GenSAS

Genome Sequence Annotation Server

Computational annotation and curation of genome sequences

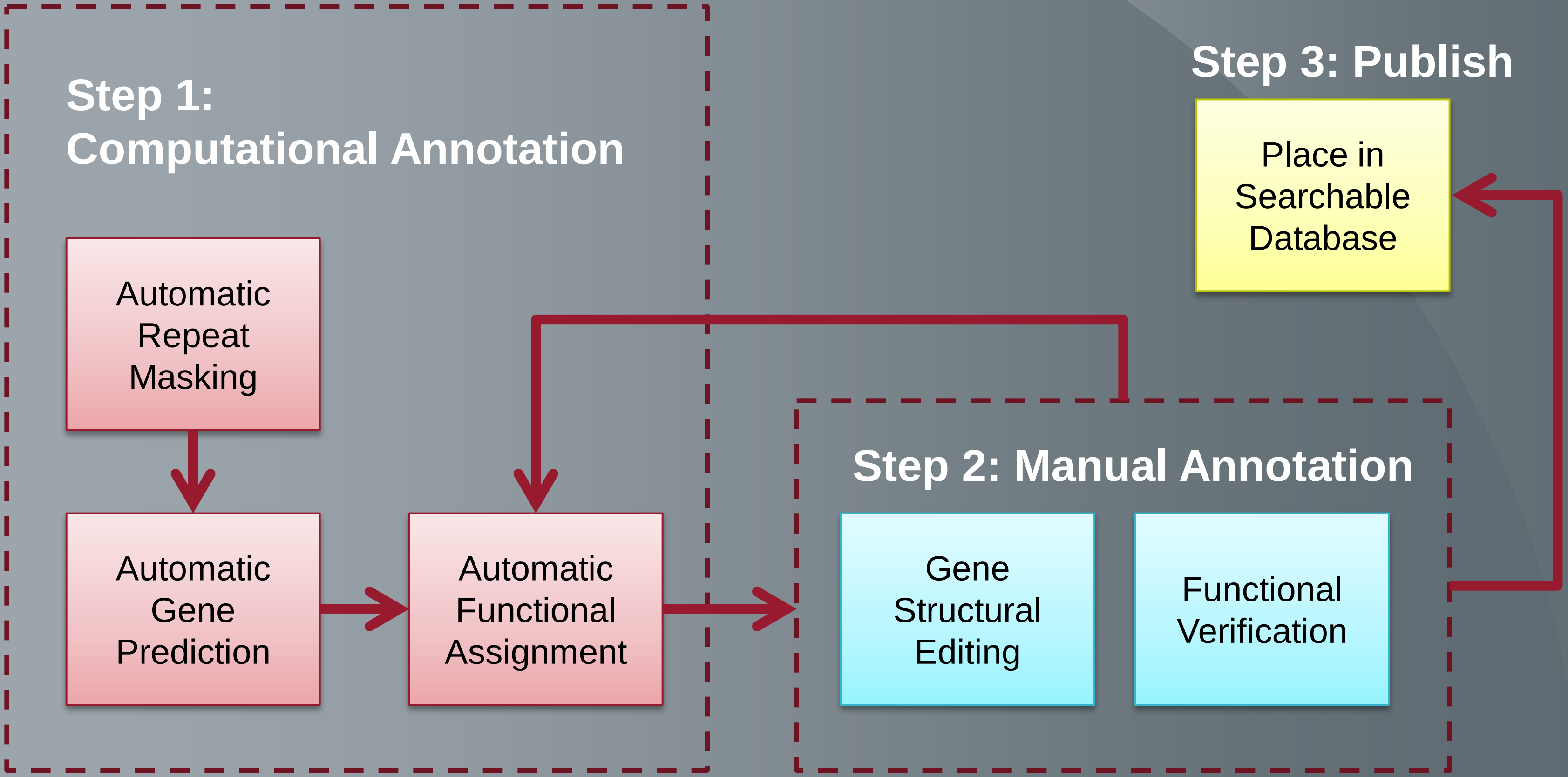
Stephen Ficklin*, Taein Lee, Jodi Humann, Chun-Huai Cheng, Jill Wegrzyn, David Neale, Dorrie Main

** Presenting author*

Pine Genome Workshop
PAG XXIII, San Diego CA. Jan 10, 2015.

WASHINGTON STATE UNIVERSITY

Genome Annotation Workflow



Existing Annotation Tools

- Computational Annotation (Automatic)
  - Maker
  - Ergatis
  - JAMg (Just Annotate My Genome)
  - iPlant DNA Subway
- Manual Annotation (Curation)
  - Apollo
  - WebApollo
  - Manatee
- These tools provide excellent annotation support but an integrated online solution to bridge all three workflow steps is lacking.

GenSAS Goals

- Streamline genomic annotation
  - Integrate all three stages of the annotation workflow
  - An easy-to-use online application
- Simplify pipeline setup for Bioinformaticist
- Support integration of final results into an online genome database for publication.
- Support analysis of very large genomes
  - Such as 22Gb *Pinus taeda* (Loblolly pine)
  - Using high-performance computing clusters
- Educate students and researchers
  - Inline instructions and best practices
  - Online videos & tutorials

GenSAS

WASHINGTON STATE UNIVERSITY

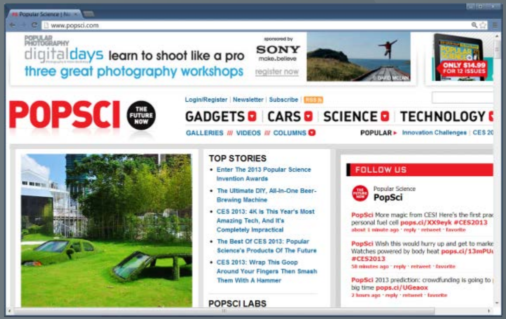
GenSAS Implementation

- A Drupal module
  - Drupal is an open-source, free Content Management System (CMS)
  - Used by millions of websites world-wide
  - High profile sites include:

THE WHITE HOUSE
BLOG
PHOTOS & VIDEO
BRIEFING ROOM
ISSUES
Relief for Middle Class Families
Thanks to a bipartisan agreement, 98 percent of Americans and 97 percent of small businesses will not see their income taxes go up.
GET THE FACTS
TOP NEWS
POPULAR TOPICS

US White House
http://www.whitehouse.gov

British Medical Journal
http://www.bmj.com/


digitaldays learn to shoot like a pro
three great photography workshops
SONY
POPSCI
GADGETS
CARS
SCIENCE
TECHNOLOGY
TOP STORIES
FOLLOW US
POPSCI LABS

Popular Science
http://www.popsci.com/

  - A Drupal site can support social and content needs for a community in addition to functionality provided by GenSAS.
  - Integrates easily with Tripal

Washington State University

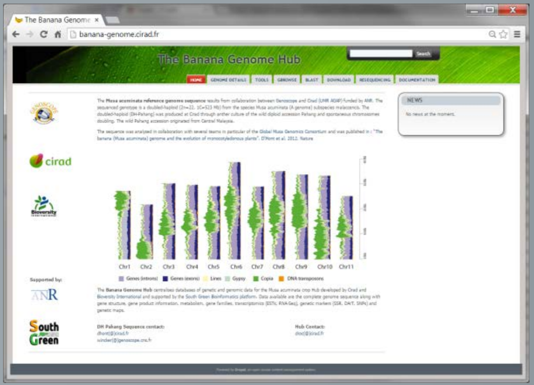
Tripal

- Construction of online genomic websites
  - A suite of Drupal modules
  - Uses the GMOD Chado schema for data storage
  - Decreases cost and man-power required to create online genomics, genetics and breeding databases
  - Provides an Application Programming Interface (API)
  - Tripal development is currently funded by:
    - NSF DIBBs Award #1443040 (2015-2018)
    - USDA NRSP10 Award (2014-2019)
    - USDA SCRI Award # 2014-51181-22376 (2014-2019)
    - Pending support from several NSF PGRP proposals (include funds to convert Dendrome/TreeGenes to Tripal)

http://tripal.info

WASHINGTON STATE UNIVERSITY

Example Sites Using Tripal



The Musa acuminata reference genome sequence results from collaboration between Genoscope and Cirad (UMR AGAP) funded by ANR. The sequenced genotype is a doubled-haploid (2n=22, 1C=523 Mb) from the species Musa acuminata (A genome) subspecies malaccensis. The doubled-haploid (DH-Pahang) was produced at Cirad through anther culture of the wild diploid accession Pahang and spontaneous chromosomes doubling. The wild Pahang accession originated from Central Malaysia.

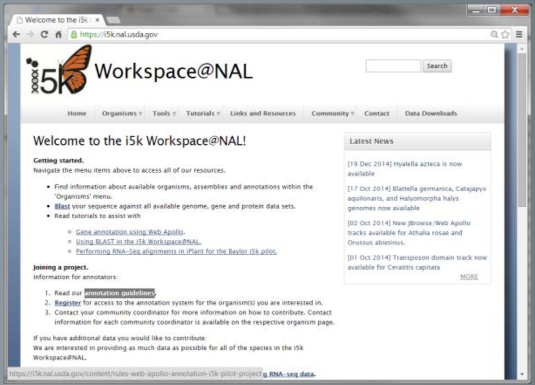
The sequence was analyzed in collaboration with several teams in particular of the Global Musa Genomics Consortium and was published in : "The banana (Musa acuminata) genome and the evolution of monocotyledonous plants". D'Hont et al. 2012. Nature

The Banana Genome Hub centralizes databases of genetic and genomic data for the Musa acuminata crop Hub developed by Cirad and Bioversity International and supported by the South Green Bioinformatics platform. Data available are the complete genome sequence along with gene structure, gene product information, metabolism, gene families, transcriptomics (ESTs, RNA-Seq), genetic markers (SSR, DArT, SNPs) and genetic maps.

DH Pahang Sequence contact:
dhont[@]cirad.fr
wincker[@]genoscope.cns.fr

Hub Contact:
droc[@]cirad.fr

Banana Genome Hub
http://banana-genome.cirad.fr/



Welcome to the i5k Workspace@NAL!

Getting started.
Navigate the menu items above to access all of our resources.

- Find information about available organisms, assemblies and annotations within the 'Organisms' menu.
- Blast your sequence against all available genome, gene and protein data sets.
- Read tutorials to assist with
  - Gene annotation using Web Apollo.
  - Using BLAST in the i5k Workspace@NAL.
  - Performing RNA-Seq alignments in iPlant for the Baylor i5k pilot.

Joining a project.
Information for annotators:

1. Read our annotation guidelines.
2. Register for access to the annotation system for the organism(s) you are interested in.
3. Contact your community coordinator for more information on how to contribute. Contact information for each community coordinator is available on the respective organism page.

If you have additional data you would like to contribute:
We are interested in providing as much data as possible for all of the species in the i5k Workspace@NAL.

[19 Dec 2014] Hyalella azteca is now available

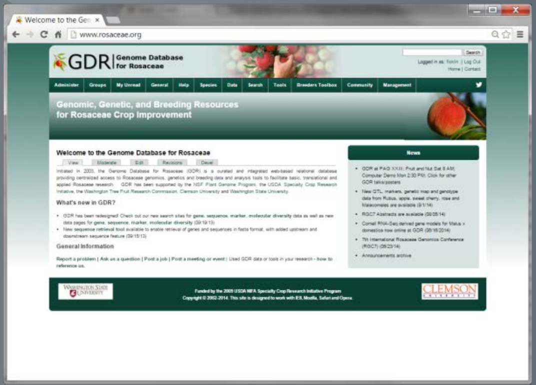
[17 Oct 2014] Blattella germanica, Catajapyx aquilonaris, and Halyomorpha halys genomes now available

[02 Oct 2014] New JBrowse/Web Apollo tracks available for Athalia rosae and Orussus abietinus.

[01 Oct 2014] Transposon domain track now available for Ceratitis capitata

https://i5k.nal.usda.gov/content/rules-web-apollo-annotation-i5k-pilot-project g RNA-seq data.

i5K Workspace@NAL
http://i5k.nal.usda.gov/



Welcome to the Genome Database for Rosaceae

Initiated in 2003, the Genome Database for Rosaceae (GDR) is a curated and integrated web-based relational database providing centralized access to Rosaceae genomics, genetics and breeding data and analysis tools to facilitate basic, translational and applied Rosaceae research. GDR has been supported by the NSF Plant Genome Program, the USDA Specialty Crop Research Initiative, the Washington Tree Fruit Research Commission, Clemson University and Washington State University.

What's new in GDR?

- GDR has been redesigned! Check out our new search sites for gene, sequence, marker, molecular diversity data as well as new data pages for gene, sequence, marker, molecular diversity (09/19/13)
- New sequence retrieval tool available to enable retrieval of genes and sequences in fasta format, with added upstream and downstream sequence feature (09/15/13)

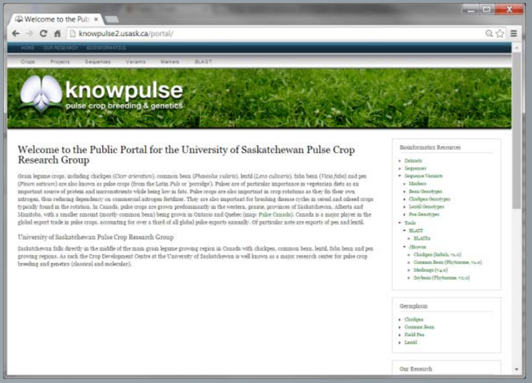
General Information

Report a problem | Ask us a question | Post a job | Post a meeting or event | Used GDR data or tools in your research - how to reference us.

- GDR at PAG XXIII: Fruit and Nut Sat 8 AM; Computer Demo Mon 2:30 PM; Click for other GDR talks/posters
- New QTL, markers, genetic map and genotype data from Rubus, apple, sweet cherry, rose and Maloideae are available (9/1/14)
- RGC7 Abstracts are available (08/28/14)
- Cornell RNA-Seq derived gene models for Malus x domestica now online at GDR (08/18/2014)
- 7th International Rosaceae Genomics Conference (RGC7) (08/23/14)
- Announcements archive

Funded by the 2009 USDA NIFA Specialty Crop Research Initiative Program
Copyright © 2002-2014. This site is designed to work with IE8, Mozilla, Safari and Opera.

Genome Database for Rosaceae
http://www.rosaceae.org



Welcome to the Public Portal for the University of Saskatchewan Pulse Crop Research Group

Grain legume crops, including chickpea (*Cicer arietinum*), common bean (*Phaseolus vularis*), lentil (*Lens culinaris*), faba bean (*Vicia faba*) and pea (*Pisum sativum*) are also known as pulse crops (from the Latin *Puls* or 'porridge'). Pulses are of particular importance in vegetarian diets as an important source of protein and micronutrients while being low in fats. Pulse crops are also important in crop rotations as they fix their own nitrogen, thus reducing dependency on commercial nitrogen fertilizer. They are also important for breaking disease cycles in cereal and oilseed crops typically found in the rotation. In Canada, pulse crops are grown predominantly in the western, prairie, provinces of Saskatchewan, Alberta and Manitoba, with a smaller amount (mostly common bean) being grown in Ontario and Quebec (map: Pulse Canada). Canada is a major player in the global export trade in pulse crops, accounting for over a third of all global pulse exports annually. Of particular note are exports of pea and lentil.

University of Saskatchewan Pulse Crop Research Group

Saskatchewan falls directly in the middle of the main grain legume growing region in Canada with chickpea, common bean, lentil, faba bean and pea growing regions. As such the Crop Development Centre at the University of Saskatchewan is well known as a major research center for pulse crop breeding and genetics (classical and molecular).

Bioinformatics Resources

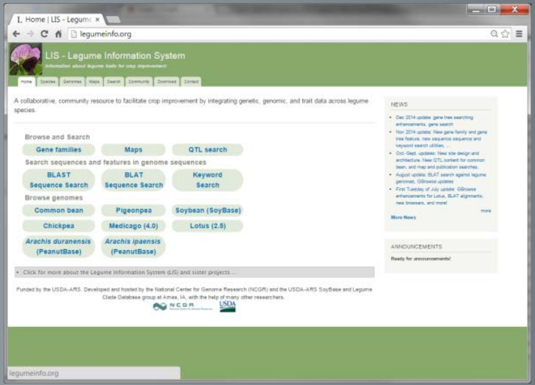
- Datasets
- Sequences
- Sequence Variants
  - Markers
  - Bean Genotypes
  - Chickpea Genotypes
  - Lentil Genotypes
  - Pea Genotypes
- Tools
  - BLAST
    - BLASTn
  - JBrowse
    - Chickpea (kabuli, v1.0)
    - Common Bean (Phytozome, v1.0)
    - Medicago (v4.0)
    - Soybean (Phytozome, v1.0)

Germplasm

- Chickpea
- Common Bean
- Field Pea
- Lentil

Our Research

Knowpulse: pulse crop genomics & breeding
http://knowpulse2.usask.ca/portal



A collaborative, community resource to facilitate crop improvement by integrating genetic, genomic, and trait data across legume species.

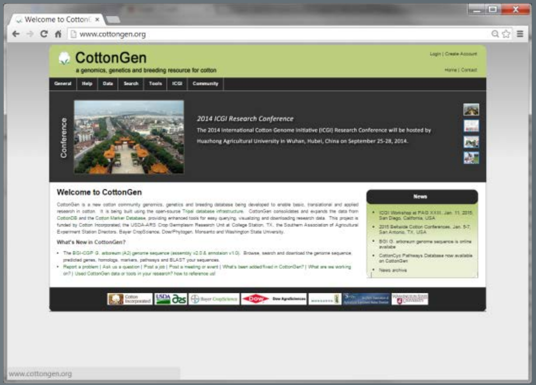
- Dec 2014 update: gene tree searching enhancements, gene search
- Nov 2014 update: New gene family and gene tree feature, new sequence sequence and keyword search utilities, ...
- Oct.-Sept. updates: New site design and architecture. New QTL content for common bean, and map and publication searches.
- August update: BLAT search against legume genomes, GBrowse updates
- First Tuesday of July update: GBrowse enhancements for Lotus, BLAT alignments, new browsers, and more!

Ready for announcements!

Click for more about the Legume Information System (LIS) and sister projects ...

Funded by the USDA-ARS. Developed and hosted by the National Center for Genome Research (NCGR) and the USDA-ARS SoyBase and Legume Clade Database group at Ames, IA, with the help of many other researchers.

Legume Information System
http://legumeinfo.org/



2014 ICGI Research Conference
The 2014 International Cotton Genome Initiative (ICGI) Research Conference will be hosted by Huazhong Agricultural University in Wuhan, Hubei, China on September 25-28, 2014.

Welcome to CottonGen

CottonGen is a new cotton community genomics, genetics and breeding database being developed to enable basic, translational and applied research in cotton. It is being built using the open-source Tripal database infrastructure. CottonGen consolidates and expands the data from CottonDB and the Cotton Marker Database, providing enhanced tools for easy querying, visualizing and downloading research data. This project is funded by Cotton Incorporated, the USDA-ARS Crop Germplasm Research Unit at College Station, TX, the Southern Association of Agricultural Experiment Station Directors, Bayer CropScience, Dow/Phytogen, Monsanto and Washington State University.

What's New in CottonGen?

- The BGI-CGP G. arboreum (A2) genome sequence (assembly v2.0 & annotation v1.0). Browse, search and download the genome sequence, predicted genes, homologs, markers, pathways and BLAST your sequences.
- Report a problem | Ask us a question | Post a job | Post a meeting or event | What's been added/fixed in CottonGen? | What are we working on? | Used CottonGen data or tools in your research? how to reference us!

- ICGI Workshop at PAG XXIII, Jan. 11, 2015, San Diego, California, USA
- 2015 Beltwide Cotton Conferences, Jan. 5-7, San Antonio, TX, USA
- BGI G. arboreum genome sequence is online available
- CottonCyc Pathways Database now available on CottonGen
- News archive

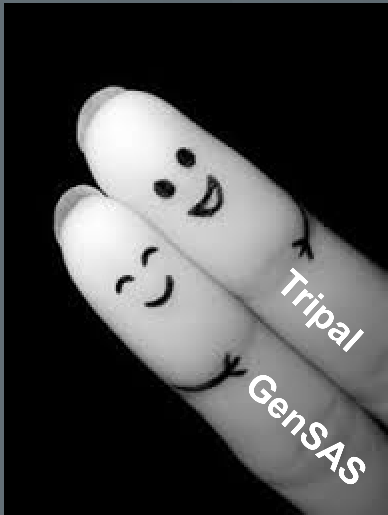
CottonGen
http://www.cottongen.org

WASHINGTON STATE UNIVERSITY

GenSAS + Tripal

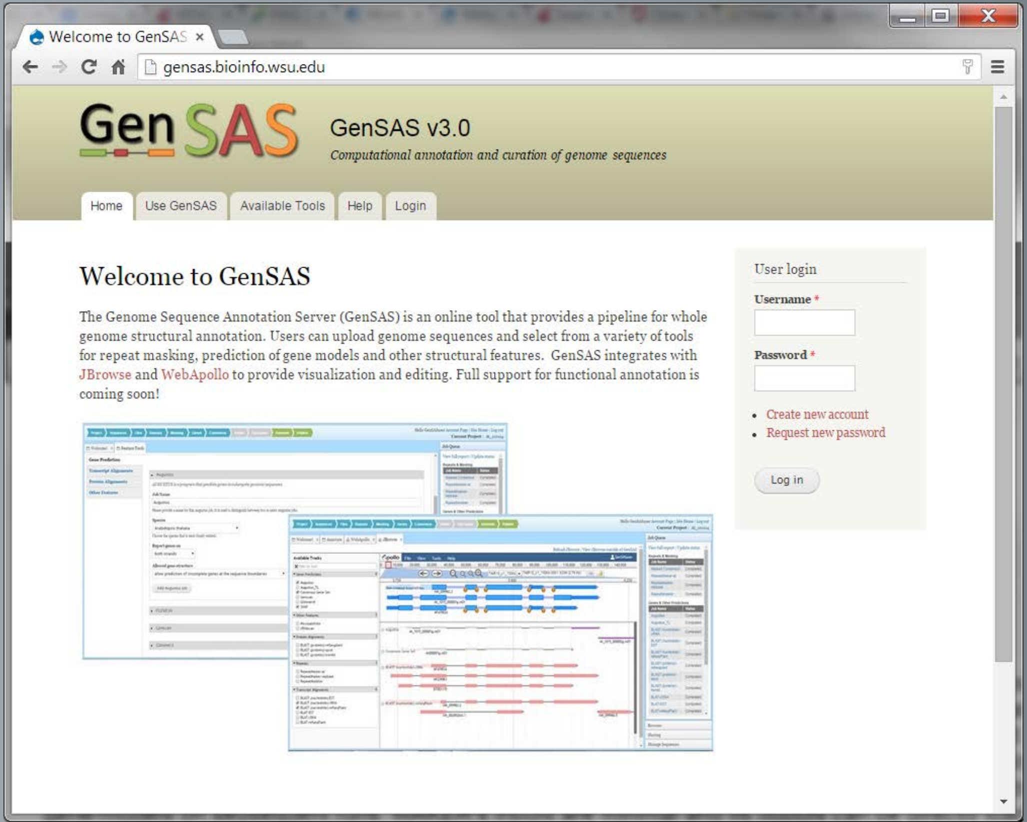
- Completes the "publish" stage
  - GenSAS fully integrates into the same Drupal site that uses Tripal
  - Results generated by GenSAS are 100% compatible
  - Versioning of results in GenSAS ensures no conflict with future annotation updates.
  - Tripal provides searching and pages (e.g. gene pages)

- Three options:
  - GenSAS expanded later with Tripal for full publication of the genomic annotations
  - GenSAS added to an existing Tripal site
  - GenSAS can be installed as a separate tool for non Tripal sites.



GenSAS Availability

- Public server:
  - http://gensas.bioinfo.wsu.edu
  - Resources are currently limited, thus accounts are given on a case-by-case basis
  - Anyone can request an account
  - Currently houses *Pinus taeda* Maker annotations



GenSAS v3.0

Computational annotation and curation of genome sequences

Welcome to GenSAS

The Genome Sequence Annotation Server (GenSAS) is an online tool that provides a pipeline for whole genome structural annotation. Users can upload genome sequences and select from a variety of tools for repeat masking, prediction of gene models and other structural features. GenSAS integrates with JBrowse and WebApollo to provide visualization and editing. Full support for functional annotation is coming soon!

- Source Code
  - http://drupal.org/project/gensas
  - Available soon…
  - download and install on local computational resources.

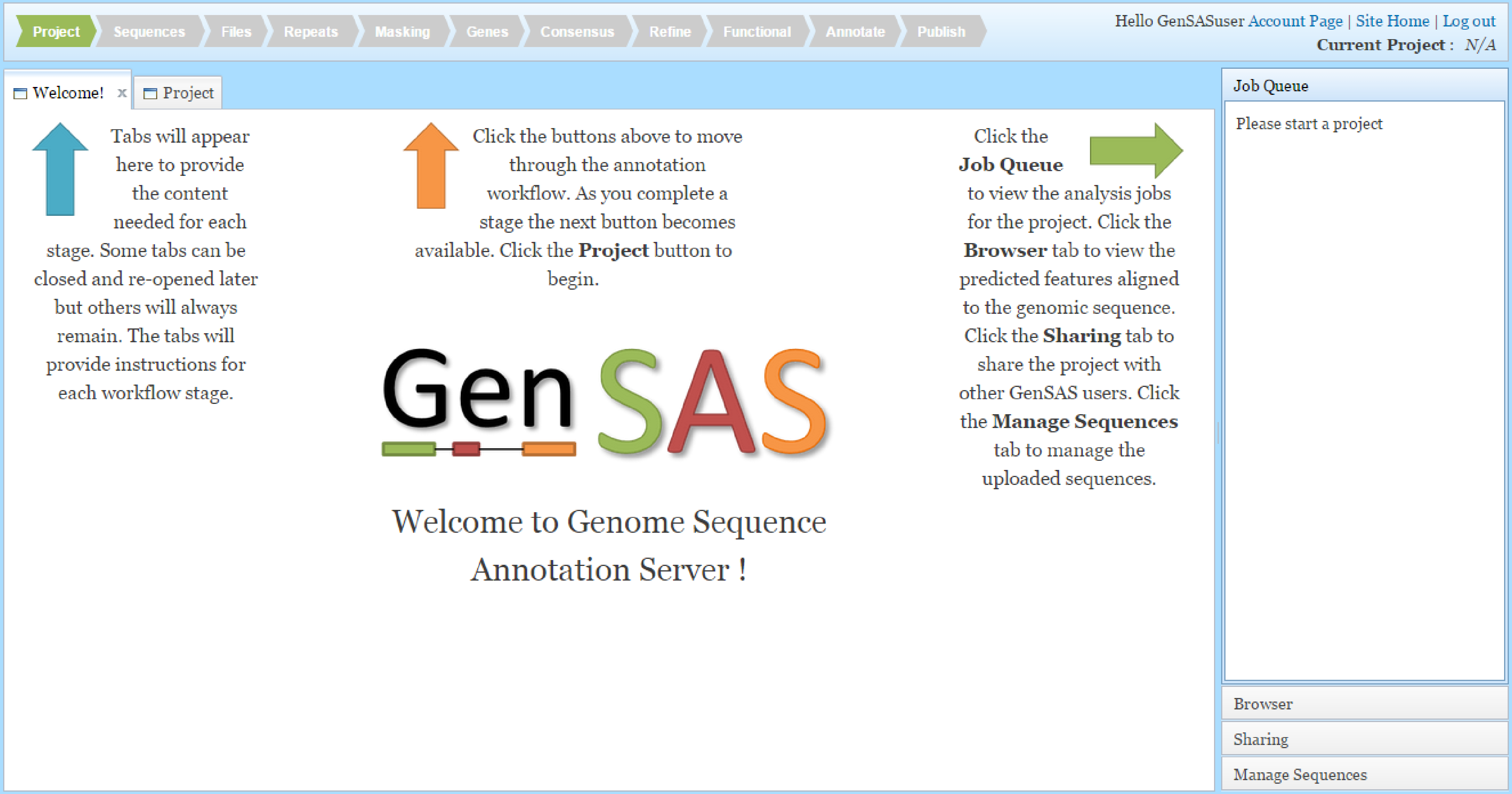
WASHINGTON STATE UNIVERSITY

Functionality Overview

- GenSAS v3.0 provides:
  - Repeat Finding & Masking: repeat libraries and *de novo*
    - *RepeatMasker, RepeatModeler*
  - Gene Prediction
    - Intrinsic
      - using tools the use heuristics and the genomic sequence
      - *Augustus, FgeneSH (parsing only), Genscan, Glimmer3, GlimmerM, SNAP, tRNAScan, getorf*
    - Extrinsic
      - using tools that use transcript or protein libraries to assist with gene prediction.
      - *BLAT, BLAST*
  - Gene Consensus Prediction
    - *EvidenceModeler*
  - Gene Visualization and Curation
    - *WebApollo, Jbrowse*
  - Publish
    - Generates GFF3 & FASTA files with properly versioned naming.

WASHINGTON STATE UNIVERSITY

Using GenSAS



Current Project : *N/A*

Please start a project

Tabs will appear here to provide the content needed for each stage. Some tabs can be closed and re-opened later but others will always remain. The tabs will provide instructions for each workflow stage.

Click the buttons above to move through the annotation workflow. As you complete a stage the next button becomes available. Click the **Project** button to begin.

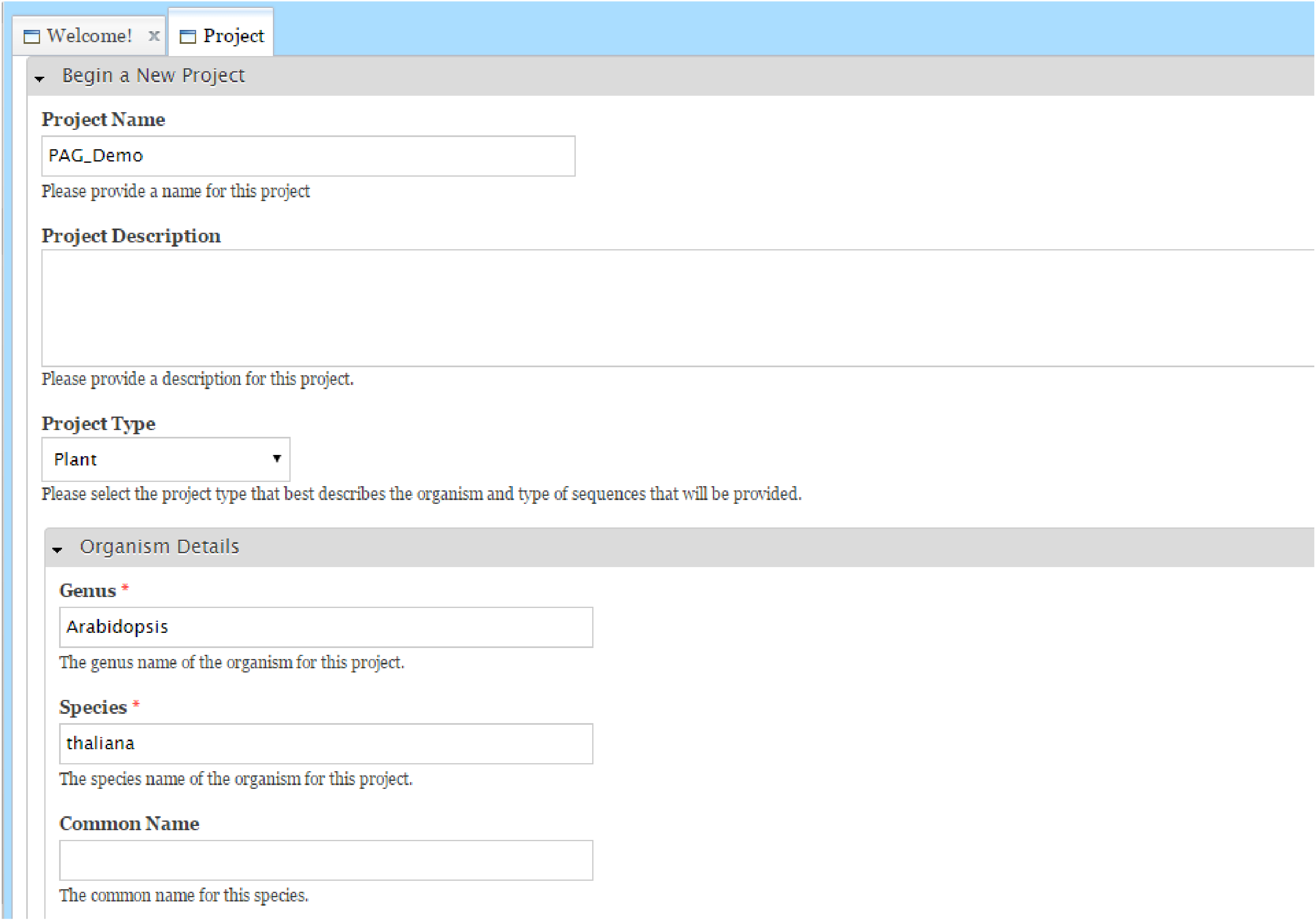
Click the **Job Queue** to view the analysis jobs for the project. Click the **Browser** tab to view the predicted features aligned to the genomic sequence. Click the **Sharing** tab to share the project with other GenSAS users. Click the **Manage Sequences** tab to manage the uploaded sequences.

Welcome to Genome Sequence Annotation Server !

- View the GenSAS Computer Demo C03 at 2:10pm today in California Room for further detail, plus Poster # P1153

WASHINGTON STATE UNIVERSITY

Step 1: Start a project



Begin a New Project

Project Name

PAG_Demo

Please provide a name for this project

Project Description

Please provide a description for this project.

Project Type

Plant ▾

Please select the project type that best describes the organism and type of sequences that will be provided.

Organism Details

Genus *

Arabidopsis

The genus name of the organism for this project.

Species *

thaliana

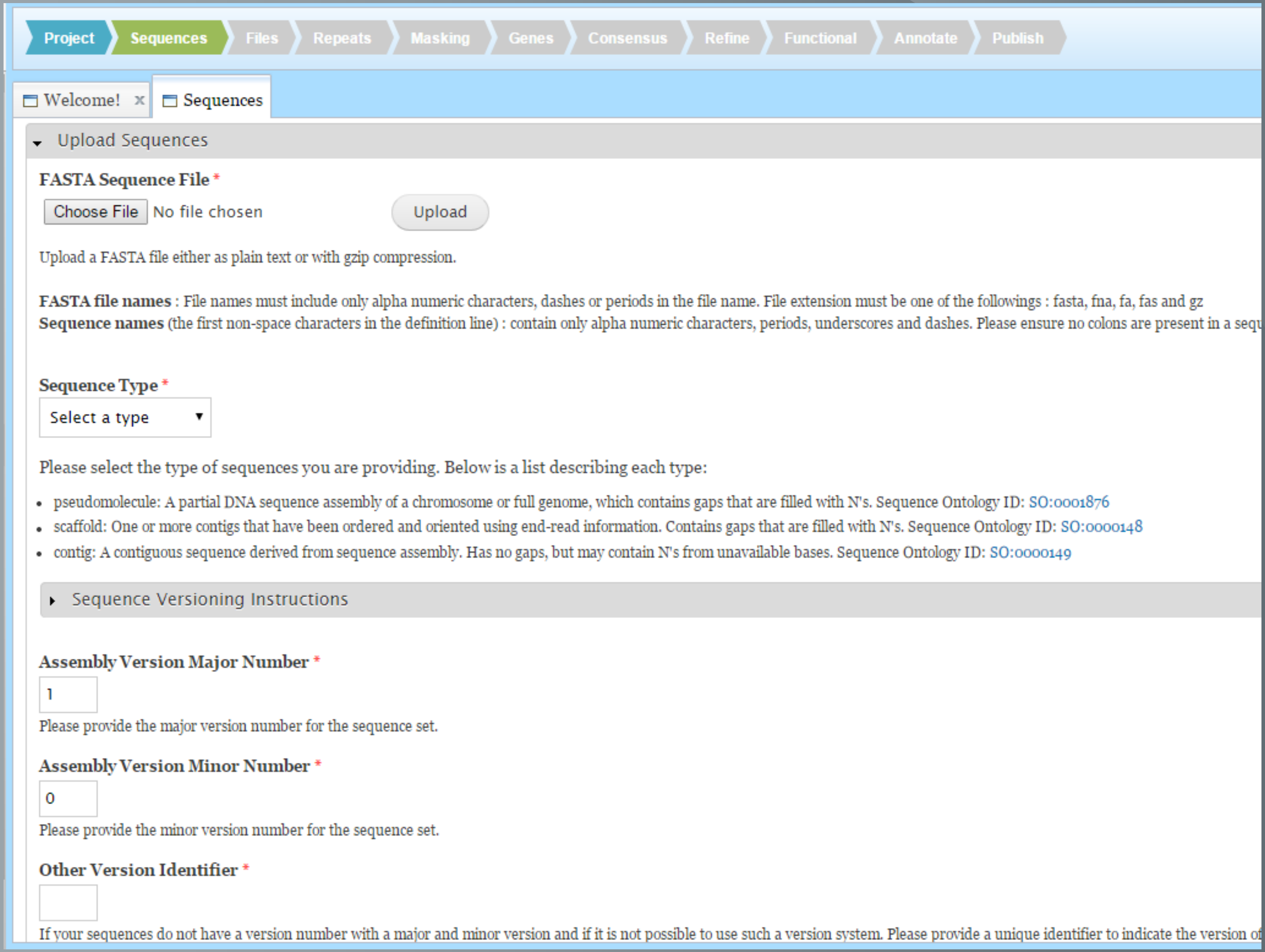
The species name of the organism for this project.

Common Name

The common name for this species.

WASHINGTON STATE UNIVERSITY

Step 2: Upload Sequences



Upload Sequences

FASTA Sequence File *

Choose File No file chosen   Upload

Upload a FASTA file either as plain text or with gzip compression.

FASTA file names : File names must include only alpha numeric characters, dashes or periods in the file name. File extension must be one of the followings : fasta, fna, fa, fas and gz
Sequence names (the first non-space characters in the definition line) : contain only alpha numeric characters, periods, underscores and dashes. Please ensure no colons are present in a sequ

Sequence Type *

Select a type ▾

Please select the type of sequences you are providing. Below is a list describing each type:

- pseudomolecule: A partial DNA sequence assembly of a chromosome or full genome, which contains gaps that are filled with N's. Sequence Ontology ID: SO:0001876
- scaffold: One or more contigs that have been ordered and oriented using end-read information. Contains gaps that are filled with N's. Sequence Ontology ID: SO:0000148
- contig: A contiguous sequence derived from sequence assembly. Has no gaps, but may contain N's from unavailable bases. Sequence Ontology ID: SO:0000149

Sequence Versioning Instructions

Assembly Version Major Number *

1

Please provide the major version number for the sequence set.

Assembly Version Minor Number *

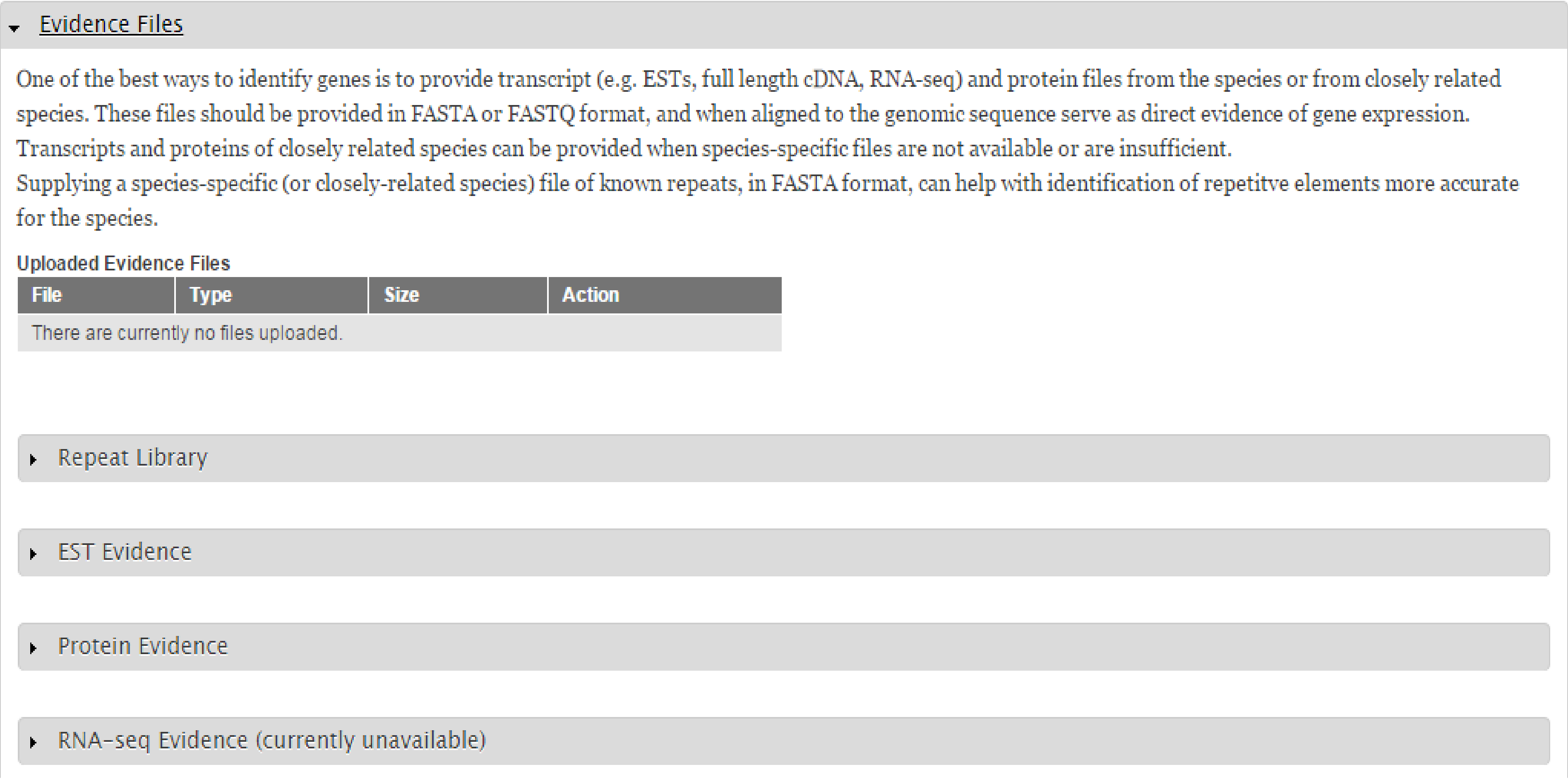
0

Please provide the minor version number for the sequence set.

Other Version Identifier *

If your sequences do not have a version number with a major and minor version and if it is not possible to use such a version system. Please provide a unique identifier to indicate the version of

Step 3: Upload Supporting Files



▾ Evidence Files

One of the best ways to identify genes is to provide transcript (e.g. ESTs, full length cDNA, RNA-seq) and protein files from the species or from closely related species. These files should be provided in FASTA or FASTQ format, and when aligned to the genomic sequence serve as direct evidence of gene expression.
Transcripts and proteins of closely related species can be provided when species-specific files are not available or are insufficient.
Supplying a species-specific (or closely-related species) file of known repeats, in FASTA format, can help with identification of repetitve elements more accurate for the species.

Uploaded Evidence Files

File	Type	Size	Action
There are currently no files uploaded.			

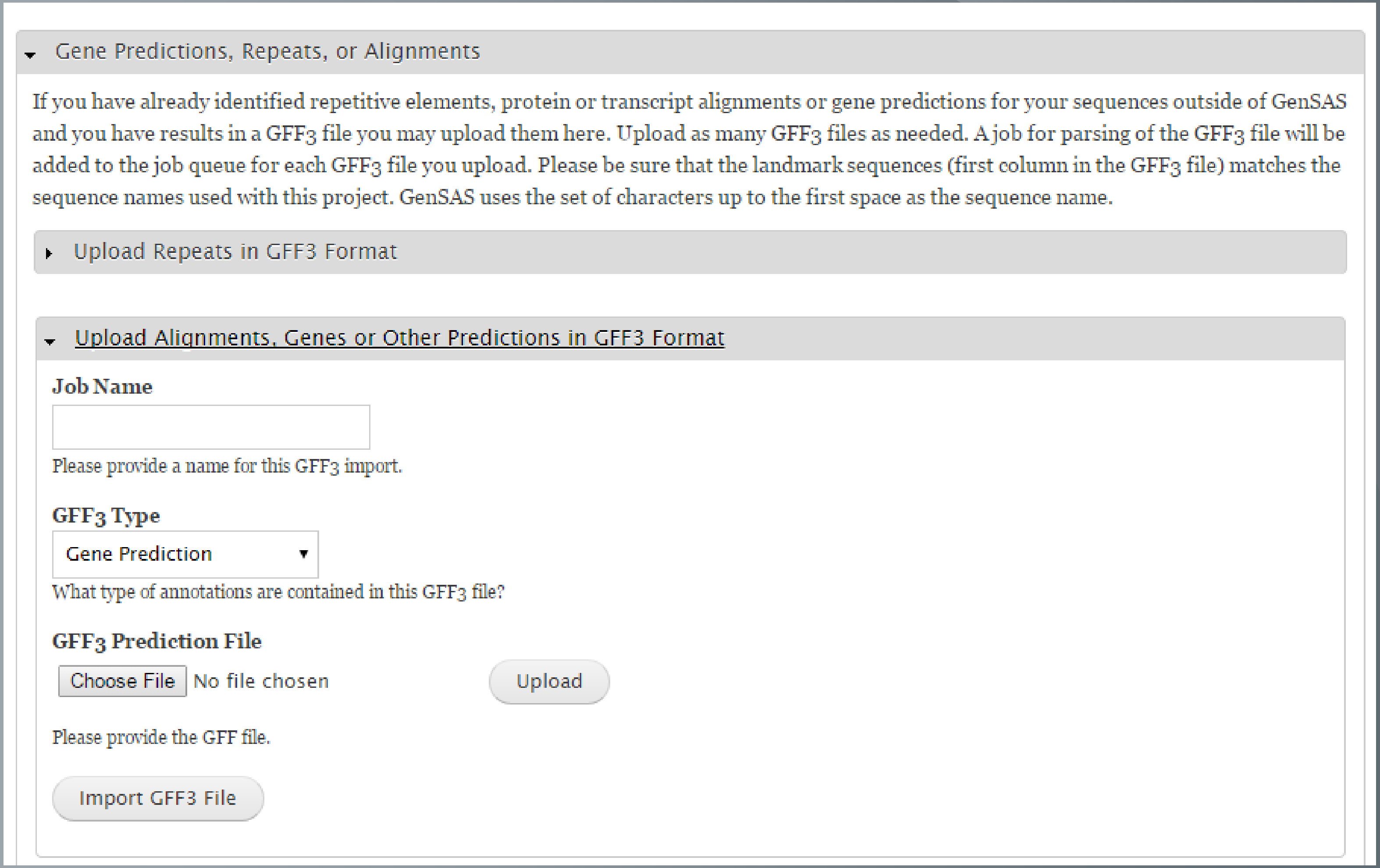
▸ Repeat Library

▸ EST Evidence

▸ Protein Evidence

▸ RNA-seq Evidence (currently unavailable)

Step 3: Upload Supporting Files



If you have already identified repetitive elements, protein or transcript alignments or gene predictions for your sequences outside of GenSAS and you have results in a GFF3 file you may upload them here. Upload as many GFF3 files as needed. A job for parsing of the GFF3 file will be added to the job queue for each GFF3 file you upload. Please be sure that the landmark sequences (first column in the GFF3 file) matches the sequence names used with this project. GenSAS uses the set of characters up to the first space as the sequence name.

▸ Upload Repeats in GFF3 Format
▾ Upload Alignments, Genes or Other Predictions in GFF3 Format

Job Name

Please provide a name for this GFF3 import.

GFF3 Type

Gene Prediction ▾

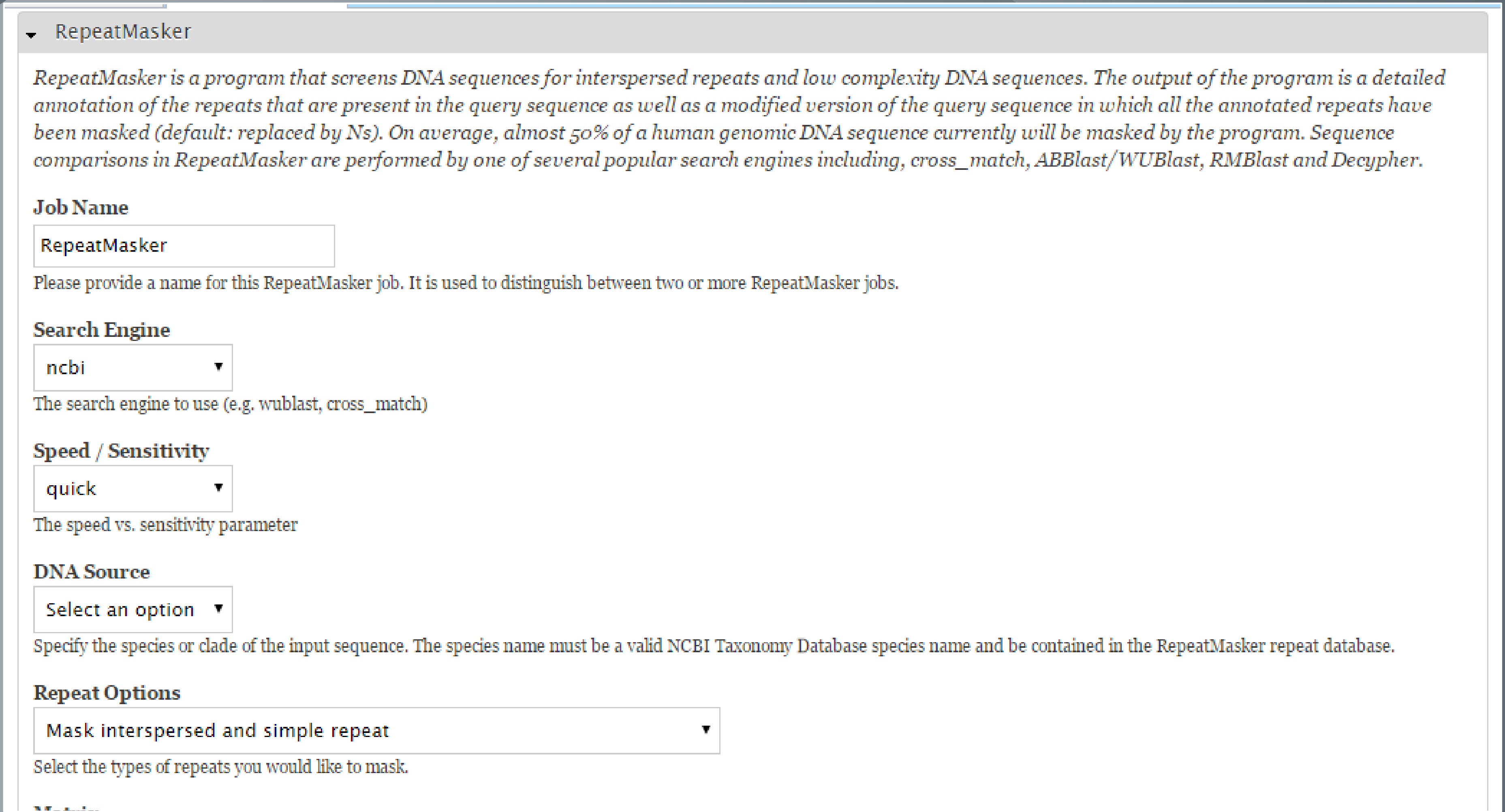
What type of annotations are contained in this GFF3 file?

GFF3 Prediction File

Please provide the GFF file.

Current *Pinus taeda* Maker predictions loaded in this way

Step 4: Repeat Finding



RepeatMasker is a program that screens DNA sequences for interspersed repeats and low complexity DNA sequences. The output of the program is a detailed annotation of the repeats that are present in the query sequence as well as a modified version of the query sequence in which all the annotated repeats have been masked (default: replaced by Ns). On average, almost 50% of a human genomic DNA sequence currently will be masked by the program. Sequence comparisons in RepeatMasker are performed by one of several popular search engines including, cross_match, ABBlast/WUBlast, RMBlast and Decypher.

Job Name

RepeatMasker

Please provide a name for this RepeatMasker job. It is used to distinguish between two or more RepeatMasker jobs.

Search Engine

ncbi ▾

The search engine to use (e.g. wublast, cross_match)

Speed / Sensitivity

quick ▾

The speed vs. sensitivity parameter

DNA Source

Select an option ▾

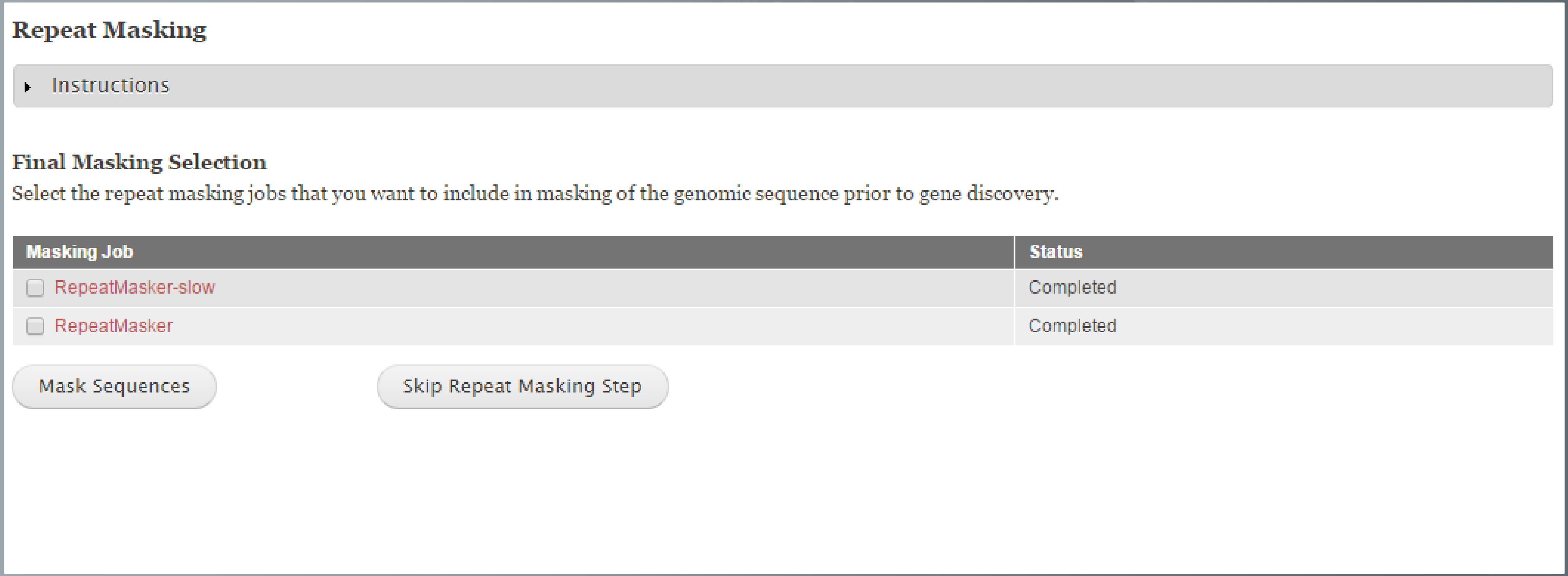
Specify the species or clade of the input sequence. The species name must be a valid NCBI Taxonomy Database species name and be contained in the RepeatMasker repeat database.

Repeat Options

Mask interspersed and simple repeat ▾

Select the types of repeats you would like to mask.

Step 5: Masking



Repeat Masking

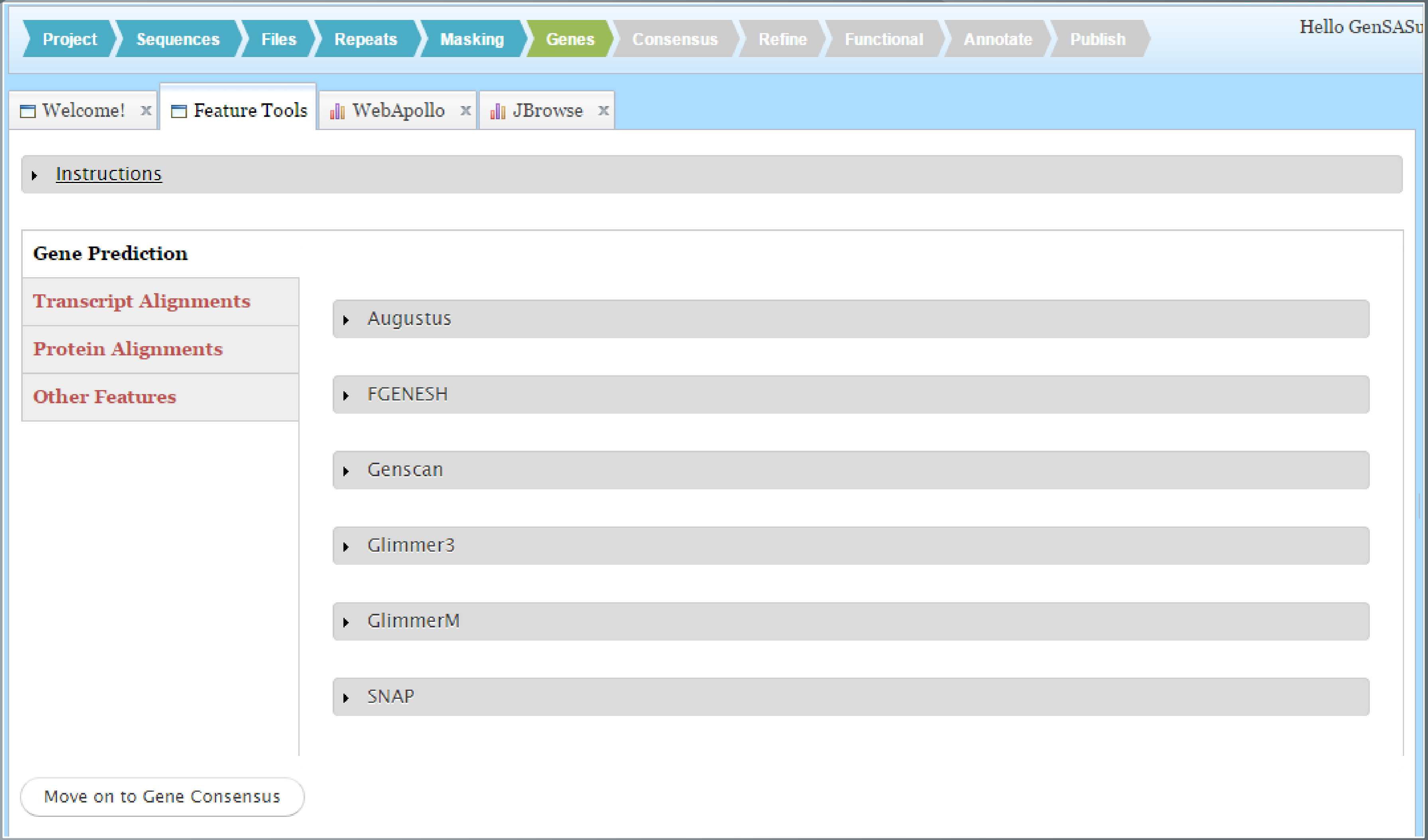
▸ Instructions

Final Masking Selection

Select the repeat masking jobs that you want to include in masking of the genomic sequence prior to gene discovery.

Masking Job	Status
□ RepeatMasker-slow	Completed
□ RepeatMasker	Completed

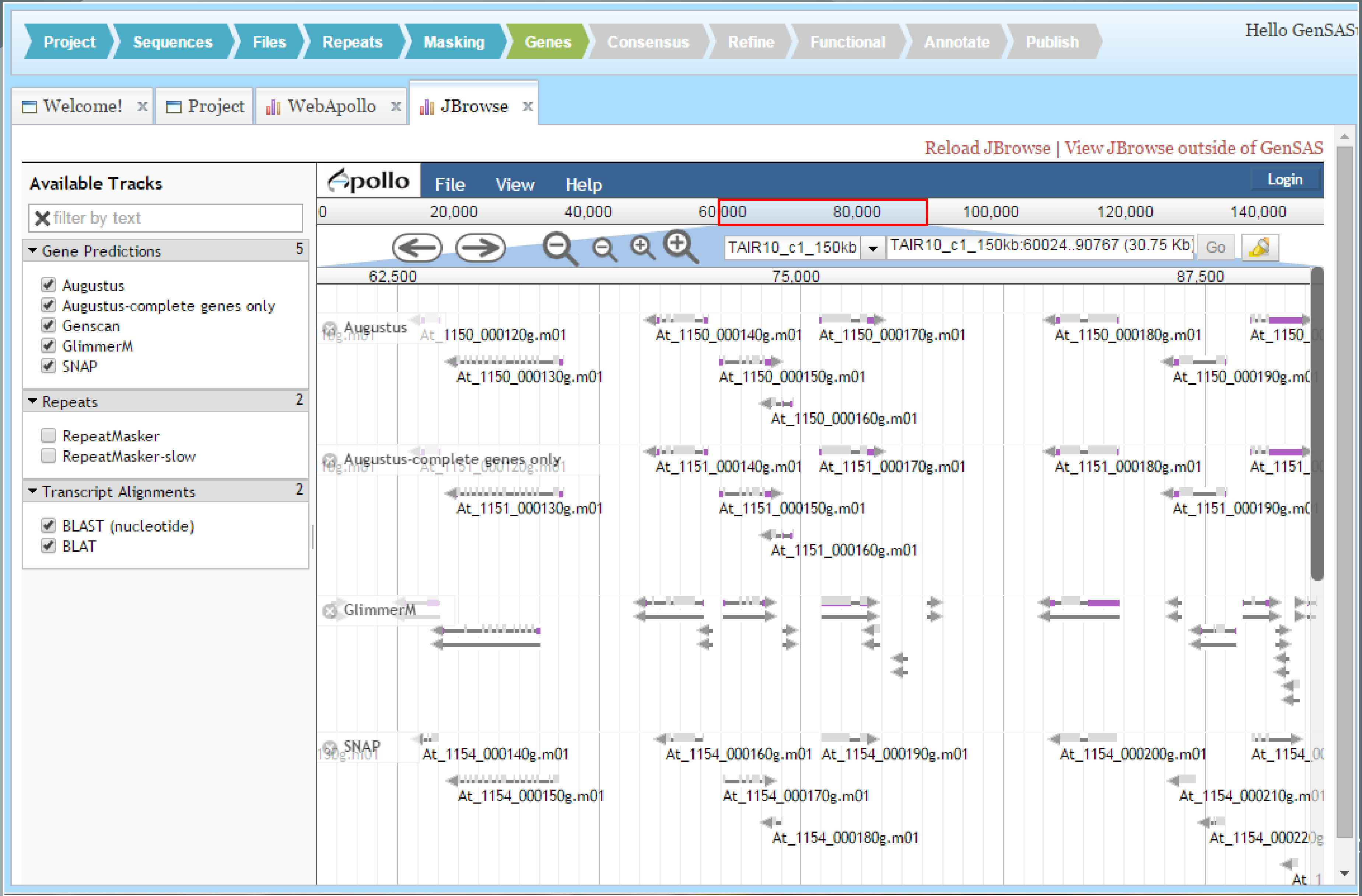
Step 6: Gene Prediction



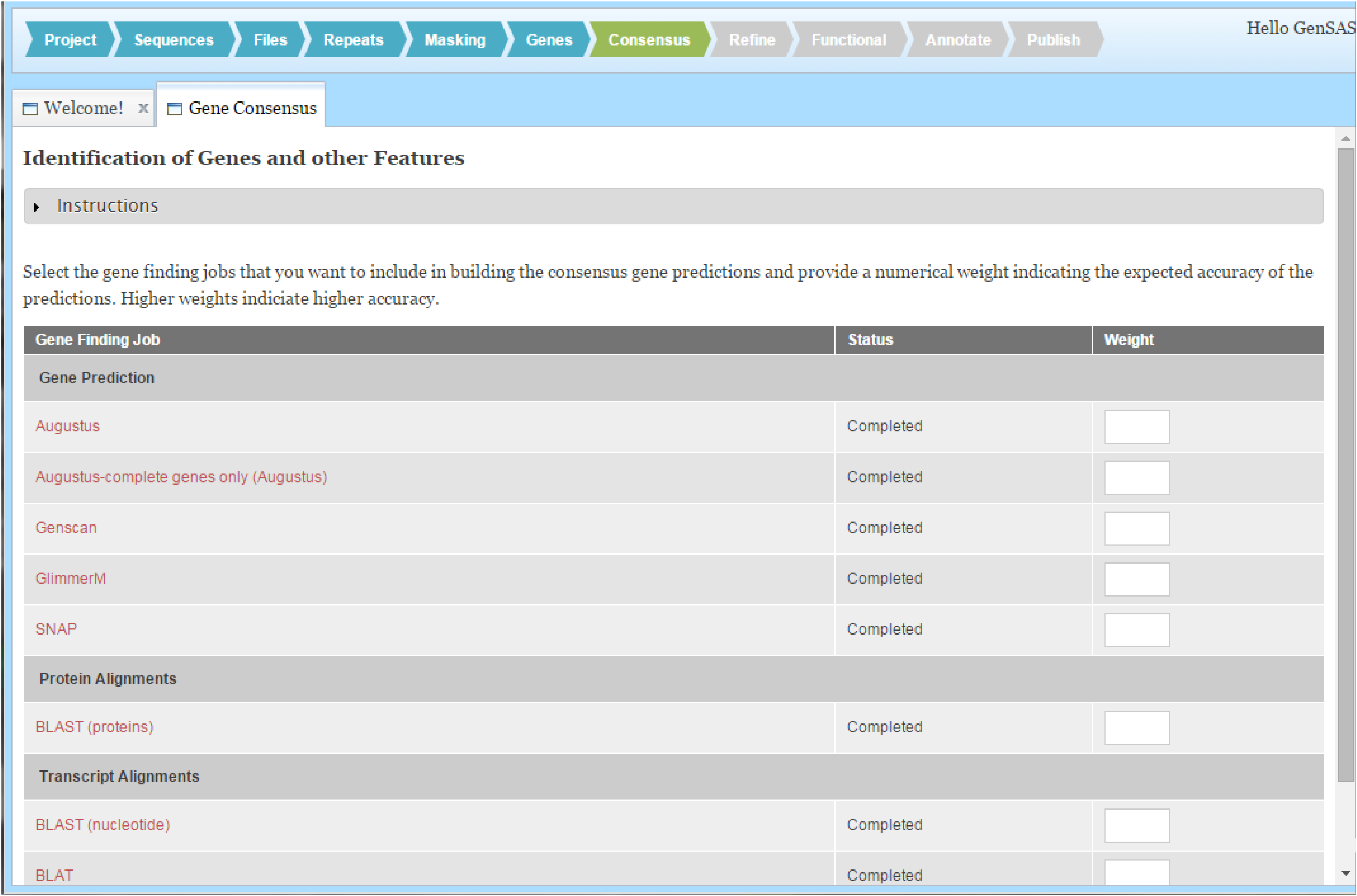
Gene Prediction

WASHINGTON STATE UNIVERSITY

View Results



Step 7: Consensus

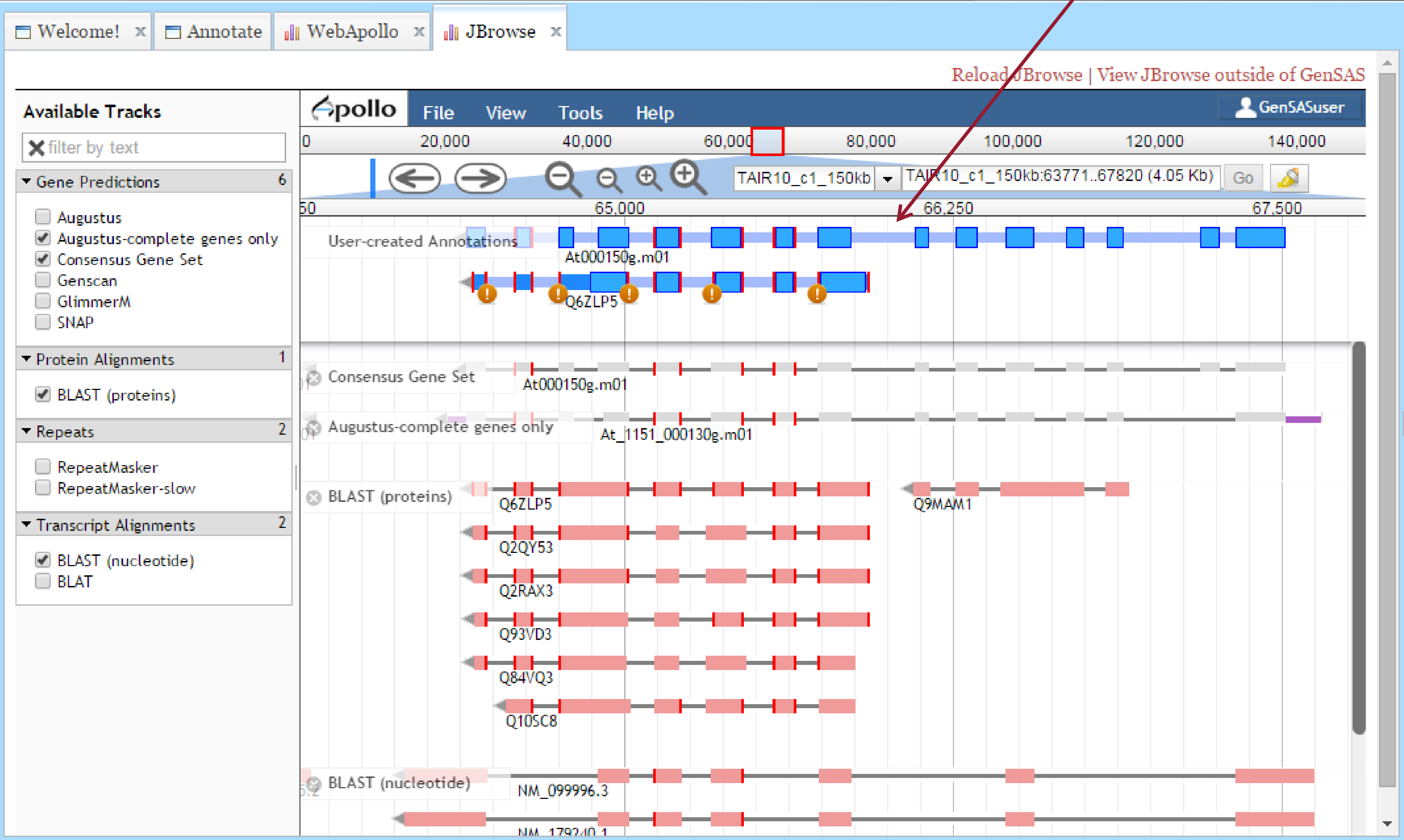


Identification of Genes and other Features

Select the gene finding jobs that you want to include in building the consensus gene predictions and provide a numerical weight indicating the expected accuracy of the predictions. Higher weights indiciate higher accuracy.

Gene Finding Job	Status	Weight
Gene Prediction		
Augustus	Completed	
Augustus-complete genes only (Augustus)	Completed	
Genscan	Completed	
GlimmerM	Completed	
SNAP	Completed	
Protein Alignments		
BLAST (proteins)	Completed	
Transcript Alignments		
BLAST (nucleotide)	Completed	
BLAT	Completed	

Step 8: Manual Curation



Curation Track

Reload JBrowse | View JBrowse outside of GenSAS

TAIR10_c1_150kb:63771..67820 (4.05 Kb)

Available Tracks

▼ Gene Predictions 6

- ☐ Augustus
- ☑ Augustus-complete genes only
- ☑ Consensus Gene Set
- ☐ Genscan
- ☐ GlimmerM
- ☐ SNAP

▼ Protein Alignments 1

- ☑ BLAST (proteins)

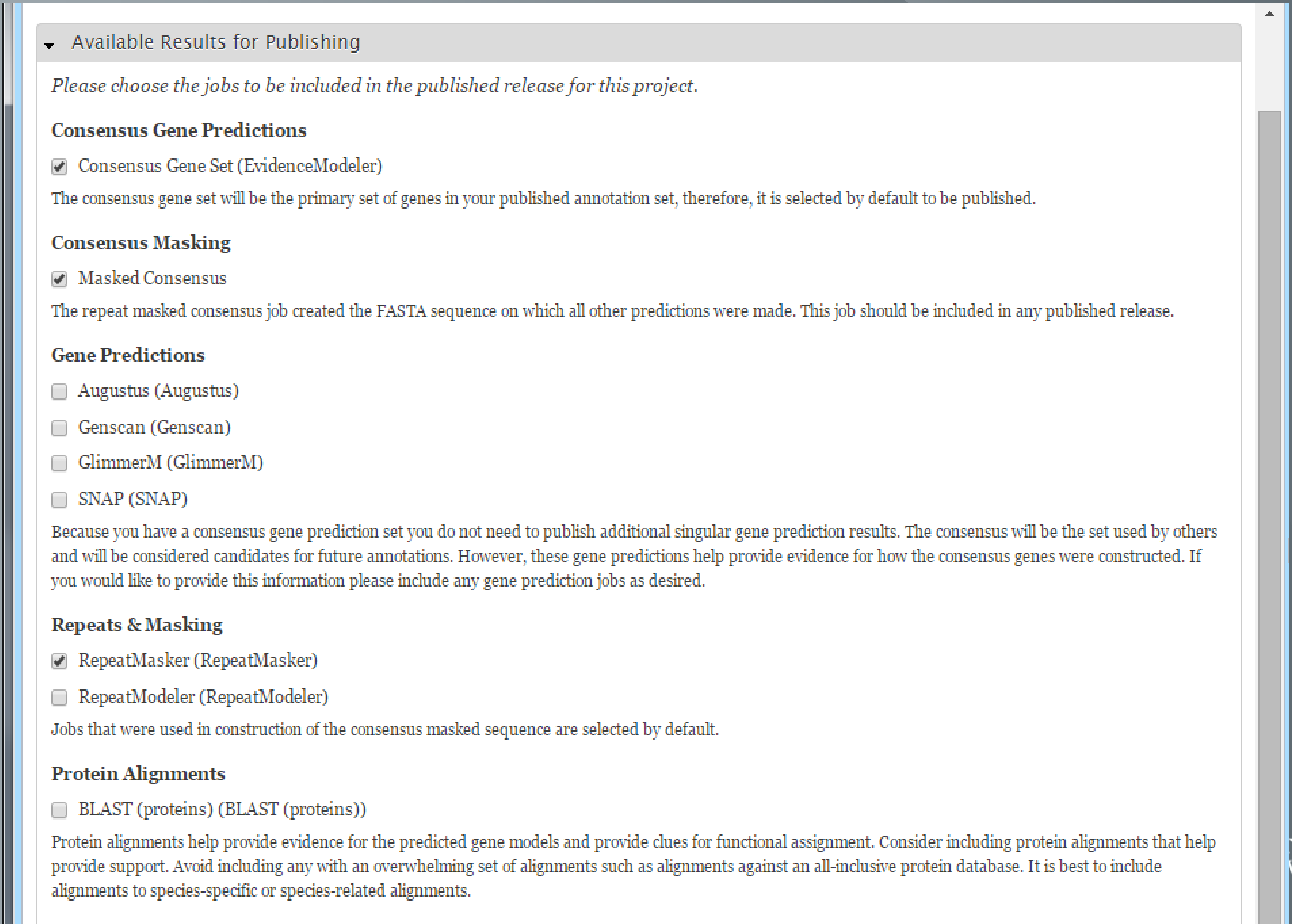
▼ Repeats 2

- ☐ RepeatMasker
- ☐ RepeatMasker-slow

▼ Transcript Alignments 2

- ☑ BLAST (nucleotide)
- ☐ BLAT

Step 9: Publish



▾ Available Results for Publishing

Please choose the jobs to be included in the published release for this project.

Consensus Gene Predictions

☑ Consensus Gene Set (EvidenceModeler)

The consensus gene set will be the primary set of genes in your published annotation set, therefore, it is selected by default to be published.

Consensus Masking

☑ Masked Consensus

The repeat masked consensus job created the FASTA sequence on which all other predictions were made. This job should be included in any published release.

Gene Predictions

□ Augustus (Augustus)

□ Genscan (Genscan)

□ GlimmerM (GlimmerM)

□ SNAP (SNAP)

Because you have a consensus gene prediction set you do not need to publish additional singular gene prediction results. The consensus will be the set used by others and will be considered candidates for future annotations. However, these gene predictions help provide evidence for how the consensus genes were constructed. If you would like to provide this information please include any gene prediction jobs as desired.

Repeats & Masking

☑ RepeatMasker (RepeatMasker)

□ RepeatModeler (RepeatModeler)

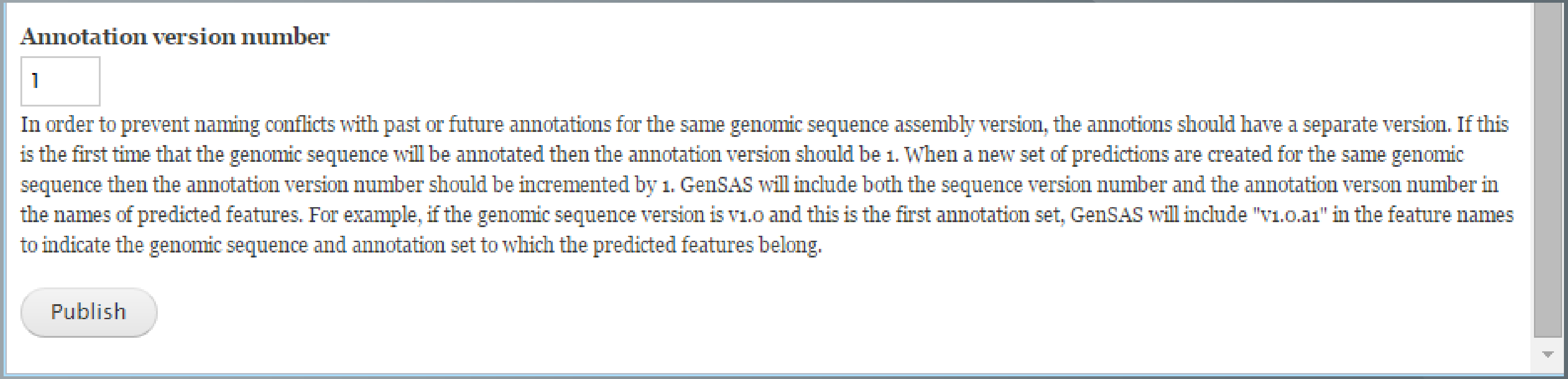
Jobs that were used in construction of the consensus masked sequence are selected by default.

Protein Alignments

□ BLAST (proteins) (BLAST (proteins))

Protein alignments help provide evidence for the predicted gene models and provide clues for functional assignment. Consider including protein alignments that help provide support. Avoid including any with an overwhelming set of alignments such as alignments against an all-inclusive protein database. It is best to include alignments to species-specific or species-related alignments.

Step 9: Publish



Annotation version number

1

In order to prevent naming conflicts with past or future annotations for the same genomic sequence assembly version, the annotions should have a separate version. If this is the first time that the genomic sequence will be annotated then the annotation version should be 1. When a new set of predictions are created for the same genomic sequence then the annotation version number should be incremented by 1. GenSAS will include both the sequence version number and the annotation verson number in the names of predicted features. For example, if the genomic sequence version is v1.0 and this is the first annotation set, GenSAS will include "v1.0.a1" in the feature names to indicate the genomic sequence and annotation set to which the predicted features belong.

Ongoing Work

- Functional Annotation
  - Does not yet fully support functional annotation.
  - Requires addition of functional tools and result viewers.
  - Currently under development
  - Will be available in a future release of GenSAS
- Support of High-Performance Computing
  - Currently executes on stand-alone server
  - Will integrate with two types of HPC schedulers: PBS, GE
  - Currently under development
  - Will be available in a future release of GenSAS
- Support of RNA-Seq Datasets
- Integration with PASA for gene refinements
- Full Integration with Tripal
  - Currently only exported files are compatible with Tripal loaders
  - Integrate so that publish button can automatically import into Tripal.

WASHINGTON STATE UNIVERSITY

Funding

- Current Funding:
  - The work presented herein was funded by the SDA/NIFA (2011-67009-30030) subaward to Dorrie Main at Washington State University (PI: David Neale at University of California, Davis)

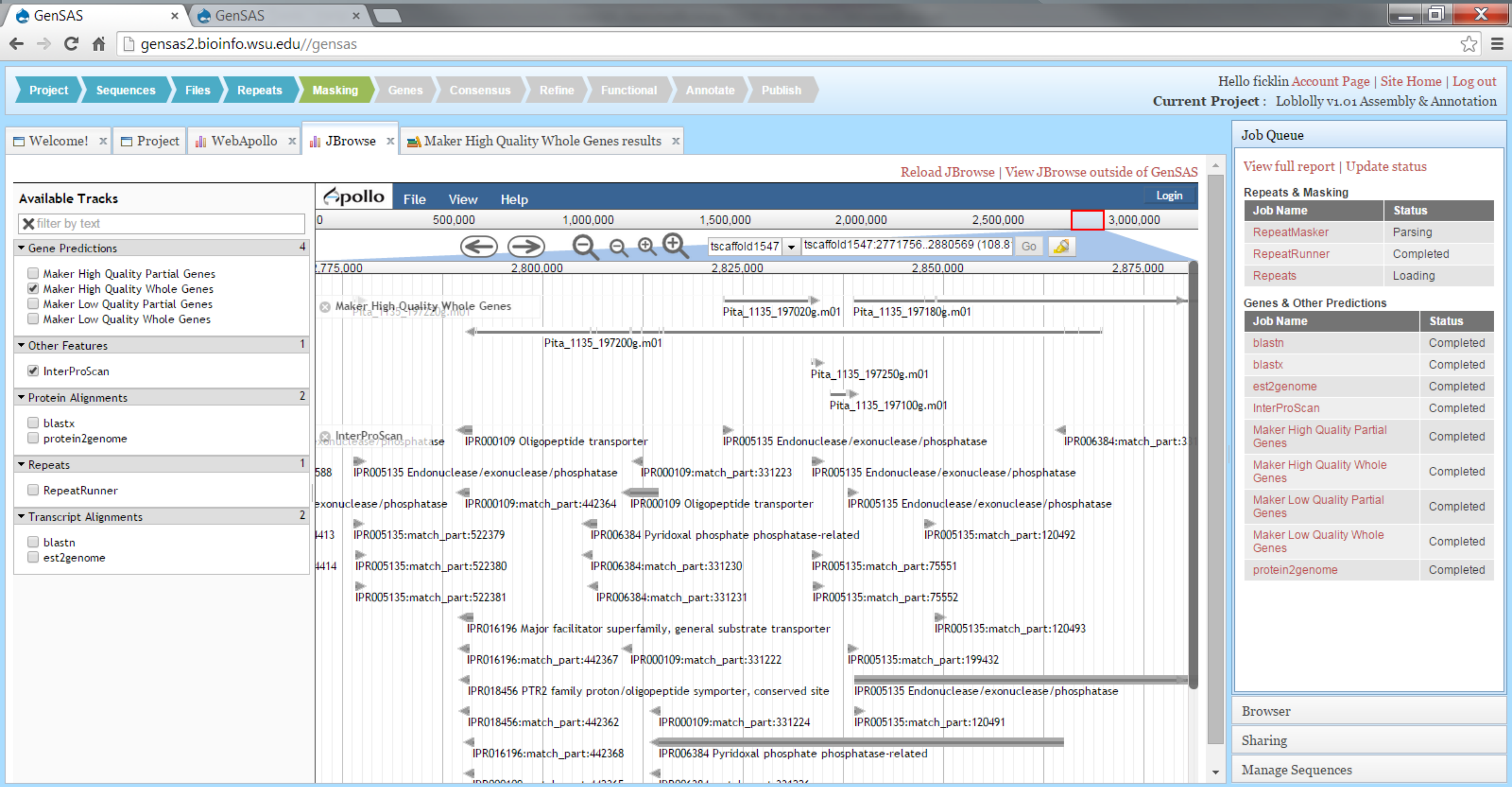
- Continued Funding
  - Further funding for development and implementation of GenSAS is provided by an USDA National Research Support Project (NRSP10) to WSU (Dorrie Main PI) through 2019.

GenSAS for Pine Annotation

- Current capabilities
  - *Pinus Taeda v1.01* Maker annotations loaded
  - Pine community can annotate the predicted genes using integrated WebApollo
  - WebApollo manages user access (via GenSAS integration)
  - Annotation training may be available via WebApollo outreach for community.
- Future capabilities
  - Functional Annotation:
    - Execute functional tools inside GenSAS
    - Result viewers for tools such as InterProScan, SignalP, TargetP, Pfam, blastp, etc.
  - Tools will execute on high-performance computing to decrease time requirements
  - Future pine genomes can be annotated directly in GenSAS

WASHINGTON STATE UNIVERSITY

Loblolly Pine in GenSAS



GenSAS | GenSAS

gensas2.bioinfo.wsu.edu//gensas

Project > Sequences > Files > Repeats > Masking > Genes > Consensus > Refine > Functional > Annotate > Publish

Hello ficklin Account Page | Site Home | Log out

Current Project : Loblolly v1.01 Assembly & Annotation

Welcome! | Project | WebApollo | JBrowse | Maker High Quality Whole Genes results

Reload JBrowse | View JBrowse outside of GenSAS

Available Tracks

filter by text

Gene Predictions 4

- ☐ Maker High Quality Partial Genes
- ☑ Maker High Quality Whole Genes
- ☐ Maker Low Quality Partial Genes
- ☐ Maker Low Quality Whole Genes

Other Features 1

- ☑ InterProScan

Protein Alignments 2

- ☐ blastx
- ☐ protein2genome

Repeats 1

- ☐ RepeatRunner

Transcript Alignments 2

- ☐ blastn
- ☐ est2genome

Apollo
File
View
Help
Login
0
500,000
1,000,000
1,500,000
2,000,000
2,500,000
3,000,000
tscaffold1547
tscaffold1547:2771756..2880569 (108.8
Go
2,775,000
2,800,000
2,825,000
2,850,000
2,875,000
Maker High Quality Whole Genes
Pita_1135_197220g.m01
Pita_1135_197020g.m01
Pita_1135_197180g.m01
Pita_1135_197200g.m01
Pita_1135_197250g.m01
Pita_1135_197100g.m01
InterProScan
exonuclease/phosphatase
IPR000109 Oligopeptide transporter
IPR005135 Endonuclease/exonuclease/phosphatase
IPR006384:match_part:3
588
IPR005135 Endonuclease/exonuclease/phosphatase
IPR000109:match_part:331223
IPR005135 Endonuclease/exonuclease/phosphatase
exonuclease/phosphatase
IPR000109:match_part:442364
IPR000109 Oligopeptide transporter
IPR005135 Endonuclease/exonuclease/phosphatase
413
IPR005135:match_part:522379
IPR006384 Pyridoxal phosphate phosphatase-related
IPR005135:match_part:120492
414
IPR005135:match_part:522380
IPR006384:match_part:331230
IPR005135:match_part:75551
IPR005135:match_part:522381
IPR006384:match_part:331231
IPR005135:match_part:75552
IPR016196 Major facilitator superfamily, general substrate transporter
IPR005135:match_part:120493
IPR016196:match_part:442367
IPR000109:match_part:331222
IPR005135:match_part:199432
IPR018456 PTR2 family proton/oligopeptide symporter, conserved site
IPR005135 Endonuclease/exonuclease/phosphatase
IPR018456:match_part:442362
IPR000109:match_part:331224
IPR005135:match_part:120491
IPR016196:match_part:442368
IPR006384 Pyridoxal phosphate phosphatase-related

Job Queue

View full report | Update status

Repeats & Masking

Job Name	Status
RepeatMasker	Parsing
RepeatRunner	Completed
Repeats	Loading

Genes & Other Predictions

Job Name	Status
blastn	Completed
blastx	Completed
est2genome	Completed
InterProScan	Completed
Maker High Quality Partial Genes	Completed
Maker High Quality Whole Genes	Completed
Maker Low Quality Partial Genes	Completed
Maker Low Quality Whole Genes	Completed
protein2genome	Completed

Browser

Sharing

Manage Sequences

Thank You!

GenSAS

More GenSAS at PAG

- Computer Demo: C03 at 2:10pm today, Jan 10th in California Room for further details
- Poster: P1153. Meet w/ Jodi Humann from 3.00-4.30 PM Monday, Jan 12th
- Contact Info: dorrie@wsu.edu, stephen.ficklin@wsu.edu, jhumann@wsu.edu

WASHINGTON STATE UNIVERSITY