Genome Annotation

Day 2

Goals of Annotation

- Find transposable elements
- Find protein coding genes
  - Exons
  - Start/Stop/Splice sites
  - Introns
- Find regulatory elements

- Find non coding RNA elements

- Name and classify results
  - Putative gene name
  - Functional identification
  - Pathway identification

- EST annotation

The first step…

Repeat Masking

- Speeds up gene predictions

BUT…

- Tandem array elements can be part of genes

THEREFORE…

- Mask only Interspersed repeats (i.e. TEs)

The next step…

Prediction Methods

- Homology based (map to genome)
  - Known proteins (or repeats)
  - ESTs

- Pattern matching

Exonerate
EST2Genome
GeneWise
GeneMapper
BLAST
RepeatMasker

Prediction Methods

- De novo
  - Prediction models
  - Conserved regions
  - ORFs

- Commonly Markov models

Zentrum für Bioinformatik
ZBH

Augustus
GeneMark
Eugene
GlimmerHMM
LTRharvest

Training Data

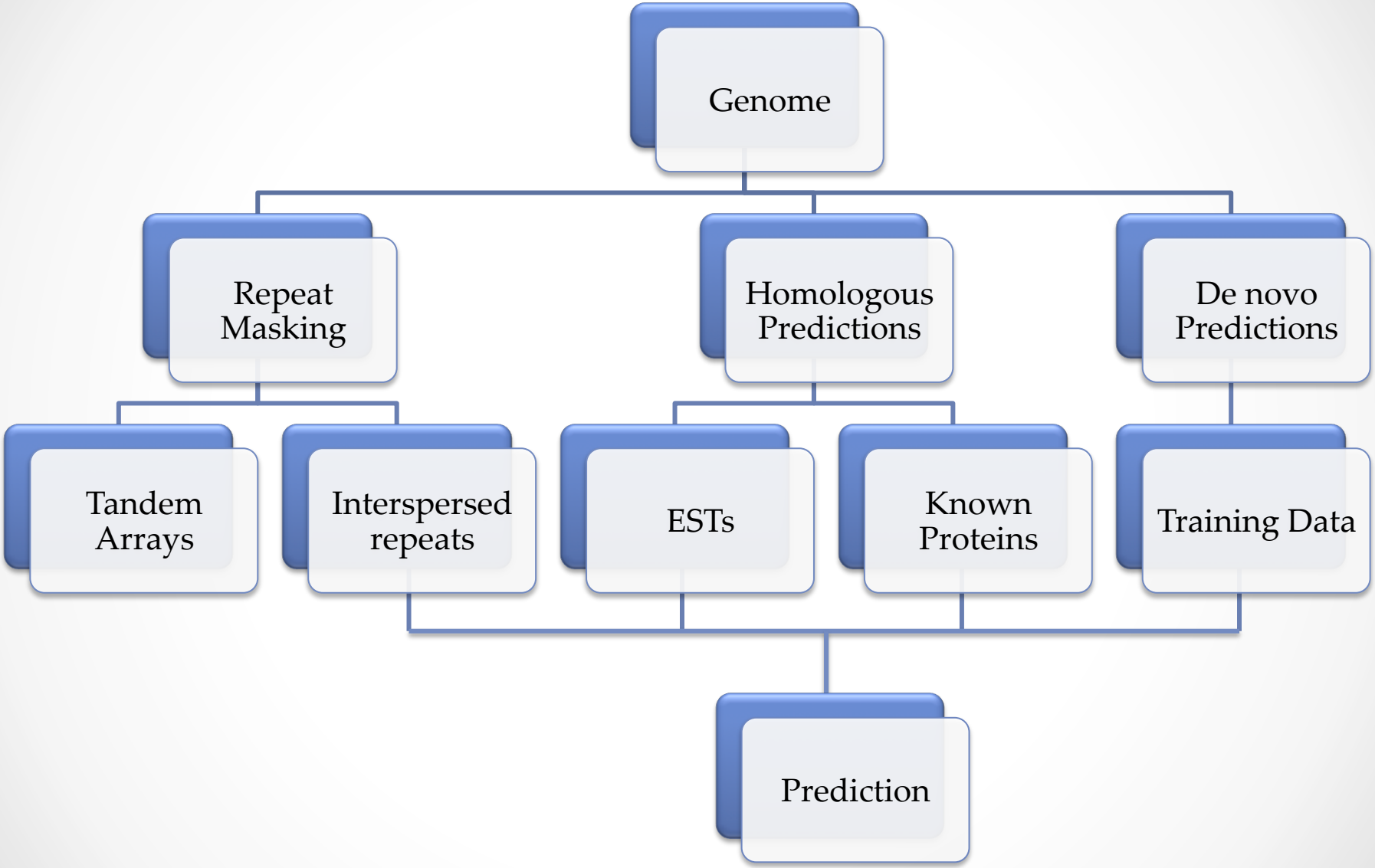
- Well annotated genes mapped to the genome
  - Homologous genes
  - Experimentally verified
  - Manually Annotated

Unfortunately…

- Novel genomes do not come with training data

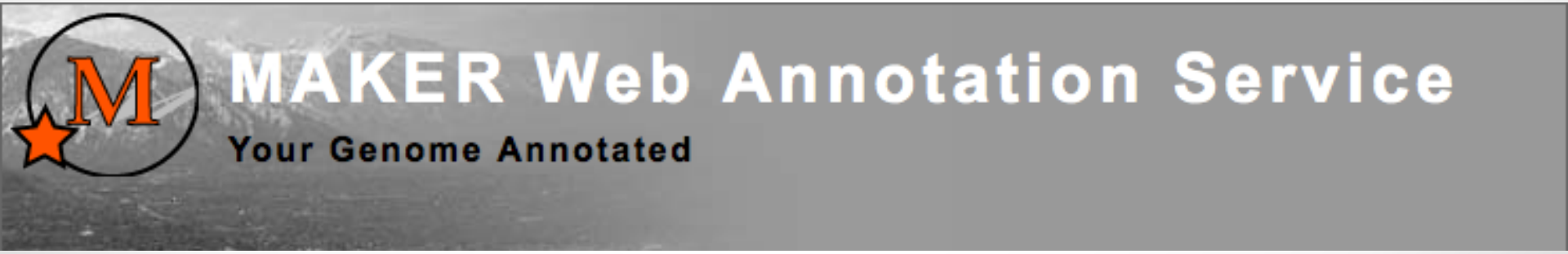
- CEGMA?

The final step…


Genome
Repeat Masking
Homologous Predictions
De novo Predictions
Tandem Arrays
Interspersed repeats
ESTs
Known Proteins
Training Data
Prediction

Automated Annotation

- Programs that combine various sources of information
- MAKER (MWAS)
- CEGMA



But first;

choose your scaffolds to annotate

MAKER Web Annotation Service

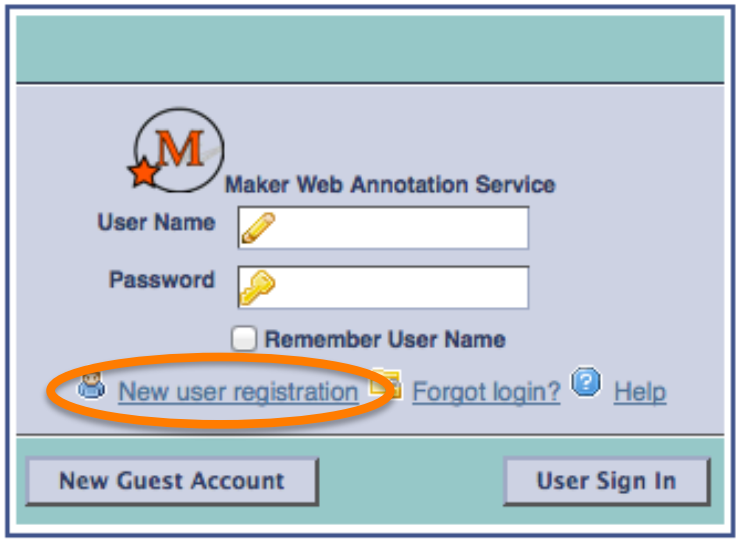
Your Genome Annotated

Home | Help | Yandell Lab

not logged-in | sign in

Welcome to the MAKER Web Annotation Service

Log into your account below, or you can access the server as a guest. While there is no login requirement for this site, users are highly encouraged to set up an account. Use the "New user registration" link to register a new account. Registration is free, and has several benefits. Registered users can submit up to 5,000,000 base pairs of sequence for each annotation job. Guest users are limited to 500,000 base pairs per annotation job submission. Registered users receive e-mail notifications as to the status of their jobs, have persistance of results on the server, and their jobs/results are protected by extra layers of security.



Valid CSS/XHTML 1.0

- Denovo Annotation
- EST Evidence
- Protein Homology Evidence
- Configure Repeat Masking
- Choose Your Gene-predictor(s)
- Already Have Your Own Gene Models/Ab-inito Predictions?
- Annotation Properties

Welcome to the MAKER Web Annotation Service (MWAS)

To get started just click on "New Job" above. You can then submit a sequence for annotation or select from a list of pre-loaded example annotation jobs. Once a job has been added to the queue you can see your job's run status as well as the results below. You can also see your jobs position in the execution queue relative to other user's jobs by clicking "Job Queue" above. For more infomation on using the MAKER Web Annotation Service, click on "Help" above.

Refresh Job Status

Your Jobs (1)

JobID	Description	Job Status	Start Time	Finish Time	Log	View Results	
6664	Sinvicta good scaffolds	results ready	06/10/2012 06:05	06/10/2012 09:38			

Good Points

- Fast
- Discover novel genes
- Allows for bulk annotation

Bad Points

- Perpetuates errors
- Low accuracy
- Don’t always know exactly what has been done

Only proteins recognizable as proteins

Predictions are frequently biased towards shorter proteins

Short exons/genes can be missed

Multiple sources of predictions are needed for some genes, others are better with fewer sources

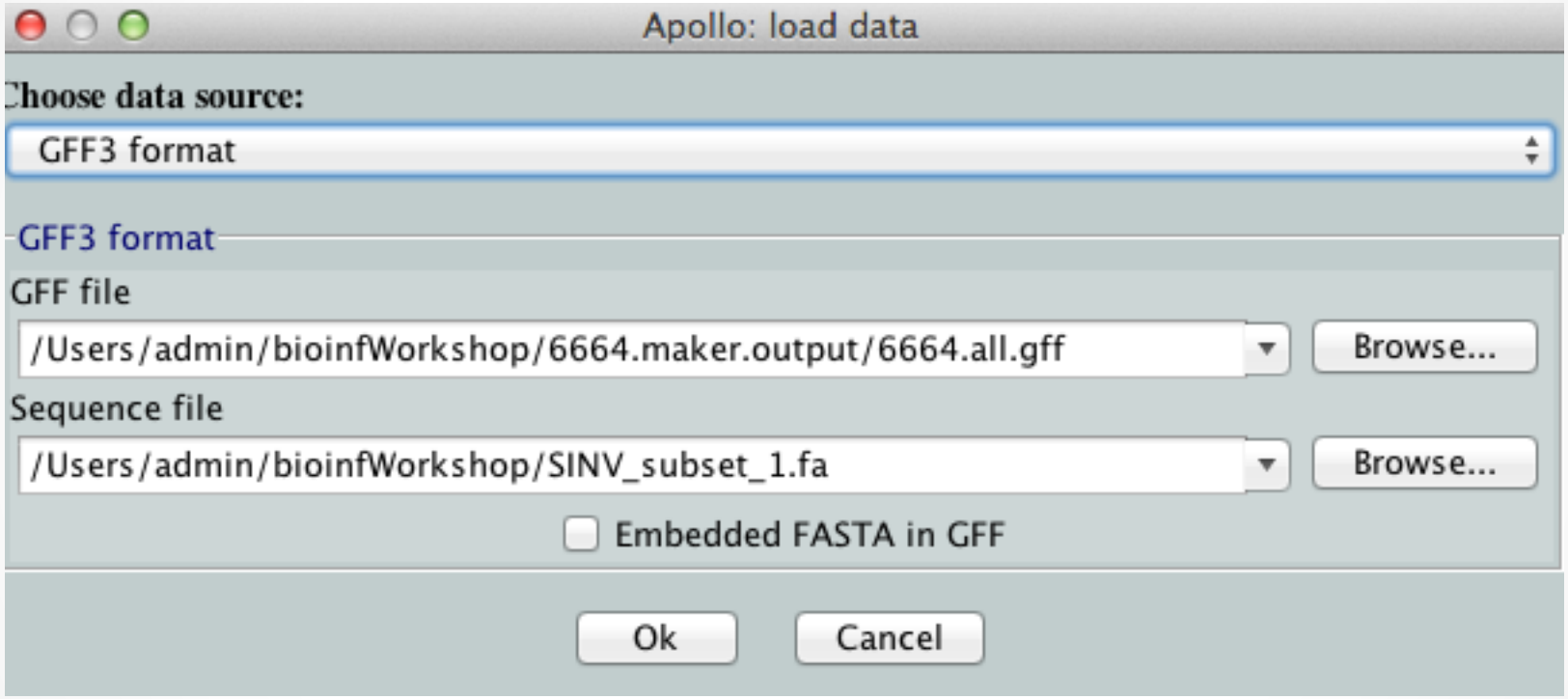
Protein coding genes only

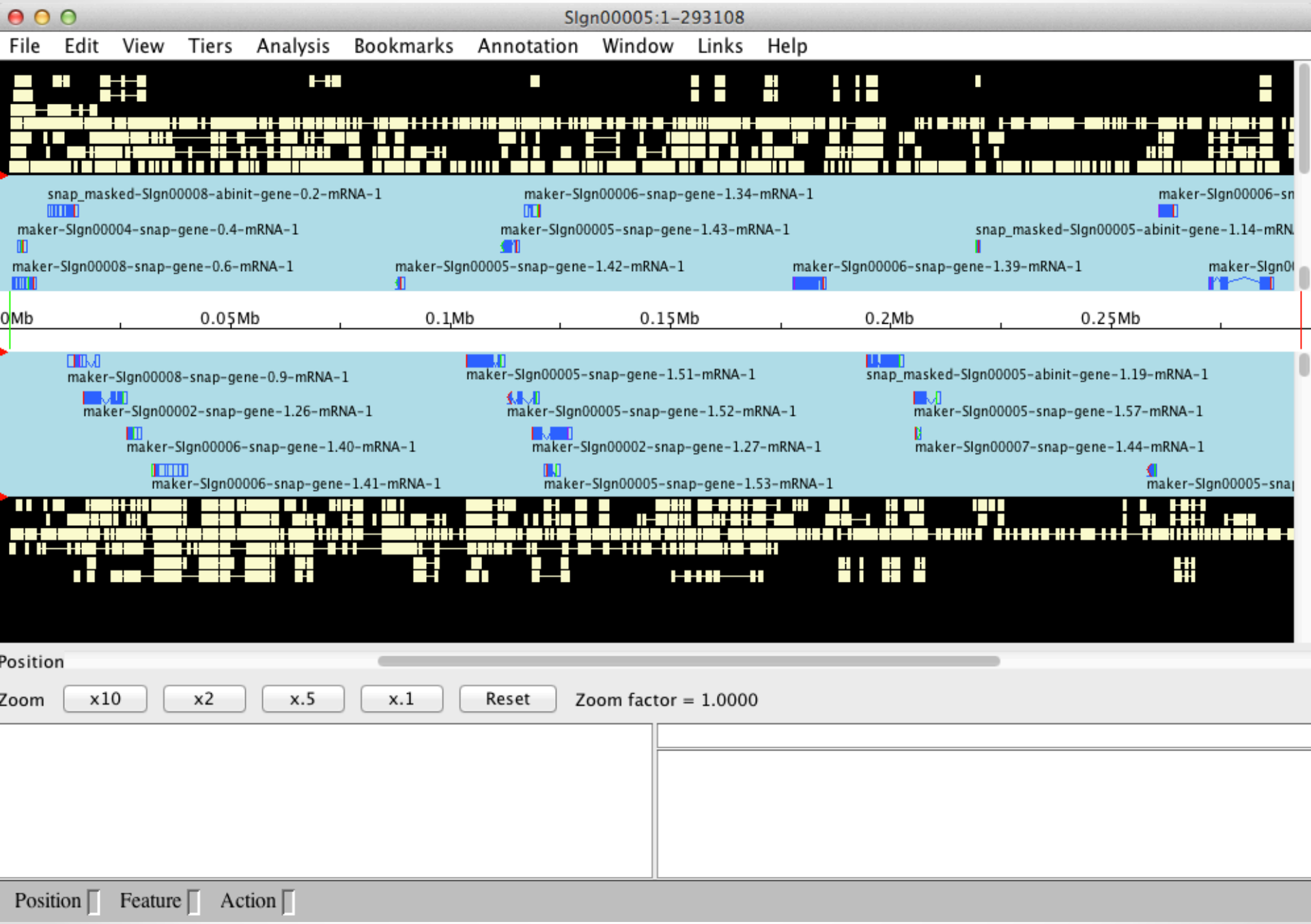
Some programs are designed for sensitivity and some specificity

The bias towards fragmented genes or chimeric genes depends on the software

Manual Annotation

- Visualization and editing tools
- Apollo
- Integrative Genomics Viewer
- Genome Browser


Apollo: load data
Choose data source:
GFF3 format
GFF3 format
GFF file
/Users/admin/bioinfWorkshop/6664.maker.output/6664.all.gff
Browse...
Sequence file
/Users/admin/bioinfWorkshop/SINV_subset_1.fa
Browse...
Embedded FASTA in GFF
Ok
Cancel


SIgn00005:1-293108
File
Edit
View
Tiers
Analysis
Bookmarks
Annotation
Window
Links
Help
snap_masked-SIgn00008-abinit-gene-0.2-mRNA-1
maker-SIgn00006-snap-gene-1.34-mRNA-1
maker-SIgn00006-sn
maker-SIgn00004-snap-gene-0.4-mRNA-1
maker-SIgn00005-snap-gene-1.43-mRNA-1
snap_masked-SIgn00005-abinit-gene-1.14-mRN
maker-SIgn00008-snap-gene-0.6-mRNA-1
maker-SIgn00005-snap-gene-1.42-mRNA-1
maker-SIgn00006-snap-gene-1.39-mRNA-1
maker-SIgn0
0Mb
0.05Mb
0.1Mb
0.15Mb
0.2Mb
0.25Mb
maker-SIgn00008-snap-gene-0.9-mRNA-1
maker-SIgn00005-snap-gene-1.51-mRNA-1
snap_masked-SIgn00005-abinit-gene-1.19-mRNA-1
maker-SIgn00002-snap-gene-1.26-mRNA-1
maker-SIgn00005-snap-gene-1.52-mRNA-1
maker-SIgn00005-snap-gene-1.57-mRNA-1
maker-SIgn00006-snap-gene-1.40-mRNA-1
maker-SIgn00002-snap-gene-1.27-mRNA-1
maker-SIgn00007-snap-gene-1.44-mRNA-1
maker-SIgn00006-snap-gene-1.41-mRNA-1
maker-SIgn00005-snap-gene-1.53-mRNA-1
maker-SIgn00005-snap
Position
Zoom
x10
x2
x.5
x.1
Reset
Zoom factor = 1.0000
Position
Feature
Action

Good Points

- Accurate
- Know exactly what has been done

Bad Points

- Very Slow
- Can’t work in bulk
- Human bias