

# GenSAS v5.1 A web-based platform for structural and functional annotation and curation

Jodi L. Humann<sup>1</sup>, Taemin Lee<sup>1</sup>, Stephen P. Ficklin<sup>1</sup>, Chun-Huai Cheng<sup>1</sup>, Heidi Hough<sup>1</sup>, Sook Jung<sup>1</sup>,  
Jill Wegrzyn<sup>2</sup>, David Neale<sup>3</sup>, and Dorrie Main<sup>1</sup>

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

[www.gensas.org](http://www.gensas.org)

## Abstract

The Genome Sequence Annotation Server v5.1 (GenSAS) is a web-based annotation and curation platform that combines several common annotation tools into one easy-to-use, integrated resource. The user-friendly interfaces, with embedded instructions, guide users through the annotation process. GenSAS has annotation tools for eukaryotes and prokaryotes and supports model and non-model organisms. Users can upload a variety of evidence files to support the annotation process for their genome sequence. These include GFF3 files of aligned features and previous annotations; FASTA files of repeat, transcript, EST, or protein sequences; and gene models from Genbank. GenSAS also allows users to upload Illumina RNA-Seq reads, align the reads to the genome using TopHat, and use the data to train the gene model prediction program Augustus, which allows for more accurate gene models for eukaryotic genomes, especially non-model organisms. JBrowse and Apollo are integrated into GenSAS allowing structural annotation results to be easily viewed and manual curation to be performed. Users can share GenSAS projects with other users enabling collaborative or community wide curation. GenSAS also has a functional annotation step to assign protein functions and identify functional domains for the official gene set. After the annotation process is complete, the final step of the GenSAS pipeline generates the required files for publication which includes merging the manual annotations from Apollo into the final annotation.

## Tools integrated into GenSAS

Supported DNA evidence	Alignment tools
ESTs and transcripts	BLAT, Nucleotide BLAST, PASA
RNA-Seq reads	TopHat
Eukaryote structural annotation	Available tools
Repeat identification	RepeatMasker, RepeatModeler
Gene model prediction	Augustus, 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
Conserved protein signatures	InterProScan
Protein family homology	Pfam
Signal peptides	SignalP
Protein localization	TargetP

**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**

**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**

**Job Queue tracks job progress and allows user to access raw output files from tools**

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

Funding provided by

