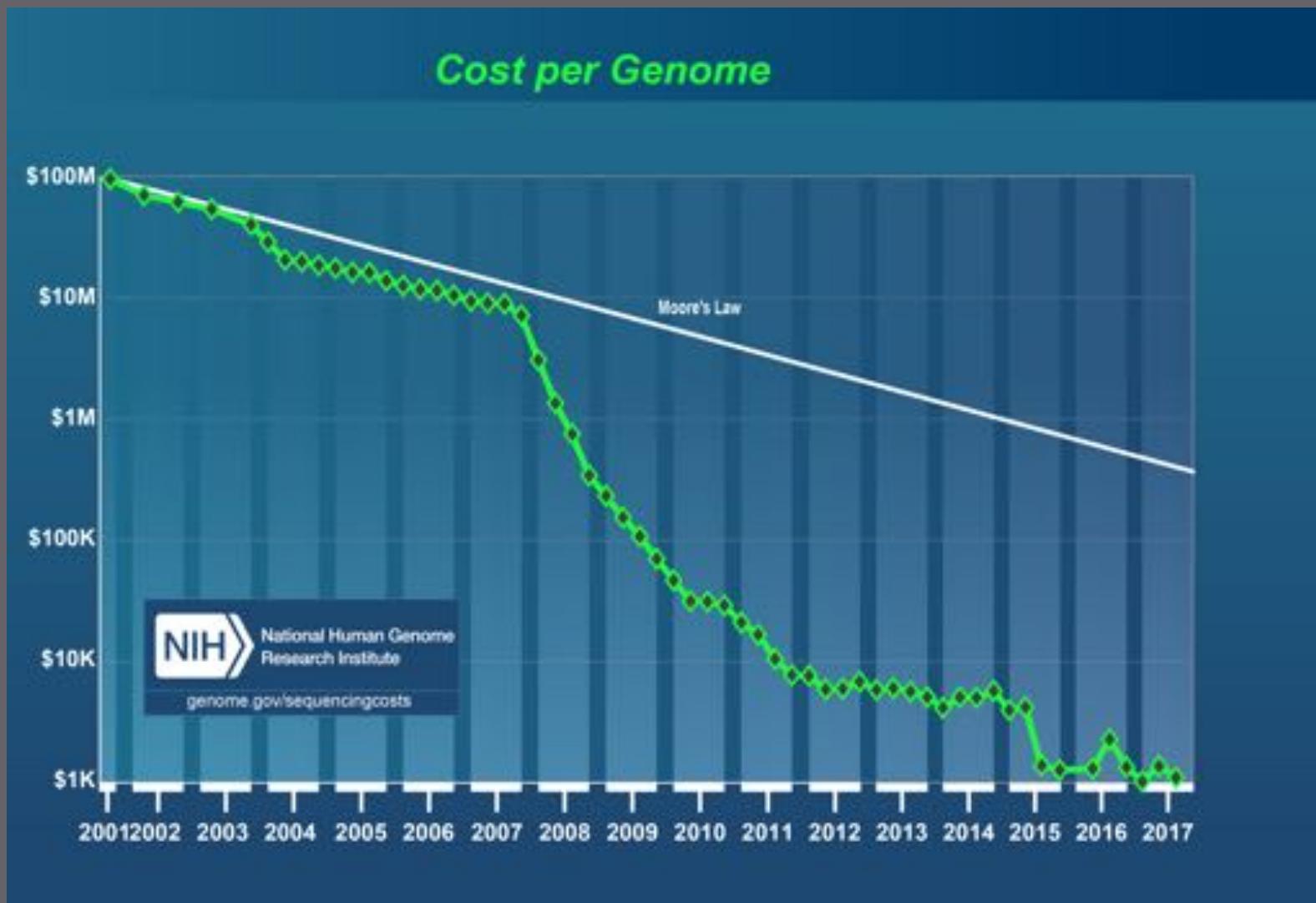


# I5K WEBINAR: AUTOMATED GENOME ANNOTATION AND ANALYSIS

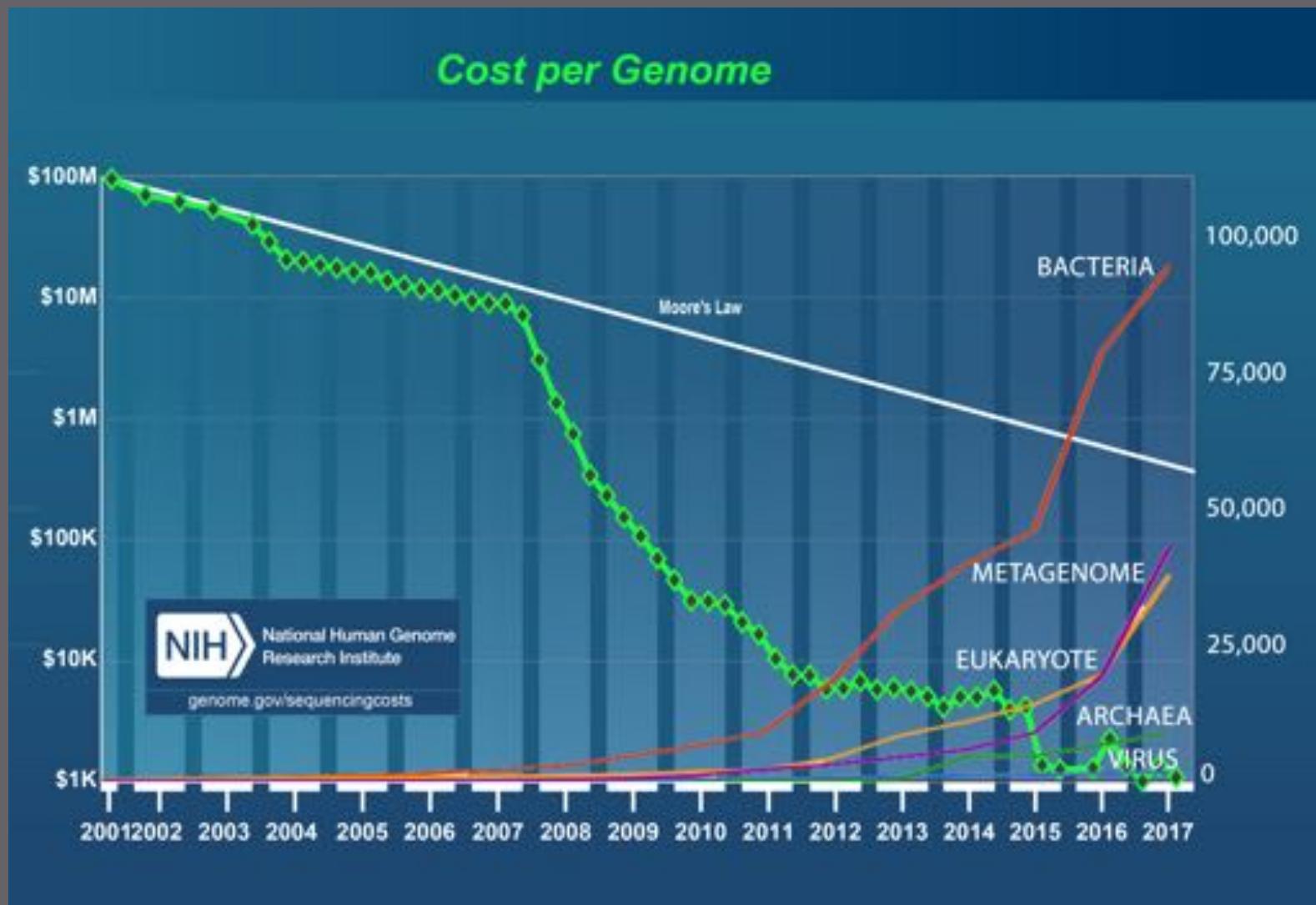
Carson Holt  
Yandell Lab

USTAR Center for Genetic Discovery  
University of Utah

# Advances in Second-Generation Technology are Making Whole Genome and Transcriptome Sequencing “Routine” Even for Small Labs



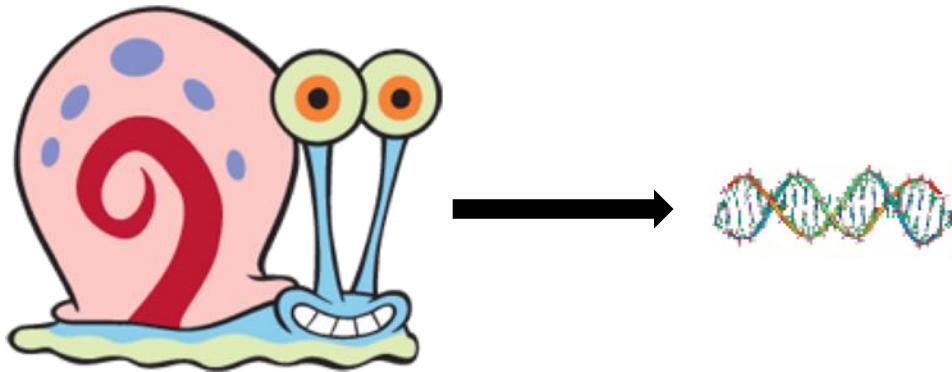
# Advances in Second-Generation Technology are Making Whole Genome and Transcriptome Sequencing “Routine” Even for Small Labs



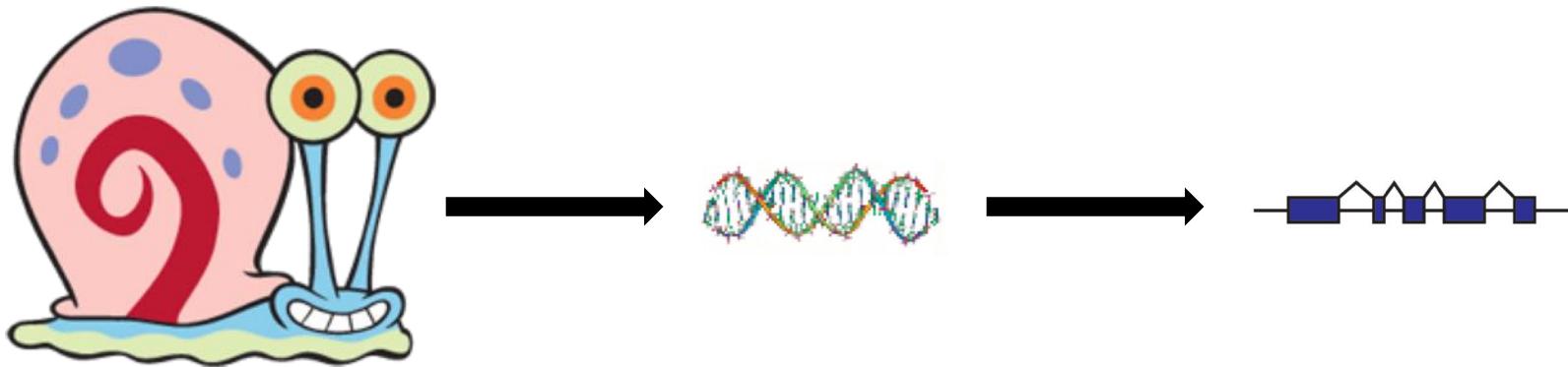
# Genome Project Overview



# Genome Project Overview

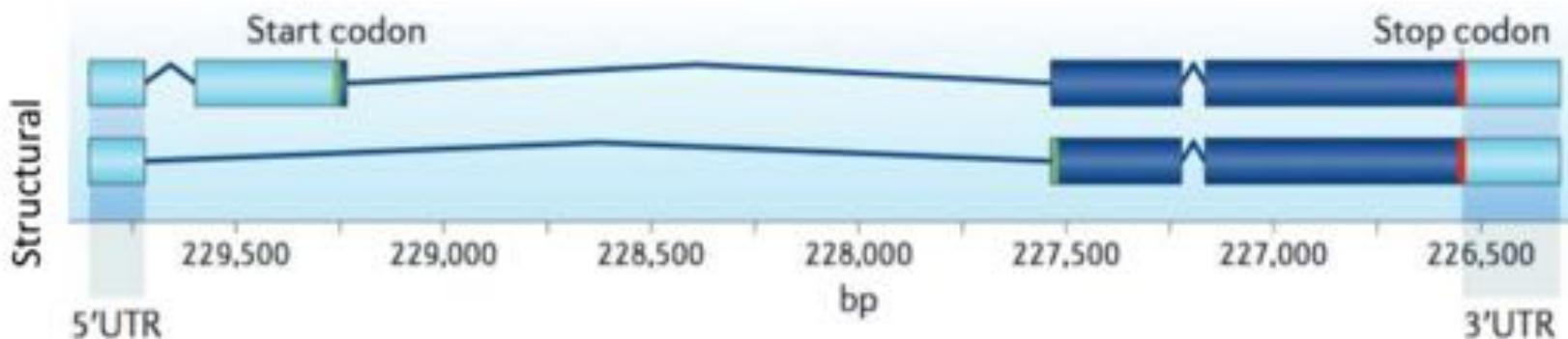


# Genome Project Overview



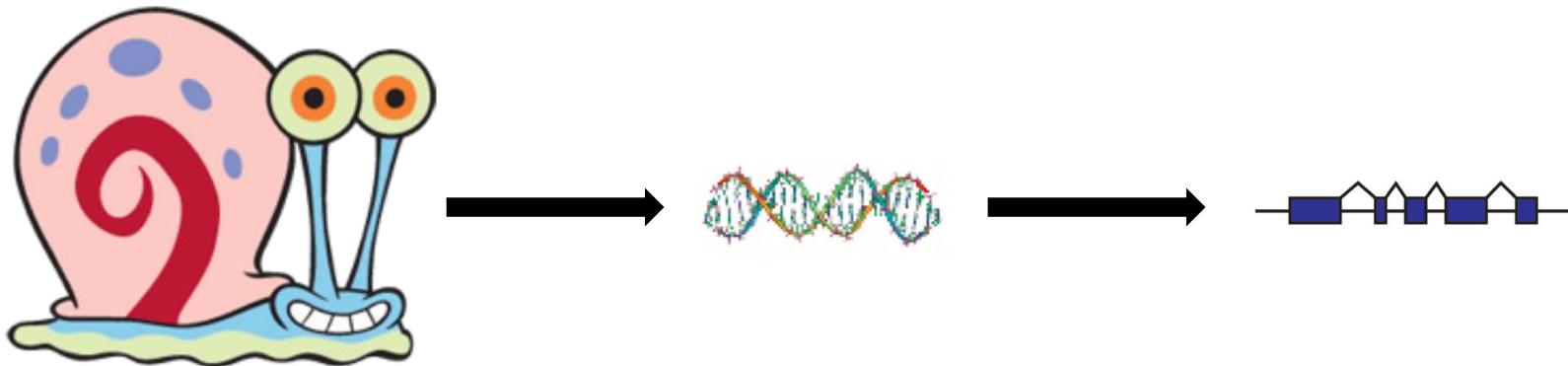
# Genome Project Overview

## What is an Annotation?

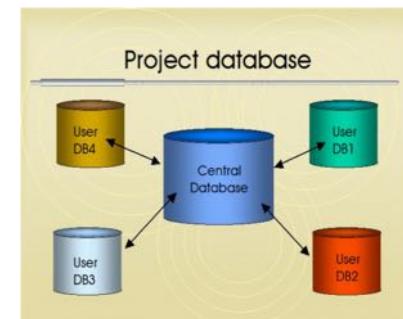
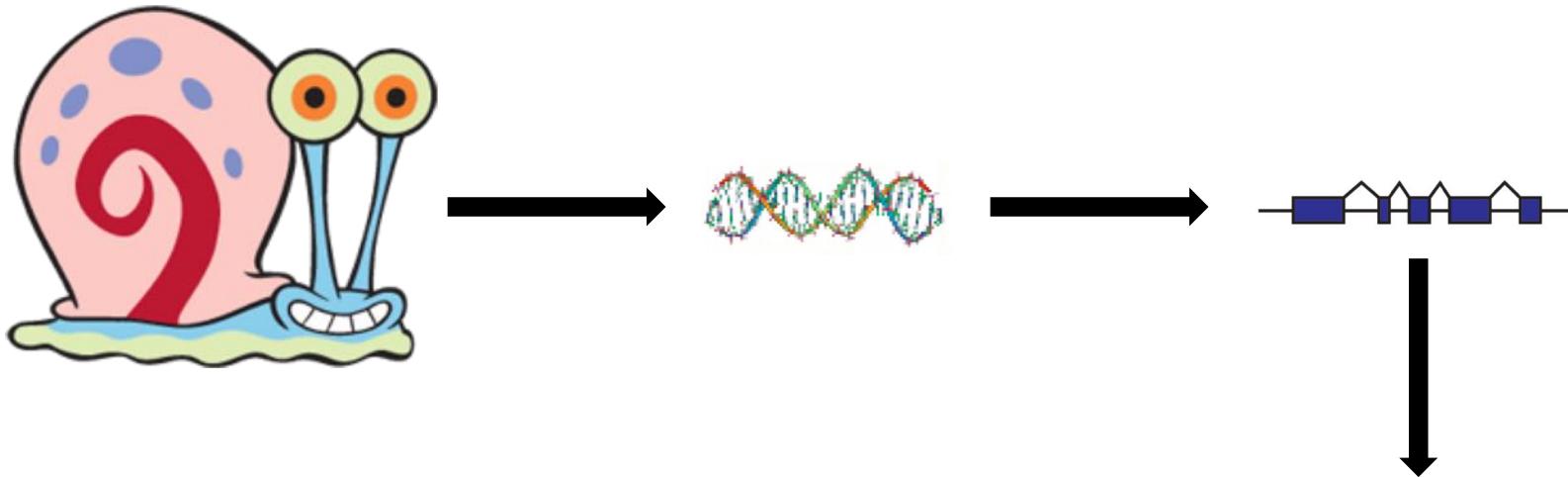


Functional	
<b>Function</b>	cAMP-dependent and sulfonylurea-sensitive anion transporter. Key gatekeeper influencing intracellular cholesterol transport.
<b>Subcellular location</b>	Membrane; Multi-pass membrane protein Ref.13 Ref.14.
<b>Domain</b>	Multifunctional polypeptide with two homologous halves, each containing a hydrophobic membrane-anchoring domain and an ATP binding cassette (ABC) domain.

# Genome Project Overview



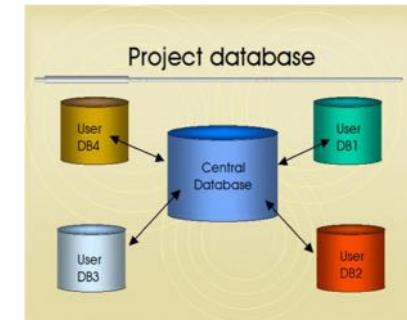
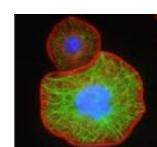
# Genome Project Overview



# Genome Project Overview



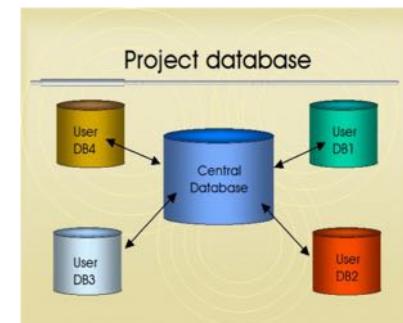
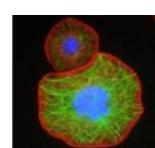
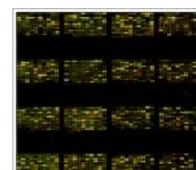
>Smg5  
MEVTFSSGGSSNASSECAIDGGTNRCGL  
EPNNGTCILSQEVKDLYRSLYTAKQLDD  
AKRNVQSVGQLFQHEIEEKRSLLVQLCKQ  
IIFKDYQSVGKKREV MWRRGYYEFIAFV



# Genome Project Overview



SUCCESS



# Genome Project Overview

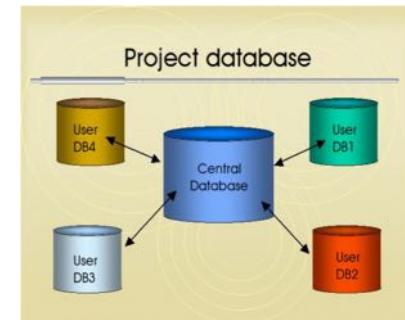
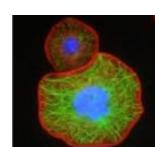
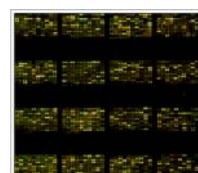


Incorrect annotations poison every experiment  
that uses them!!

FAILURE



```
>Smg5
MEVTSSGGSSNASSECAIDGGTNRCGL
EPNNGTCILSQEVKDLYRSLYTAKQLDD
AKRNVQSVGQLFQHEIEEKRSLLVQLCKQ
IIFKDYQSVGKKREV MWRRGGYYEFIAFV
```





# MAKER

an annotation pipeline and genome-database management  
tool for second-generation genome projects

# Easy-to-use by design

# Easy-to-use by design

<b>User Requirements:</b>	Can be run by a single individual with little bioinformatics experience
---------------------------	---

# Easy-to-use by design

<b>User Requirements:</b>	Can be run by a single individual with little bioinformatics experience
<b>System Requirements:</b>	Can run on laptop or desktop computers (running Linux or Mac OS X)

# Easy-to-use by design

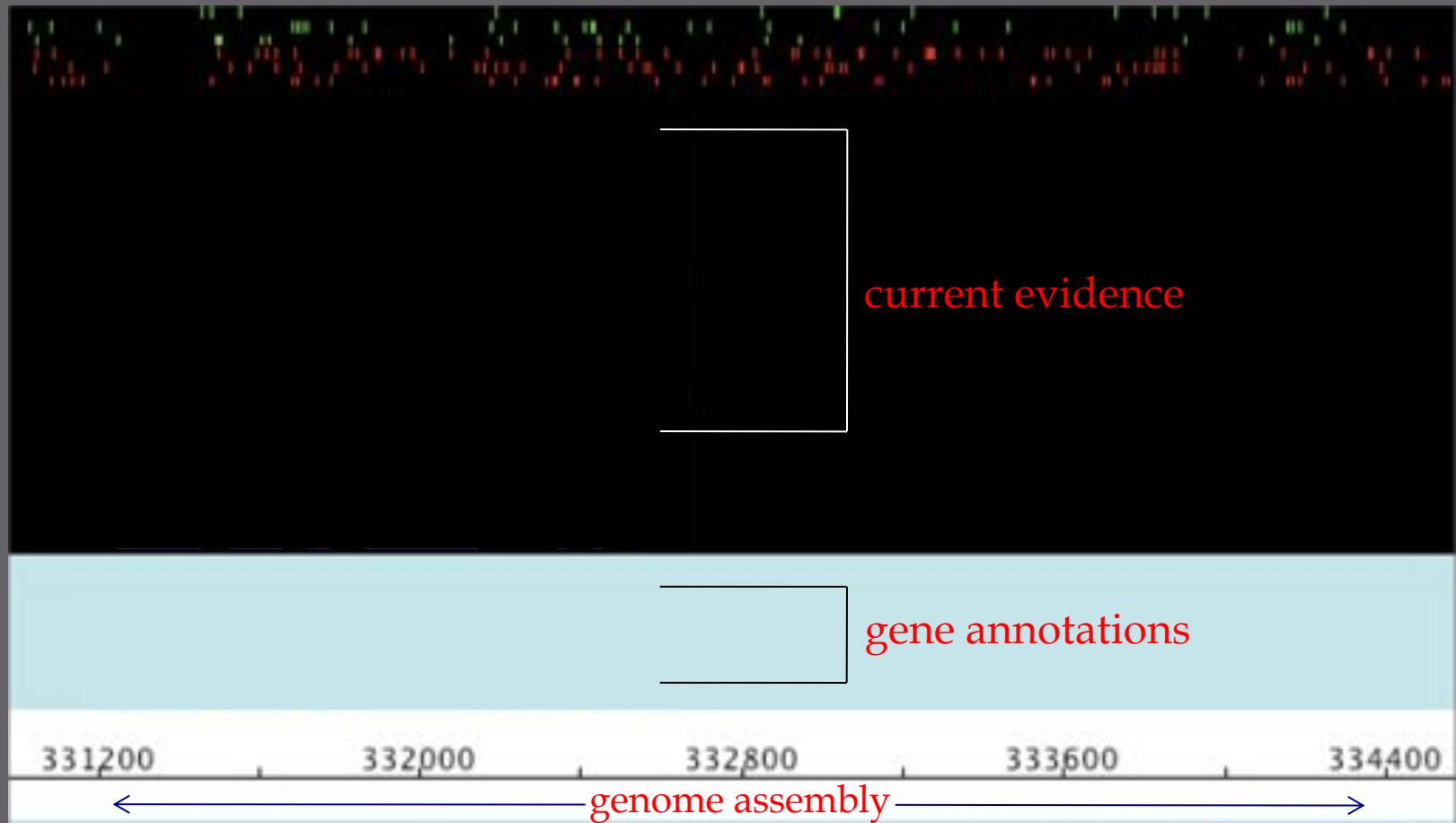
<b>User Requirements:</b>	Can be run by a single individual with little bioinformatics experience
<b>System Requirements:</b>	Can run on laptop or desktop computers (running Linux or Mac OS X)
<b>Program Output:</b>	Output is compatible with popular annotation tools like Apollo, GBrowse, and JBrowse

# Easy-to-use by design

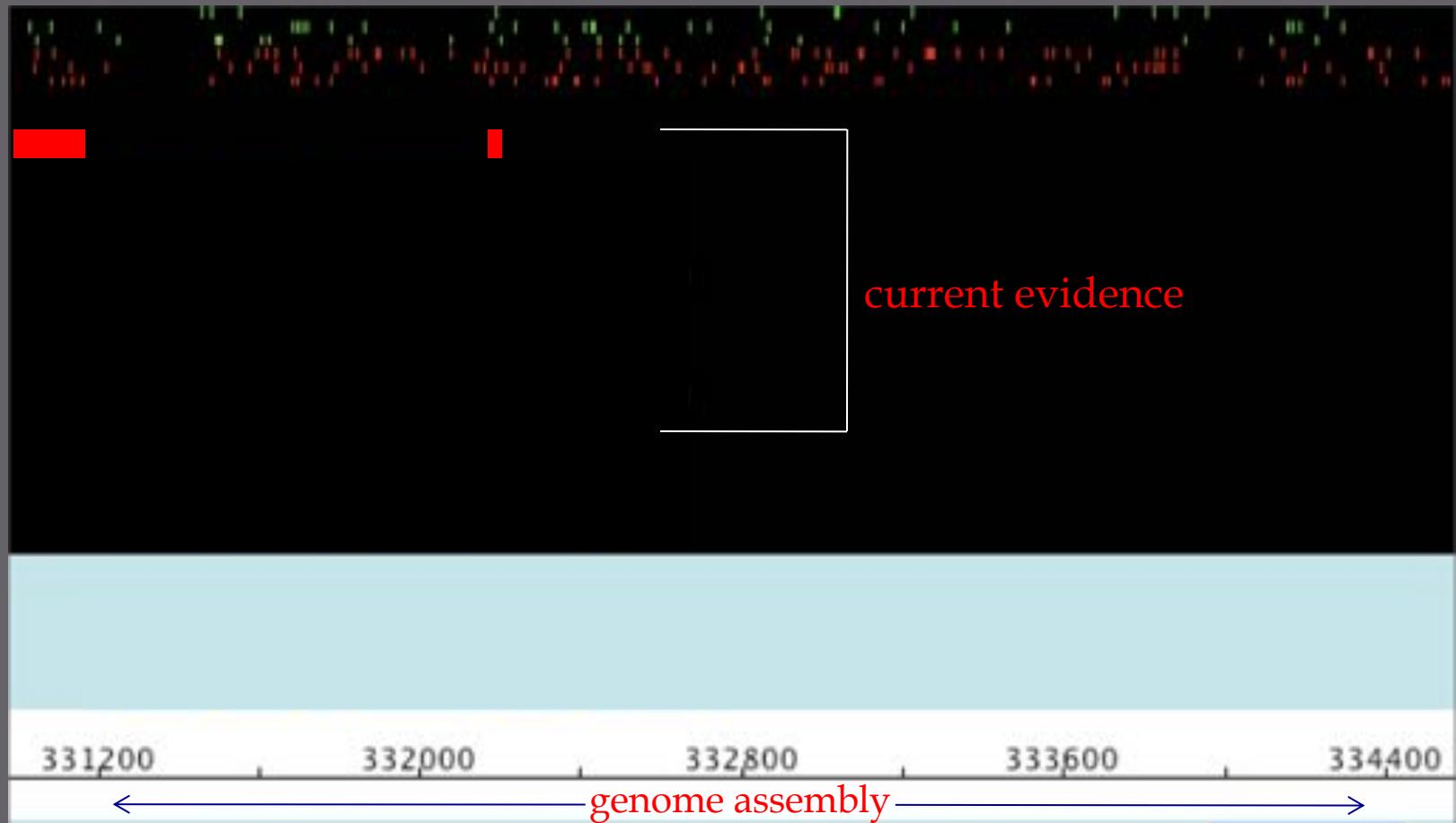
<b>User Requirements:</b>	Can be run by a single individual with little bioinformatics experience
<b>System Requirements:</b>	Can run on laptop or desktop computers (running Linux or Mac OS X)
<b>Program Output:</b>	Output is compatible with popular annotation tools like Apollo, GBrowse, and JBrowse
<b>Availability:</b>	Free open source application (for the academic community)

# How does MAKER work?

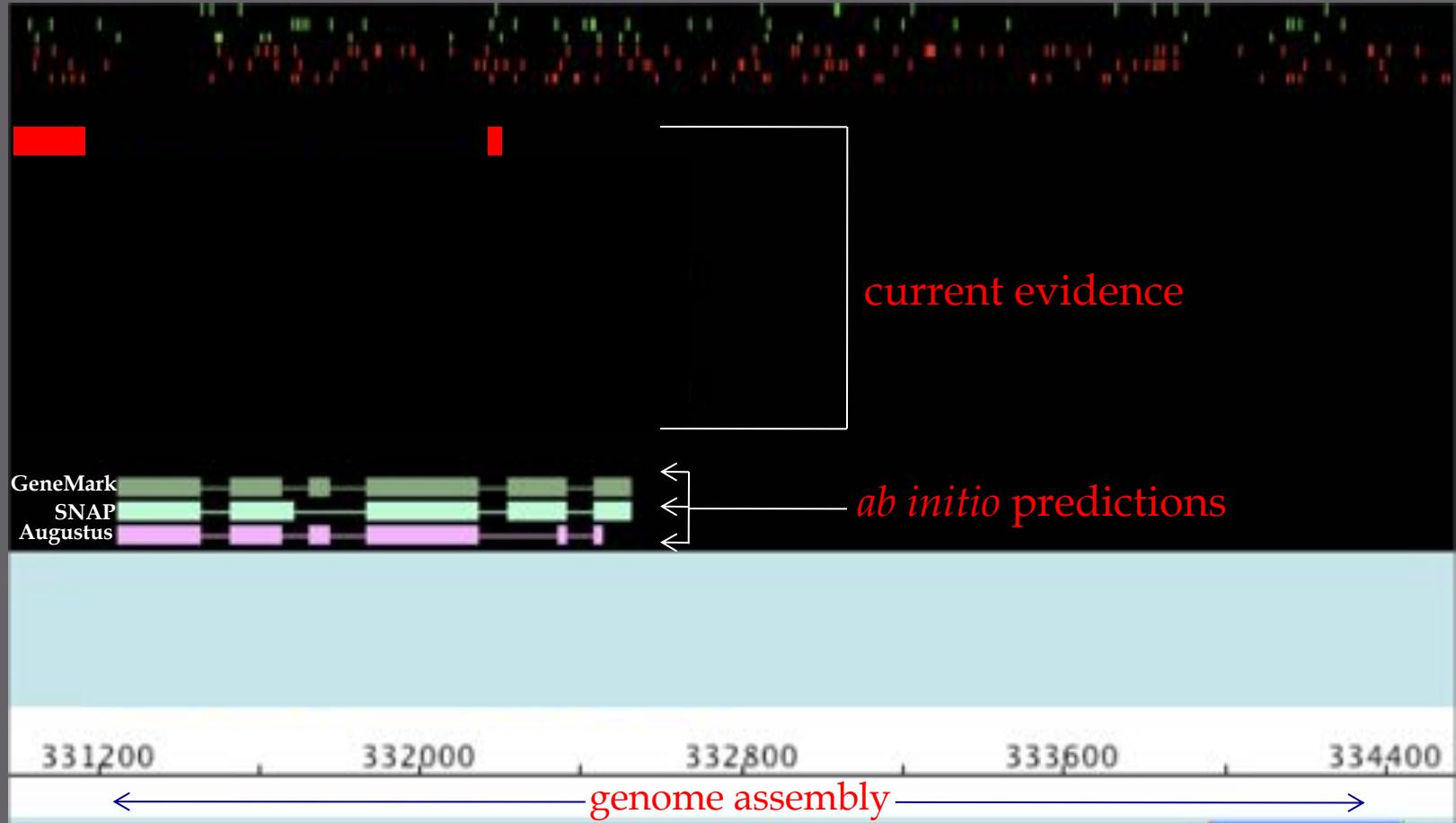
# Annotating the Genome – Apollo View



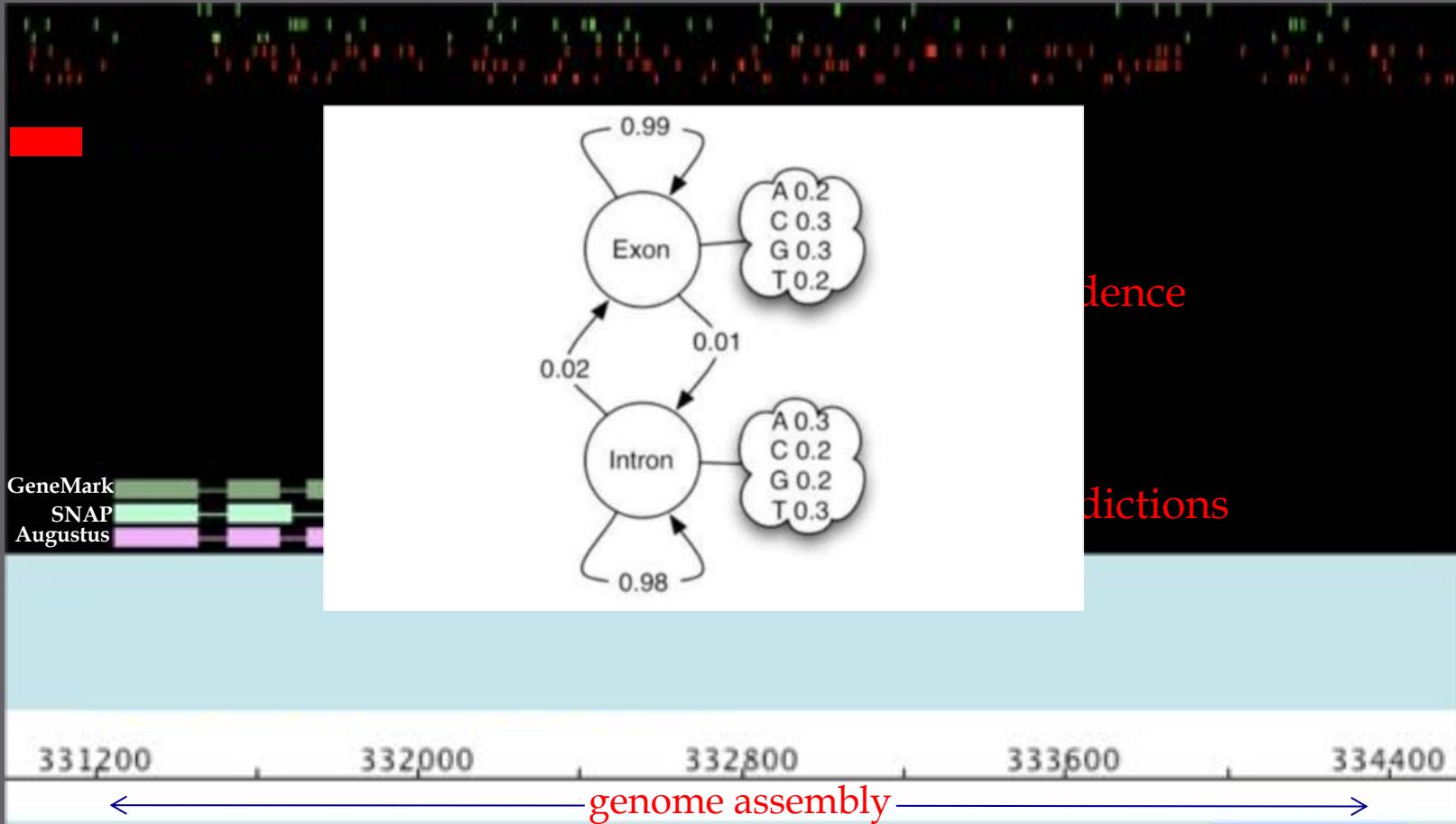
# Identify and mask repetitive elements



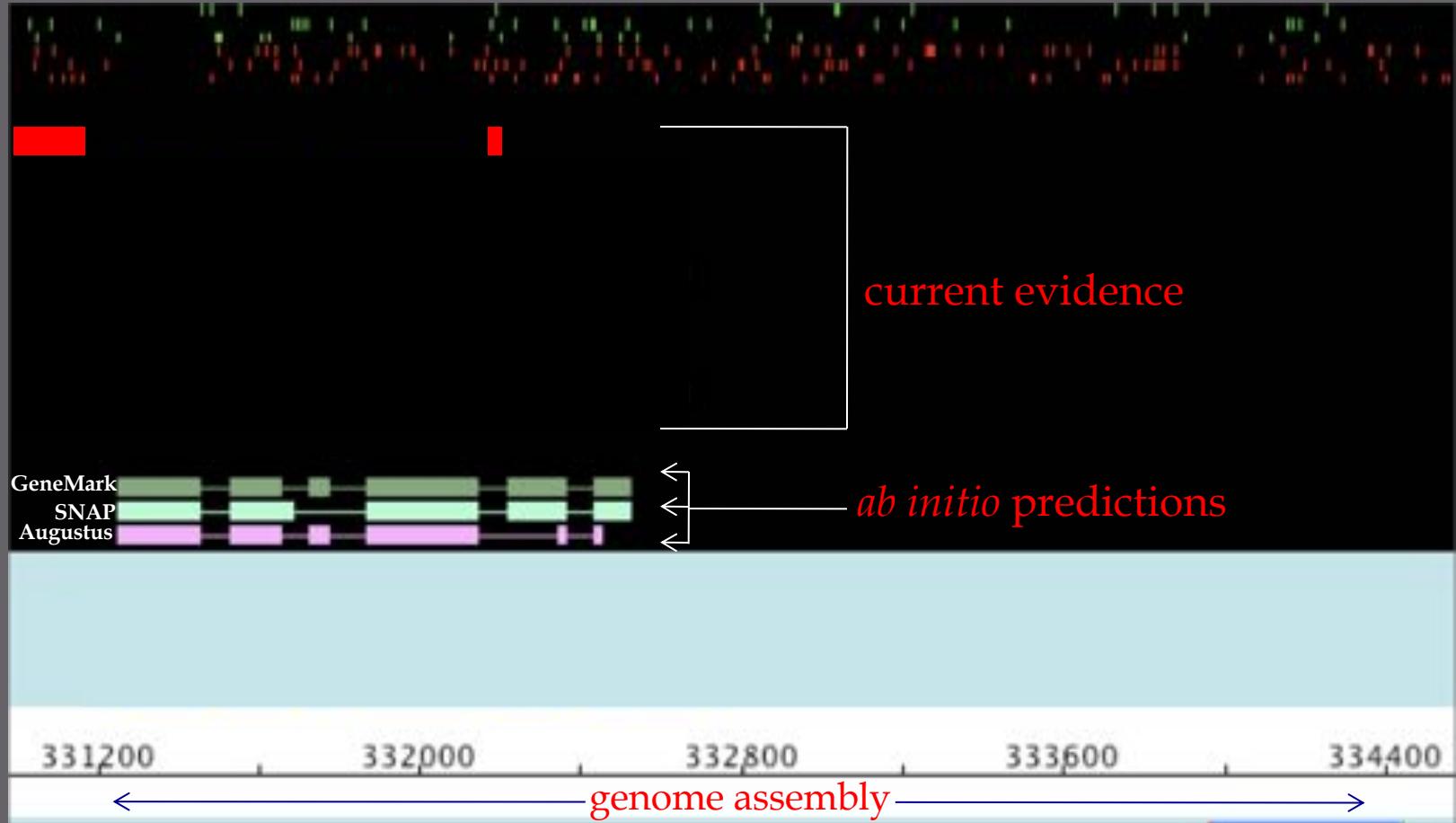
# Generate *ab initio* gene predictions



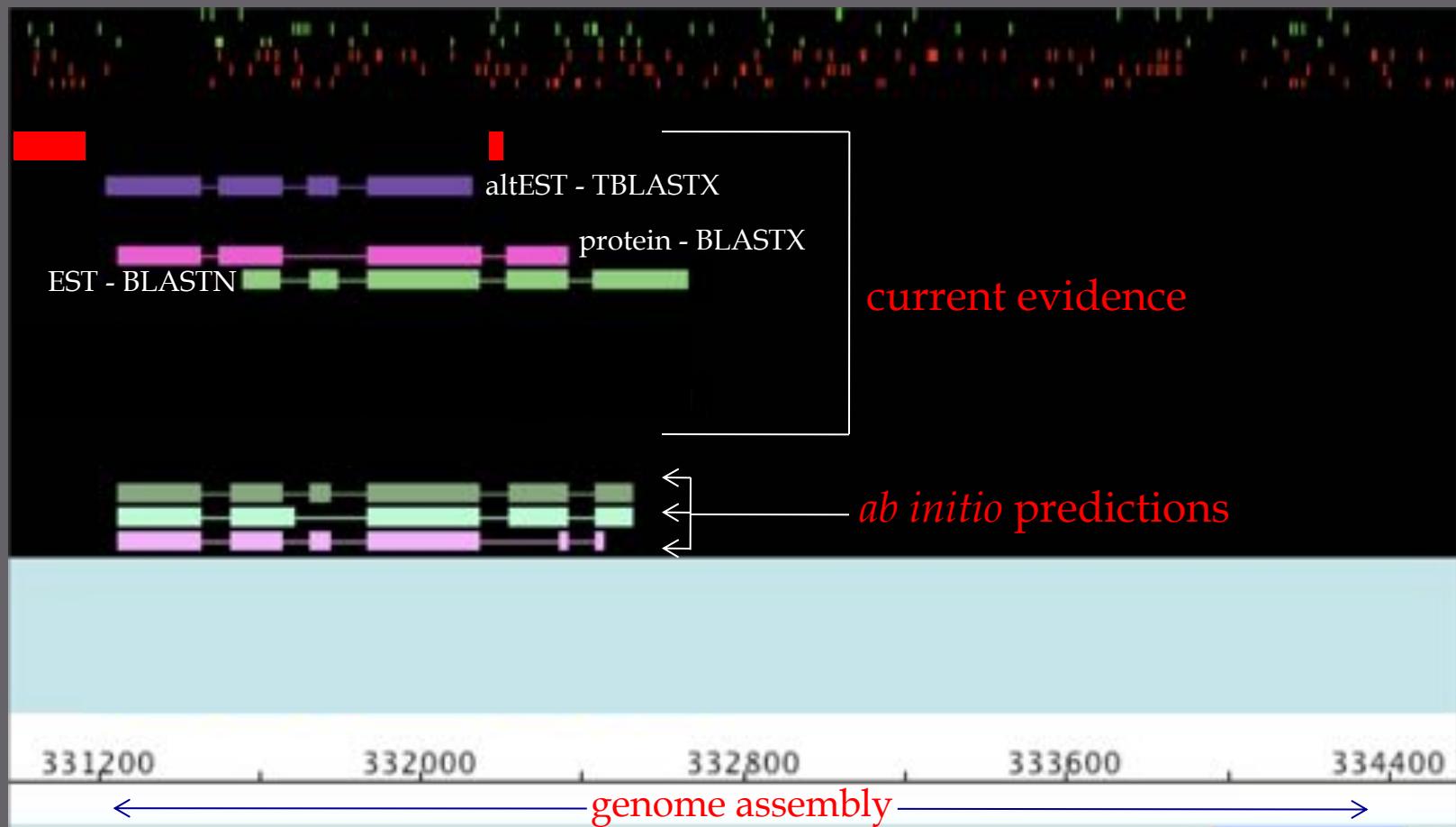
# Generate *ab initio* gene predictions



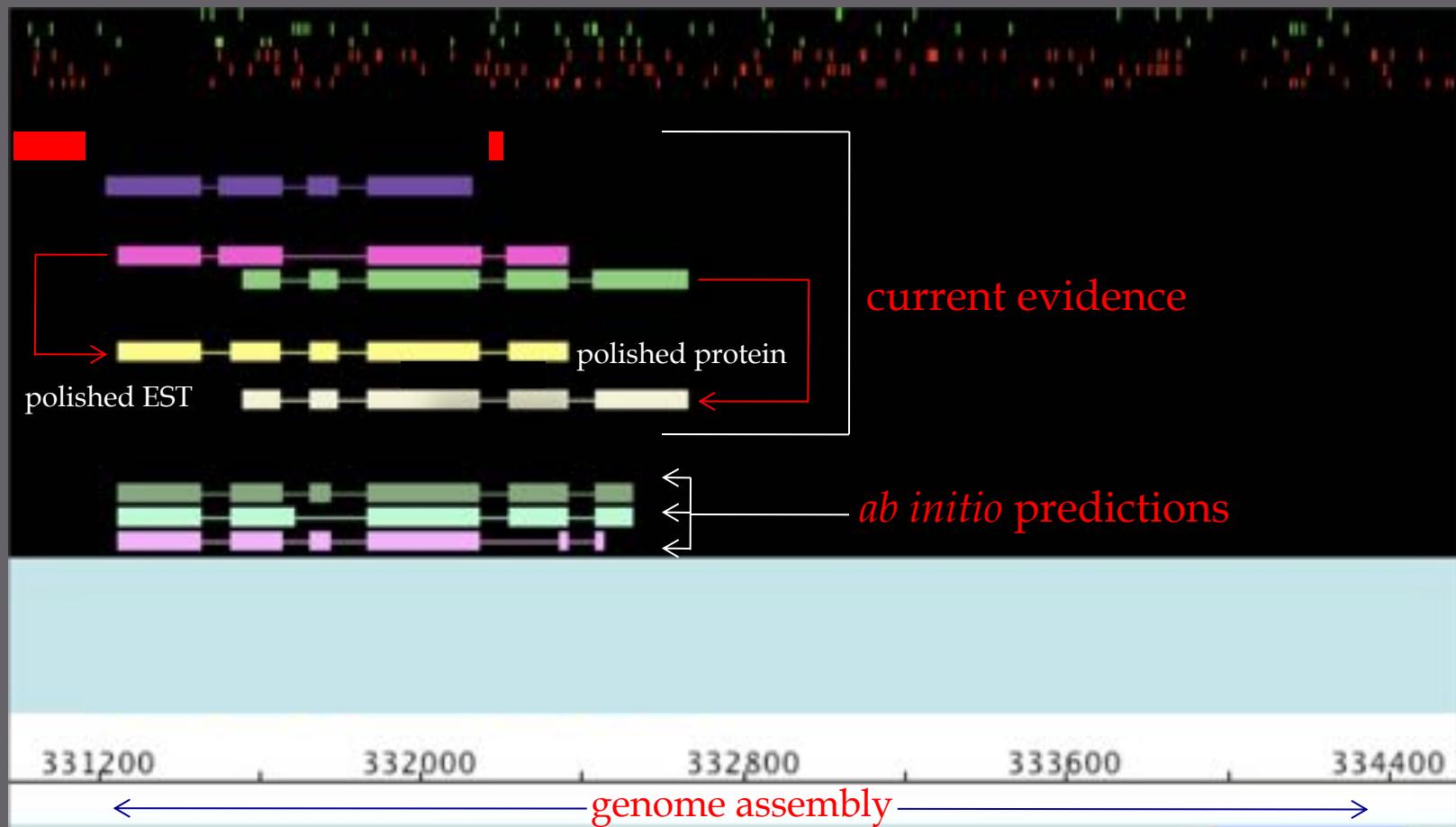
# Generate *ab initio* gene predictions



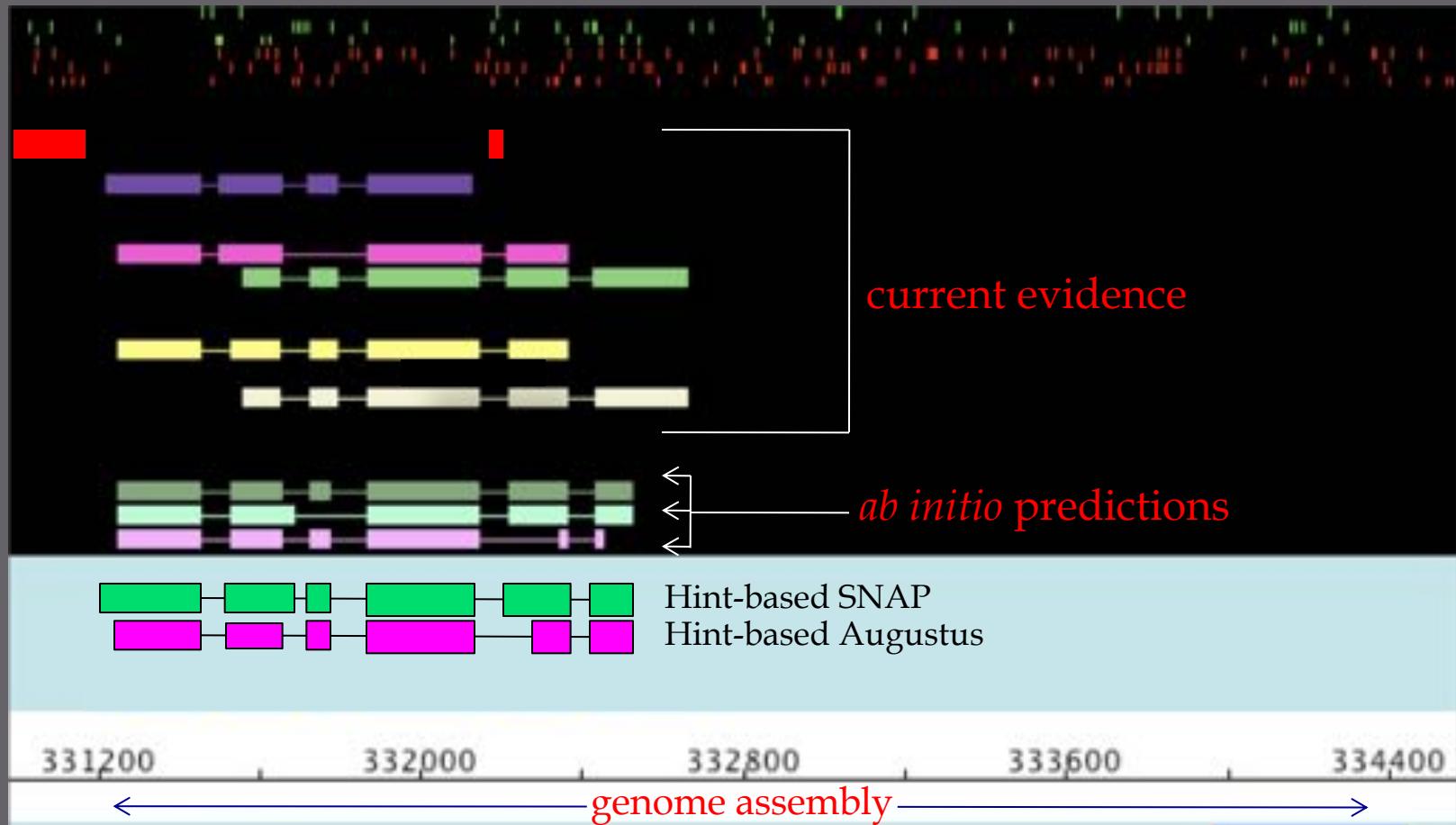
# Align EST and protein evidence



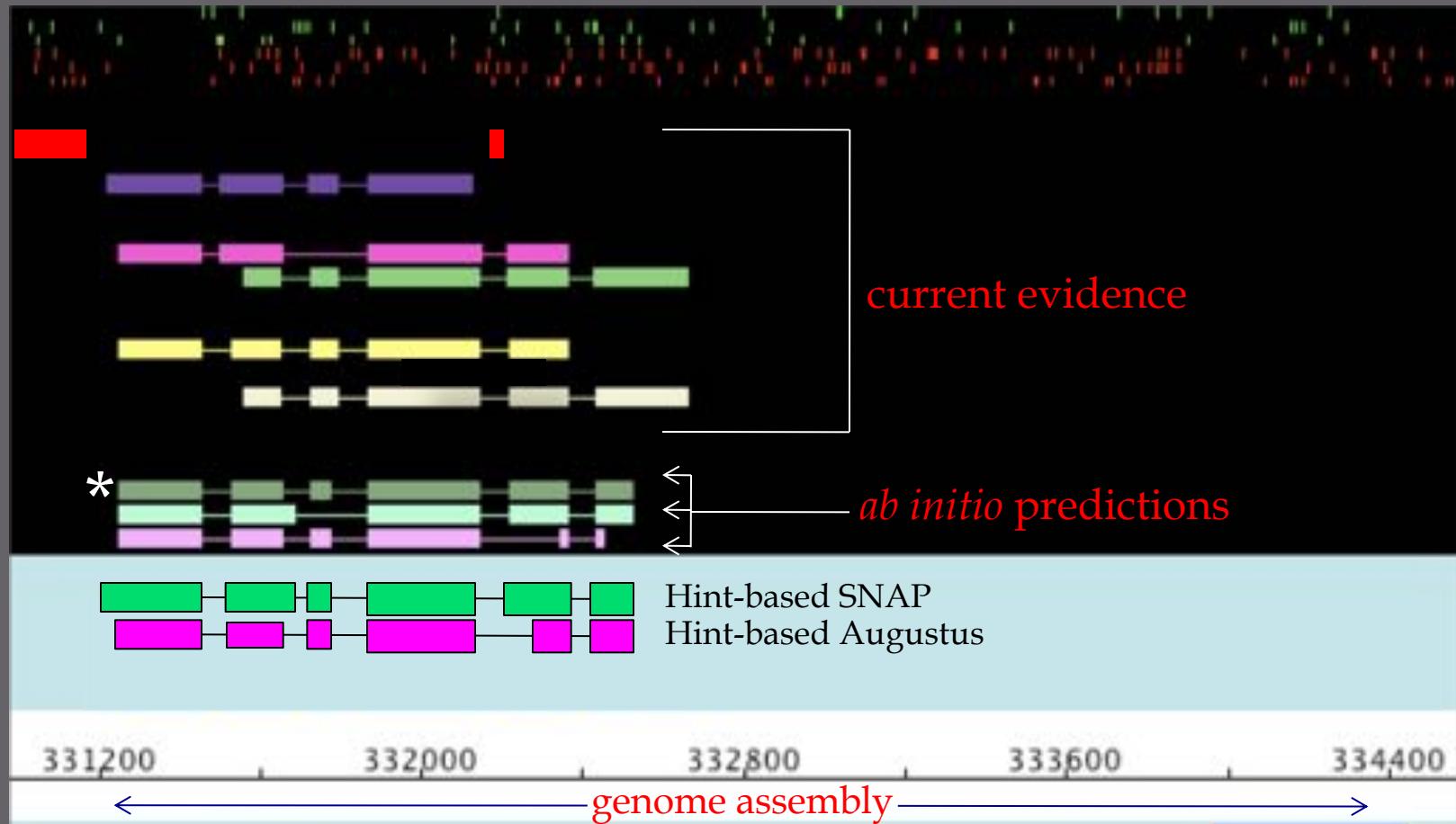
# Polish BLAST alignments with Exonerate



# Pass gene-finders evidence-based ‘hints’

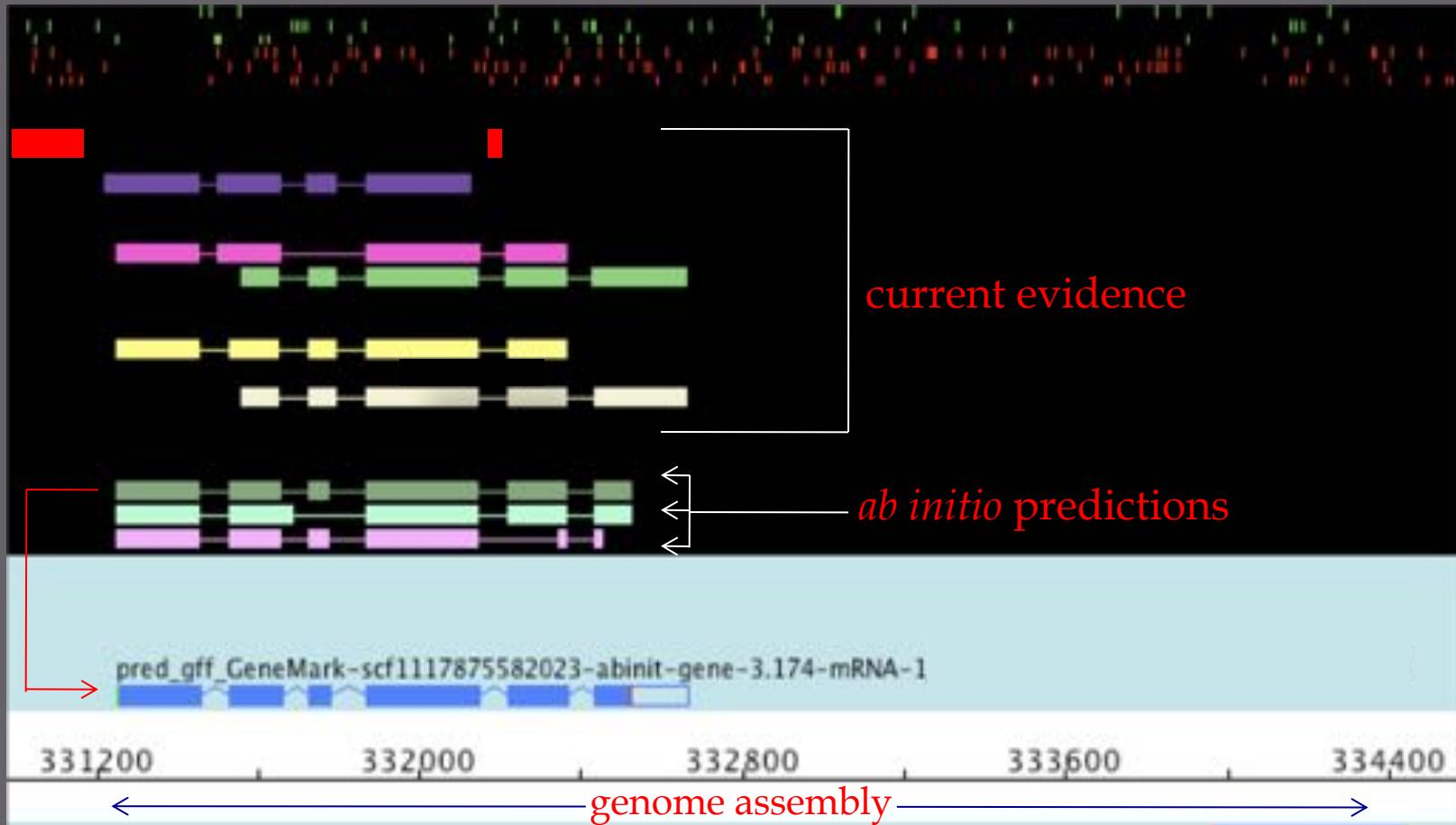


# Identify gene model most consistent with evidence

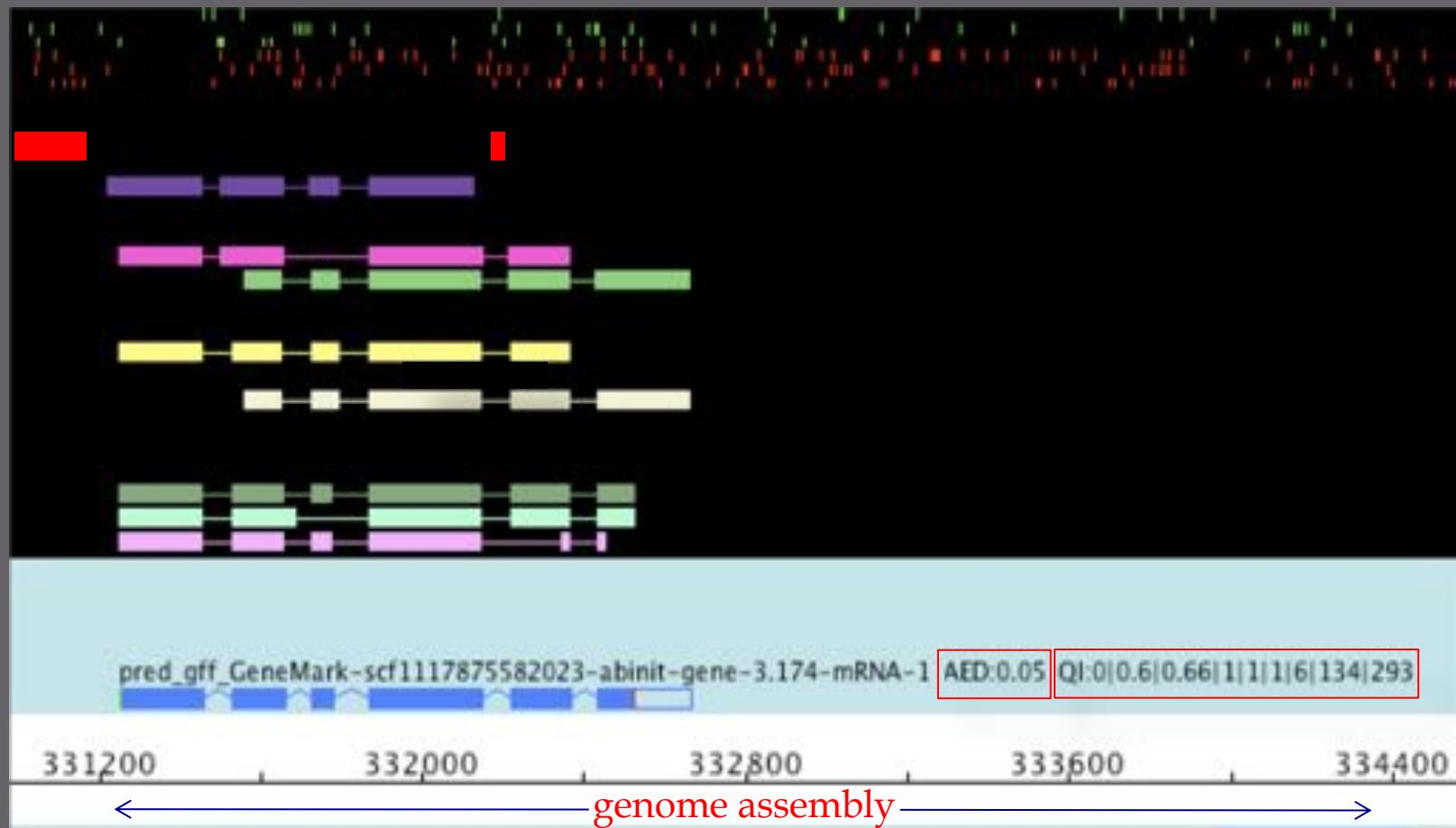


\*Quantitative Measures for the Management and Comparison of Annotated Genomes  
Karen Eilbeck , Barry Moore , Carson Holt and Mark Yandell BMC Bioinformatics 2009  
10:67doi:10.1186/1471-2105-10-67

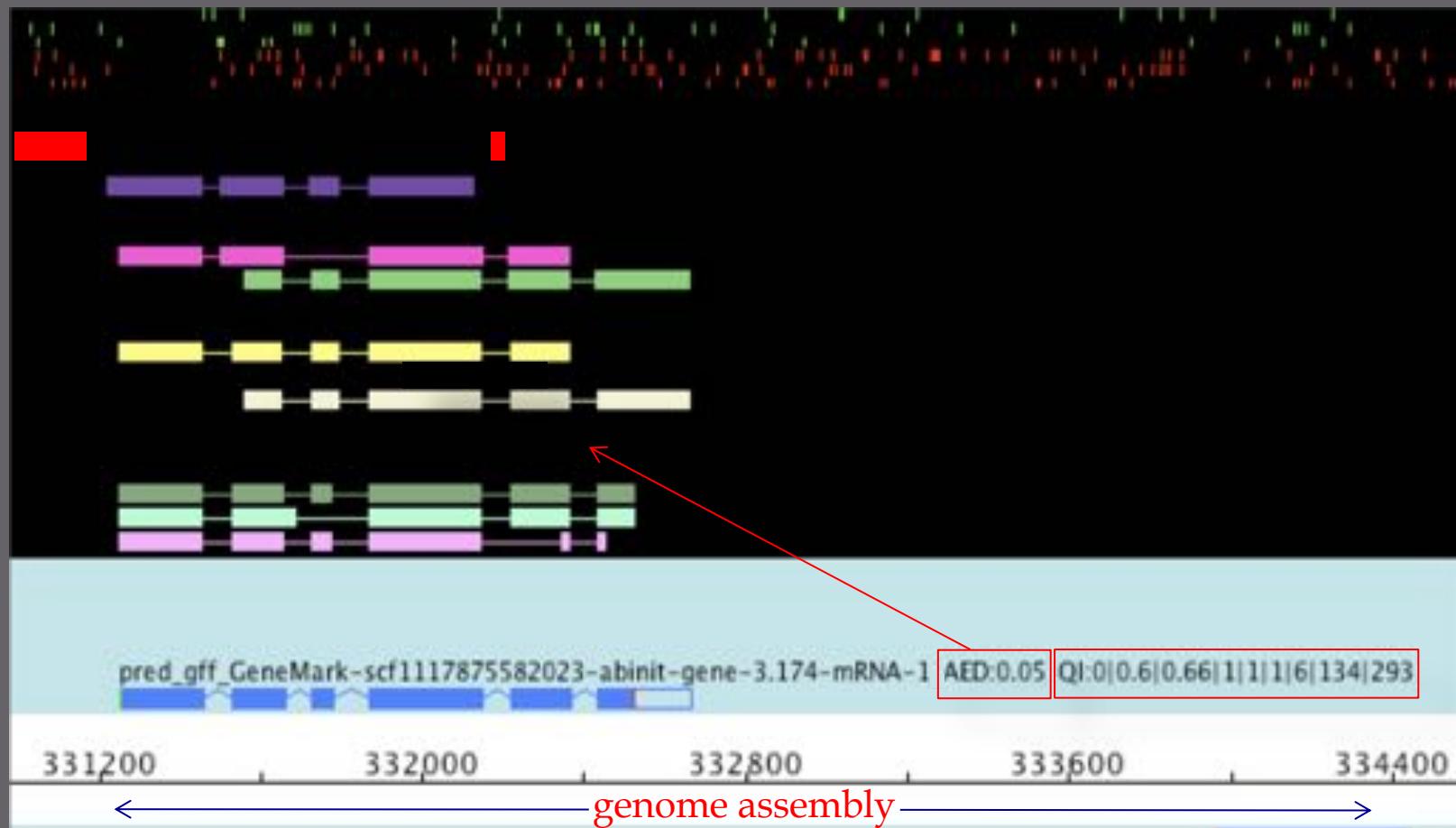
Revise it further if necessary; create new annotation



# Compute support for each portion of gene model

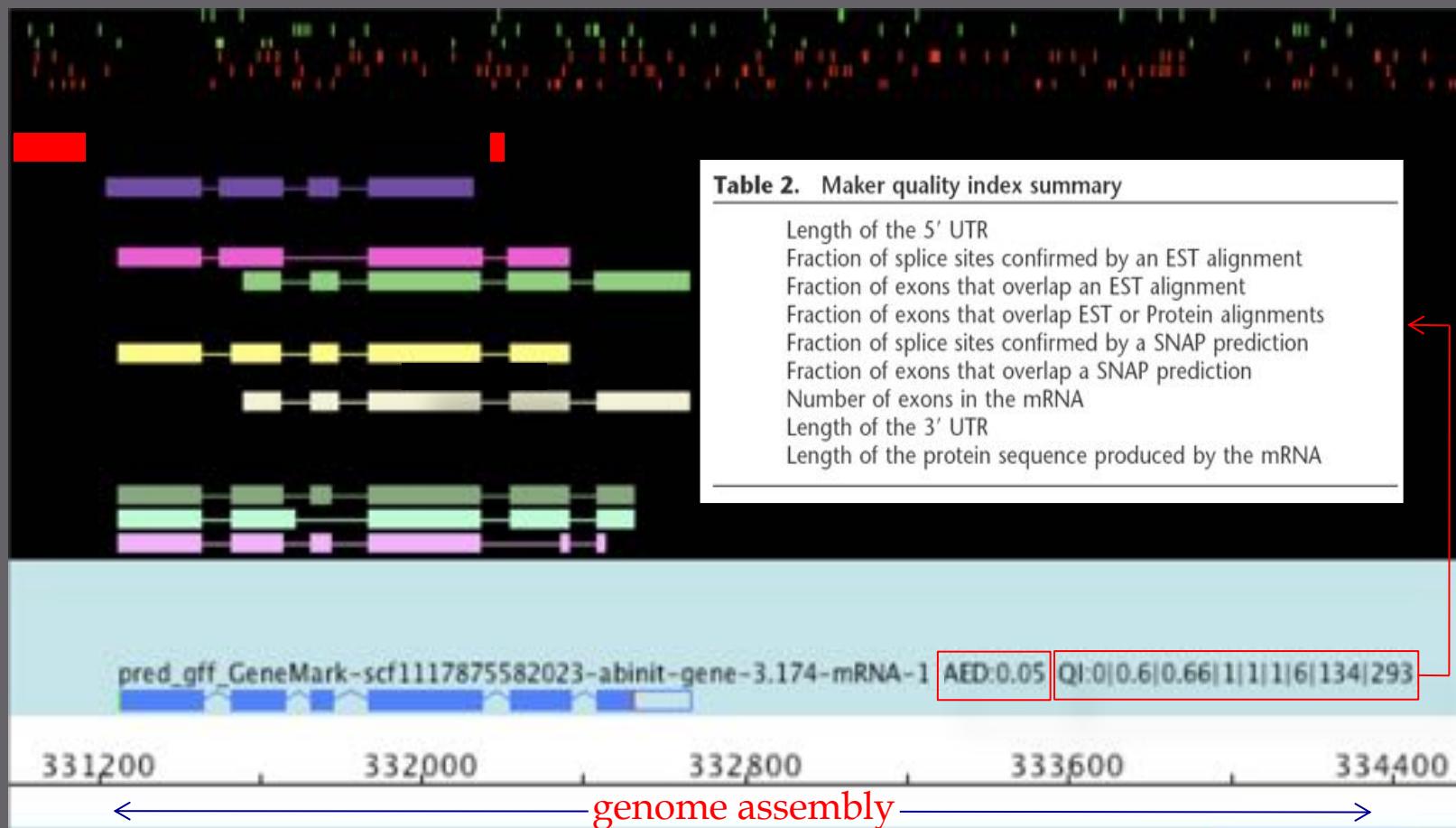


# Compute support for each portion of gene model



\*Quantitative Measures for the Management and Comparison of Annotated Genomes  
Karen Eilbeck , Barry Moore , Carson Holt and Mark Yandell BMC Bioinformatics 2009  
10:67doi:10.1186/1471-2105-10-67

# Compute support for each portion of gene model



\*Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sanchez Alvarado A, Yandell M: **MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes.** *Genome Res* 2008, 18:188-196.

scaffold00080	repeatmasker	match_part	723561	723736	914	+	.	ID=scaffold00080;hsp:4987;Par
scaffold00080	repeatmasker	match	724950	725171	495	+	.	ID=scaffold00080;hit:3405;Name=specie
scaffold00080	repeatmasker	match_part	724950	725171	495	+	.	ID=scaffold00080;hsp:4988;Par
scaffold00080	repeatmasker	match	726925	727293	724	+	.	ID=scaffold00080;hit:3406;Name=specie
scaffold00080	repeatmasker	match_part	726925	727293	724	+	.	ID=scaffold00080;hsp:4989;Par
scaffold00080	maker	gene	56197	58302	.	.	.	ID=ACEP_00015614;Name=ACEP_00015614;Alias=mak
scaffold00080	maker	mRNA	56197	58302	.	.	.	ID=ACEP_00015614-RA;Parent=ACEP_00015614;Name
scaffold00080	maker	exon	56197	56274	.	.	.	ID=ACEP_00015614-RA:exon:126;Parent=ACEP_0001
scaffold00080	maker	exon	56569	56584	.	.	.	ID=ACEP_00015614-RA:exon:127;Parent=ACEP_0001
scaffold00080	maker	exon	56797	56906	.	.	.	ID=ACEP_00015614-RA:exon:128;Parent=ACEP_0001
scaffold00080	maker	exon	57851	57941	.	.	.	ID=ACEP_00015614-RA:exon:129;Parent=ACEP_0001
scaffold00080	maker	exon	58067	58302	18726	-	.	ID=ACEP_00015614-RA:exon:130;Parent=ACEP_0001
scaffold00080	maker	CDS	56197	56274	.	+	0	ID=ACEP_00015614-RA:cds:124;Parent=ACEP_00015
scaffold00080	maker	CDS	56569	56584	.	+	0	ID=ACEP_00015614-RA:cds:125;Parent=ACEP_00015
scaffold00080	maker	CDS	56797	56906	.	+	2	ID=ACEP_00015614-RA:cds:126;Parent=ACEP_00015
scaffold00080	maker	CDS	57851	57941	.	+	1	ID=ACEP_00015614-RA:cds:127;Parent=ACEP_00015
scaffold00080	maker	CDS	58067	58302	.	+	2	ID=ACEP_00015614-RA:cds:128;Parent=ACEP_00015
scaffold00080	maker	gene	2797	3250	.	-	.	ID=ACEP_00015615;Name=ACEP_00015615;Alias=mak

# GFF3

>ACEP\_00015614-RA protein AED:0.401129943502825 QI:0101010.211115101177

MEKSODYQDHYVIQFLEVYNYFYVHEKQYYLKVLNKRITLTSNKYYPIGVIVAFFVPIVE  
IRHPDDLKTILSNPKHKKSSFYDNIKPWLGTSLTSEDCTLSHFIPMLSNRHFTIVHCI  
VHVLGTKWQLQRKILISTFHFDILNQFVEIFEKESKNMIKSLKNAEGTVVKDLSSFI

>ACEP\_00015615-RA protein AED:0.223684210526316 QI:60101010.510101210193

MOTQQKHKREIPTEPGEINSVISFQICISCCDLSYCHIESPTNATNAIYSRQRRAKSNSK  
RKRPGRNGVDAATRLYGSSLWLWPASLFYAYHC

>ACEP\_00015613-RA protein AED:0.191135958515638 QI:910.410.3310.8310.210.516101238

MEKSONSLTYQDHYSIEFGSVYVFGSREILNKKIITLSNKFLRTFYPIGIERAFFVPIVSI  
RHPDOLKSFFYDNIKPWLGTSLINEGTKWQLQRKILIPFTHFIDILNQFVEIFEKESKNH  
IKSLKNAEGTVVKDLSSFISEYTLNAICESYLIPIAGTVLHNININGVHIDPNEHLYQLOQ  
YLILIDFCPRGEIEVLTHEYHSVLDHHVIVSMLRFSKFDHERSGIVR

>ACEP\_00015612-RA protein AED:0.469714351691491 QI:1010.410.3310.8310.210.516101238

MYFVNENFYFANASVTNHRNGRTSLHTELLSSKVXREQTLRNILTAKI  
RILSSRRPRPGDIISKDPQDVLETQETRNGIYIANARTEISIKEEIPD  
INLNATGEELESRTTKTSKNELSPTTEKIEINISSIVSTPATNQKLGDMEHHRHIIYD  
ESLYNHFRPVESNIPVEDMSQFLYFGQKLQPDALNVTSSNNNSVETTSTNPNTSSRRYST

KRFTATIATPHEEINNEENNIAETLERRTVEKNQVSRIKNTFRSSGNGLYYRKPRPTAD  
VVSNTIIPGNESSPGIIGGSETKSEMNTVDDKFRSDSSAHADGSRDRVPLERKNAAHVSE  
VSMITQ1PROQSDVEARVIVVESEIGAKLSNKTITASNGIVESIQLRINDFGKDNEKKKITEA

EAGNVEKSEHKLMSVERTISTSQRESERFRESDSTETHMSLVPSPRAESSTLRIVVPPDONSFG  
ACCEATEEECERGCRHNTKQDDEPMDTAPBRRRBCGCAACCSAETTCGCCCCPDDPDDPDDP

# FASTA

249



GMOD in the Cloud toolset



Galaxy: Data analysis & integration



BioMart: Data mining system



GBrowse\_syn: Synteny viewer



CMap: Comparative map viewer



Chado: Biological database schema



GBrowse: Genome annotation viewer



JBrowse: Super-fast genome annotation viewer



MAKER  
Annotate this!

MAKER: Genome annotation pipeline



Tripal: Chado web interface



InterMine: Data warehousing



Pathway Tools: Metabolic, regulatory pathways

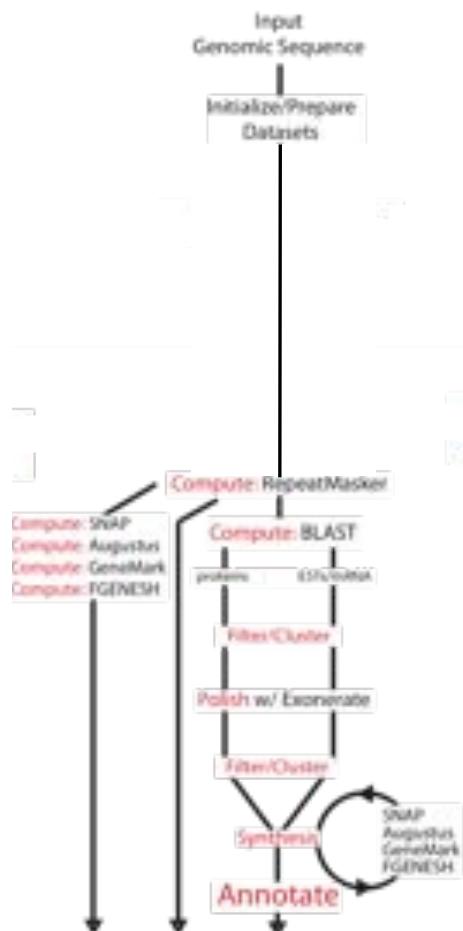


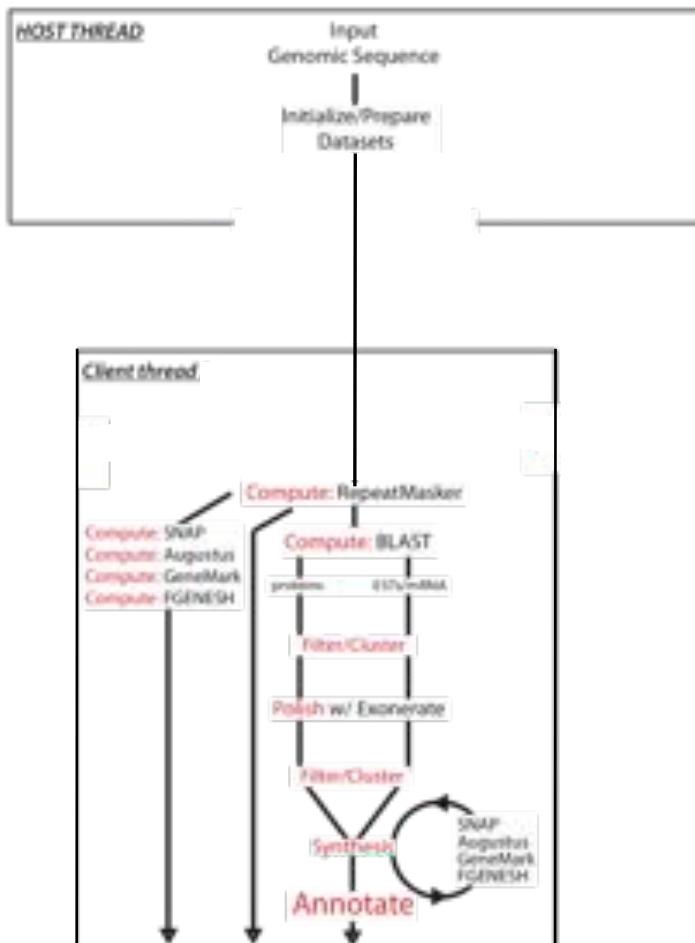
Canto: literature annotation tool

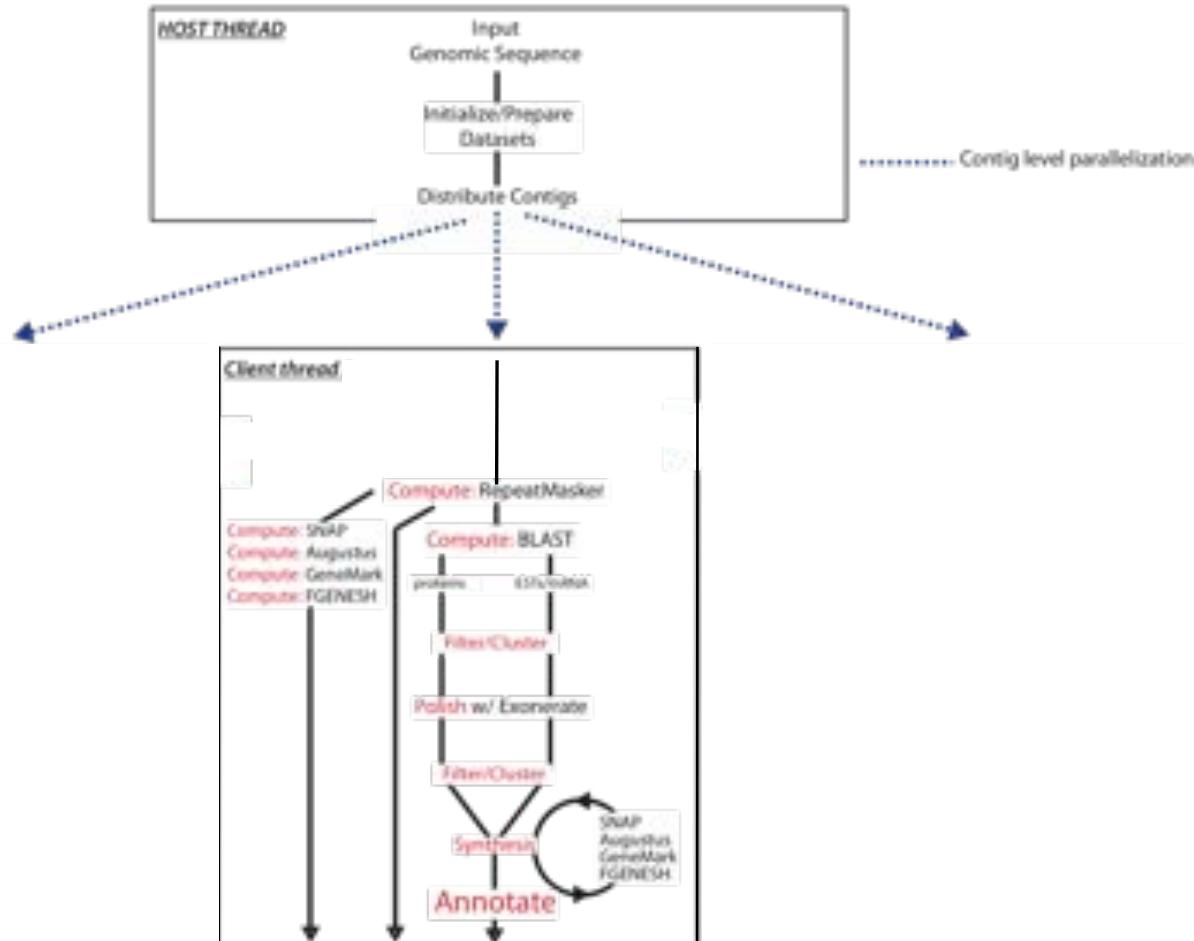
# Distributed Parallelization

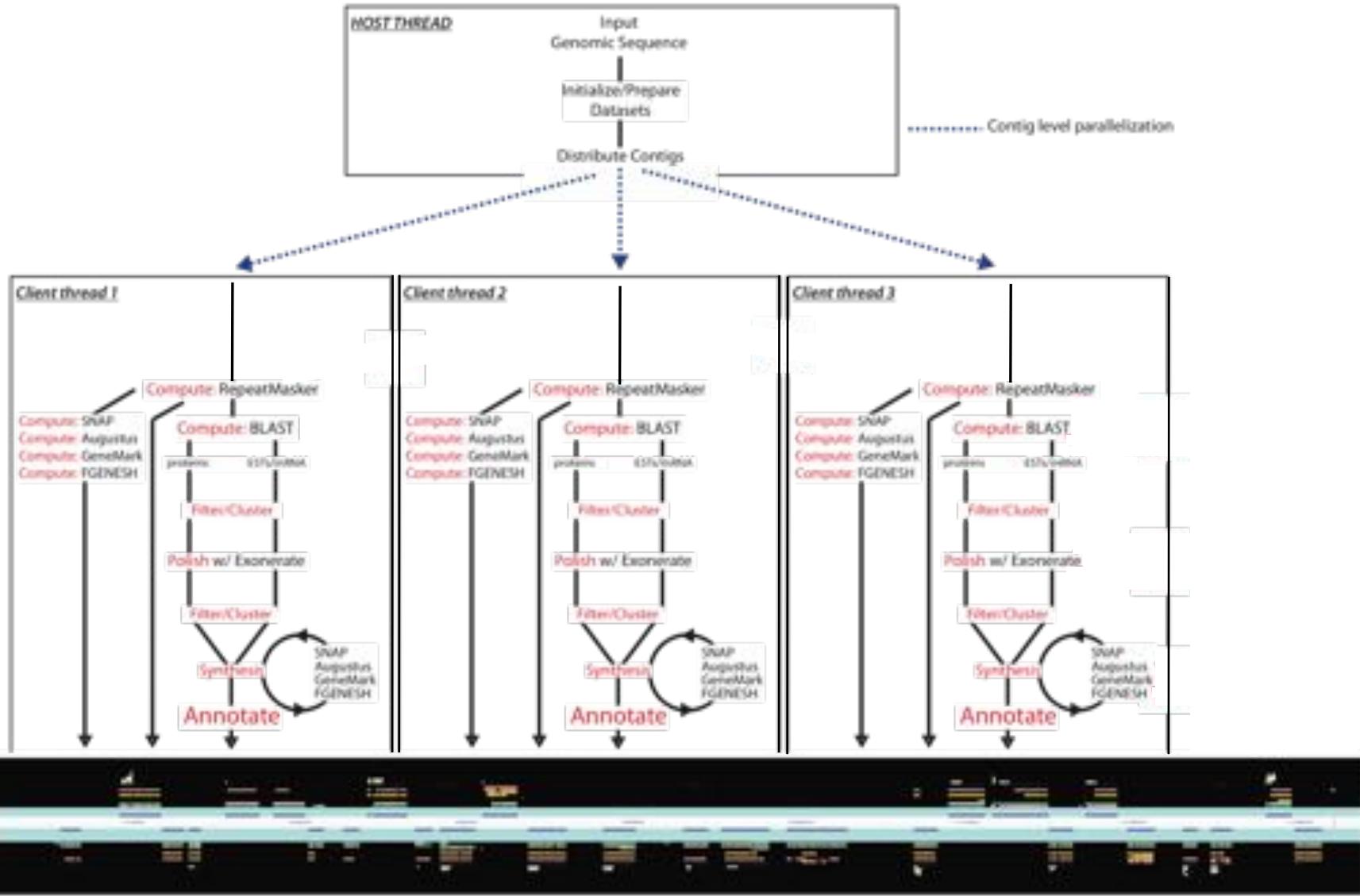
- Supports Message Passing Interface (MPI), a communication protocol for computer clusters which essentially allows multiple computers to act like a single powerful machine.

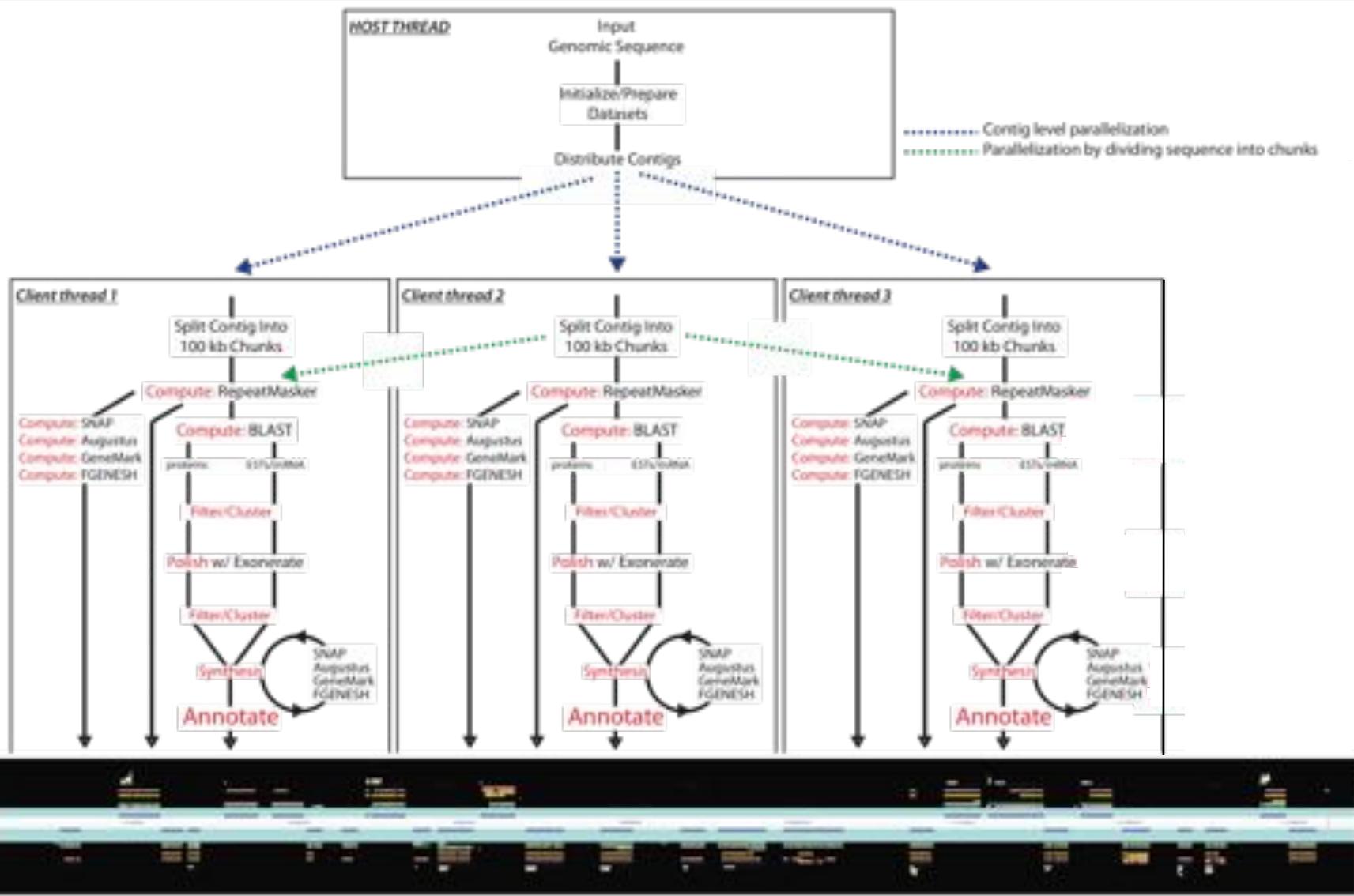


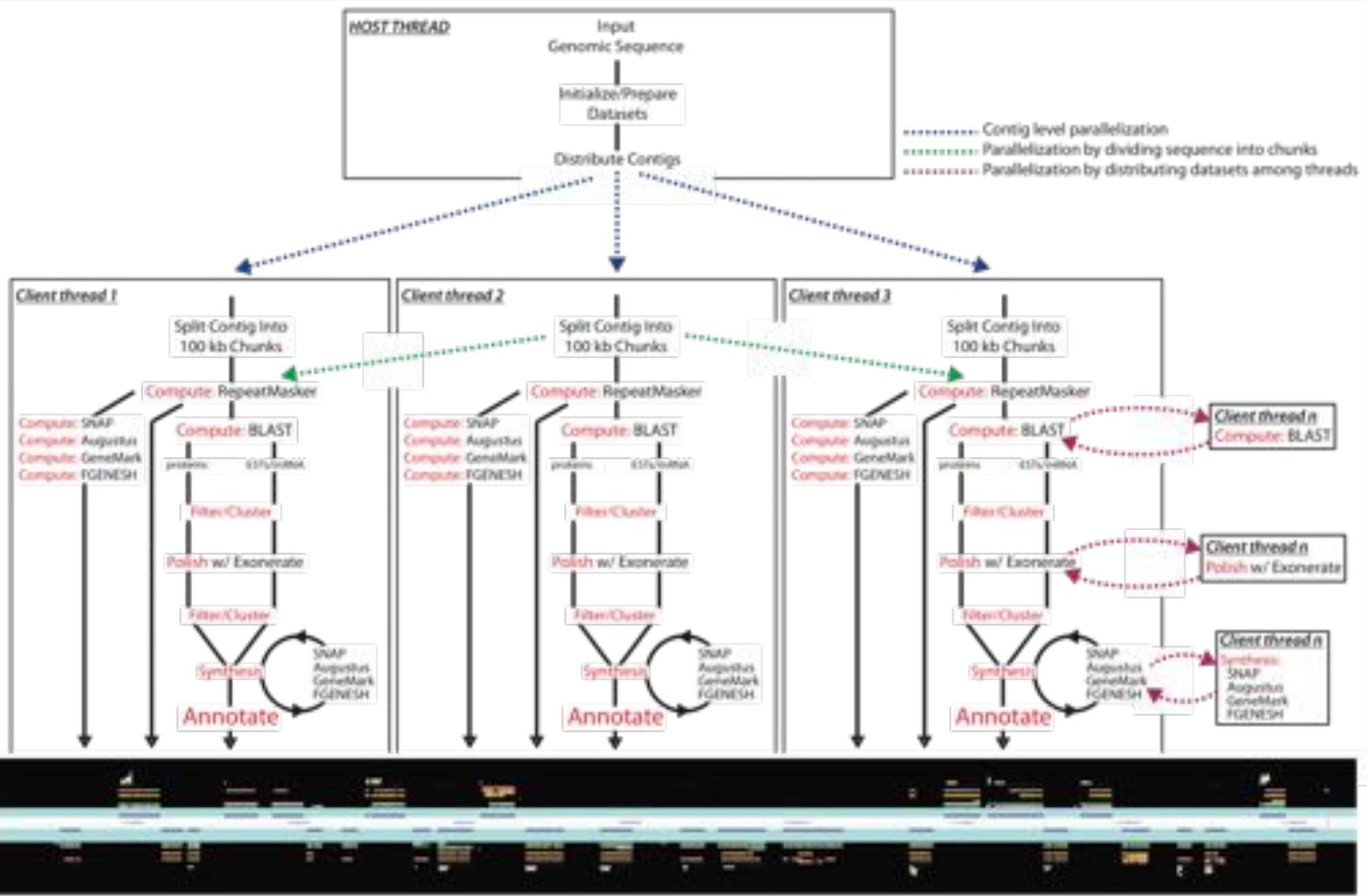




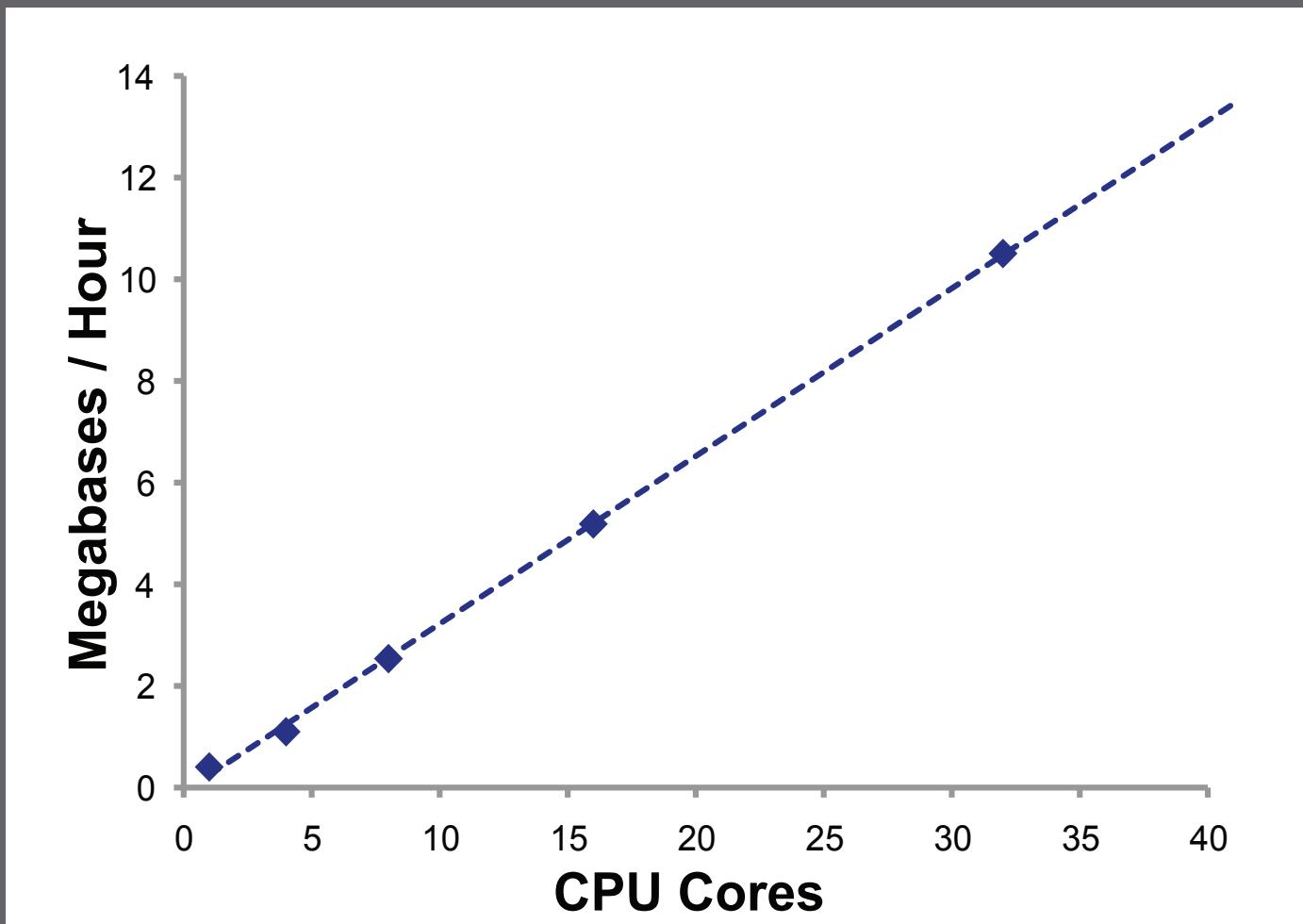






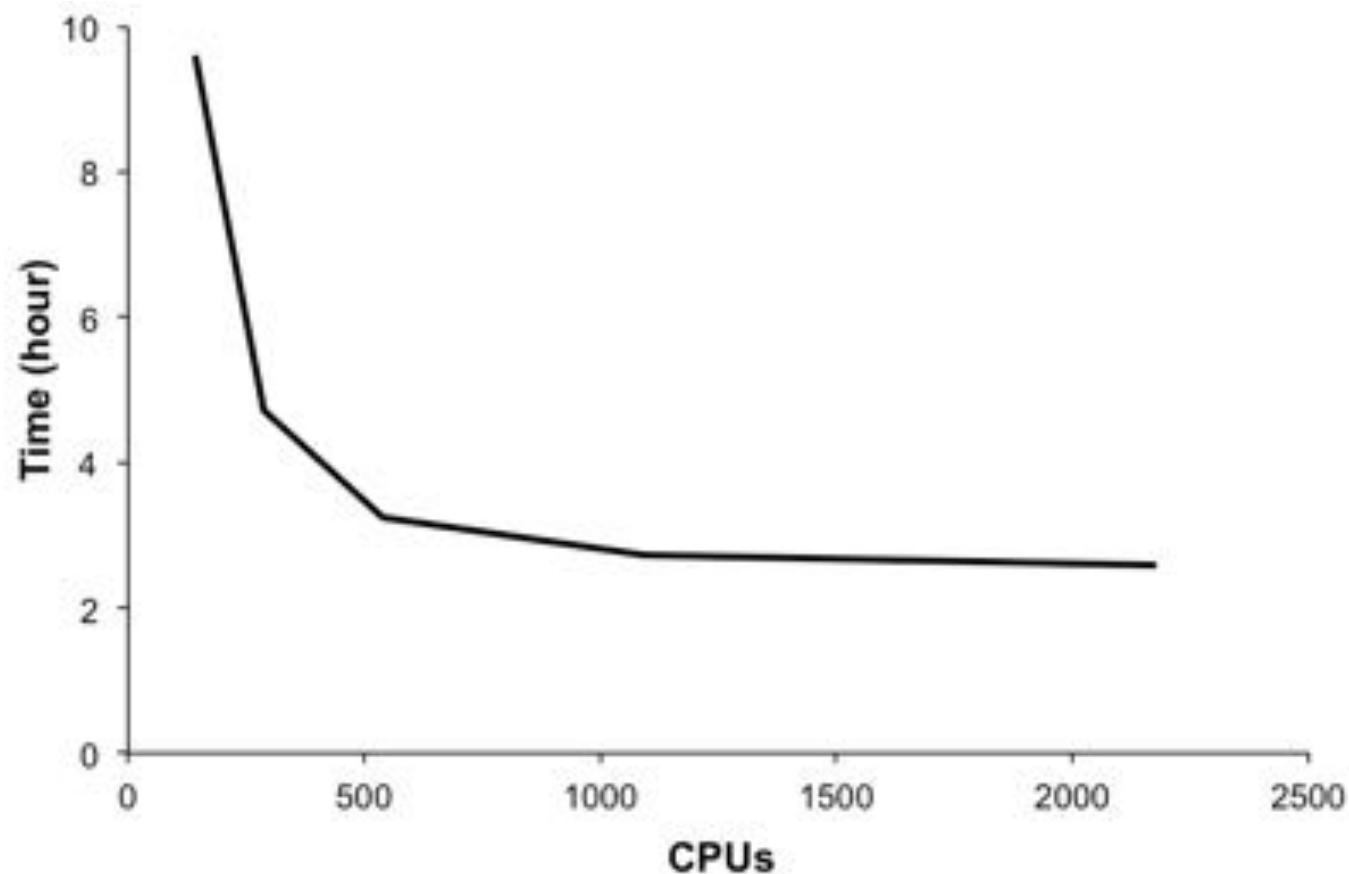


# Data throughput

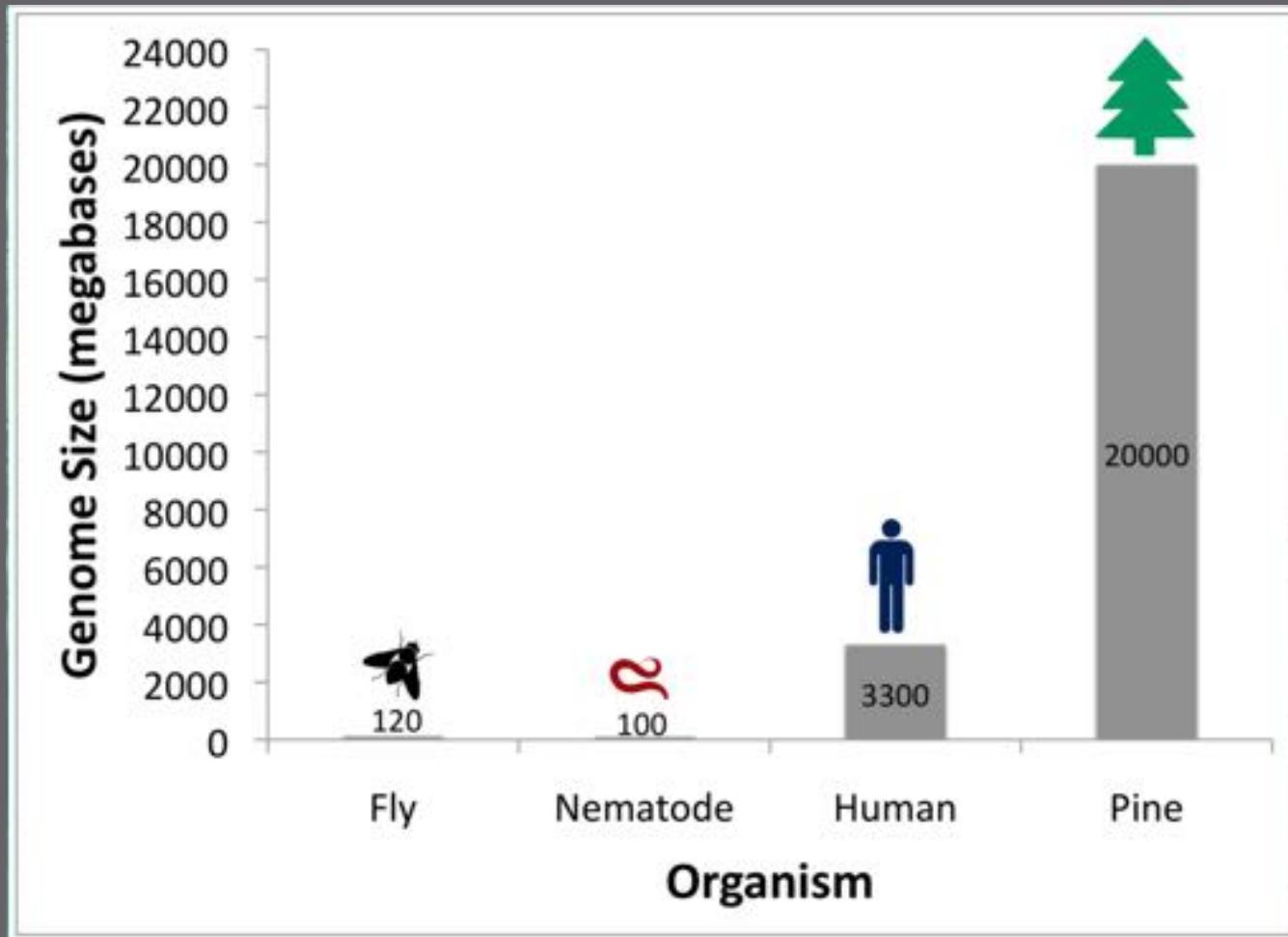


# Data throughput

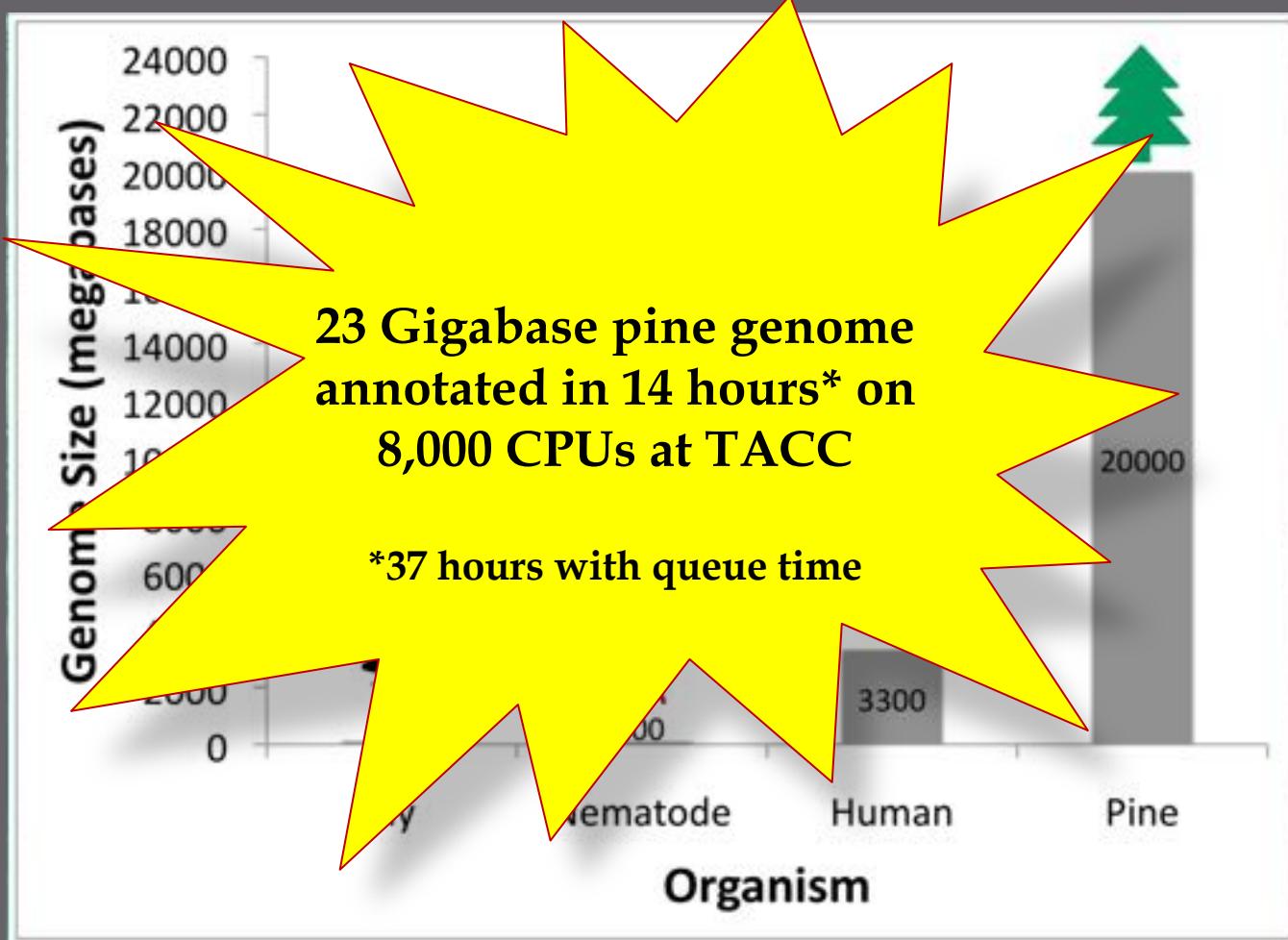
Annotation of the *Zea mays* Genome



# Genome Sizes



# Genome Sizes



# Next Generation of Genomics

“Second-generation” genome projects have different needs and characteristics than earlier works.

# **First *versus* second-generation genomes**

First-generation genomes:

- Classic experimental systems
- Large community
- Big \$
- Much prior knowledge about genome



# Annotation Evaluation

## Gene model accuracies for *ab initio* prediction and genome annotation programs

Reference Organism	Performance Category	<i>Ab Initio</i> Prediction		
		Augustus	GeneMark	SNAP
<i>Caenorhabditis elegans</i>	Nucleotide Accuracy			

# Annotation Evaluation

## Gene model accuracies for *ab initio* prediction and genome annotation programs

Reference Organism	Performance Category	<i>Ab Initio</i> Prediction		
		Augustus	GeneMark	SNAP
<i>Caenorhabditis elegans</i>	Nucleotide Accuracy	88.29%	88.09%	85.10%

# Annotation Evaluation

## Gene model accuracies for *ab initio* prediction and genome annotation programs

Reference Organism	Performance Category	Ab Initio Prediction			MAKER Annotations SNAP
		Augustus	GeneMark	SNAP	
<i>Caenorhabditis elegans</i>	Nucleotide Accuracy	88.29%	88.09%	85.10%	88.48%

# Annotation Evaluation

**Gene model accuracies for *ab initio* prediction and genome annotation programs**

Reference Organism	Performance Category	<i>Ab Initio</i> Prediction			MAKER Annotations SNAP
		Augustus	GeneMark	SNAP	
<i>Caenorhabditis elegans</i>	Nucleotide Accuracy	88.29%	88.09%	85.10%	88.48%
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	77.04%	74.68%	69.78%	80.27%

# Annotation Evaluation

**Gene model accuracies for *ab initio* prediction and genome annotation programs**

Reference Organism	Performance Category	<i>Ab Initio</i> Prediction			MAKER Annotations SNAP
		Augustus	GeneMark	SNAP	
<i>Caenorhabditis elegans</i>	Nucleotide Accuracy	88.29%	88.09%	85.10%	88.48%
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	77.04%	74.68%	69.78%	80.27%
<i>Drosophila melanogaster</i>	Nucleotide Accuracy	76.08%	66.54%	69.29%	74.33%

# Annotation Evaluation

Gene model accuracies for *ab initio* prediction and genome annotation programs

Reference Organism	Performance Category	Ab Initio Prediction			MAKER Annotations SNAP
		Augustus	GeneMark	SNAP	
<i>Caenorhabditis elegans</i>	Nucleotide Accuracy	88.29%	88.09%	85.10%	88.48%
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	77.04%	74.68%	69.78%	80.27%
<i>Drosophila melanogaster</i>	Nucleotide Accuracy	76.08%	66.54%	69.29%	74.33%

With enough training data, *ab initio* gene predictors can match or even out-perform annotation pipelines\*

\*nGASP - the nematode genome annotation assessment project Avril Coghlan , Tristan J Fiedler , Sheldon J McKay ,Paul Flicek , Todd W Harris , Darin Blasiar , The nGASP Consortium and Lincoln D Stein BMC Bioinformatics 2008, 9:549doi:10.1186/1471-2105-9-549

# *First versus second-generation genomes*

Second-generation genomes:

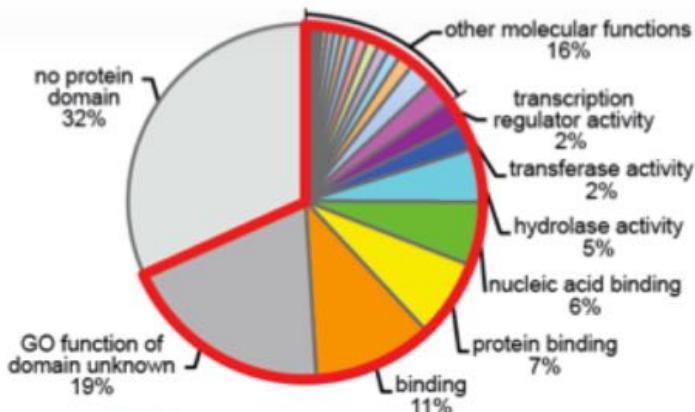
- New experimental systems
  - Genome will be the central resource for research
- Little prior knowledge about genome
  - Usually no genetics
- Small communities
- Less \$



# Evaluation of *Ab initio* gene predictors on emerging model organism genomes

# Evaluation of *Ab initio* gene predictors on emerging model organism genomes

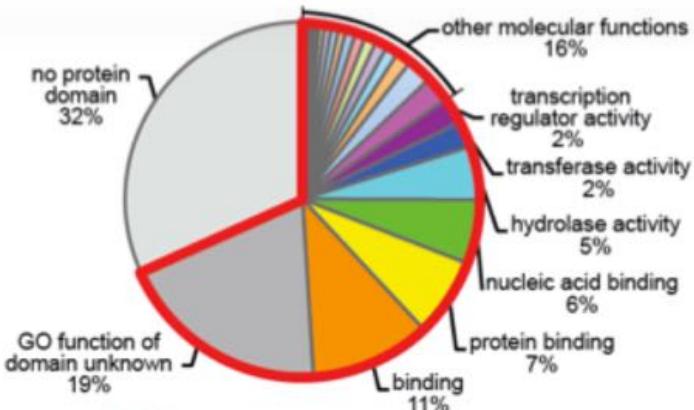
Average of Six Reference Proteomes



68% Contain Pfam Domain

# Evaluation of *Ab initio* gene predictors on emerging model organism genomes

Average of Six Reference Proteomes



68% Contain Pfam Domain

*Schmidtea mediterranea*

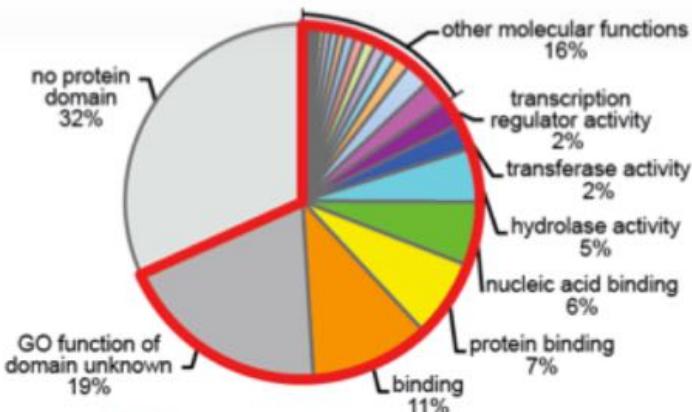
SNAP - *ab initio*



6% Contain Pfam Domain

# Evaluation of *Ab initio* gene predictors on emerging model organism genomes

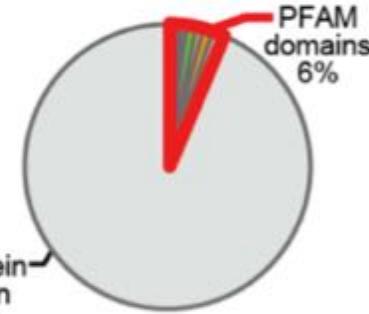
Average of Six Reference Proteomes



68% Contain Pfam Domain

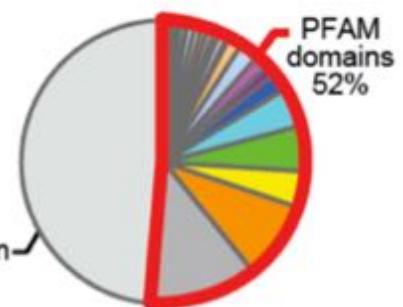
*Schmidtea mediterranea*

SNAP - *ab initio*



6% Contain Pfam Domain

MAKER - SNAP



52% Contain Pfam Domain

# Evaluation of gene models with non-matched species parameters

**Gene model accuracies for *ab initio* prediction and genome annotation programs**

Reference Organism	Performance Category	Ab Initio Prediction		
		Augustus	GeneMark	SNAP
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	77.04%	74.68%	69.78%

# Evaluation of gene models with non-matched species parameters

## Gene model accuracies for *ab initio* prediction and genome annotation programs

Reference Organism	Performance Category	<i>Ab Initio Prediction</i>		
		Augustus	GeneMark	SNAP
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	77.04%	74.68%	69.78%

## Gene model accuracies when using unmatched species parameter files

Reference Organism	Performance Category	<i>Ab Initio Prediction</i>		
		Augustus	GeneMark	SNAP
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	57.85%	48.62%	43.84%

# Evaluation of gene models with non-matched species parameters

## Gene model accuracies for *ab initio* prediction and genome annotation programs

Reference Organism	Performance Category	<i>Ab Initio</i> Prediction		
		Augustus	GeneMark	SNAP
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	77.04%	74.68%	69.78%
	Exon Accuracy	67.03%	61.31%	56.40%

## Gene model accuracies when using unmatched species parameter files

Reference Organism	Performance Category	<i>Ab Initio</i> Prediction		
		Augustus	GeneMark	SNAP
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	57.85%	48.62%	43.84%
	Exon Accuracy	30.71%	16.51%	18.58%

# Evaluation of gene models with non-matched species parameters

## Gene model accuracies for *ab initio* prediction and genome annotation programs

Reference Organism	Performance Category	Ab Initio Prediction		
		Augustus	GeneMark	SNAP
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	77.04%	74.68%	69.78%
	Exon Accuracy	67.03%	61.31%	56.40%

## Gene model accuracies when using unmatched species parameter files

Reference Organism	Performance Category	Ab Initio Prediction		
		Augustus	GeneMark	SNAP
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	57.85%	48.62%	43.84%
	Exon Accuracy	30.71%	16.51%	18.58%

# Evaluation of gene models with non-matched species parameters

## Gene model accuracies for *ab initio* prediction and genome annotation programs

Reference Organism	Performance Category	Ab Initio Prediction		
		Augustus	GeneMark	SNAP
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	77.04%	74.68%	69.78%
	Exon Accuracy	67.03%	61.31%	56.40%

## Gene model accuracies when using unmatched species parameter files

Reference Organism	Performance Category	Ab Initio Prediction			MAKER Annotations SNAP
		Augustus	GeneMark	SNAP	
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	57.85%	48.62%	43.84%	73.77% 60.11%
	Exon Accuracy	30.71%	16.51%	18.58%	

# Evaluation of gene models with non-matched species parameters

## Gene model accuracies for *ab initio* prediction and genome annotation programs

Reference Organism	Performance Category	Ab Initio Prediction		
		Augustus	GeneMark	SNAP
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	77.04%	74.68%	69.78%
	Exon Accuracy	67.03%	61.31%	56.40%
<i>Drosophila Melanogaster</i>	Nucleotide Accuracy	76.08%	66.54%	69.29%
	Exon Accuracy	61.37%	47.31%	47.01%

## Gene model accuracies when using unmatched species parameter files

Reference Organism	Performance Category	Ab Initio Prediction			MAKER Annotations SNAP
		Augustus	GeneMark	SNAP	
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	57.85%	48.62%	43.84%	73.77%
	Exon Accuracy	30.71%	16.51%	18.58%	60.11%
<i>Drosophila Melanogaster</i>	Nucleotide Accuracy	67.47%	66.51%	48.92%	74.44%
	Exon Accuracy	30.62%	26.25%	19.94%	53.69%

# Evaluation of gene models with non-matched species parameters

Gene model accuracies for *ab initio* prediction and genome annotation programs

Reference Organism	Performance Category	Ab Initio Prediction		
		Augustus	GeneMark	SNAP
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	77.04%	74.68%	69.78%
	Exon Accuracy	67.03%	61.31%	56.40%
<i>Drosophila Melanogaster</i>	Nucleotide Accuracy	76.08%	66.54%	69.29%
	Exon Accuracy	61.37%	47.31%	47.01%
<i>Caenorhabditis elegans</i>	Nucleotide Accuracy	88.29%	88.09%	85.10%
	Exon Accuracy	74.62%	68.88%	61.38%

Gene model accuracies when using unmatched species parameter files

Reference Organism	Performance Category	Ab Initio Prediction			MAKER Annotations SNAP
		Augustus	GeneMark	SNAP	
<i>Arabidopsis thaliana</i>	Nucleotide Accuracy	57.85%	48.62%	43.84%	73.77%
	Exon Accuracy	30.71%	16.51%	18.58%	60.11%
<i>Drosophila Melanogaster</i>	Nucleotide Accuracy	67.47%	66.51%	48.92%	74.44%
	Exon Accuracy	30.62%	26.25%	19.94%	53.69%
<i>Caenorhabditis elegans</i>	Nucleotide Accuracy	66.18%	67.26%	68.24%	85.02%
	Exon Accuracy	28.33%	30.01%	35.44%	63.14%

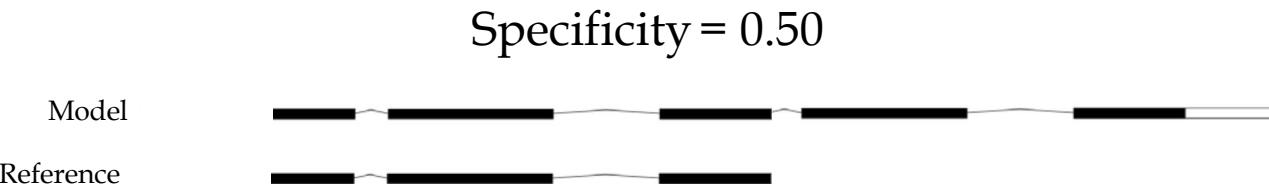
# Beyond *de novo* annotation

- Quality control and data prioritization

# Beyond *de novo* annotation

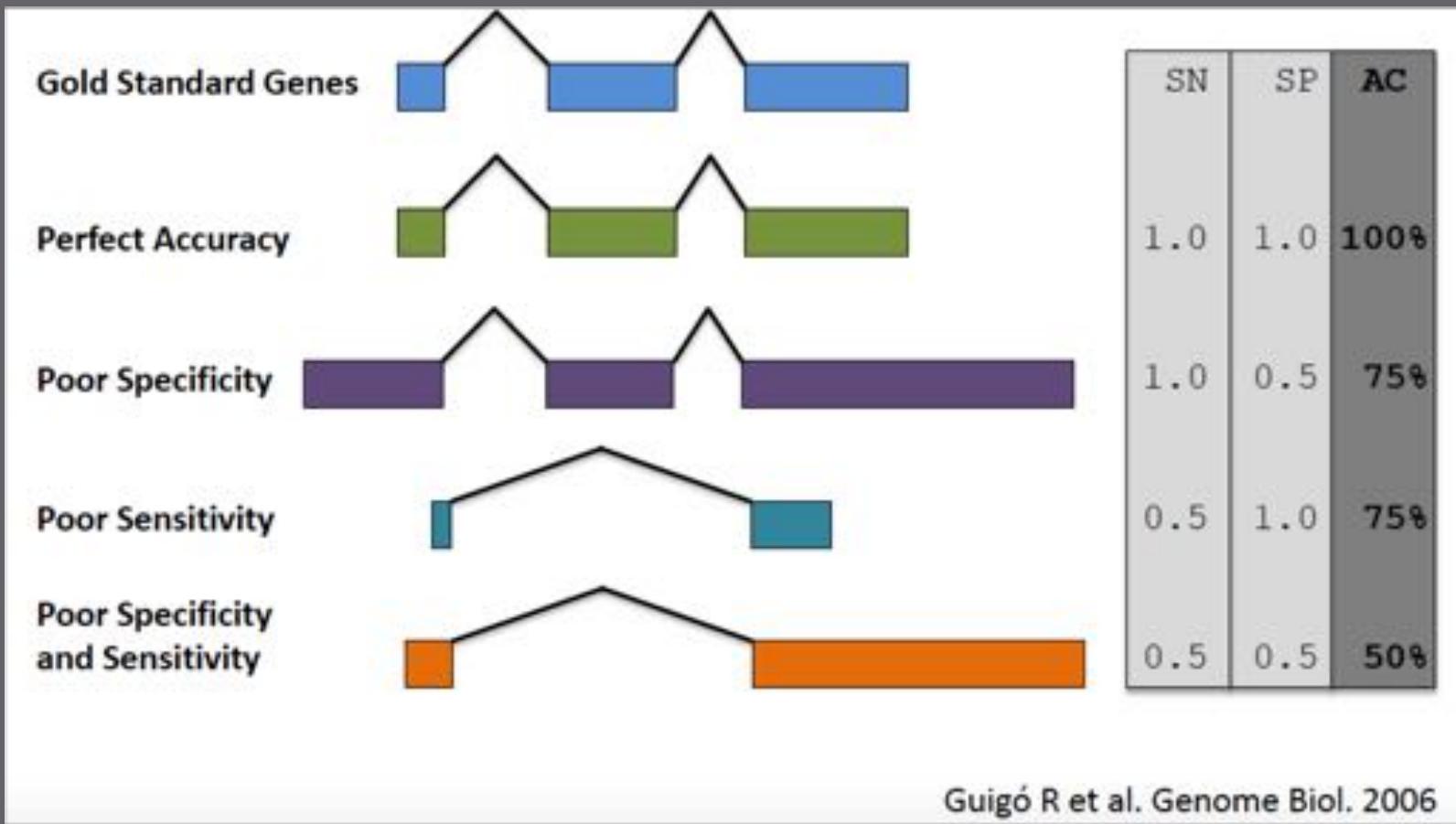
- Quality control and data prioritization
- Update/revise legacy annotation sets
- Integrating new evidence into existing databases

# Quality control and data prioritization



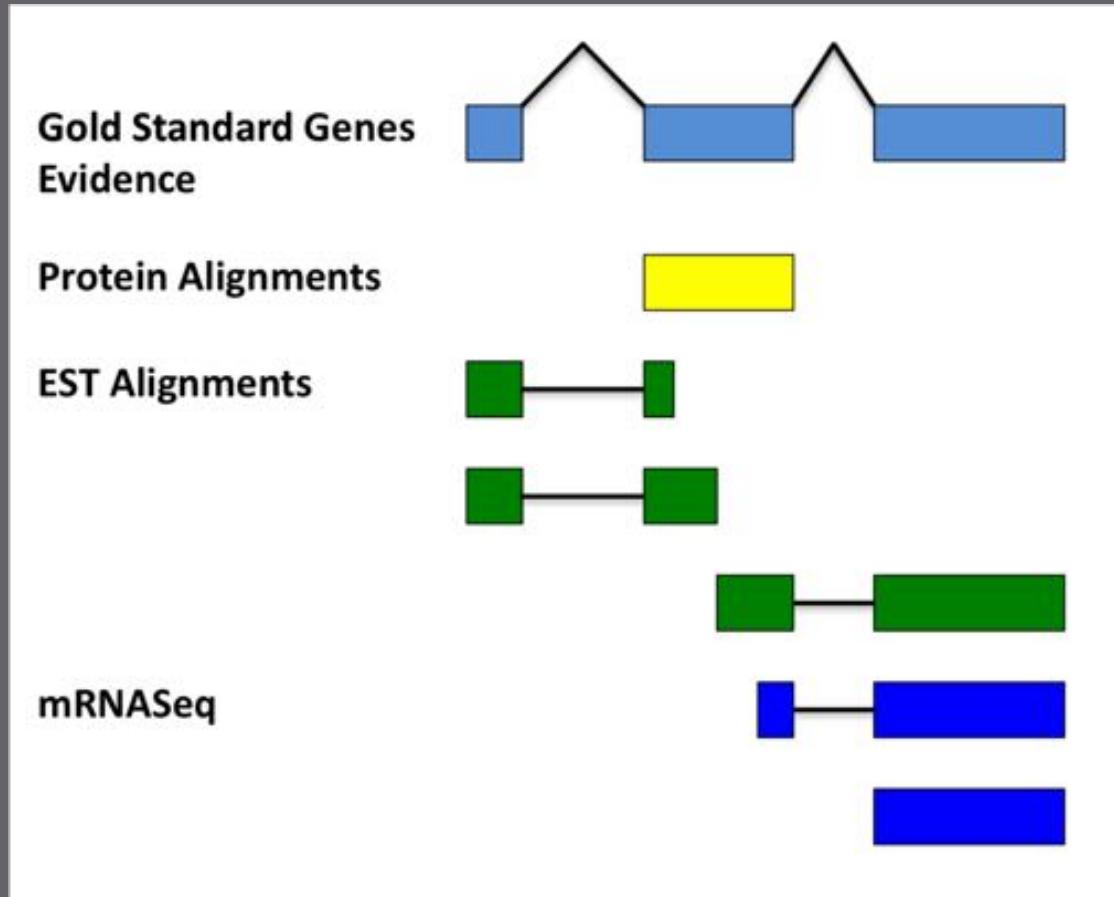
Guigó R et al. Genome Biol. 2006

# Quality control and data prioritization



Guigó R et al. Genome Biol. 2006

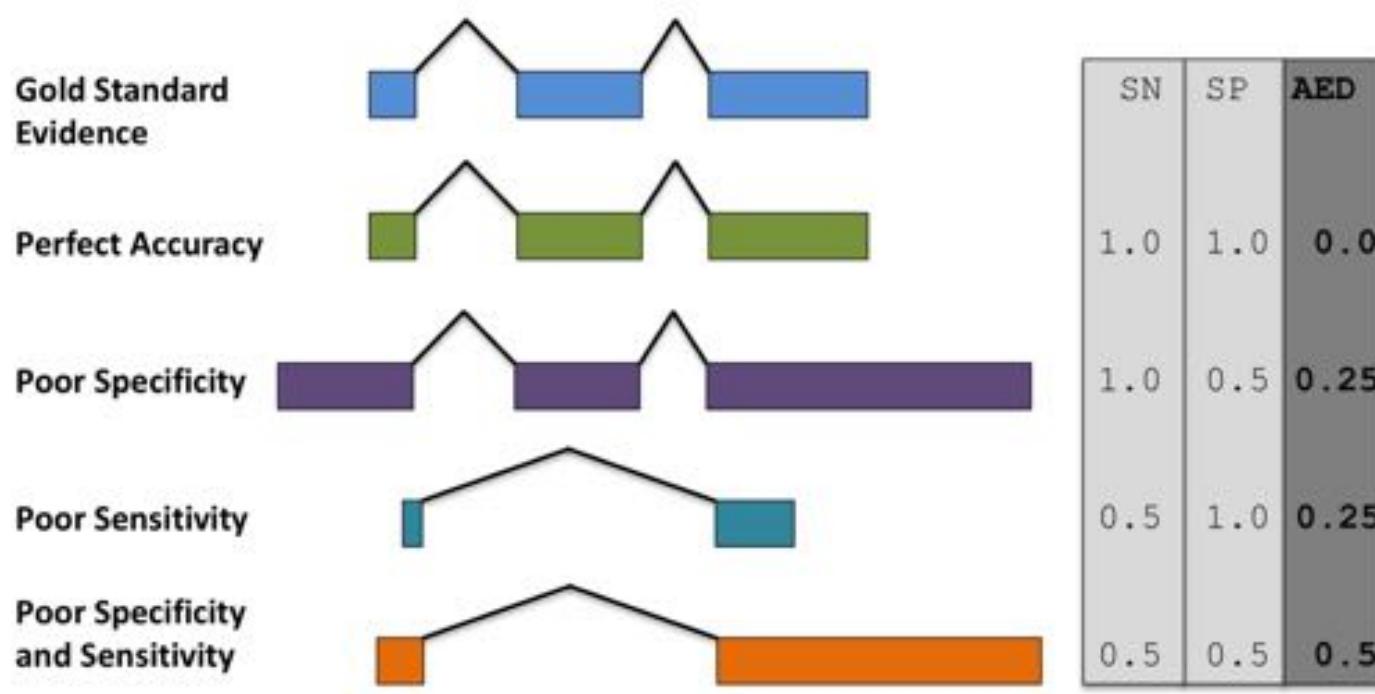
# Quality control and data prioritization



\*Quantitative Measures for the Management and Comparison of Annotated Genomes  
Karen Eilbeck , Barry Moore , Carson Holt and Mark Yandell BMC Bioinformatics 2009  
10:67doi:10.1186/1471-2105-10-67

# Quality control and data prioritization

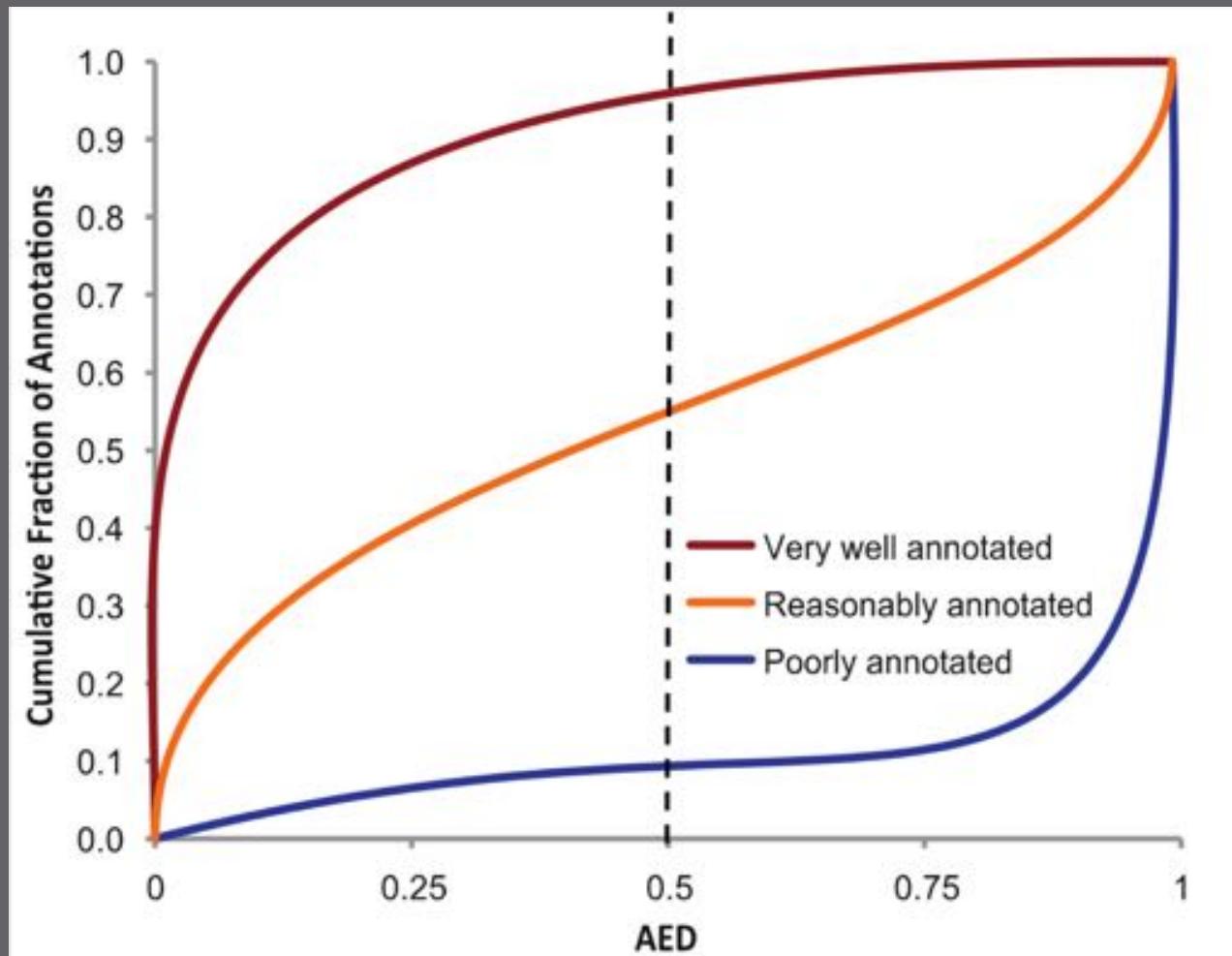
$$AED = 1 - \frac{(SN + SP)}{2}$$



Eilbeck et al BMC Bioinformatics 2009

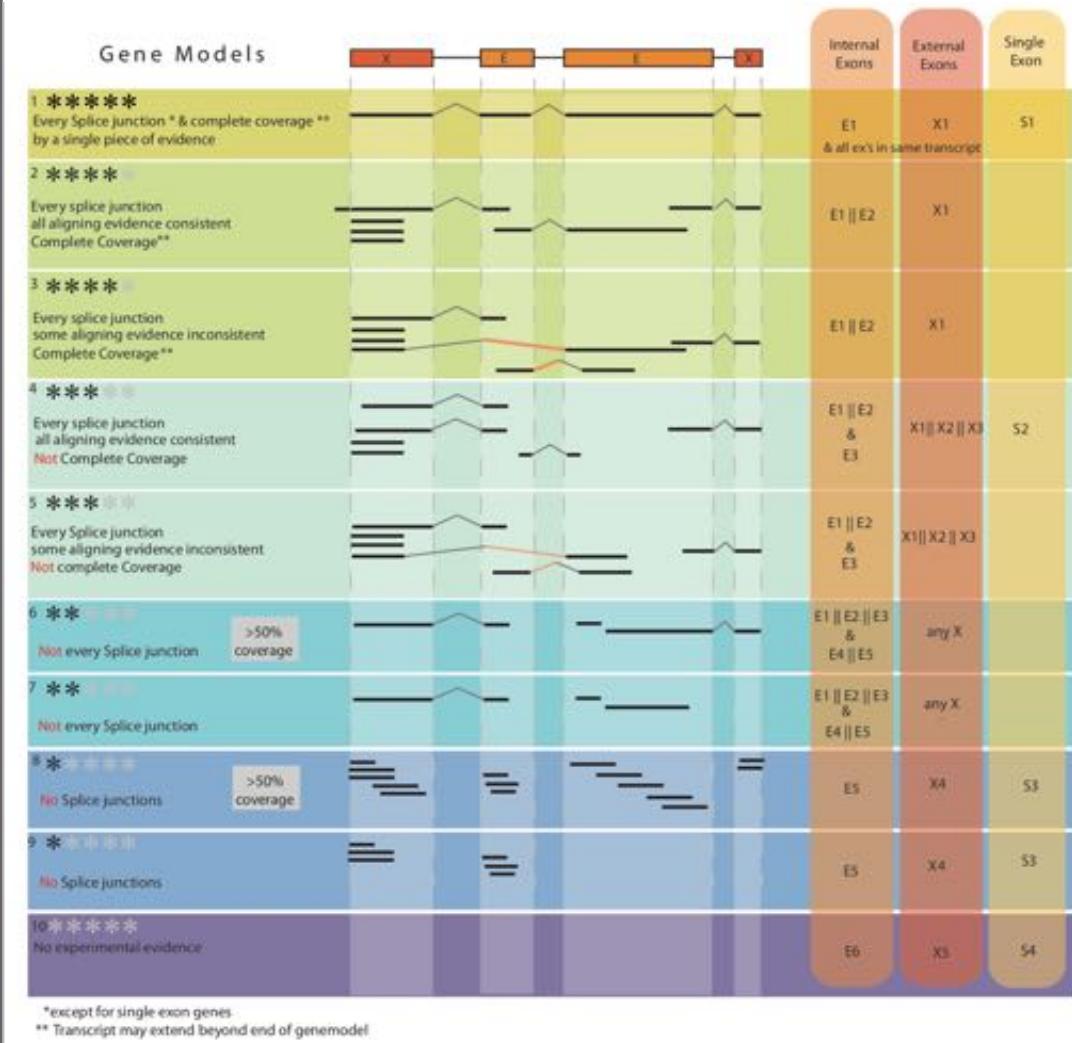
\*Quantitative Measures for the Management and Comparison of Annotated Genomes  
Karen Eilbeck , Barry Moore , Carson Holt and Mark Yandell BMC Bioinformatics 2009  
10:67doi:10.1186/1471-2105-10-67

# Quality control and data prioritization

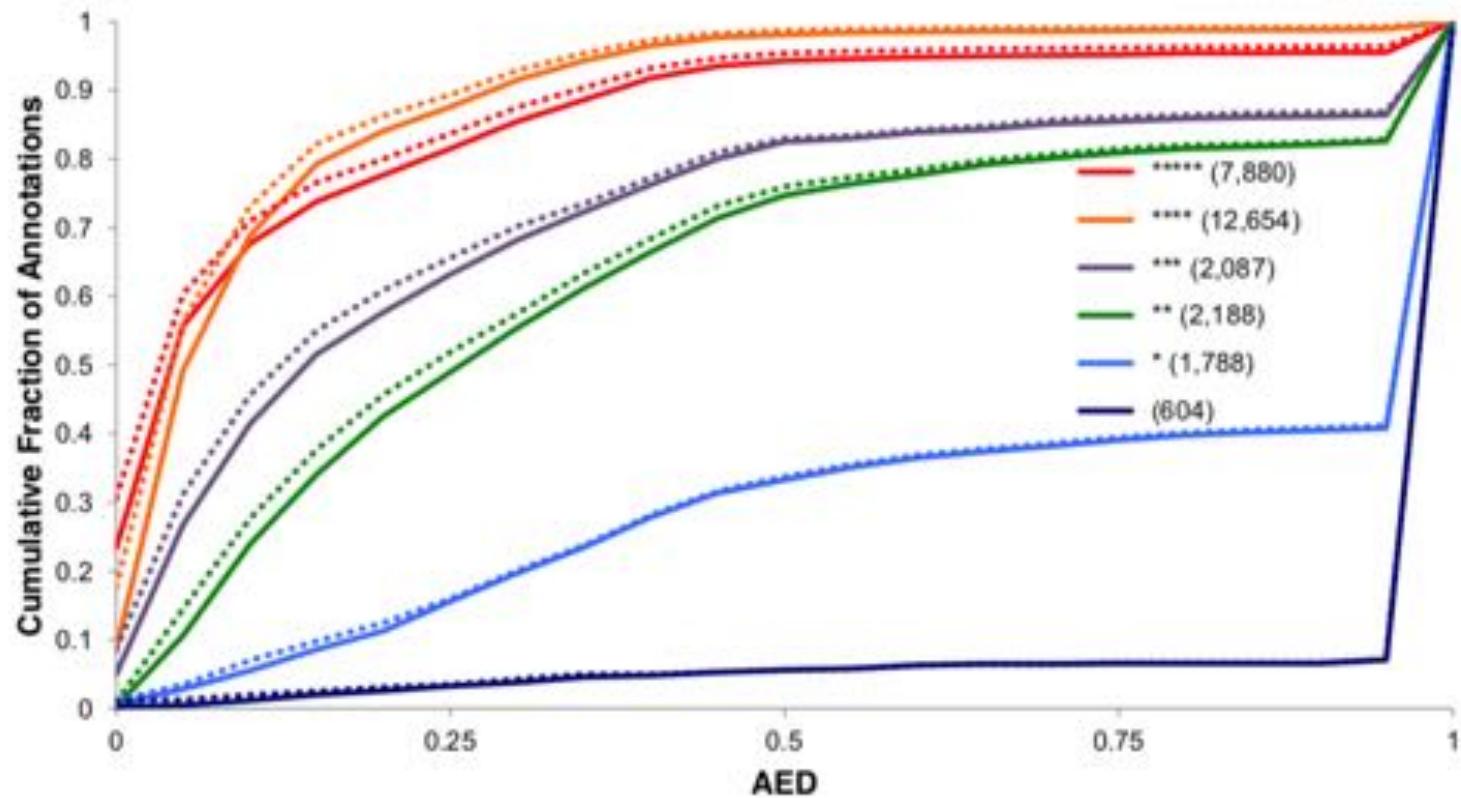


# Evaluation of AED on Arabidopsis

## TAIR Star Rating System

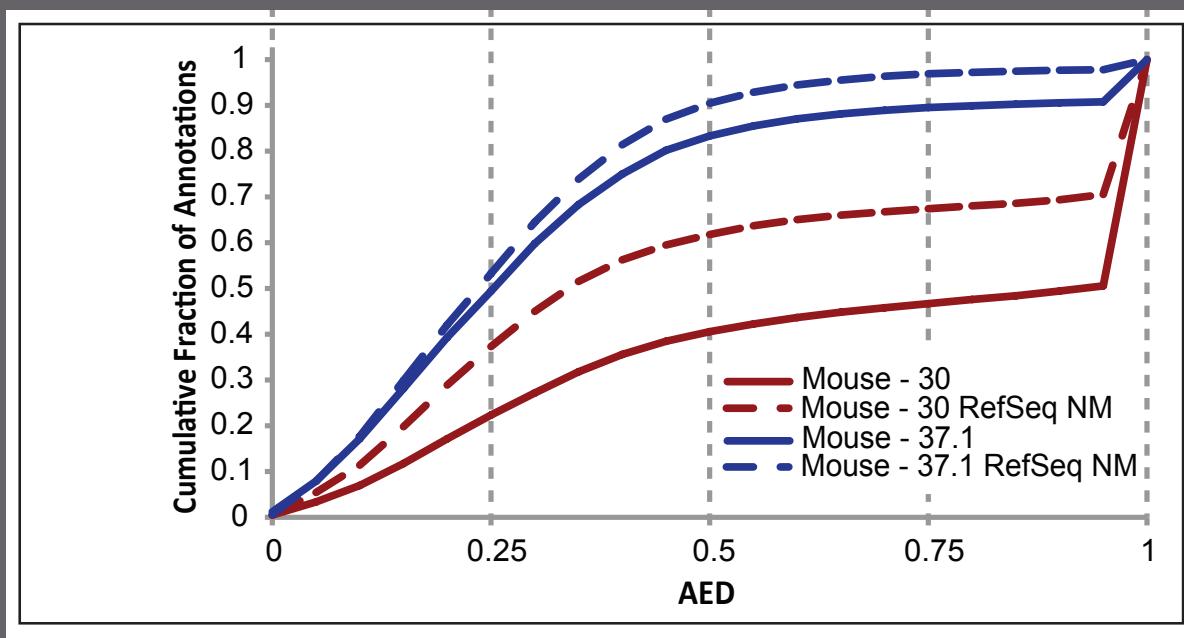


# Evaluation of AED on Arabidopsis

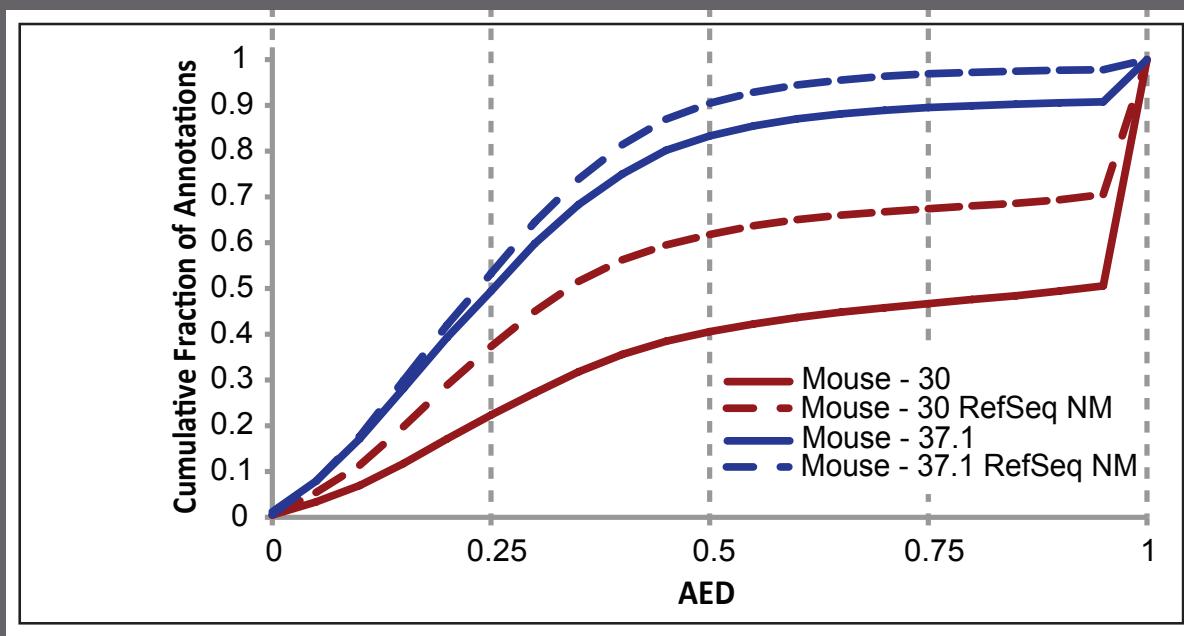


Evidence: mRNA-seq (17 experiments), ESTs, full length cDNAs, Swiss-Prot (minus Arabidopsis)

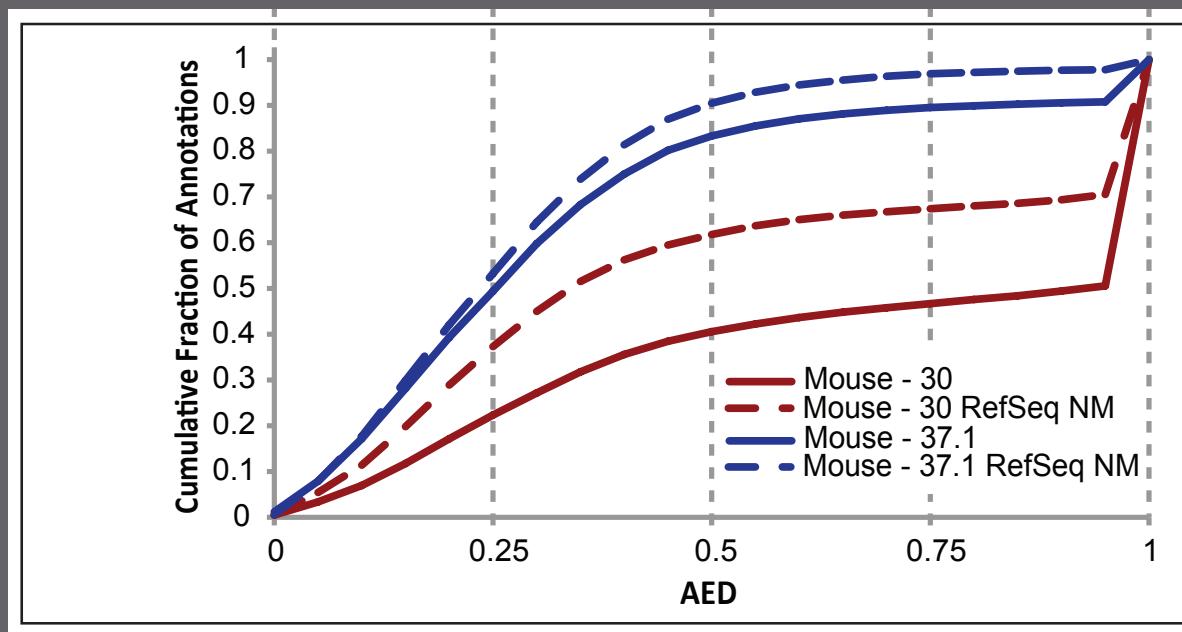
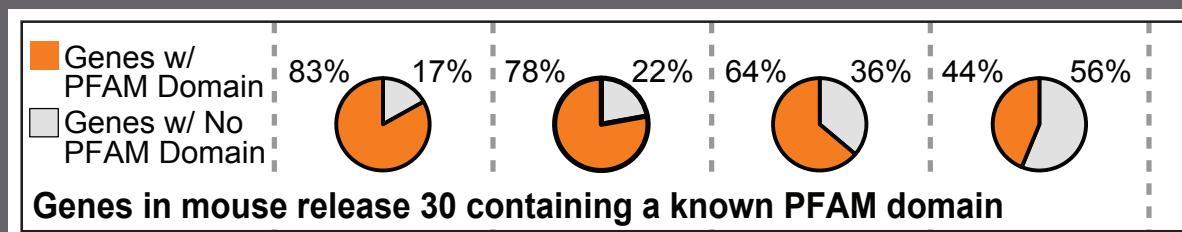
# Evaluation of AED on the mouse genome



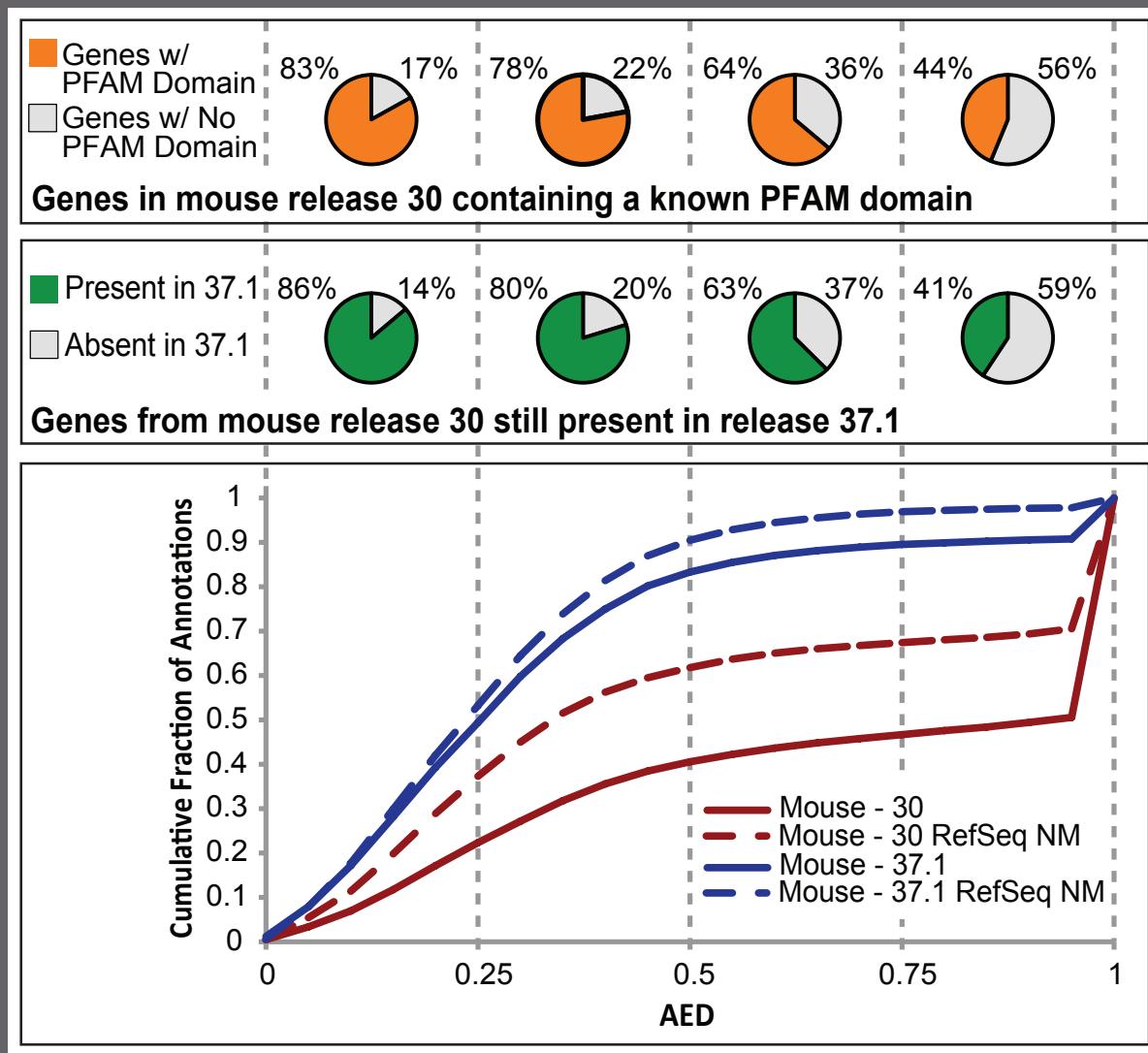
# Evaluation of AED on the mouse genome



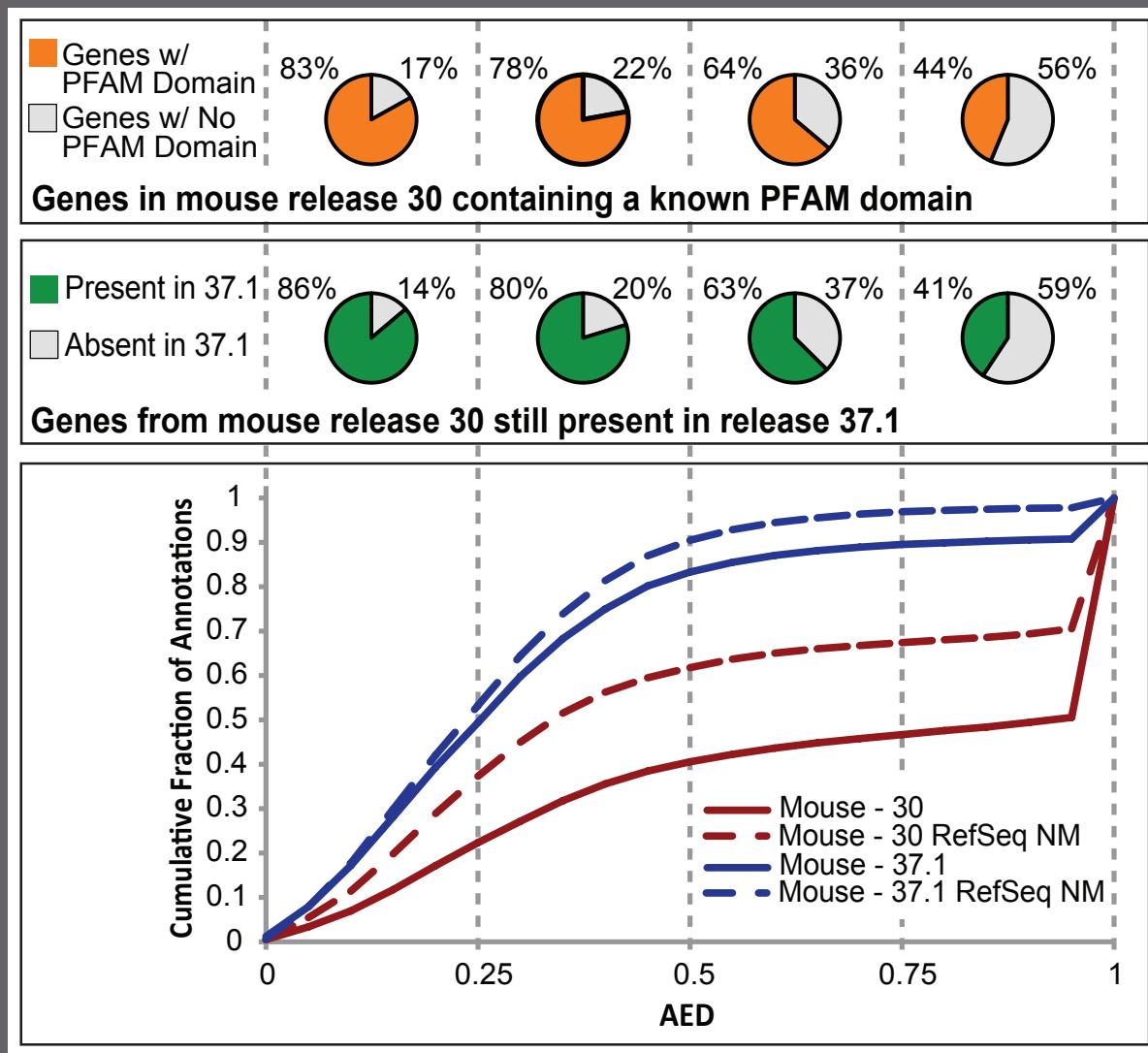
# Evaluation of AED on the mouse genome



# Evaluation of AED on the mouse genome

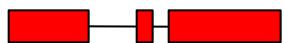


# Evaluation of AED on the mouse genome

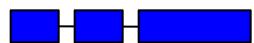


# Update/revise legacy annotation sets

# Update/revise legacy annotation sets



Legacy Annotation Set 1

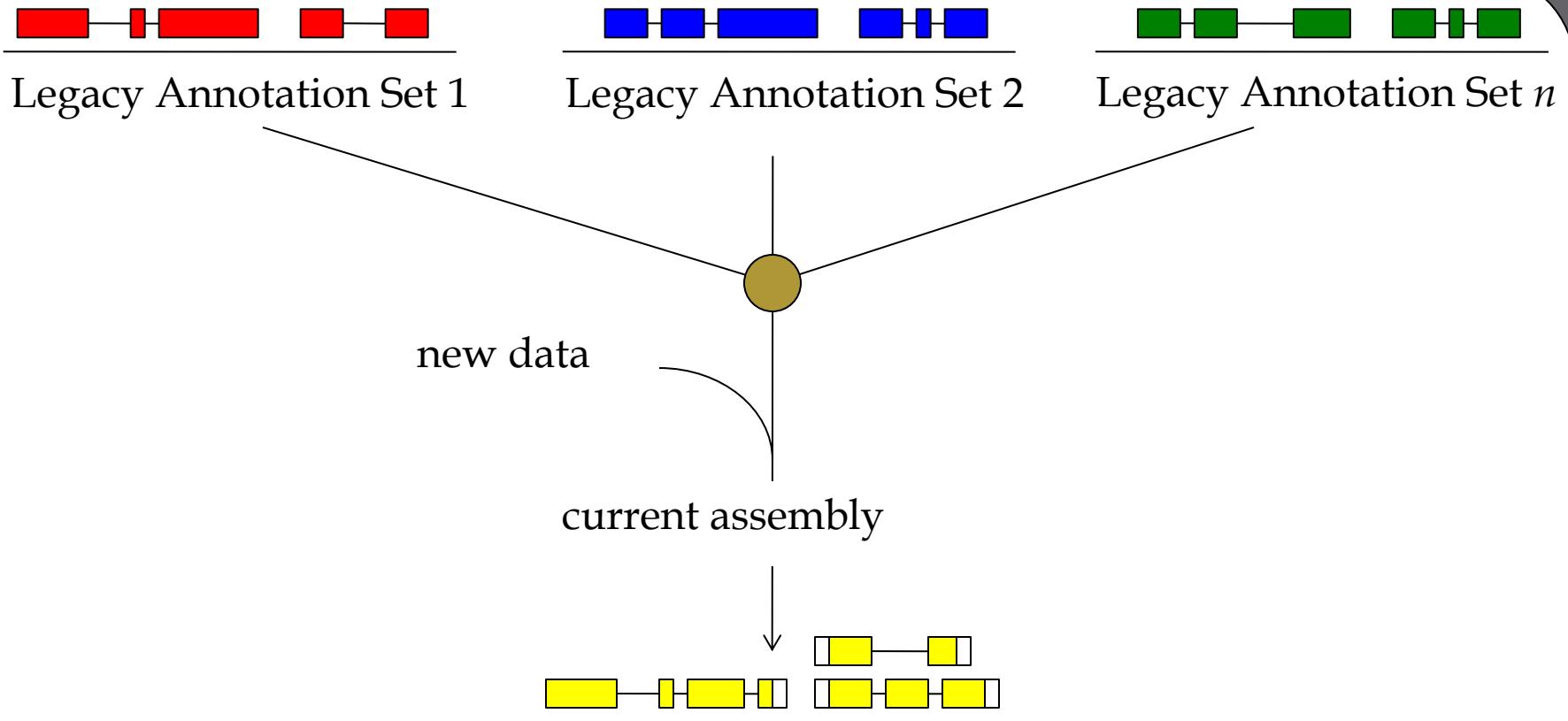


Legacy Annotation Set 2



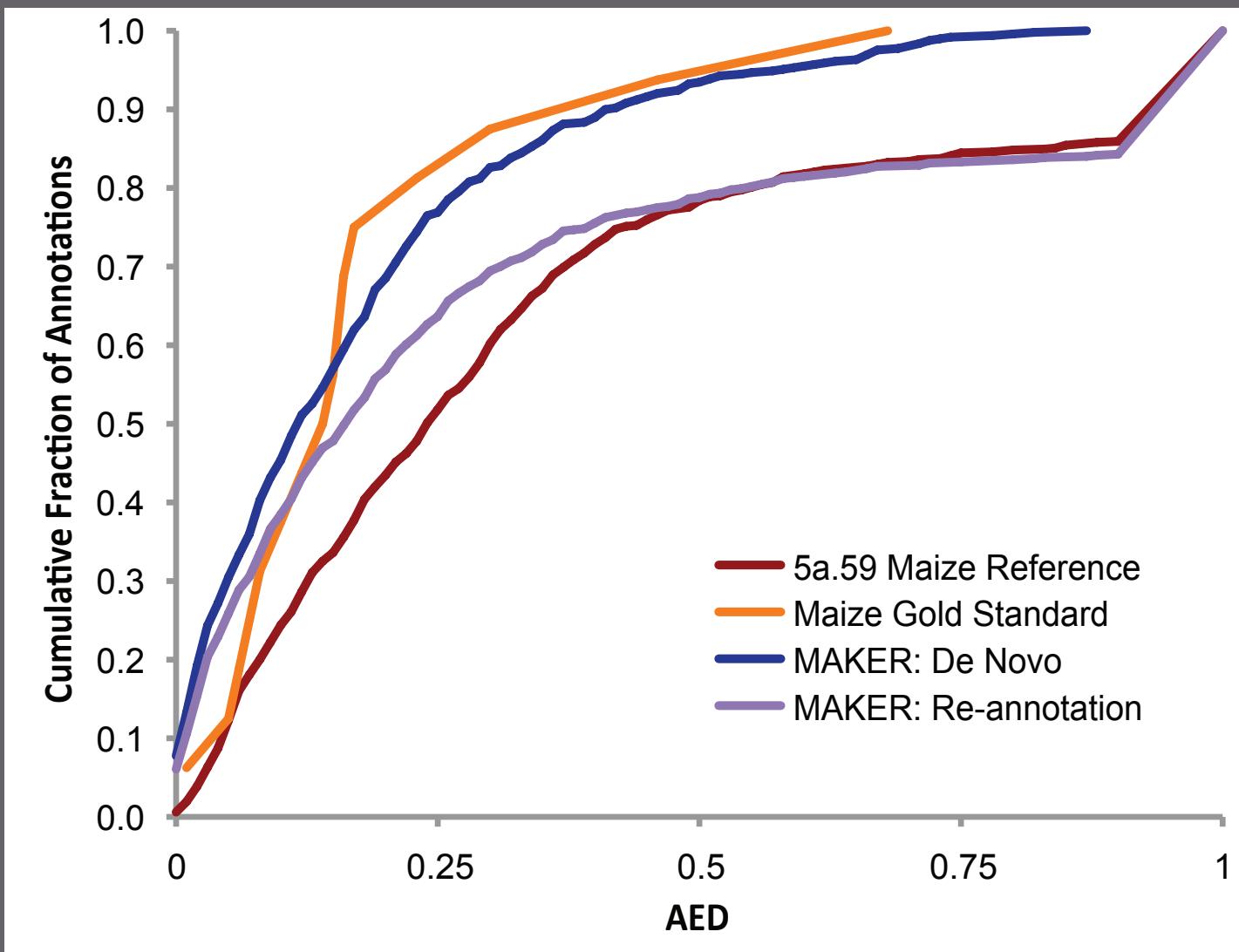
Legacy Annotation Set  $n$

# Update/revise legacy annotation sets



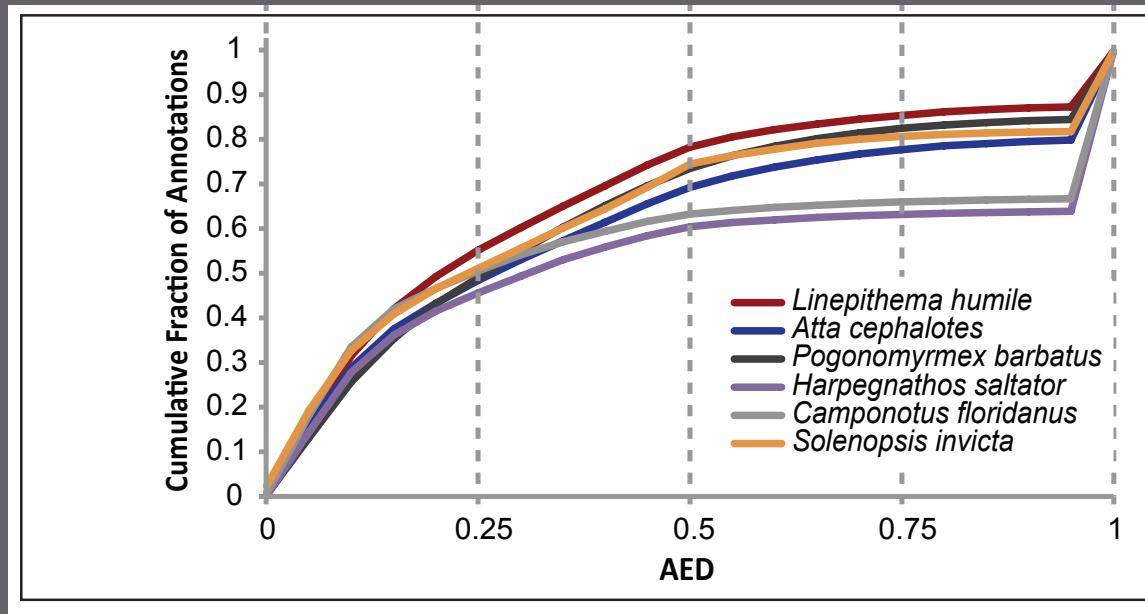
- Identify legacy annotation most consistent with new data
- Automatically revise it in light of new data
- If no existing annotation, create new one

# Re-annotation of the maize genome

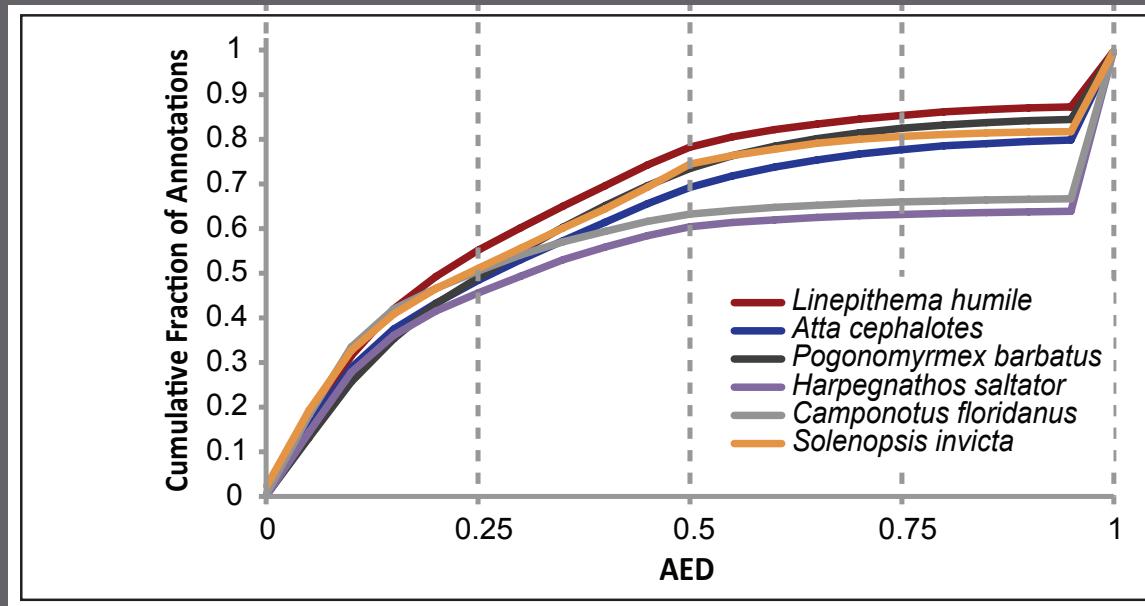
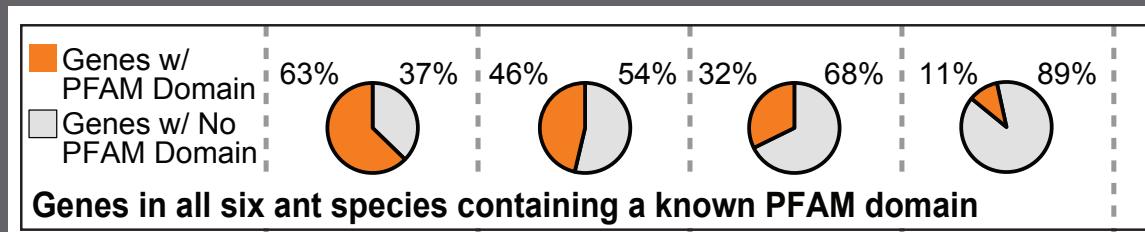


# Integrating new evidence into existing databases

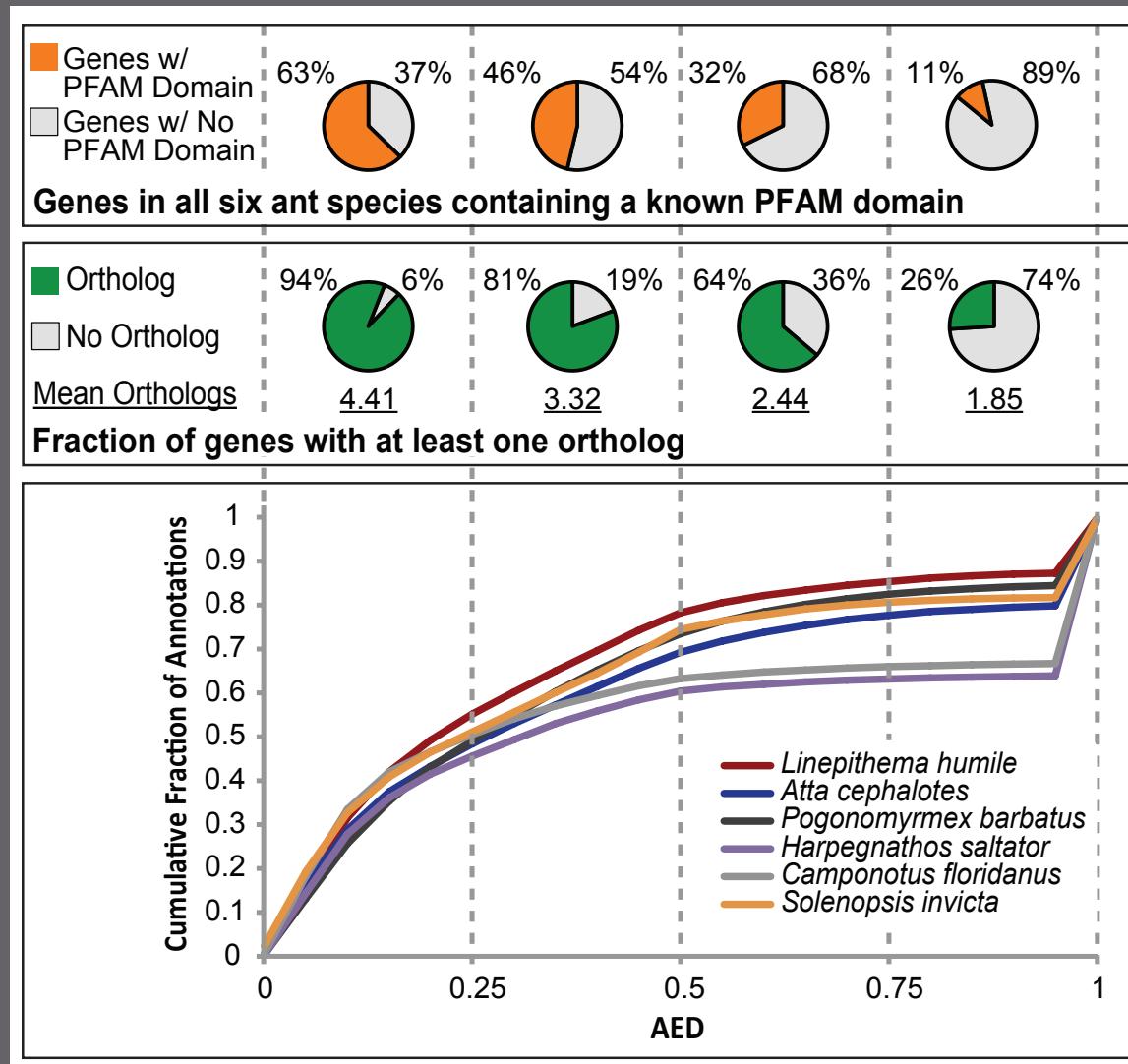
# Integrating new evidence into existing databases



# Integrating new evidence into existing databases



# Integrating new evidence into existing databases



Firefox File Edit View History Bookmarks Tools Window Help 02:38 Tue 1:01 AM

MAKER Web Annotation Service

http://derminger.genetics.utah.edu/cgi-bin/MWAS/maker.cgi

Google

Most Visited - Yandell Lab Reading Group - Ya... BioPerl BioPerl-Raw CHPC at the Univer... Notes ...

MAKER Web Annotation Service

 MAKER Web Annotation Service  
Your Genome Annotated

Home Help Yandell Lab

http://yandell-lab.org:10000

Welcome to the MAKER Web Annotation Service

Log into your account before, 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

<http://www.yandell-lab.org/>

MAKER Web Annotation Service

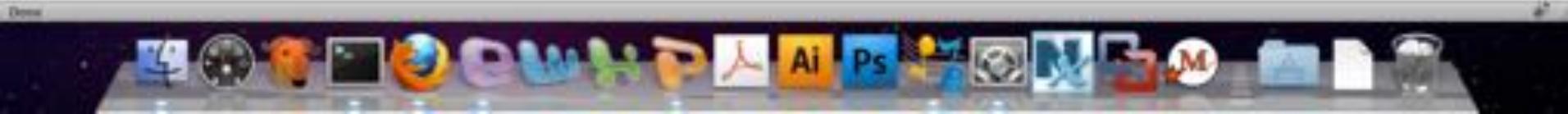
User Name:   
Password:   
 Remember User Name

New Guest Account [Create account](#) [Log In](#)

Forgot Password

© 2007-2010 Mark Yandell

YANDELL-LAB.ORG



www.technoegg.com

Example jobs are meant to be used with the [step-by-step tutorial](#) hosted by the Generic Model Organism Database Project. After clicking load for an example job, the fields below will be filled out for you. You can then review them or edit them. You will need to click on "Add Job to Queue" at the bottom of the page before the example job will start.

Carrie Ann

© Microsoft Corporation. All rights reserved.

14

10 AUGUST 2008 • JOURNAL OF CLIMATE

Assessment Set 686

ESTs from the same source as your genomic sequence; and (2) (optionally) ESTs from a closely related organism, for example if your genomic sequence is human, this second set of ESTs might be from mouse.

**Upload** a multi-FASTA file of ESTs to be aligned from the same source.

Select a file below



Most Visited - Yandell Lab Reading Group - Ya... BioPerl BioPerl-Bug CHPC at the Universit... Notes joneses  
MAKER Web Annotation Service

# MAKER Web Annotation Service

Your Genome Annotated

Home New Job Manage Files Running Jobs Edit Account Contact Us Help Yandell Lab

You are logged in as guest\_276 | Logout

Copy the URL [http://derminger.genetics.utah.edu/cgi-bin/MWAS/maker.cgi?username=guest\\_276](http://derminger.genetics.utah.edu/cgi-bin/MWAS/maker.cgi?username=guest_276) to netv.

Welcome to the MAKER

To get started just click on "the pre-loaded example annotation" as the results below. You can "Job Queue" above. For more

[Refresh Job Status](#)

Your Jobs (1)

JobID	Description
1323	Chromosome 17

© 2007-2009 Mark Yandell

## Annotation Status Summary:

TOTAL Configs: 1  
FINISHED: 1  
INCOMPLETE: 0  
FAILED: 0  
SKIPPED: 0

[Download All Data](#)

View config individually: [NT\\_011983.15](#)

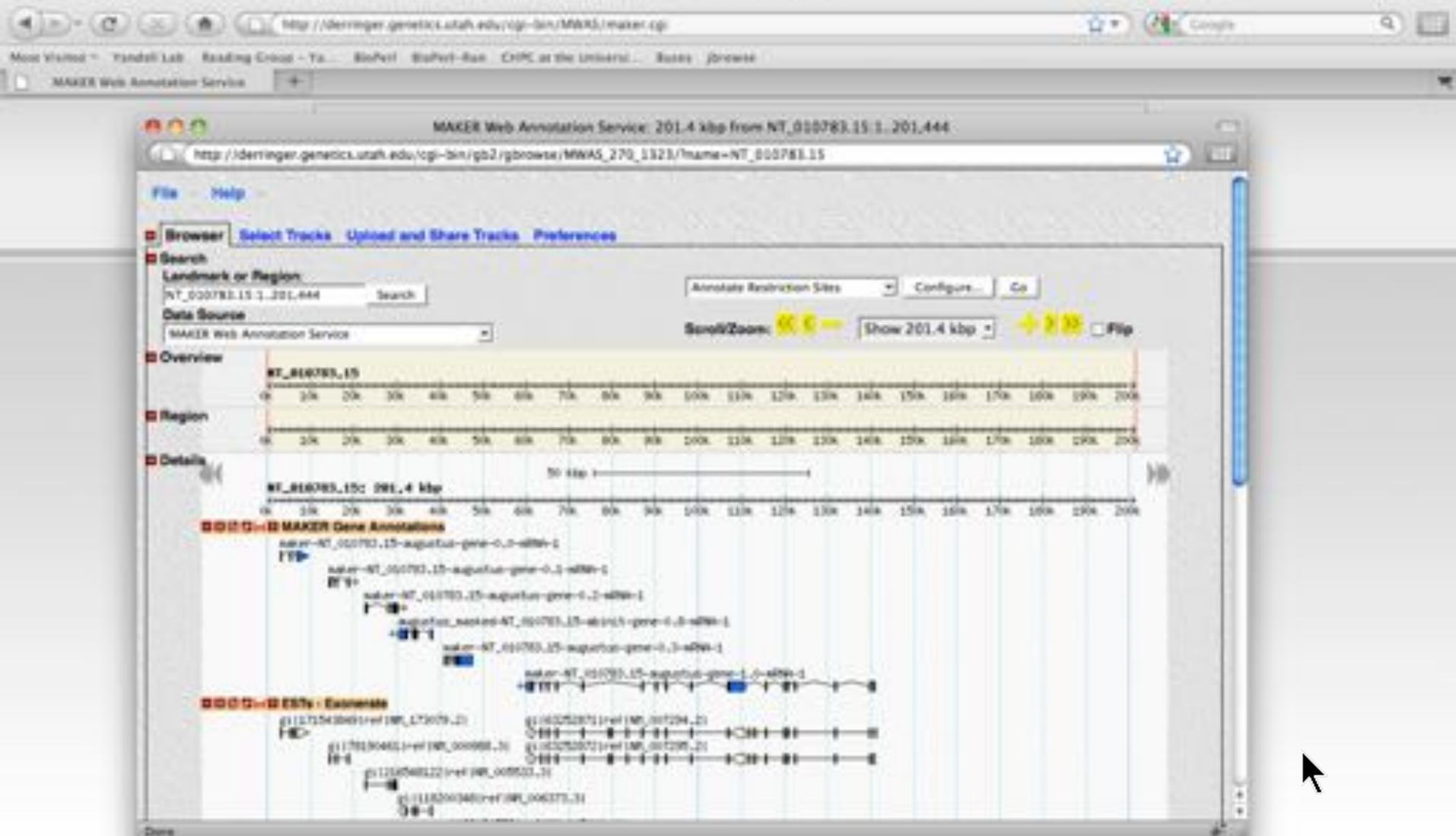
[View in Chimpse](#) [View in JBrowse](#)  
[View in Apollo](#) [SOBA Statistics](#)

Clicking on "Launch in Apollo" will install a Java Web Start version of Apollo if not already installed. If for some inexplicable reason the program doesn't start [click here](#).

[View Results](#)



VIEW CSS/HTML, 1.0



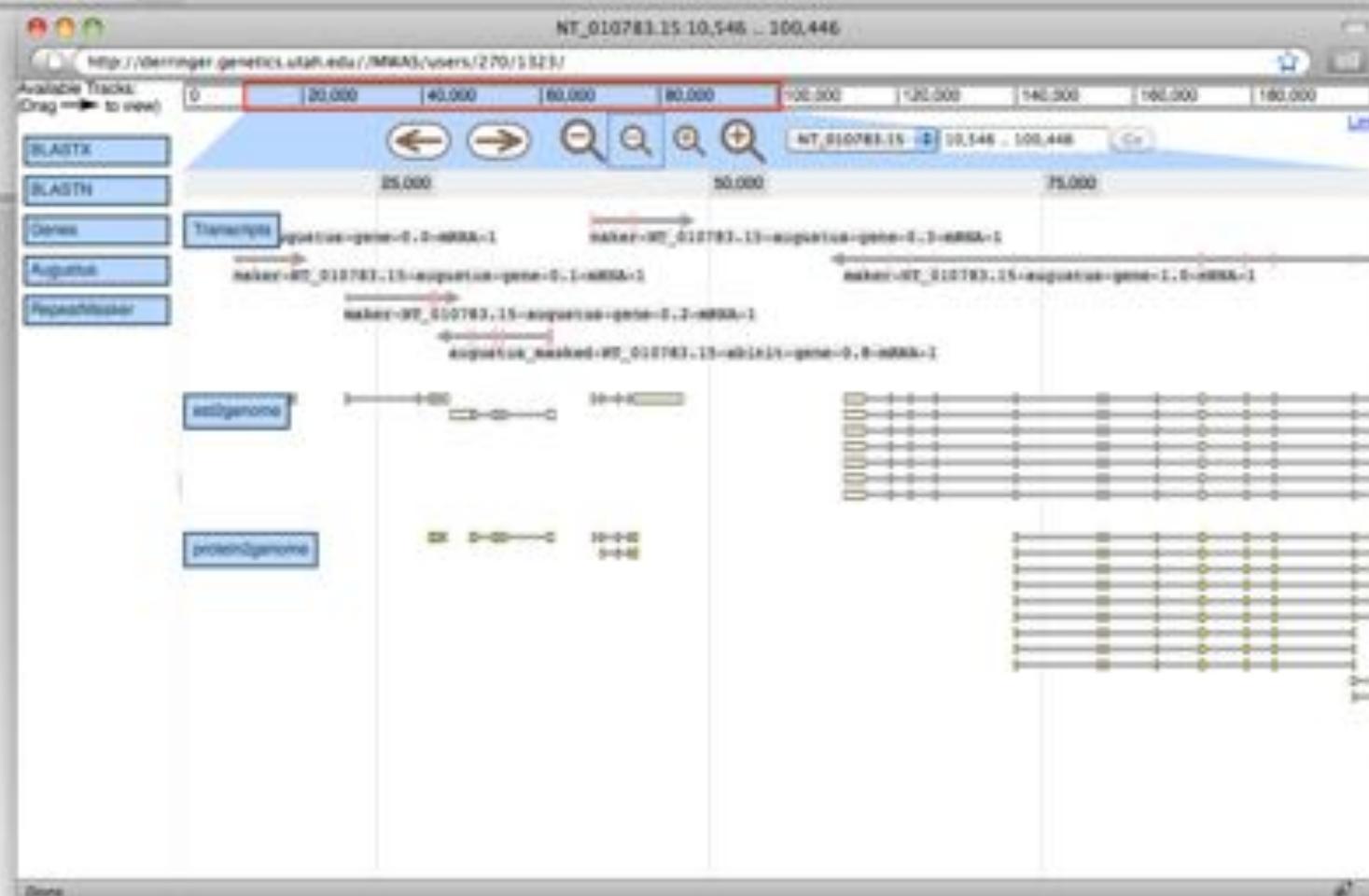
MANAGERIAL Accounting Services

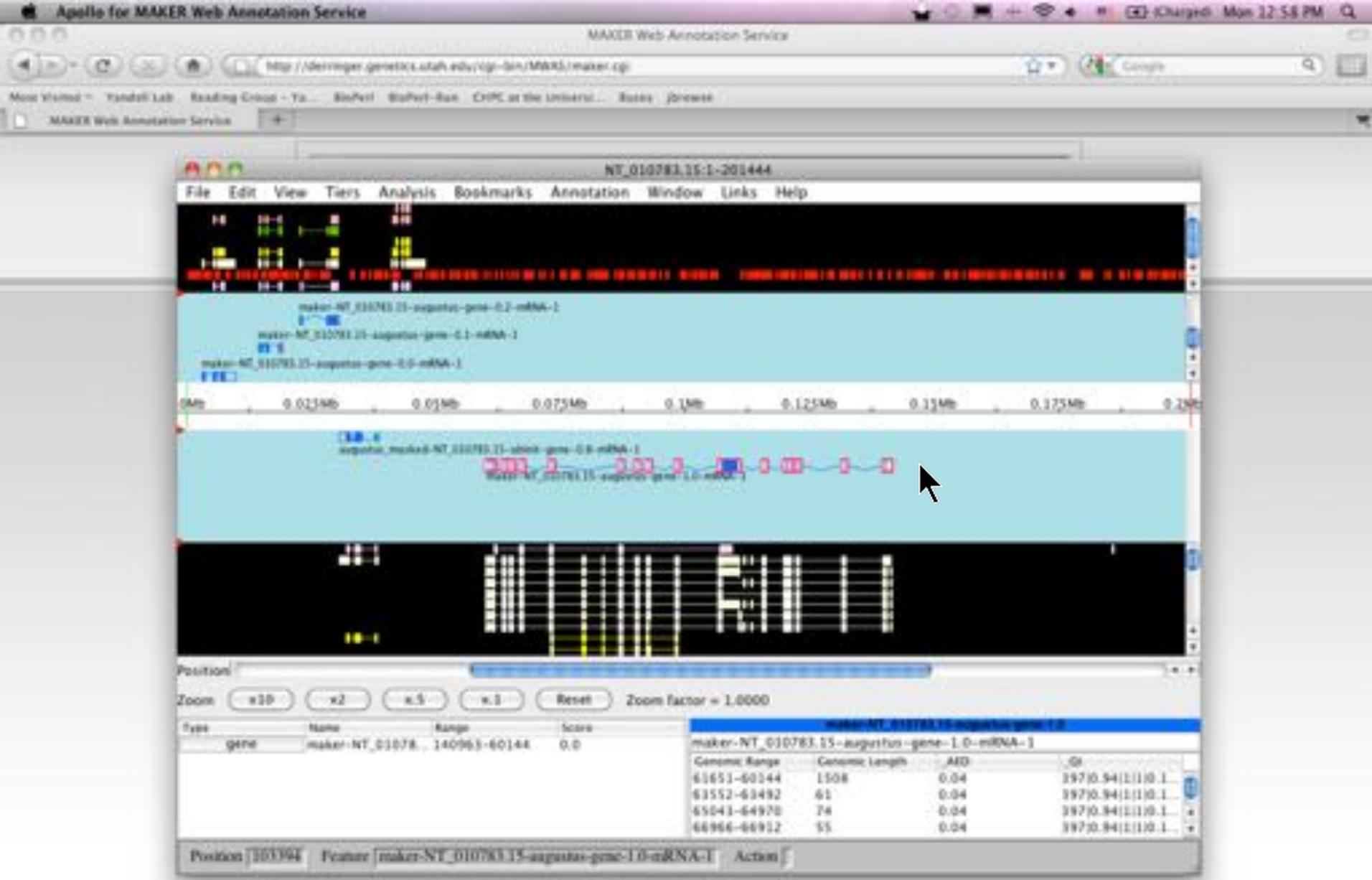
<http://www.elsevier.com/locate/jmca> | journal info

白山 81

10

Most Visited = [Yandell Lab](#) [Reading Group - Ya...](#) [BioPerl](#) [BioPerl-RNA](#) [CRPC at the Universi...](#) [Notes](#) [journals](#)





Most Visited: Randall Lab Reading Group - Ta... BioPerl BioPerl-Raw CHPC at the Universit... Biotools jbrowse

MAKER Web Annotation Service

SOBA



# SOBA

## Sequence Ontology Bioinformatic Analysis

[Home](#) [Browser](#) [Wiki](#) [GFF3](#) [GVF](#) [Resources](#) [About](#) [Request A Term](#) [Site Map](#)

## Genome Summary

For FileID: NT\_010783.15.gff

Below are the feature types and sources for the GFF3 file you uploaded. You must select at least one feature type and source and then click on the headers below to view the analyses.

## Feature Types

[Select All](#) [Unselect All](#) [Invert Selection](#)

sfCDS  
sfcontig  
sfexon  
sfexpressed\_sequence\_match  
sfgene  
sfmatch  
sfmatch\_part  
sfmRNA  
sfprotein\_match

[Select All](#) [Unselect All](#) [Invert Selection](#)

## Term Validation:

Term usage is OK!

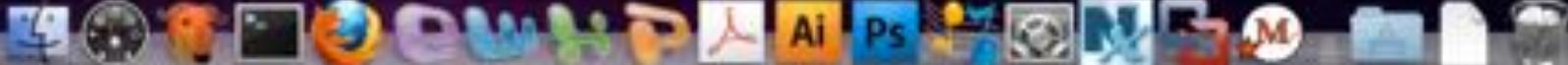
## Sources

[Select All](#) [Unselect All](#) [Invert Selection](#)

sf  
sfaugustus\_masked  
sfblastn  
sfblastx  
sfest2genome  
sfmaker  
sfprotein2genome  
sfrepearmasker

[Select All](#) [Unselect All](#) [Invert Selection](#)

Done



Most Visited - Yandell Lab Reading Group - Ya... BioPerl BioPerl-Bug CHPC at the Universit... Notes jonesse MAKER Web Annotation Service



# MAKER Web Annotation Service

Your Genome Annotated

[Home](#)[New Job](#)[Manage Files](#)[Running Jobs](#)[Edit Account](#)[Contact Us](#)[Help](#)[Yandell Lab](#)

Logged in as guest\_279 | Logout

You are logged in as a guest user. You may become a registered user at any time by clicking on "Edit Account".

Copy the URL [http://derminger.genetics.utah.edu/cgi-bin/MWAS/maker.cgi?guest\\_id=279](http://derminger.genetics.utah.edu/cgi-bin/MWAS/maker.cgi?guest_id=279) or login with the username [guest\\_279](#) to return to your jobs after exiting.



MAKER Job Details

Assigned Id:

1655

## Annotation Post Processing

Indicate the post processing you would like to perform on the annotations. You can add functional annotations such as putative gene functions, Gene Ontology terms, and protein domains to the gene models. You can also change long MAKER assigned names into shorter gene names.

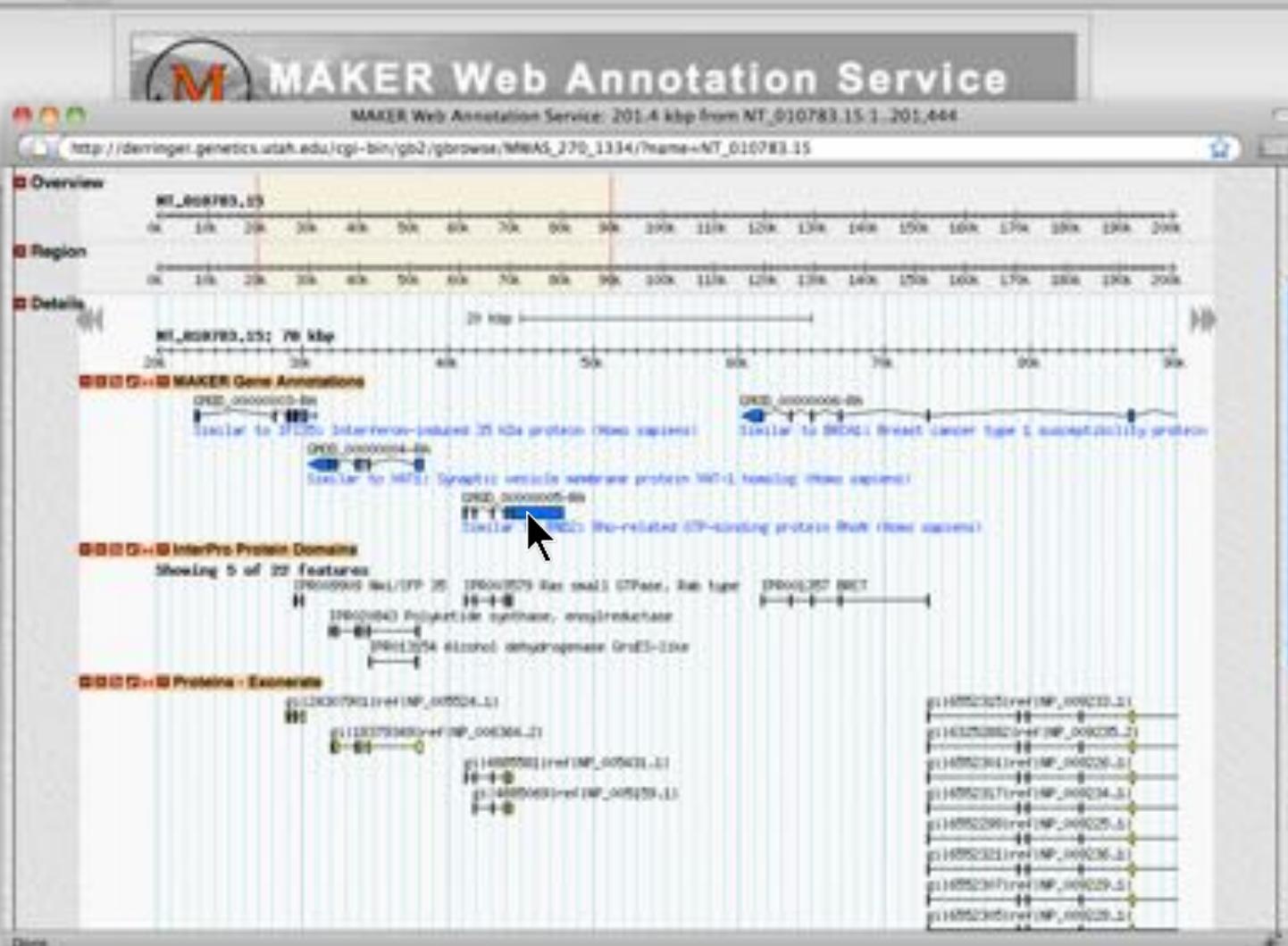
Add putative gene functions via comparison to the UniProt/Swiss-Prot protein database

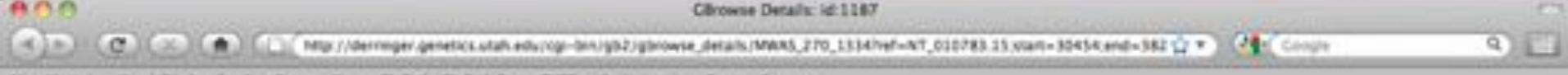
Add protein domains and GO terms via InterProScan

Choose a prefix for making shorter gene names (blank to ignore)

CMOD..

[Add Job to Queue](#)[Save/Come Back Later](#)[Clear](#)





## GMD0\_0000004-RA Details

Name: GMD0\_0000004-RA  
Type: mRNA

Description: Similar to VKT1: Synaptic vesicle membrane protein VAT-1 homolog (Homo sapiens)

Source: maker

Position: NT\_010783.15:30454..306291 (- strand)

Length: 7838

Aliases: augustus\_masked-NT\_010783.15-ab init-gene-0.8-mRNA-1

Discret: Gene3D G3DSA/3.40/50.725

InterPro:IPR002085

InterPro:IPR002364

InterPro:IPR011232

InterPro:IPR013149

InterPro:IPR013154

InterPro:IPR016040

InterPro:IPR020843

PANTHER:PTHR11695

PANTHER:PTHR11695\_SF29

PIR:PF00107

PIR:PF00245

Prosite:PS01162

SMART:SM00629

superfamily:SSF51739

superfamily:SSF51736

Note: Similar to VKT1: Synaptic vesicle membrane protein VAT-1 homolog (Homo sapiens)

Ontology term: GO:0003824

GO:0005488

GO:0008190

GO:0008270

GO:0016491

GO:0055114

Done



# MAKER Versions

## □ MAKER

- 2008. Based on earlier annotation pipelines developed by Mark Yandell

## □ MAKER 2 / MAKER-P

- 2011. Introduction of MPI parallelization, support for multiple gene predictors, GFF3 pass-through, and quality metrics like AED (Annotation Edit Distance) from the Sequence Ontology consortium.
- 2015. Support for tRNA and snoRNA annotation. Improved parallelization on large plant genomes.

## □ MAKER 3

- 2016. EVM (Evidence Modeler) support for improved annotation and user defined evidence probability weighting.

# MAKER Wiki

<http://weatherby.genetics.utah.edu/MAKER/wiki>

# Acknowledgements

- Mark Yandell
- Barry Moore
- Michael Campbell
- Daniel Ence