An easy to use, web-based platform for individual and collaborative structural and functional genome annotation

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

dorrie@wsu.edu
What is DNA annotation and why do it?

- Getting the DNA sequence is only the first step
- Need to know the biological relevance of the DNA sequence
- Annotated sequence can be used to find putative genes of interest for study
Step 1: Nucleotide level
Genes, ORFs, genetic markers, tRNA, rRNA, ncRNA, repeats, regulatory elements

Step 2: Protein level
Translate genes and ORFs into proteins, search for homologs, assign putative function

Step 3: Process level
Assign GO terms, do lab experiments (mutagenesis, transcriptomics, RNA silencing, etc.)
What scientists want

• Current annotation tools:
  • Many tools available, but run independently of each other
  • Most of the tools are run via the command line and require server access

• Scientists want a platform that:
  • Is a single location for DNA annotation
  • Does not require management of computing equipment and software tools
  • Is easy to use and can be adapted to a variety of DNA sequences
What is GenSAS?

• A single website that combines numerous annotation tools into one interface

• User accounts keep data private and secure as well as allow users to share data for collaborative annotation

• Easy-to-use interfaces, with integrated instructions allow researchers at all skill levels to annotate DNA
Welcome to GenSAS

The Genome Sequence Annotation Server (GenSAS) is an online tool 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. Please see our video tutorial under the "Help" tab.

Beta-testing GenSAS v5.0, to be released August
GenSAS welcome tab provides users with a quick overview of what each of the three screen sections do.
• **Sequence Tab:**
  - Single sequence or multi-sequence FASTA file
  - Please make sure your assembly is good quality
  - New to v5.0, users can upload a multi-sequence FASTA file and create a subset based on sequence names or minimum size
  - If no subset is created all sequences in multi-sequence FASTA file are analyzed with the same parameters

• **Project Tab:**
  - Open existing project or shared project
  - Create new project
• GFF3 Tab (optional):
  • Previous annotations
  • Output from other tools
• Evidence Tab (optional):
  • EST, mRNA sequences
  • Repeat motifs
  • Protein sequences
  • NCBI gene structures for organism
  • Pre-processed Illumina RNA-Seq reads

The more organism specific data you have, the better the annotation will be
• Repeats Tab:
  • RepeatMasker – evidence based repeat finder
  • RepeatModeler – *de novo* repeat finder
  • Can run each tool multiple times with different parameters by changing job name

• Masking Tab:
  • Look at the results in JBrowse and choose which set(s) to use in the consensus
  • Masked consensus is then used as input for the annotation tools unless the user elects to skip repeat masking
• Align Tab:
  • New in v5.0, Align step has been added to allow users to align RNA-Seq data for training the gene prediction programs
  • Also can align full-length transcripts and proteins

• Structural Tab:
  • Gene prediction programs
  • SSR Finder
  • tRNAscanSE

All tabs have an Instructions section that can be opened and collapsed
• **OGS Tab:**
  • New to v5.0, Official Gene Set tab allows user to designate gene model set for manual annotation process and final publication
  • Use previous annotation, output from single gene predictor or generate consensus using EvidenceModeler

• **Refine Tab:**
  • Use PASA and RNA evidence to refine OGS gene models
- Job status can be monitored through Job Queue
- Progress through GenSAS is automatically saved
- Users can log off GenSAS and jobs will continue running
- While jobs are running, users can look at the completed results in Apollo/JBrowse
- Once the project has results, users can share the project with other GenSAS users for collaborative annotation
• Functional Tab:
  • OGS gene models and other user selected gene models are functionally annotated
• Manual annotation from Apollo are automatically merged into OGS at Publish Step

GenSAS exports data in GFF3 and FASTA formats