

# First-pass Brassica BAC annotation

Martin Trick

Computational Biology Group

John Innes Centre, UK

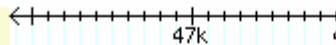
# Multinational Brassica Genome Sequencing Project

- Expect a full annotation from TIGR, to be published after project completion
- Community requirement for “real-time”, first-pass annotation of finished BAC sequences:
  - Microsynteny with Arabidopsis sequence, *ab initio* gene prediction (GeneZilla, GlimmerHMM, FGenesH), EST evidence, BLASTX vs UniProt, “novel” genes
  - Web accessible: data and annotation tools. Use via web forms
  - Shotgun phase sequences and trace files to be publicly available
- Possible delivery mechanisms:
  - ENSEMBL annotation pipeline, ENSEMBL browser
  - Bespoke TIGR annotation pipeline, GBrowse browser
  - Dual Arabidopsis and Brassica-centric views

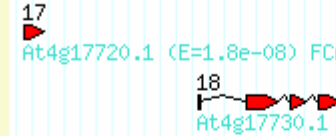
Landmark or Region

23:45860..55859

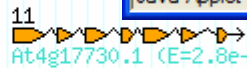
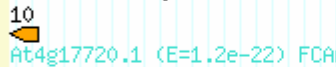
Overview of 23



GeneZilla predictions



GlimmerHMM predictions



EST alignments (est2genome)



EST alignments (BLAT)

**Jalview alignment editor**

File Edit Font View Colour Calculate Align Help

40 50 60 70 80 90 100 110 120

23 48034 48903/1-896 GTGGAATCAAAAAGGTAAAGTTTATACAG&CAAAAAGTACTGAGATGGTTTATCTTTGATTGGTTTGTGTTGTTGTTACAG&CAAGAA

EM\_CN727251/1-896 GTGGAATCAAAAAG-----ACAAGAA

Quality/1-896

done Redraw time = 0 ms

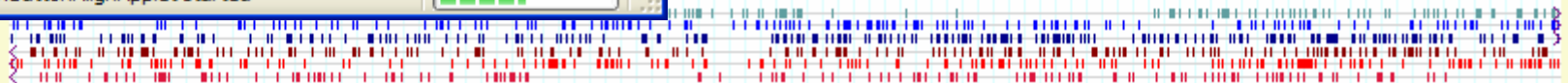
Java Applet Window

http://jililn4 - EM:CN727251 (23:48034..48903)...

Close window

View EST alignment, EM:CN727251 vs 23:48034,48903 JalView

Applet jalview.ButtonAlignApplet started



**Jalview alignment editor**

File Edit Font View Colour Calculate Align Help

10 20 30 40 50 60 70 80 90

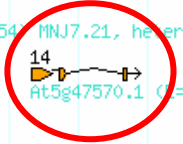
GeneZilla predict 26 46344 47705/1-150 MAGRLSGVASRIMGNGGVARSAAASLQRAGMGLPVGKHIUVPDKPVSQISLRFENSFRFLRLFPLIASQLSUNDELMQDNGTAFPEPCID  
 17 18  
 At5g47570.1/1-150 MAGRLSGVASRIMGNGGVARSVUGSSLRQRAGMGLPVGKHIUVPDKP-----LSUNDELMQDNGTAFPEPCID  
 19  
 At  
 Quality/1-150

done Redraw time = 0 ms

Java Applet Window

**GlimmerHMM predictions**

12  
 At5g47620.3 (E=1.5e-154) MNJ7.21, he  
 14  
 At5g47570.1 (E=...



**EST alignments (est2genome)**

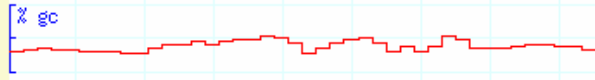
EM:CD825228  
 EM:CD818666  
 EM:CD829261  
 EM:CD829901

**EST alignments (BLAT)**

EM:CD825228  
 EM:CD818666  
 EM:CD829261  
 EM:CD829901

**EST BLAST matches**

EM:CD825228  
 EM:CD818666  
 EM:CD829261  
 EM:CD829901



**6-frame translation**

```

...
...
...
...
...
...

```

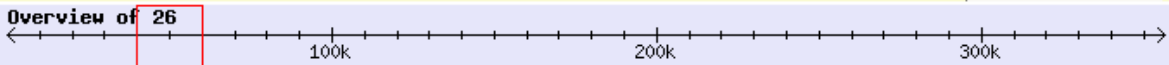
**http://jiilin4 - glimmer 14 (26:46344..47705) - Mozilla Firefox**

Add feature Close window

View protein alignment **JalView**

Strand	Feature	Start	Stop	Phase	BLASTX	Translation...
+	Initial	46344	46643	0	At5g47570.1 7.1e-20	MAGRLSGVASRIMGNGGVARSAAASLQRAGMGLPVGKHIUVPDKPVSQISLRFENSFRFLRLFPLIASQLSUNDELMQDNGTAFPEPCID
+	Internal	46729	46824	0	At5g47570.1 2.0e-08	YEALGNLCCGLSFFAALGMLAVLNDKASKVFP
+	Terminal	47649	47705	0	At5g47570.1 2.0e-08	SFRVYPYDNLRLVELGGEP+

Applet jalview.ButtonAlignApplet started



**GeneZilla predictions**  
 17 18 19 21 22 23 27 28  
 At5g47620.2 (E=3.7e-30) MNJ7.21, heterogeneous nuclear ribonucleoprotein, putative / hnRNP, putative, GO:0003723 [RM]

**Jalview alignment editor**

File Edit Font View Colour Calculate Align Help

10 20 30 40 50 60 70 80 90

26 46344 47907/1-125 MAGRLSGVASRIMGGNGVARSAASSLRQRAGMGLPVGKHIVPDKPLSVNDELMDNDGTAFPEPCIDRIADTVGKYEALGMLCGGLSFFAAL

At5g47570.1/1-125 MAGRLSGVASRIMGGNGVARSVCSLSLRQRAGMGLPVGKHIVPDKPLSVNDELMDNDGTAFPEPCIDRIADTVGKYEALGMLSGGLGFTVGL

Quality/1-125

done

Java Applet Window

**GlimmerHMM predic**

13

done

**EST alignments (est2genome)**

EM:CD825228

EM:CD818666

EM:CD829261

EM:CD829901

**EST alignments (BLAT)**

EM:CD825228

EM:CD818666

EM:CD829261

EM:CD829901

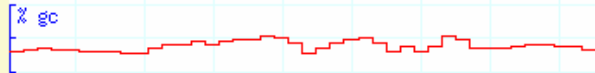
**EST BLAST matches**

EM:CD825228

EM:CD818666

EM:CD829261

EM:CD829901



**6-frame translation**

```

...
...
...
...
...
...

```

http://jilin4 - genezilla 21 (26:46344..47907) - Mozilla Firefox

Add feature Close window

View protein alignment **JalView**

Strand	Feature	Start	Stop	Phase	BLASTX	Translation...
+	initial-exon	46344	46481	0	At5g47570.1 3.2e-17	MAGRLSGVASRIMGGNGVARSAASSLRQRAGMGLPVGKH
+	internal-exon	46557	46643	0	At5g47570.1 7.0e-11	LSVNDELMDNDGTAFPEPCIDRIADTVGK
+	internal-exon	46729	46824	0	At5g47570.1 2.0e-08	YEALGMLCGGLSFFAALGMLAVLNDKASKVFP
+	final-exon	47649	47705	0	At5g47570.1 2.0e-08	SPRVYPYDINLRVELGSEP*
+	poly-A-signal	47901	47907			