More details on these parameters are available from the Samtools specification: http://samtools.github.io/hts-specs/SAMv1.pdf



8.2.3.3 Variant

The information panel for a variant shows the position, sequence accession and information about the sequence changes. This information includes reference, alternate, quality, an info line, and the sample genotype. The data is displayed directly from the VCF file. More information on this format is available from the 1000 genomes project:

http://www.1000genomes.org/wiki/Analysis/Variant%20Call%20Format/vcf-variant-call-format-version-41

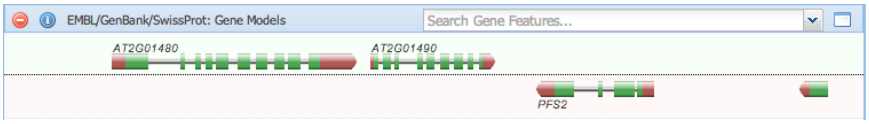


8.3 Display

The large central window is used to display data from active tracks. Each track has a toolbar along the top containing the track name and controls to customize the display. This toolbar always has a red minus icon on the left to close the track followed and a blue information icon to open more details about the track. Right clicking on a track will open a configuration menu with controls to manage the track. Each track type is described in more detail below.

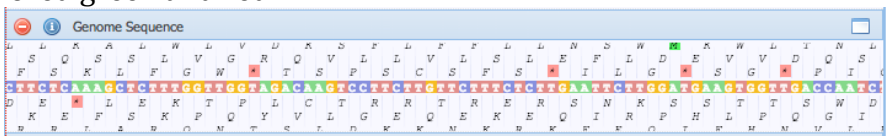
8.3.1 Gene Models

The gene models track displays a combination of Gene, CDS and mRNA features. These features are combined into gene models during the load process. A gene model is rendered with CDS coding sequence in green, UTR in red (exons overhanging CDS from mRNA), and introns as a thin gray bar (gene annotation in between CDS/mRNA). Each of these models is clickable to get more details. The toolbar contains a search box to quickly jump to a new gene of interest. Typing text into this field will search all of the gene models for the current assembly and display the results.

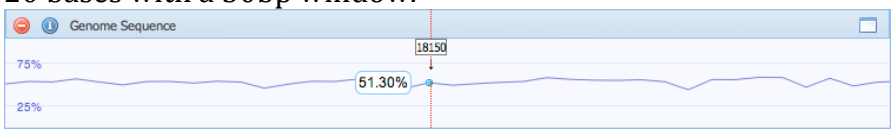


8.3.2 Sequence

The sequence track displays different data depending on the current zoom. With a small region each individual sequence base is rendered and colored along with a six-frame protein translation above and below the sequence. The start and stop codons are colored green and red.

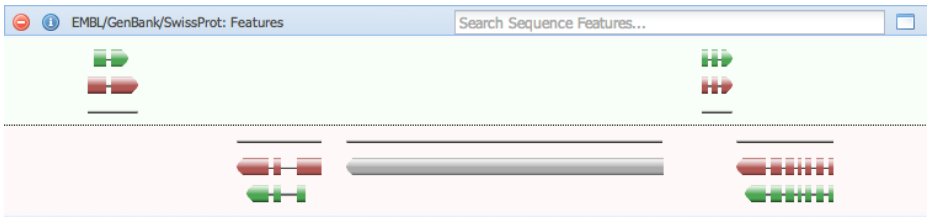


At further zoom levels the sequence is replaced by a GC content line graph. Hovering over the line graph will popup the percent in that region. GC content is computed every 20 bases with a 50bp window.



8.3.3 Features

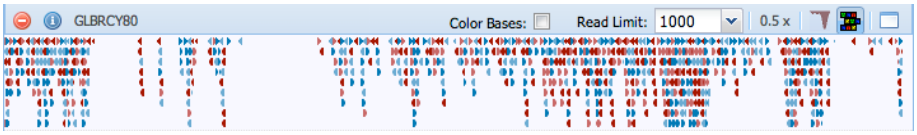
All genomic features are rendered in the Features track. This track shows CDS, Gene and mRNA from the gene models along with RepeatRegion, matPeptide and other miscellaneous features. The Feature track also contains a search field that will match any feature in the current assembly.



8.3.4 Reads

The Reads can render two different displays depending on the toggle button setting. The default setting is to display a histogram of read depth for the region. The y-scale of this histogram is displayed in a drop down field. By default the scale updates dynamically to match the maximum value of all samples allowing visual comparison. Typing into this field will change the scale for the track.

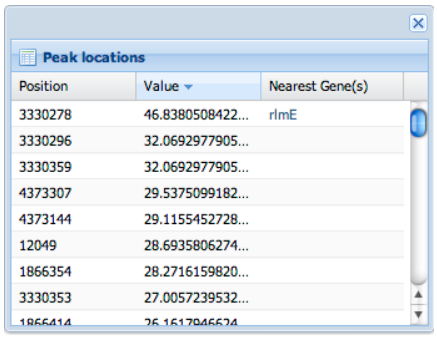
Clicking the read toggle button will switch to rendering the actual aligned reads. In this view reads are rendered as boxes at the alignment position. The number of reads displayed is controlled by the Read Limit dropdown. This value is used to take a random sampling from the current region. At closer zoom levels the actual bases along with mismatches and gaps can be seen for each read. The Reads track right click menu contains controls to change the color of the track. The Color value is used for the histogram and forward reads. The Reverse Color value is used for reverse reads.



8.3.5 Density

The density track renders a histogram of data from the sample. It also uses a dynamic y-scale based on the current maximum sample. The right click menu has an item to change the track color.

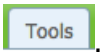
If peak data is assigned to the sample controls are added to the tool bar. Forward and Reverse icons allow navigating to the next and previous peaks positions, and the table icon will open a peak browser window with the position, y-value and nearest Gene for every peak.



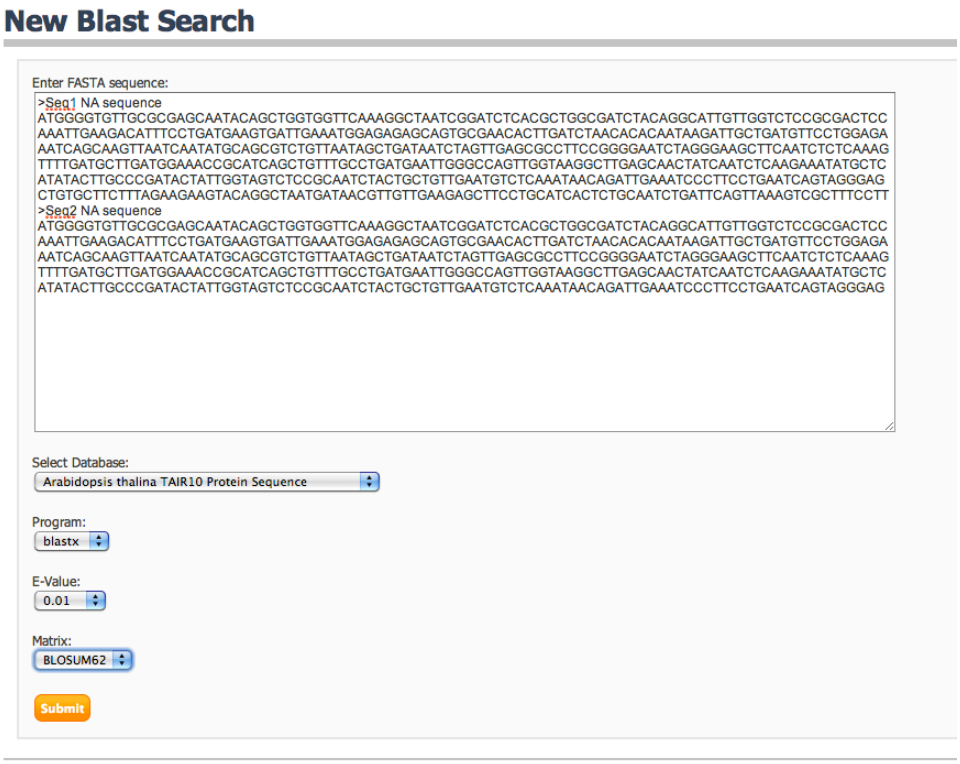
8.3.6 Variants

The variant track displays SNP's (single nucleotide polymorphism), Insertions and Deletions with blue, green and red boxes respectively. Matching sequence will render as a pale blue box. These boxes are interactive. Clicking one will open the information panel with details on the variant. The track is based on a diploid organism and renders two rows of variants, one above and one below to represent each allele. Heterozygous variants will have a one colored box above and one pale blue box below. Homozygous variants will have a colored box above and below.

9 Local Blast Tool

To open the Blast tool, first navigate to the tools tab 

. Then select Blast from the list of tools. The Blast tool can be used to align sequence against databases stored on the system. These databases are managed by system administrators separately from assembly data and protected by group membership. You must be a member of the same group as the Blast Database in order to run blast queries against it. The local blast tool uses the legacy blastall program. BLAST+ is not supported by the system. Blast Sequence can be supplied in multiline Fasta or raw sequence format. The sequence can be either nucleotide or protein depending on the chosen BLAST program.



9.1 Options

9.1.1 Database

You must choose a database to blast against. The database could be protein or nucleic. You must choose the correct program and database based on your input sequence.

9.1.2 Program

The BLAST program used to align sequence will vary based on the type of sequence and database used.

9.1.2.1 blastn

Search a **nucleotide** database using a **nucleotide** query

9.1.2.2 tblastn

Search a **translated nucleotide** database using a **protein** query

9.1.2.3 blastx

Search a **protein** database using a **translated nucleotide** query

9.1.2.4 tblastx

Search a **translated nucleotide** database using a **translated nucleotide** query

9.1.2.5 blastp

Search a **protein** database using a **protein** query

9.1.3 E-value

Sets the expectation value cutoff used to filter sequence alignments. A lower e-value is more restrictive than a higher e-value.

9.1.4 Matrix

Select the substitution matrix used to score base pair alignments. The following table is provided by NCBI for choosing a matrix based on protein sequence length.

Query Length	Substitution Matrix	Gap Costs
<35	PAM-30	(9,1)
35-50	PAM-70	(10,1)
50-85	BLOSUM-80	(10,1)
85	BLOSUM-62	(10,1)

9.2 Viewing Blast Reports

Reports generated by the BLAST program are stored in the system for future browsing. Administrators can also load blast reports into the site and attach them to features such as Genes. When using the BLAST tool a report is immediately displayed after submitting sequence. Each account records the last 15 reports generated. These results are accessible from the Recent Blasts table described in section 2.3. When viewing genomic features the BLAST format page described in section 5.3 will display reports. Blast report pages include a tabular listing of hits for each query and a visual display of HSP alignments.

9.2.1 Hit Listing

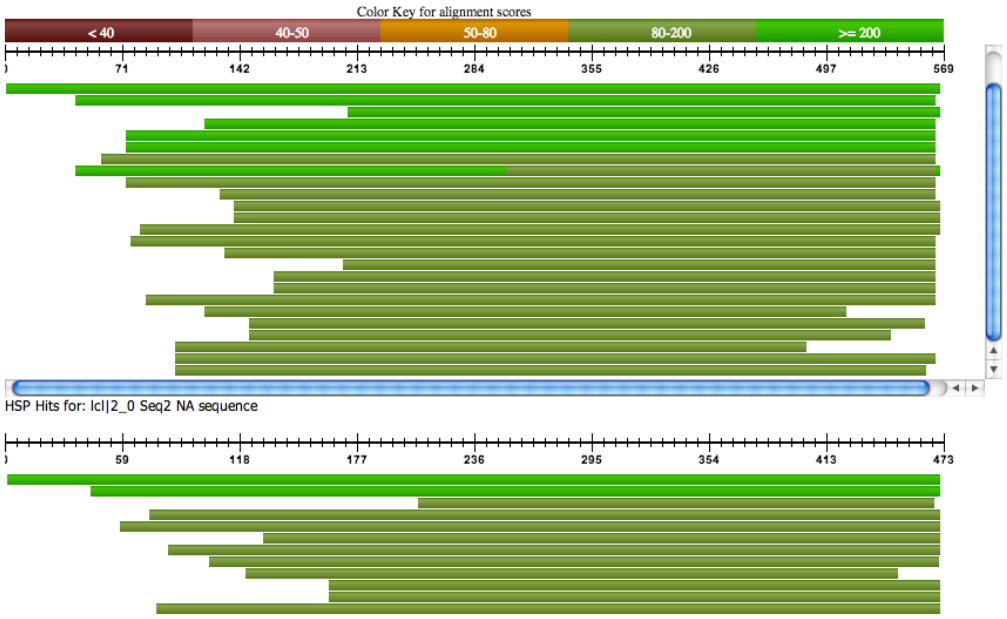
A tabular listing of each database sequence matched by the supplied query sequence is displayed. This table includes information about the original query sequence along with details for each database hit. The query row includes a query ID, feature ID if the hit is connected with a stored feature, query description, query length, number of database hits and a link to show or hide the table of hits. Each hit is listed below the query row in a nested table with columns for Accession, Definition, Bit score and E-value. There is also a link to view details about the hit Alignment

Blast results for 2 queries

ID	Feature ID	Description	Query Length	Hits	Options
lcl|1_0		Seq1 NA sequence	569	25	Hide
AT5G07910.1	[+]	Symbols: | Leucine-rich repeat (LRR) family protein | chr5:2521937-2523769 REVERSE LENGTH=262		941	1.41793e-130 [Alignment]
AT2G30105.1	[+]	Symbols: | CONTAINS InterPro DOMAIN/s: Leucine-rich repeat, typical subtype (InterPro:IPR003591), Leucine-rich repeat (InterPro:IPR001611),		320	1.45716e-35 [Alignment]
AT1G12970.1	[+]	Symbols: PIRL3 | plant intracellular ras group-related LRR 3 | chr1:4423727-4425632 FORWARD LENGTH=464		227	6.46248e-22 [Alignment]
AT2G17440.1	[+]	Symbols: PIRL5 | plant intracellular ras group-related LRR 5 | chr2:7571331-7573406 FORWARD LENGTH=526		224	1.95051e-21 [Alignment]
AT3G15410.1	[+]	Symbols: | Leucine-rich repeat (LRR) family protein | chr3:5203380-5207279 FORWARD LENGTH=584		216	2.45926e-20 [Alignment]
AT3G15410.2	[+]	Symbols: | Leucine-rich repeat (LRR) family protein | chr3:5203380-5207279 FORWARD LENGTH=590		216	2.46878e-20 [Alignment]
AT3G26500.1	[+]	Symbols: PIRL2 | plant intracellular ras group-related LRR 2 | chr3:9708195-9709944 REVERSE LENGTH=471		552	2.84e-20 [Alignment]
AT4G35470.1	[+]	Symbols: PIRL4 | plant intracellular ras group-related LRR 4 | chr4:16846531-16848448 FORWARD LENGTH=549		583	1.11713e-19 [Alignment]
AT3G11330.1	[+]	Symbols: PIRL9 | plant intracellular ras group-related LRR 9 | chr3:3552330-3554695 REVERSE LENGTH=499		566	2.72088e-17 [Alignment]
AT5G05850.1	[+]	Symbols: PIRL1 | plant intracellular ras group-related LRR 1 | chr5:1762691-1764609 REVERSE LENGTH=506		357	5.07059e-17 [Alignment]
AT1G69550.1	[+]	Symbols: | disease resistance protein (TIR-NBS-LRR class) | chr1:26148836-26153374 REVERSE LENGTH=1400		1054	1.54662e-16 [Alignment]
AT1G71400.1	[+]	Symbols: AtRLP12, RLP12 | receptor like protein 12 | chr1:26909905-26912448 FORWARD LENGTH=847		180	1.68136e-15 [Alignment]
AT4G29880.1	[+]	Symbols: PIRL7 | plant intracellular ras group-related LRR 7 | chr4:14607078-14608379 REVERSE LENGTH=373		174	7.42542e-15 [Alignment]
AT5G62230.2	[+]	Symbols: ERL1 | ERECTA-like 1 | chr5:24996433-25002130 FORWARD LENGTH=918		165	1.67039e-13 [Alignment]
AT3G24240.1	[+]	Symbols: | Leucine-rich repeat receptor-like protein kinase family protein | chr3:8780551-8784150 FORWARD LENGTH=1141		629	4.29482e-13 [Alignment]
AT5G17680.1	[+]	Symbols: | disease resistance protein (TIR-NBS-LRR class), putative | chr5:5822999-5827153 FORWARD LENGTH=1294		162	4.35399e-13 [Alignment]
AT1G27170.2	[+]	Symbols: | transmembrane receptors;ATP binding | chr1:9433577-9439219 FORWARD LENGTH=1384		321	5.94585e-13 [Alignment]
AT1G27170.1	[+]	Symbols: | transmembrane receptors;ATP binding | chr1:9434718-9439219 FORWARD LENGTH=1384		321	5.94585e-13 [Alignment]
AT5G63930.1	[+]	Symbols: | Leucine-rich repeat protein kinase family protein | chr5:25583006-25586392 FORWARD LENGTH=1102		304	1.06786e-12 [Alignment]
AT1G68780.1	[+]	Symbols: | RNI-like superfamily protein | chr1:25831881-25833335 REVERSE LENGTH=432		158	1.16854e-12 [Alignment]
AT2G24130.1	[+]	Symbols: | Leucine-rich receptor-like protein kinase family protein | chr2:10258148-10261220 FORWARD LENGTH=980		158	1.42673e-12 [Alignment]
AT5G22320.1	[+]	Symbols: | Leucine-rich repeat (LRR) family protein | chr5:7388175-7390426 REVERSE LENGTH=452		156	2.19692e-12 [Alignment]
AT5G58150.1	[+]	Symbols: | Leucine-rich repeat protein kinase family protein | chr5:23530216-23532573 REVERSE LENGTH=785		156	2.53509e-12 [Alignment]
AT1G73066.1	[+]	Symbols: | Leucine-rich repeat family protein | chr1:27481785-27483581 FORWARD LENGTH=598		155	3.24765e-12 [Alignment]
AT5G06940.1	[+]	Symbols: | Leucine-rich repeat receptor-like protein kinase family protein | chr5:2148078-2150771 REVERSE LENGTH=872		154	4.74497e-12 [Alignment]
lcl|2_0		Seq2 NA sequence	473	12	Hide
AT5G07910.1	[+]	Symbols: | Leucine-rich repeat (LRR) family protein | chr5:2521937-2523769 REVERSE LENGTH=262		785	2.18002e-107 [Alignment]
AT2G30105.1	[+]	Symbols: | CONTAINS InterPro DOMAIN/s: Leucine-rich repeat, typical subtype (InterPro:IPR003591), Leucine-rich repeat (InterPro:IPR001611),		260	2.53432e-27 [Alignment]
AT1G12970.1	[+]	Symbols: PIRL3 | plant intracellular ras group-related LRR 3 | chr1:4423727-4425632 FORWARD LENGTH=464		182	3.16348e-16 [Alignment]
AT3G11330.1	[+]	Symbols: PIRL9 | plant intracellular ras group-related LRR 9 | chr3:3552330-3554695 REVERSE LENGTH=499		343	8.25045e-16 [Alignment]
AT3G26500.1	[+]	Symbols: PIRL2 | plant intracellular ras group-related LRR 2 | chr3:9708195-9709944 REVERSE LENGTH=471		492	1.49872e-15 [Alignment]

9.2.2 HSP visualization

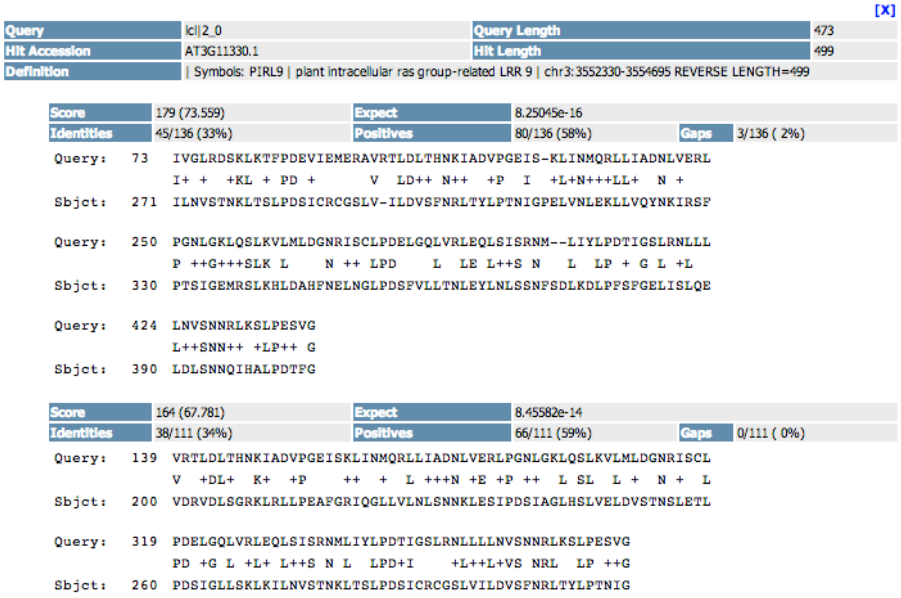
A visualization of HSP alignments is rendered below the hit listing. One chart is displayed per query in the report. This chart is color-coded based on bit score and shows the HSP locations aligned to the query. HSP's are rendered in the same order as the hit listing.



9.2.3 Alignment details

Selecting the [Alignment] link will display a window containing detailed information about the sequence alignment. This window displays ID and length for the query and hit along with statistics for each HSP. Alignment details include bit score, e-

value, gaps, positives and identities. The aligned sequence is also displayed for visual inspection of alignment quality.



[X]

Query	lcl|2_0	Query Length	473
Hit Accession	AT3G11330.1	Hit Length	499
Definition	| Symbols: PIRL9 | plant intracellular ras group-related LRR 9 | chr3:3552330-3554695 REVERSE LENGTH=499		

Score	179 (73.559)	Expect	8.25045e-16		
Identities	45/136 (33%)	Positives	80/136 (58%)	Gaps	3/136 (2%)

```
Query:   73  IVGLRDSKLKTFPDEVIEMERAVRTLDLTHNKIADVPGEIS-KLINMQRLLIADNLVERL
             I+ +  +KL + PD +       V  LD++ N++  +P  I  +L+N+++LL+  N +
Sbjct:  271  ILNVSTNKLTSLPDSICRCGSLV-ILDVSFNRLTYLPTNIGPELVNLEKLLVQYNKIRSF

Query:  250  PGNLGKLQSLKVLMLDGNRISCLPDELGQLVRLEQLSISRNM--LIYLPDTIGSLRNLLL
             P ++G+++SLK L    N ++ LPD    L  LE L++S N   L  LP + G L +L
Sbjct:  330  PTSIGEMRSLKHLDAHFNELNGLPDSFVLLTNLEYLNLSSNFSDLKDLPFSFGELISLQE

Query:  424  LNVSNNRLKSLPESVG
             L++SNN++ +LP++ G
Sbjct:  390  LDLSNNQIHALPDTFG
```

Score	164 (67.781)	Expect	8.45582e-14		
Identities	38/111 (34%)	Positives	66/111 (59%)	Gaps	0/111 (0%)

```
Query:  139  VRTLDLTHNKIADVPGEISKLINMQRLLIADNLVERLPGNLGKLQSLKVLMLDGNRISCL
             V  +DL+  K+  +P    ++  +  L +++N +E +P ++  L SL  L +  N +  L
Sbjct:  200  VDRVDLSGRKLRLLPEAFGRIQGLLVLNLSNNKLESIPDSIAGLHSLVELDVSTNSLETL

Query:  319  PDELGQLVRLEQLSISRNMLIYLPDTIGSLRNLLLLNVSNNRLKSLPESVG
             PD +G L +L+ L++S N L  LPD+I    +L++L+VS NRL  LP ++G
Sbjct:  260  PDSIGLLSKLKILNVSTNKLTSLPDSICRCGSLVILDVSFNRLTYLPTNIG
```