

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 user-friendly 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 the completeness of the annotation.

Tools integrated into GenSAS

Supported DNA evidence	Alignment tools
ESTs and transcripts	BLAT, Nucleotide BLAST, PASA
RNA-Seq reads	HISAT2, TopHat
Eukaryote structural annotation	Available tools
Repeat identification	RepeatMasker, RepeatModeler
Gene model prediction	Augustus, BRAKER2, GeneMarkES, Genscan, GlimmerM, SNAP
Gene model consensus	EvidenceModeler
Prokaryote structural annotation	Available tools
Gene model prediction	GeneMarkS, Glimmer3
Other structural features	Available tools
Open reading frames (ORFs)	Getorf
Simple sequence repeats (SSRs)	SSR Finder
tRNAs	tRNAScan-SE
rRNAs	RNAmmer
Functional annotation	Available tools
Protein homology	Protein BLAST, DIAMOND
Conserved protein signatures	InterProScan
Protein family homology	Pfam
Signal peptides	SignalP
Protein localization	TargetP



User accounts ensure data is kept private, progress is saved automatically and allows projects to be shared with other GenSAS users for collaborative annotation

Jobs are submitted to computational cluster and user can log out of GenSAS and come back later

Flowchart tracks progress through annotation project

Embedded instructions and detailed User's Guide help lead user through annotation project

Completely web-based with no software to install

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

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

Integrated JBrowse and Apollo to view results and perform manual curation of gene models

