A web-based platform for structural and GenSAS functional annotation of model and non-model v6.0 organisms

Jodi L. Humann¹, Taein Lee¹, Stephen P. Ficklin¹, Chun-Huai Cheng¹, Heidi Hough¹, Sook Jung¹, Jill Wegrzyn², David Neale³, and Dorrie Main¹

¹Washington State University, Pullman, WA; ²University of Connecticut, Storrs, CT; ³University of California, Davis, CA Contact email: jhumann@wsu.edu, dorrie@wsu.edu

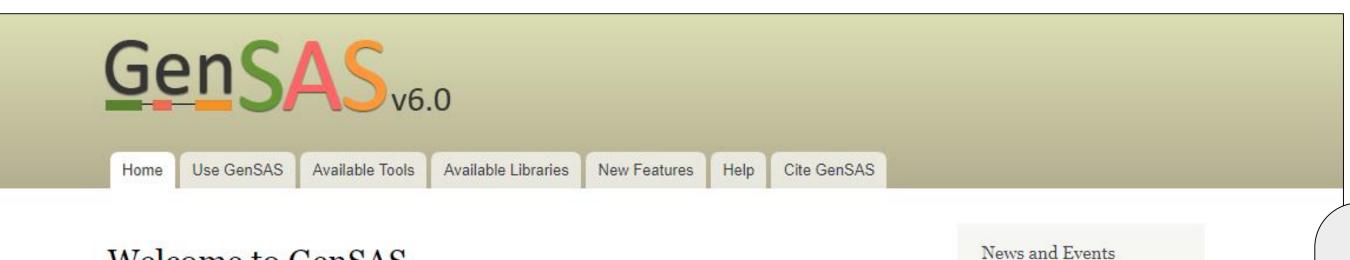
www.gensas.org

Abstract

The Genome Sequence Annotation Server v6.0 (GenSAS) is a web-based annotation platform that combines several common annotation tools into one easy-to-use, integrated resource. The annotation process is carried out in userfriendly interfaces with embedded instructions and only requires a user account and internet access. GenSAS has annotation tools for eukaryotes and prokaryotes and supports model and non-model organisms. GenSAS allows users to upload Illumina RNA-Seq reads (or specify datasets from the NCBI SRA database), align the reads to the genome using HISAT2 or TopHat2, and use the aligned data to train the gene model prediction programs AUGUSTUS and BRAKER2, which allows for more accurate gene models for eukaryotic genomes, especially non-model organisms. JBrowse and Apollo are integrated into GenSAS allowing for structural annotation results to be easily viewed and manual annotation curation to be performed. Users can share GenSAS projects with other users and manual curation can easily be a collaborative project. GenSAS also has a functional annotation step to assign protein functions and identify functional domains to the official gene set. After the annotation process is complete, the final step of the GenSAS pipeline is to generate the required files for publication and allows the user to run BUSCO on the predicted proteins to assess he completeness of the annotation.

Supported DNA evidence	Alignment tools
ESTs and transcripts	BLAT, Nucleotide BLAST, PASA
RNA-Seq reads	HISAT2, TopHat

Tools integrated into GenSAS



Going to PAG 2019? Learn

more about GenSAS

User login

Jsername

Available tools		
RepeatMasker, RepeatModeler		
Augustus, BRAKER2, GeneMarkES, Genscan, GlimmerM, SNAP		
EVidenceModeler		
Available tools		
GeneMarkS, Glimmer3		
Available tools		
Getorf		
SSR Finder		
tRNAScan-SE		
RNAmmer		
Available tools		
Protein BLAST, DIAMOND		
InterProScan		
Pfam		
SignalP		
TargetP		

User accounts ensure data is kept private,

Jobs are submitted to

computational cluster and user can log out of GenSAS and come back later

Welcome to GenSAS

The Genome Sequence Annotation Server (GenSAS) is an online platform that provides a pipeline for whole genome structural and functional annotation. Users can upload genome sequences and select from a variety of tools for repeat masking, prediction of gene models and other structural features as well as functional annotation tools. GenSAS integrates with JBrowse and Apollo to provide visualization and editing. For more details about GenSAS, please see our User's Guide.

4004-Dis.001022528.00

4804-0h-006523508-m01

Please see the information about user account and project limits before starting a GenSAS project.

progress is saved automatically and allows projects to be shared with other GenSAS users for collaborative annotation

Password *				
Create new Request ne Log in Current Project : Oryza sativa Sequences Project GF	w password	n Structural Consensus OGS Refine Functional Annotate Publish	Flowchart tracks progress through annotation project	Embedded instructions and detailed User's Guide help lead
🗆 Welcome! 🗴 🗖 Structural	Tools		Job Queue	user through
Gene Prediction			 View full report Update status 	
Other Features	✓ <u>Augustus</u>		Repeats & Masking Job Name Status	annotation project
	AUGUSTUS is a program that predicts g	enes in eukaryotic genomic sequences.	Masked Repeat Consensus Completed RepeatMasker Completed	
	Job Name		RepeatMasker Completed RepeatModeler Completed	
	Augustus		Genes & Other Predictions	
		used to distinguish between two or more Augustus jobs.	Job Name Status Augustus-At Completed	
	Species		Augustus-At Completed Augustus-proteins Completed	
	Arabidopsis thaliana Choose the species that is most closely related. *	denotes older versions for that species.	Augustus- trainedRNAseg Complete	b Queue tracks job progress and
	Report genes on		BLAST nucleotide (blastn) Complete	a Queue tracks job progress and
	both strands 🔻		BLAT Complet	llows user to access raw output
can be	Allowed gene structure		EvidenceModeler Complete	•
	allow prediction of incomplete genes at the sequence boundaries		EvidenceModeler-2 Complete GeneMarkES Complete	files from tools
iles or	Options for training Augustus		GlimmerM Complet	
	Options for training Augustus		MSU-release 7 annotation Completed	
files for	Add Augustus Job		NCBI RefSeq Completed	
	Current Project : Oryza sativa Chri			Hello GenSASuser Account Page Site Home Log out
ments)	FORMERIN	Sequences Project GFF3 Evidence Repeats Mas	sking Align Structural Consensus OGS Refine Functional A	nnotate Publish
	FGENESH	🗖 Welcome! 🛪 🗖 Project 📲 Apollo 🗙		Job Queue
	▶ CeneMarkES	t 50o57oddocof7 x Ella View Testa Usla		NC 020256 1 Completed

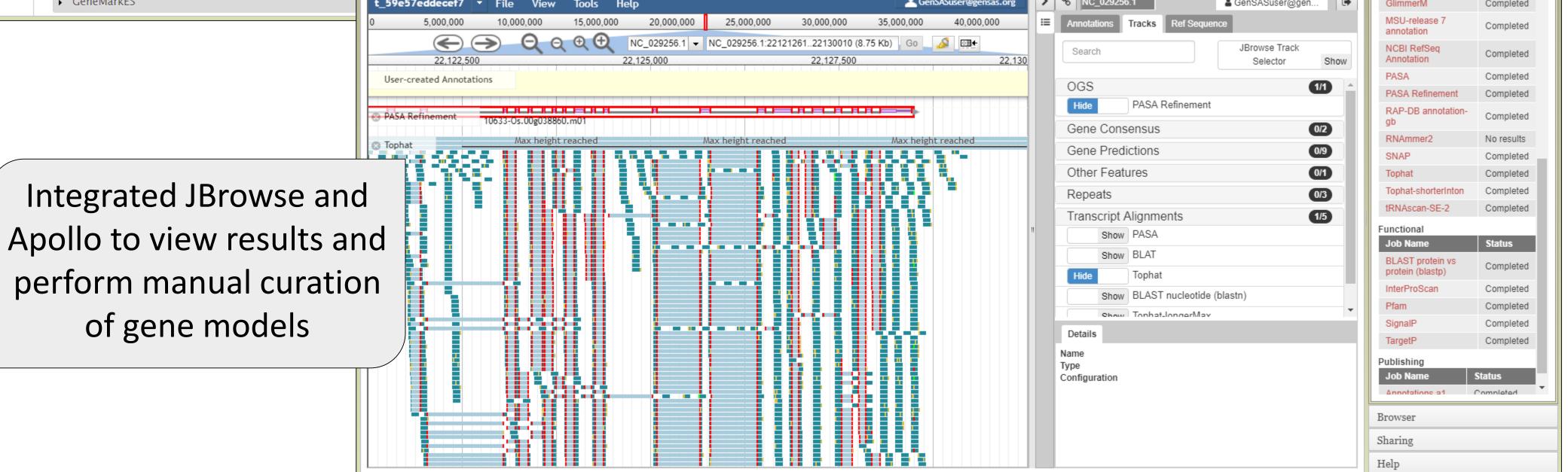
Completely web-based with no software to install

Add Auguston (siz

Tools settings are customizable and can used with user uploaded evidence files with GenSAS provided global library files repeat, transcript and protein alignmer

Manual curations from Apollo are integrated into final gene set during Publish Step

Results are exported in GFF3 and FASTA file formats which can be used with many downstream analysis programs and for publication



Funding provided by







