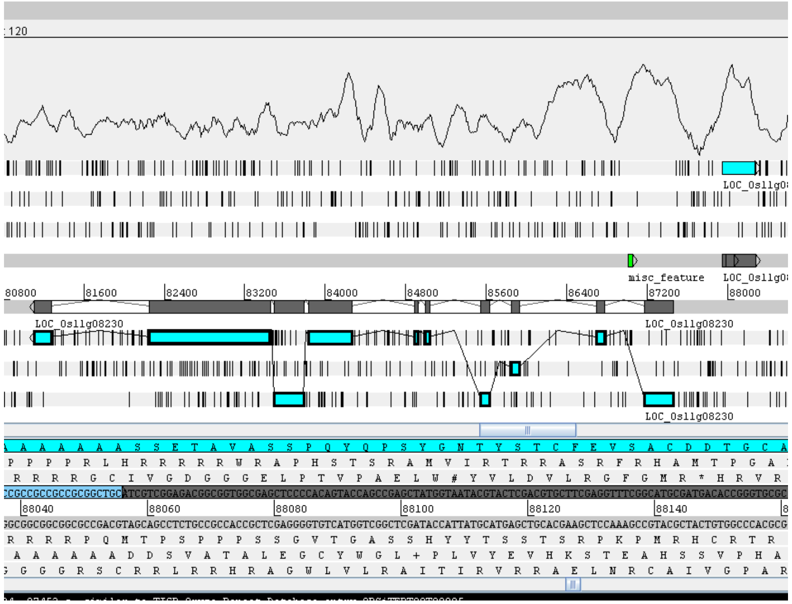
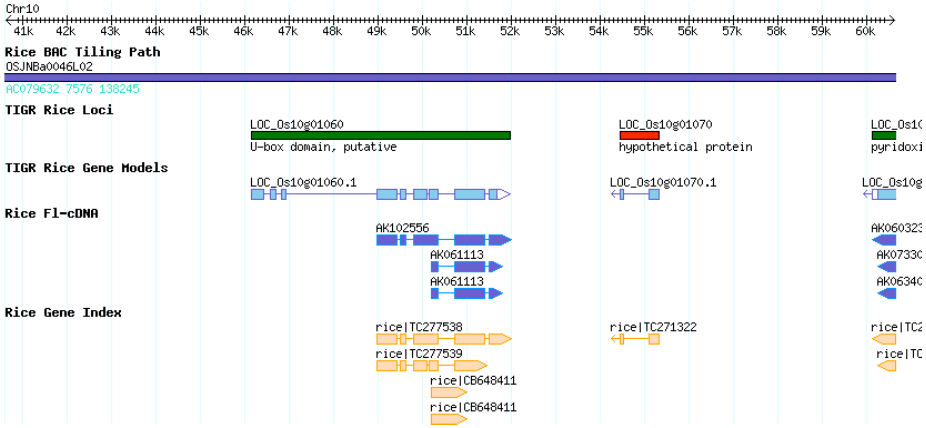
# Viewing Tools For Annotation Data





## TIGR Plant Genome Annotation Workshop
## May 2007

# Viewing Tools For Annotation

- Annotation viewing tools allow you to visualize the location of features in respect to each other.
- They can also display evidence alignments and other information relevant to the process of annotation
- Annotation editors allow the user to make changes in the annotation of a sequence and save these changes to a database.
- This talk will focus on open source annotation viewers and editors, however many commercial solutions are available also.

# Artemis

http://www.sanger.ac.uk/Software/Artemis

- Reads in Genbank, EMBL and GFF3 files and sequence in FASTA format.
- Displays annotation derived from a feature table.
- Has a selection of tools such as a G/C graph.
- Allows the direct display of blast results.
- Artemis is Java based, easy to run and install.
- Allows changing of the annotation and export in GenBank, EMBL, and GFF3 format

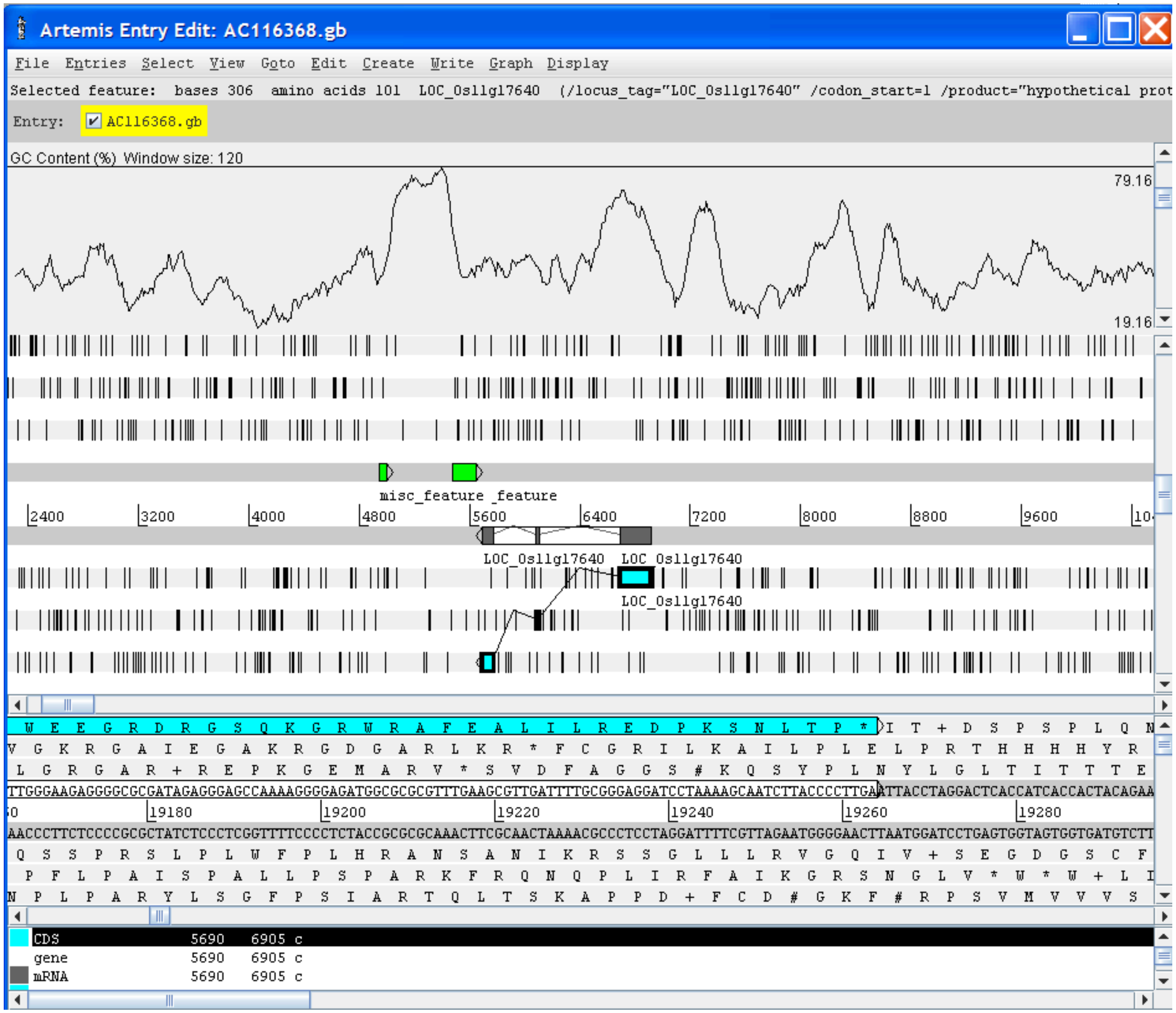
# Artemis Display – BAC AC116368

GC Plot

Sequence
And
Feature
Overview

DNA View
And
6 Frame
Translation

Feature Table



Selected feature: bases 306 amino acids 101 LOC_Os11g17640 (/locus_tag="LOC_Os11g17640" /codon_start=1 /product="hypothetical prot

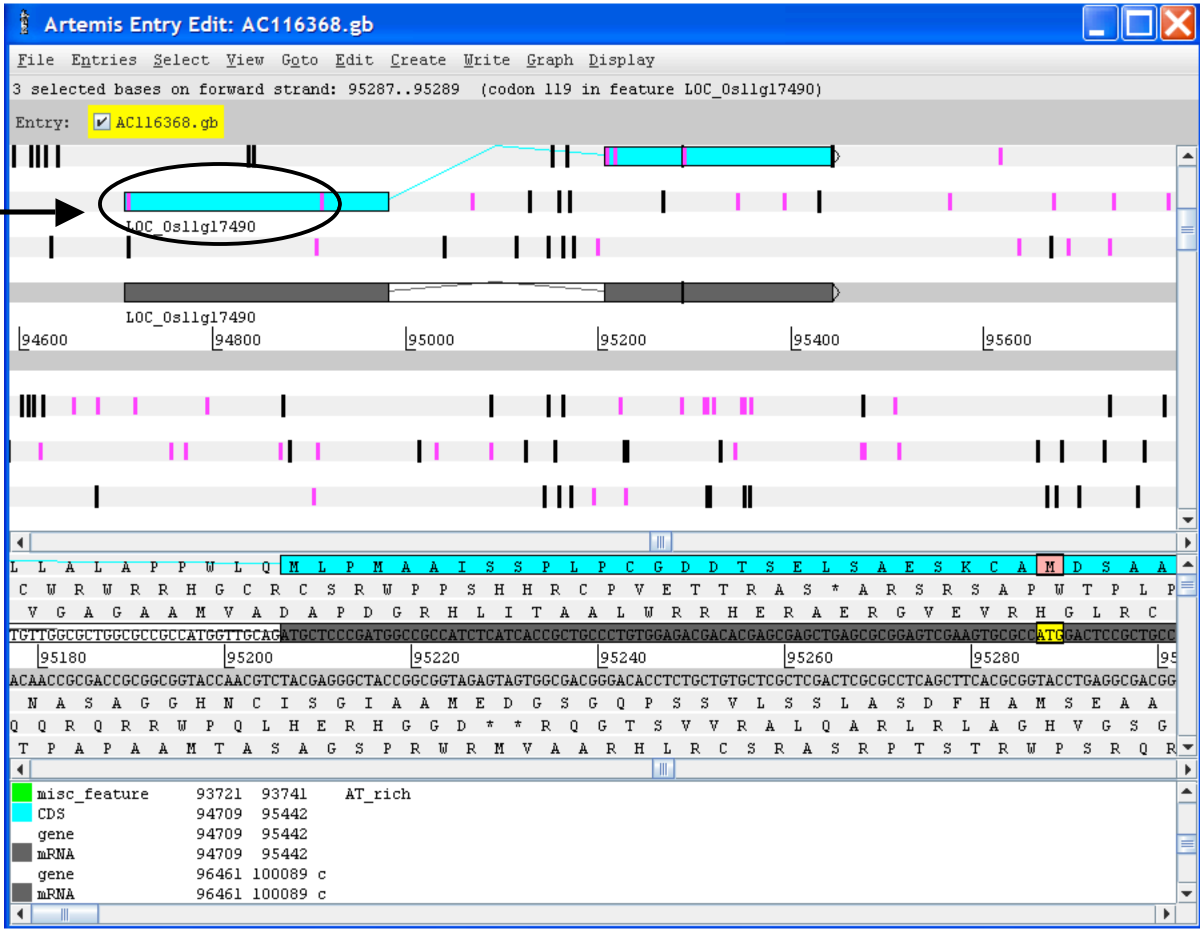
Entry: ☑ AC116368.gb

```
 W E E G R D R G S Q K G R W R A F E A L I L R E D P K S N L T P * I T + D S P S P L Q N
V G K R G A I E G A K R G D G A R L K R * F C G R I L K A I L P L E L P R T H H H H Y R
 L G R G A R + R E P K G E M A R V * S V D F A G G S # K Q S Y P L N Y L G L T I T T T E
TTGGGAAGAGGGGCGCGATAGAGGGAGCCAAAAGGGGAGATGGCGCGTTTGAAGCGTTGATTTTGCGGGAGGATCCTAAAAGCAATCTTACCCCTTGAATTACCTAGGACTCACCATCACCACTACAGAA
0         19180       19200       19220       19240       19260       19280
AACCCTTCTCCCCGCGCTATCTCCCTCGGTTTTCCCCTCTACCGCGCGCAAACTTCGCAACTAAAACGCCCTCCTAGGATTTTCGTTAGAATGGGGAACTTAATGGATCCTGAGTGGTAGTGGTGATGTCTT
 Q S S P R S L P L W F P L H R A N S A N I K R S S G L L L R V G Q I V + S E G D G S C F
  P F L P A I S P A L L P S P A R K F R Q N Q P L I R F A I K G R S N G L V * W * W + L I
N P L P A R Y L S G F P S I A R T Q L T S K A P P D + F C D # G K F # R P S V M V V V S
```

|  |  |  |  |  |
|---|---|---|---|---|
| 🟦 | CDS | 5690 | 6905 | c |
|  | gene | 5690 | 6905 | c |
| ■ | mRNA | 5690 | 6905 | c |

# Artemis as an Annotation Tool

Evidence shows that this region should be annotated as a 5'-UTR (hypothetically)



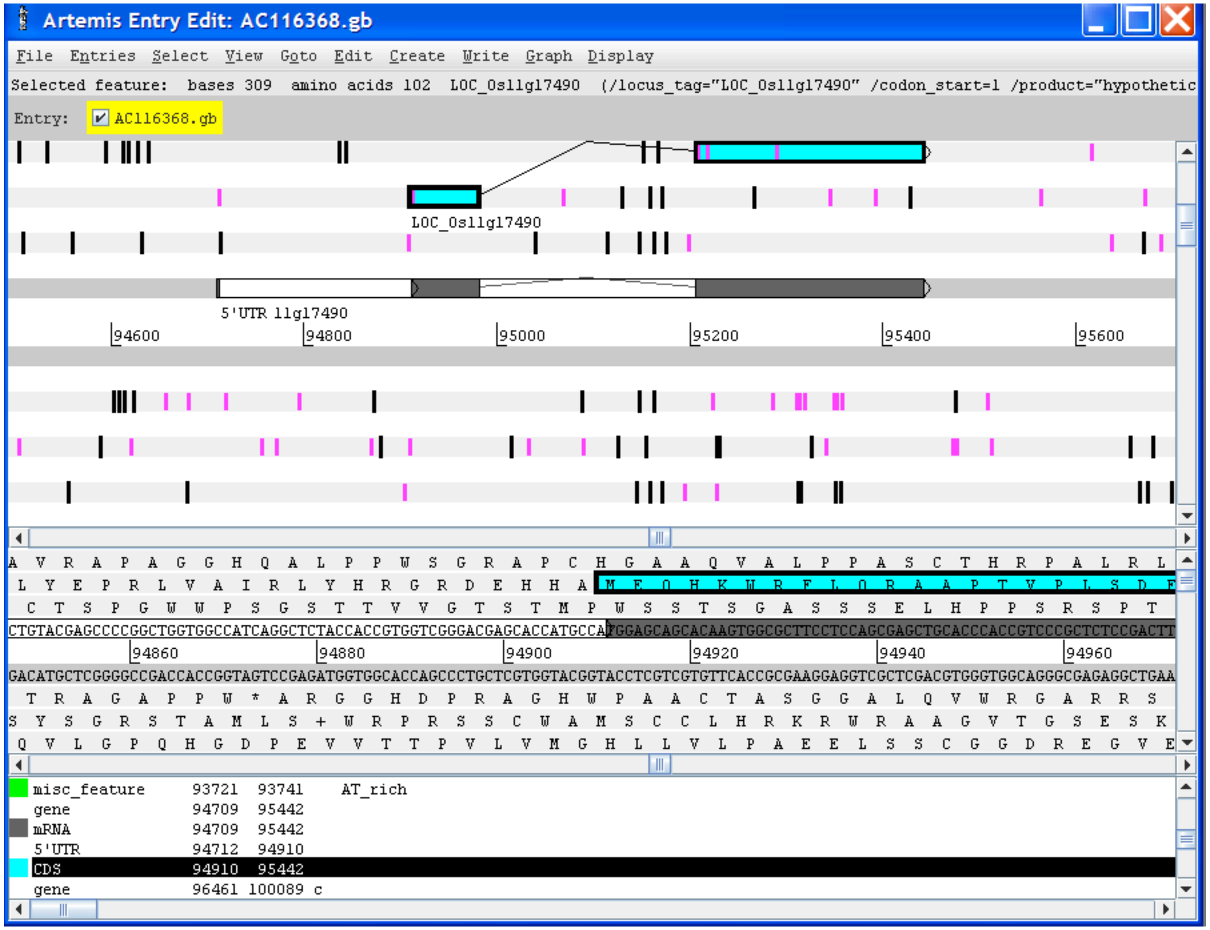
3 selected bases on forward strand: 95287..95289 (codon 119 in feature LOC_Os11g17490)

```
L  L  A  L  A  P  P  W  L  Q  M  L  P  M  A  A  I  S  S  P  L  P  C  G  D  D  T  S  E  L  S  A  E  S  K  C  A  M  D  S  A  A
 C  W  R  W  R  R  H  G  C  R  C  S  R  W  P  P  S  H  H  R  C  P  V  E  T  T  R  A  S  *  A  R  S  R  S  A  P  W  T  P  L  P
  V  G  A  G  A  A  M  V  A  D  A  P  D  G  R  H  L  I  T  A  A  L  W  R  R  H  E  R  A  E  R  G  V  E  V  R  H  G  L  R  C
TGTTGGCGCTGGCGCCGCCATGGTTGCAGATGCTCCCGATGGCCGCCATCTCATCACCGCTGCCCTGTGGAGACGACACGAGCGAGCTGAGCGCGGAGTCGAAGTGCGCCATGGACTCCGCTGCC
   95180               95200               95220               95240               95260               95280               95
ACAACCGCGACCGCGGCGGTACCAACGTCTACGAGGGCTACCGGCGGTAGAGTAGTGGCGACGGGACACCTCTGCTGTGCTCGCTCGACTCGCGCCTCAGCTTCACGCGGTACCTGAGGCGACGG
  N  A  S  A  G  G  H  N  C  I  S  G  I  A  A  M  E  D  G  S  G  Q  P  S  S  V  L  S  S  L  A  S  D  F  H  A  M  S  E  A  A
Q  Q  R  Q  R  R  W  P  Q  L  H  E  R  H  G  G  D  *  *  R  Q  G  T  S  V  V  R  A  L  Q  A  R  L  R  L  A  G  H  V  G  S  G
 T  P  A  P  A  A  M  T  A  S  A  G  S  P  R  W  R  M  V  A  A  R  H  L  R  C  S  R  A  S  R  P  T  S  T  R  W  P  S  R  Q  R
```

```
misc_feature      93721   93741     AT_rich
CDS               94709   95442
gene              94709   95442
mRNA              94709   95442
gene              96461  100089 c
mRNA              96461  100089 c
```

# Artemis as an Annotation Tool

5’ UTR
inserted
in Gene
model



Selected feature: bases 309 amino acids 102 LOC_Os11g17490 (/locus_tag="LOC_Os11g17490" /codon_start=1 /product="hypothetic

Entry: ☑ AC116368.gb

```
A  V  R  A  P  A  G  G  H  Q  A  L  P  P  W  S  G  R  A  P  C  H  G  A  A  Q  V  A  L  P  P  A  S  C  T  H  R  P  A  L  R  L
 L  Y  E  P  R  L  V  A  I  R  L  Y  H  R  G  R  D  E  H  H  A  M  E  Q  H  K  W  R  F  L  Q  R  A  A  P  T  V  P  L  S  D  F
  C  T  S  P  G  W  W  P  S  G  S  T  T  V  V  G  T  S  T  M  P  W  S  S  T  S  G  A  S  S  S  E  L  H  P  P  S  R  S  P  T
CTGTACGAGCCCCGGCTGGTGGCCATCAGGCTCTACCACCGTGGTCGGGACGAGCACCATGCCATGGAGCAGCACAAGTGGCGCTTCCTCCAGCGAGCTGCACCCACCGTCCCGCTCTCCGACTT
             94860               94880               94900               94920               94940               94960
GACATGCTCGGGGCCGACCACCGGTAGTCCGAGATGGTGGCACCAGCCCTGCTCGTGGTACGGTACCTCGTCGTGTTCACCGCGAAGGAGGTCGCTCGACGTGGGTGGCAGGGCGAGAGGCTGAA
  T  R  A  G  A  P  P  W  *  A  R  G  G  H  D  P  R  A  G  H  W  P  A  A  C  T  A  S  G  G  A  L  Q  V  W  R  G  A  R  R  S
S  Y  S  G  R  S  T  A  M  L  S  +  W  R  P  R  S  S  C  W  A  M  S  C  C  L  H  R  K  R  W  R  A  A  G  V  T  G  S  E  S  K
 Q  V  L  G  P  Q  H  G  D  P  E  V  V  T  T  P  V  L  V  M  G  H  L  L  V  L  P  A  E  E  L  S  S  C  G  G  D  R  E  G  V  E
```

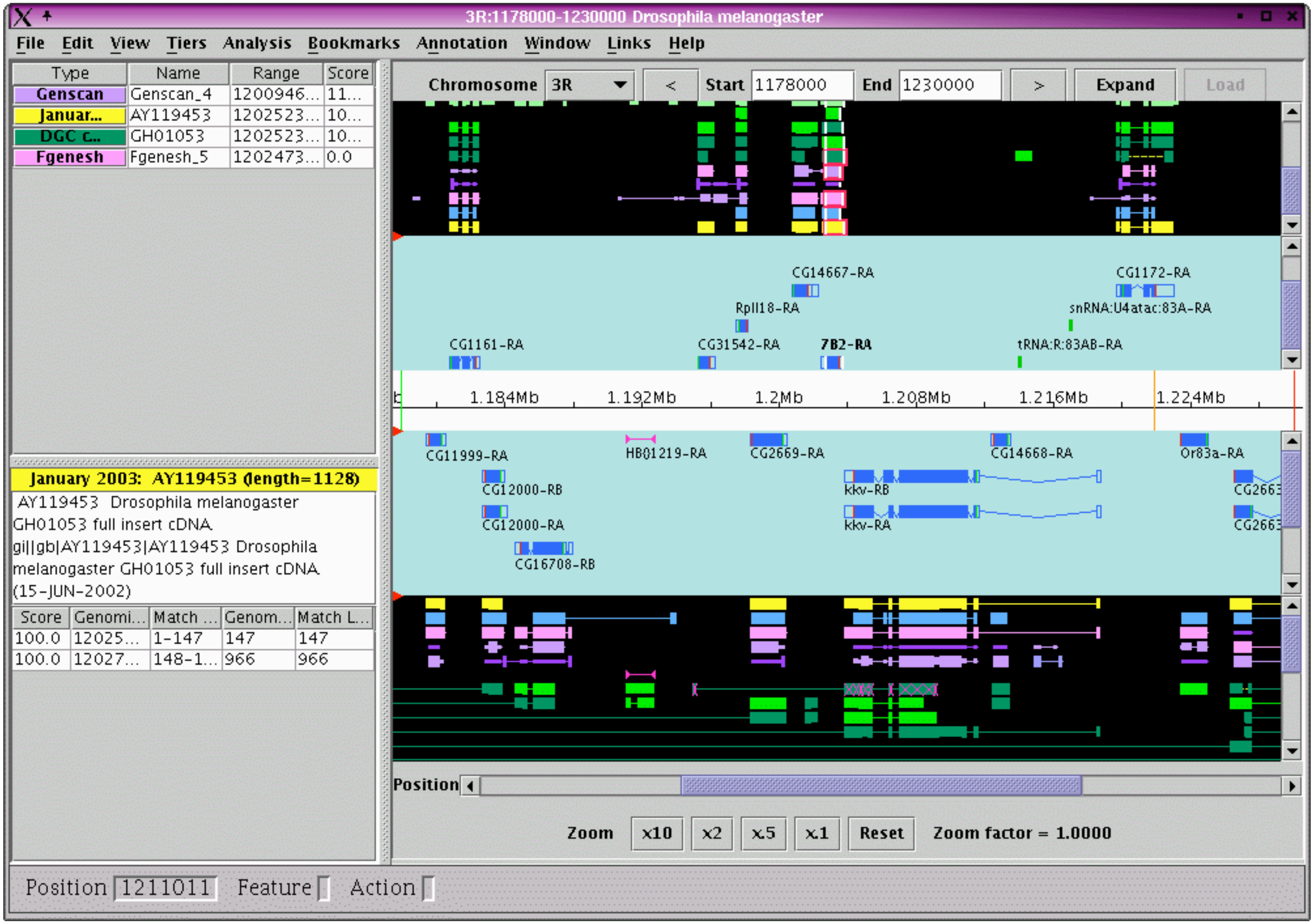
```
misc_feature    93721  93741    AT_rich
gene            94709  95442
mRNA            94709  95442
5'UTR           94712  94910
CDS             94910  95442
gene            96461 100089 c
```

# Apollo Genome Editor

http://www.gmod.org/apollo.shtml

- Apollo is a more heavy duty genome editor than Artemis
- Gene predictions, EST, FL-cDNA, protein alignments and splice sites can be displayed in tiers
- Edge matches and start/stop codons are shown
- Various annotation helpers are available such as gene reports, alignments of evidence using Jalview, exon details
- A variety of input formats can be used to load into Apollo (GAME-XML, GFF, Chado XML)
- A chado schema database can also be used as a data source

# Apollo Screen Shot



3R:1178000-1230000 Drosophila melanogaster

Chromosome 3R

Start 1178000

End 1230000

| Type | Name | Range | Score |
|---|---|---|---|
| Genscan | Genscan_4 | 1200946... | 11... |
| Januar... | AY119453 | 1202523... | 10... |
| DGC c... | GH01053 | 1202523... | 10... |
| Fgenesh | Fgenesh_5 | 1202473... | 0.0 |

**January 2003: AY119453 (length=1128)**

AY119453 Drosophila melanogaster GH01053 full insert cDNA. gi||gb|AY119453|AY119453 Drosophila melanogaster GH01053 full insert cDNA. (15-JUN-2002)

| Score | Genomi... | Match ... | Genom... | Match L... |
|---|---|---|---|---|
| 100.0 | 12025... | 1-147 | 147 | 147 |
| 100.0 | 12027... | 148-1... | 966 | 966 |

Zoom factor = 1.0000

Position 1211011 Feature Action

# Gbrowse – Generic Genome Browser

- Part of the Generic Model Organism Database (GMOD) project (www.gmod.org/?q=node/71).
- A web based annotation viewer
- Annotation data is read directly from GFF files in memory, or the GFF file and sequence are stored in a database
- Various plugins are available for enhanced functionality
- Simple to set up on your own local server or desktop computer

# Gbrowse – Generic Genome Browser

- Annotation data is displayed in user selectable tracks as ‘glyphs’.
- A range of shapes and colors for the glyphs are available.
- Perl subroutines can be used to change the glyphs on the fly.
- DNA sequence can be loaded into the database and displayed in a track along with the annotation.
- Annotation provided by the user can also be displayed in a track. The annotation data can be in a file on a remote server.

# TIGR Rice Genome Browser

http://www.tigr.org/tigr-scripts/osa1_web/gbrowse/rice

Chromosome Selection

Overview Bar

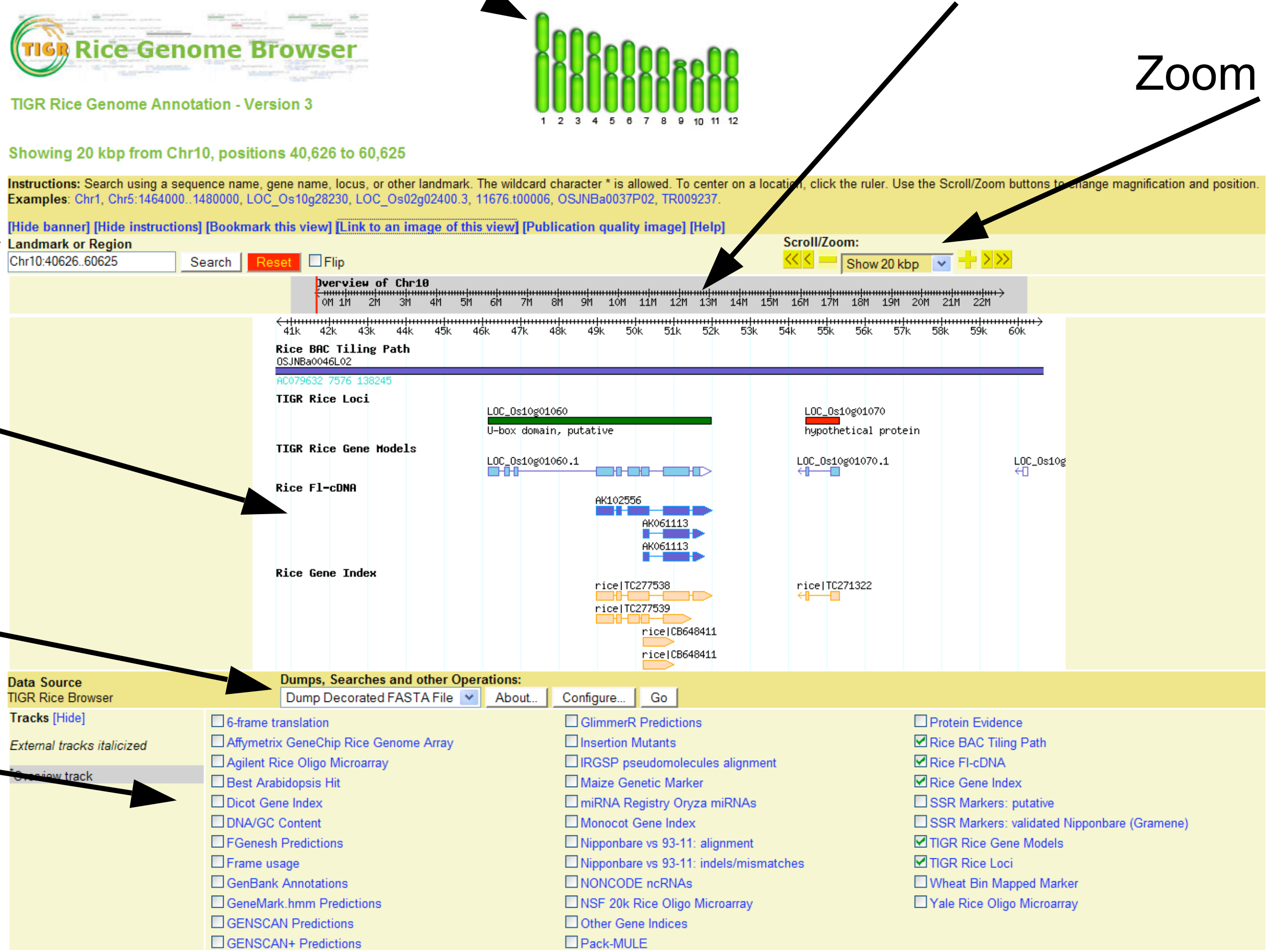
Zoom Controls

Name or Region Search

Tracks Displaying Annotation

Plugin Controls

Track Selection



TIGR Rice Genome Annotation - Version 3

**Showing 20 kbp from Chr10, positions 40,626 to 60,625**

**Instructions:** Search using a sequence name, gene name, locus, or other landmark. The wildcard character * is allowed. To center on a location, click the ruler. Use the Scroll/Zoom buttons to change magnification and position.
**Examples**: Chr1, Chr5:1464000..1480000, LOC_Os10g28230, LOC_Os02g02400.3, 11676.t00006, OSJNBa0037P02, TR009237.

[Hide banner] [Hide instructions] [Bookmark this view] [Link to an image of this view] [Publication quality image] [Help]

**Landmark or Region**
Chr10:40626..60625 Search Reset ☐ Flip

**Scroll/Zoom:**
Show 20 kbp

**Data Source**
TIGR Rice Browser

**Dumps, Searches and other Operations:**
Dump Decorated FASTA File About... Configure... Go

**Tracks** [Hide]

*External tracks italicized*

Overview track

☐ 6-frame translation
☐ Affymetrix GeneChip Rice Genome Array
☐ Agilent Rice Oligo Microarray
☐ Best Arabidopsis Hit
☐ Dicot Gene Index
☐ DNA/GC Content
☐ FGenesh Predictions
☐ Frame usage
☐ GenBank Annotations
☐ GeneMark.hmm Predictions
☐ GENSCAN Predictions
☐ GENSCAN+ Predictions
☐ GlimmerR Predictions
☐ Insertion Mutants
☐ IRGSP pseudomolecules alignment
☐ Maize Genetic Marker
☐ miRNA Registry Oryza miRNAs
☐ Monocot Gene Index
☐ Nipponbare vs 93-11: alignment
☐ Nipponbare vs 93-11: indels/mismatches
☐ NONCODE ncRNAs
☐ NSF 20k Rice Oligo Microarray
☐ Other Gene Indices
☐ Pack-MULE
☐ Protein Evidence
☑ Rice BAC Tiling Path
☑ Rice Fl-cDNA
☑ Rice Gene Index
☐ SSR Markers: putative
☐ SSR Markers: validated Nipponbare (Gramene)
☑ TIGR Rice Gene Models
☑ TIGR Rice Loci
☐ Wheat Bin Mapped Marker
☐ Yale Rice Oligo Microarray

# TIGR Rice Genome Browser

http://www.tigr.org/tigr-scripts/osa1_web/gbrowse/rice

Image Width
in Pixels

Key positions
(Description of Track)

Change Order
Of Track Selection
Table

Set Track Order
In Display Panel

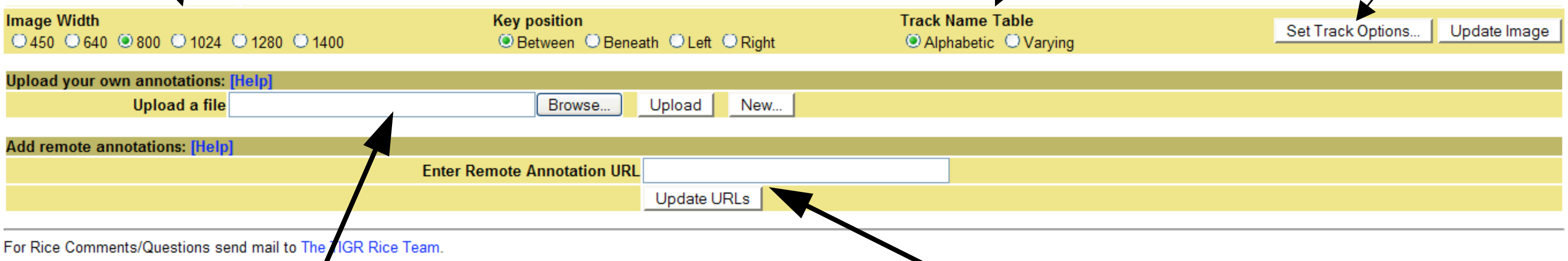


**Image Width**
○ 450 ○ 640 ● 800 ○ 1024 ○ 1280 ○ 1400

**Key position**
● Between ○ Beneath ○ Left ○ Right

**Track Name Table**
● Alphabetic ○ Varying

**Upload your own annotations:** [Help]

**Upload a file**

**Add remote annotations:** [Help]

**Enter Remote Annotation URL**

For Rice Comments/Questions send mail to The TIGR Rice Team.

Display User Annotations
in a Track

Access GFF file on
Remote server

# Comparison Between Viewers

| | Artemis | Apollo | Genome Browser |
|---|---|---|---|
| Data Source | Flat file | Flat file or database | Flat file or database |
| Sharing | Not designed for sharing | Users can share databases | Convenient via the web, possible to share databases |
| Annotating | Designed for annotating sequence | Designed for annotating sequence | Users may add “private” annotation |

# Closing words …

- This talk was a general overview of some of the open source annotation viewers/editors available
- These programs have many more features than we have time to cover
- Download and evaluate to see if they meet your needs