

# The Gramene Genome Browser;

## Visualisation of Annotation on Plant Genomes

W. Spooner (whs@ebi.ac.uk), C. Liang, S. Wei, L. Stein, and D. Ware

### Browsable Species

Sequenced genomes;

- **Oryza sativa**
- **Arabidopsis thaliana**

Partially sequenced genomes;

- **Zea Mays**

Chromosome-anchored FPC maps;

- **Zea Mays**
- **Oryza Rufipogon**

Other species are in the pipeline...

### Gene Model Structures

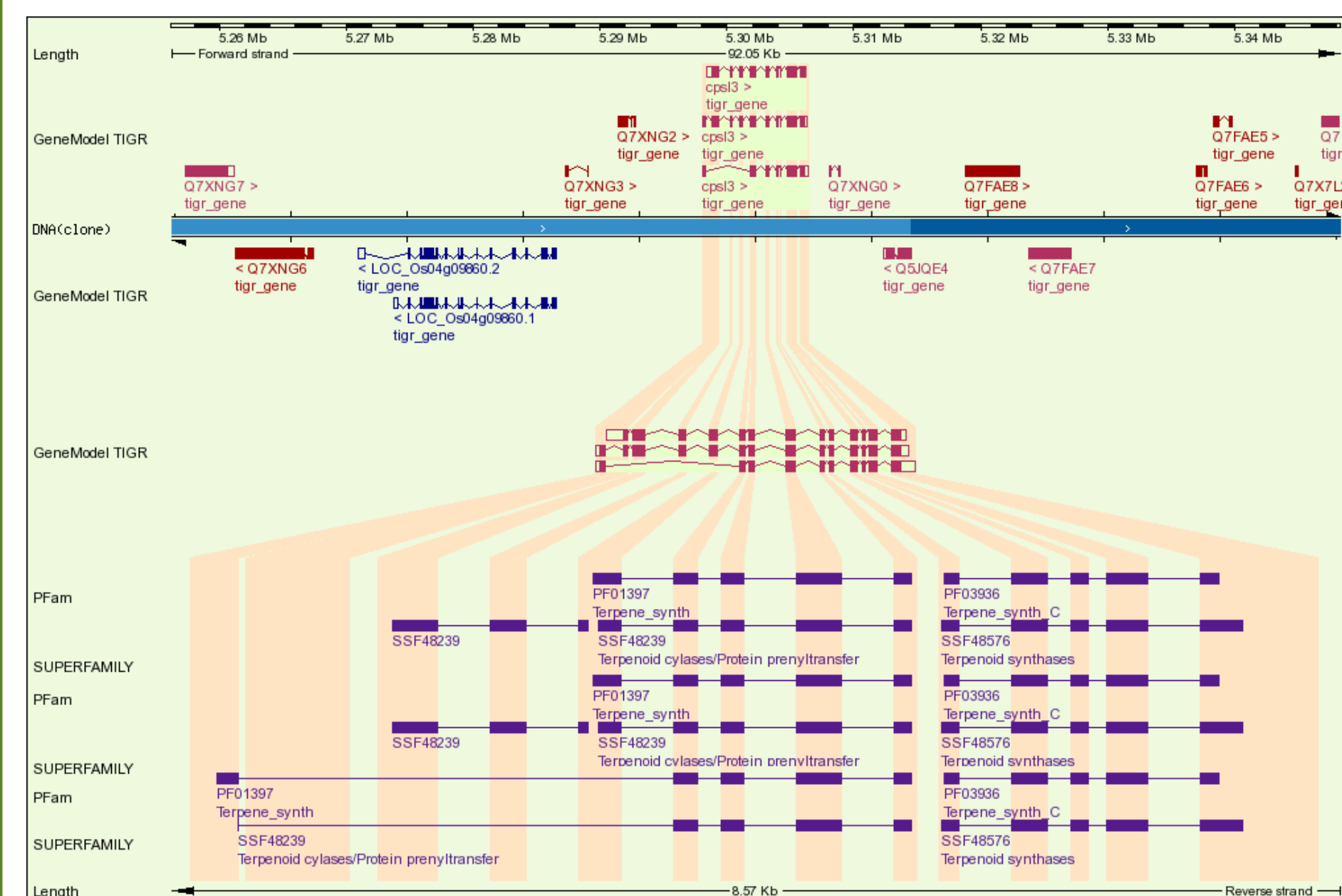
Various classes of gene model are used;

- Ab-initio predictions, e.g. FgenesH,
- Evidence-based, e.g. Ensembl,
- Imported data sets, e.g. TIGR rice loci.

### Gene Model Annotations

Gene models are further annotated by;

- Protein domains from InterProScan,
- GO terms from InterPro annotations,
- Cross-references via sequence homology;
  - Gramene Genes, Proteins, Pathways,
  - NCBI RefSeq, Genes and UniGenes,
  - TIGR Gene Indices.



Genomic context of alternative splice forms of the *cps13* gene prediction (TIGR LOC\_Os01g66100) overlaid with constituent protein domains.

### Aligned Features

All 10,000,000 GenBank DNA sequences from Poaceae species are aligned to the sequenced genomes using BLAT, and classified into 80 browser tracks.

### Repeat Features

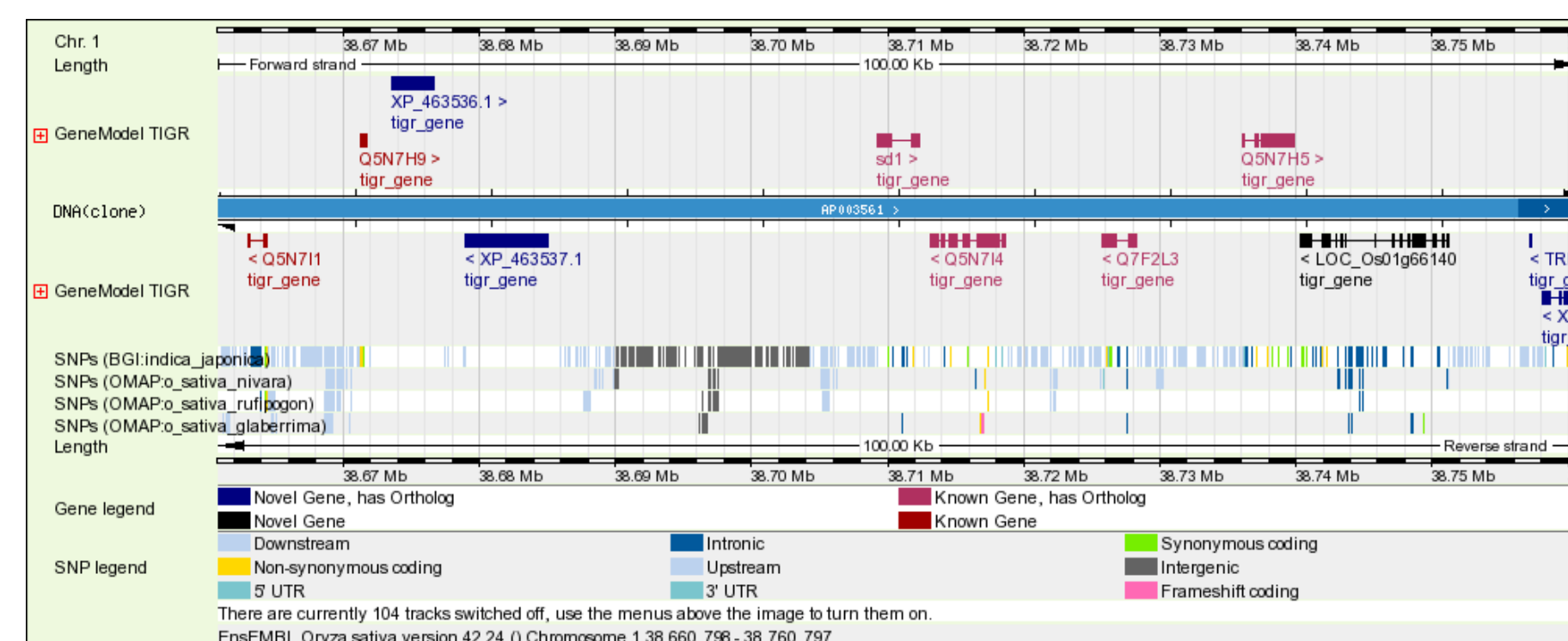
Repeat sequences from the MIPS REcat library are aligned using RepeatMasker.

### Genetic Markers and QTLs

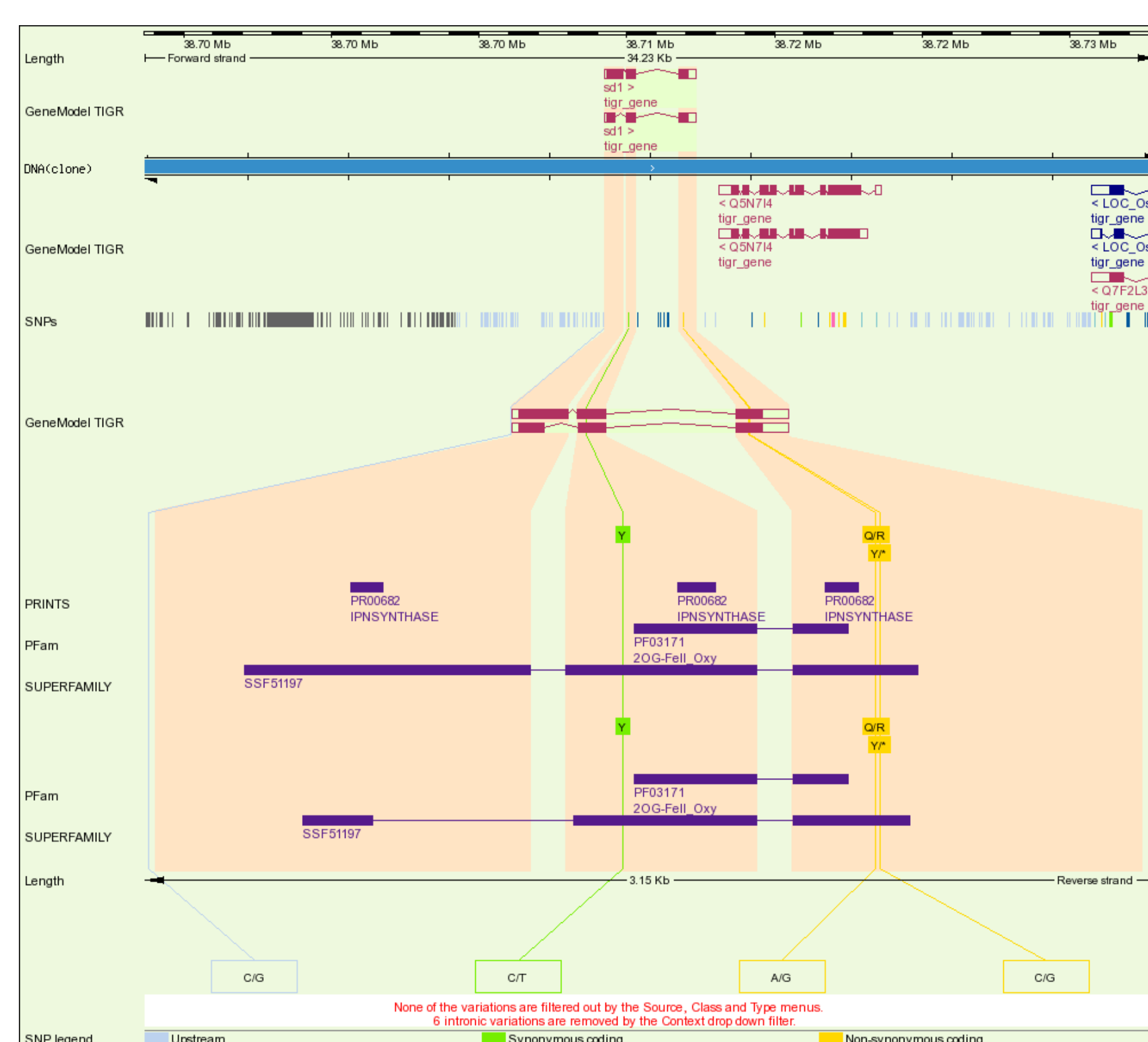
The alignment of DNA sequences are used to infer the locations of sequence-based genetic markers (RFLPs, SSRs etc). The locations of QTLs can also be estimated based on locations of associated markers.

### Variations (SNPs)

Variations from NCBI dbSNP and OMAP are mapped to the genome via alignment of flanking sequences.



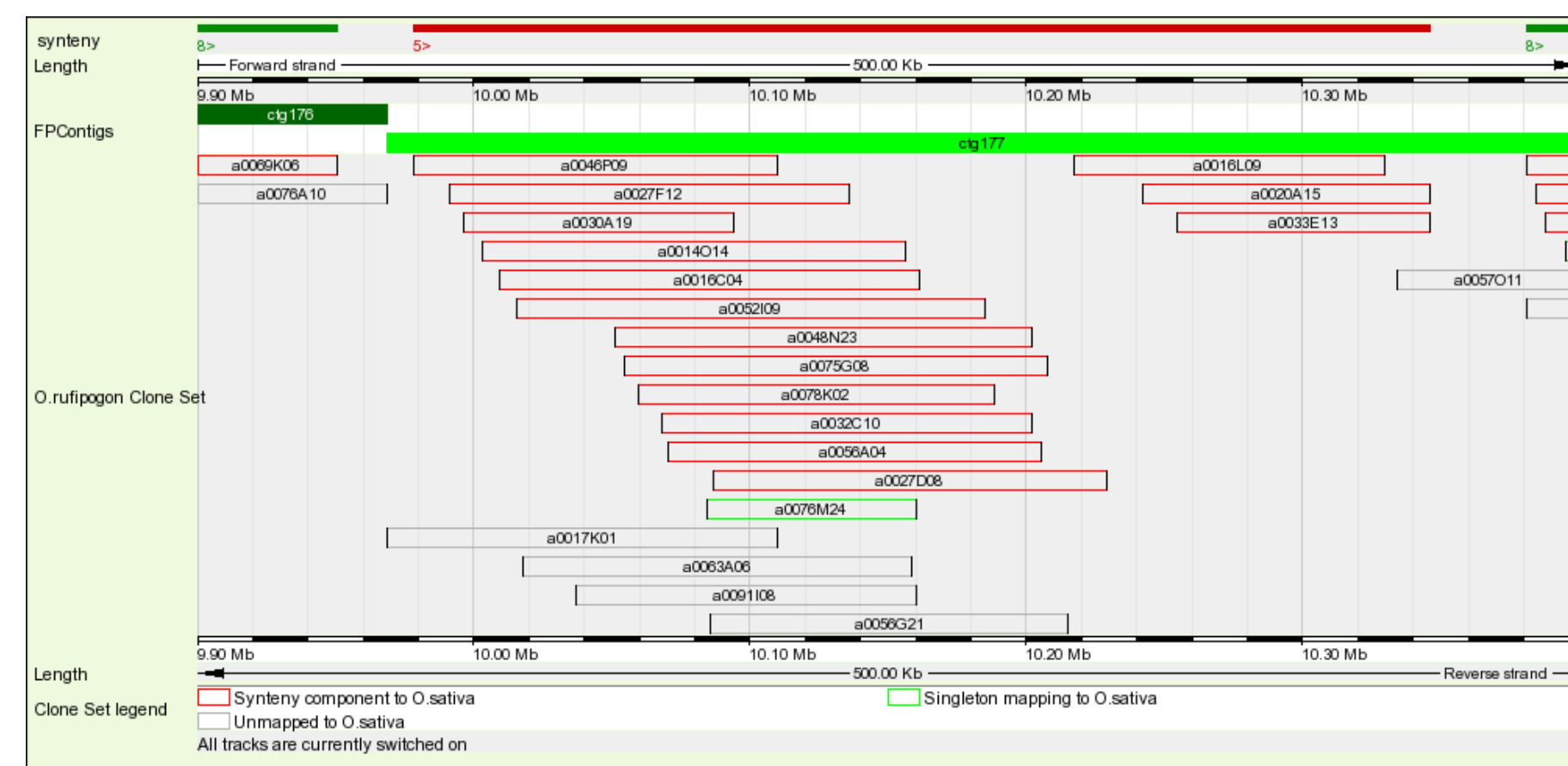
Putative Variations/SNPs in a 100 Kb region of rice Chr1 for four 'populations'. The genetic consequence of each SNP is indicated by colour coding.



Putative SNPs in the region of the *Sd1* gene (TIGR LOC\_Os01g66100) and their genetic consequence in the context of protein domains.

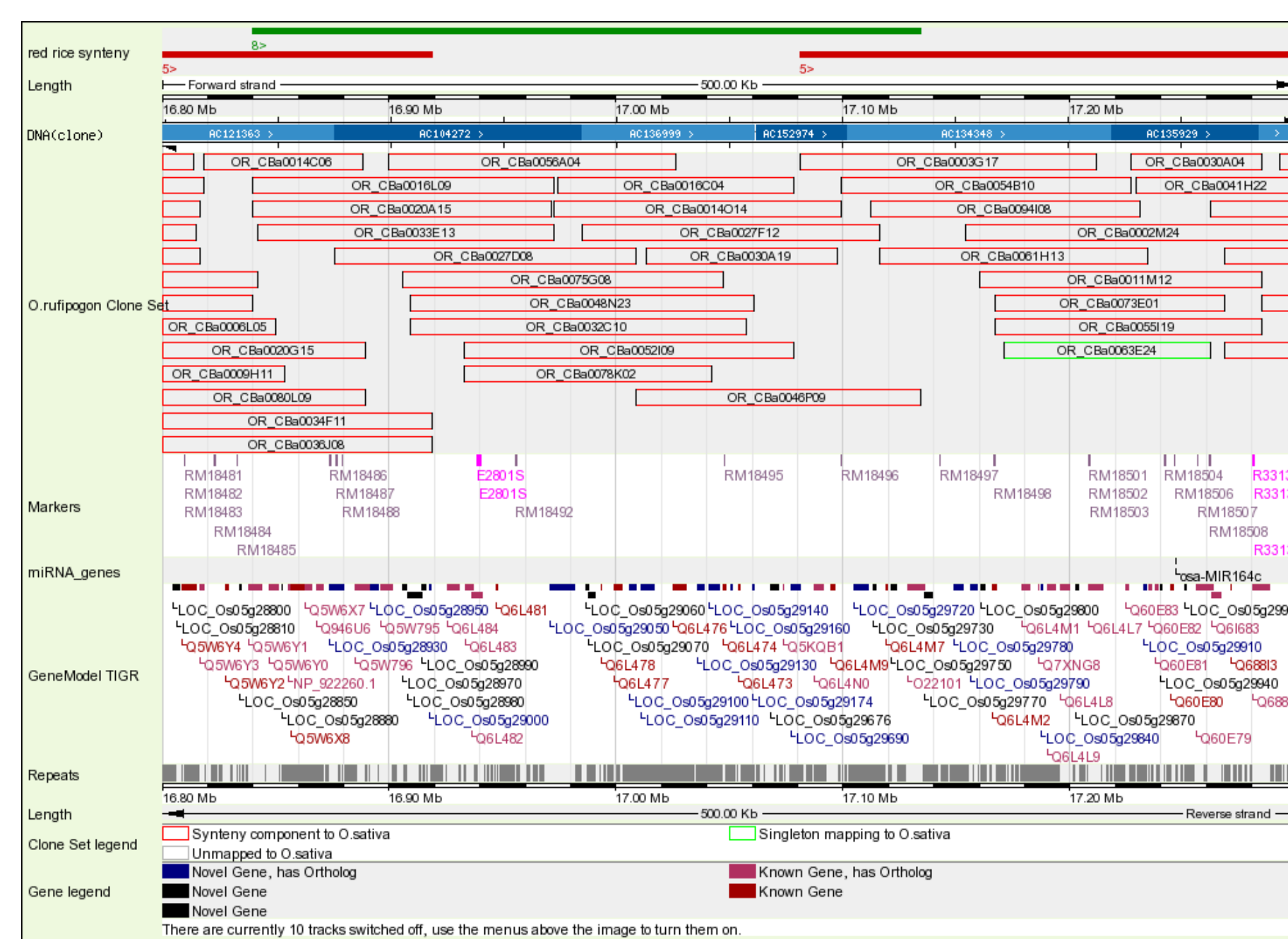
### FPContig assemblies

The Genome browser can also be used to visualise FingerPrint Contig assemblies.



A 5Mbase region of the FPC map for *O. rufipogon* Chr 8 showing the Contig and BAC clone tiling paths. Colour coding shows the syntenic relationship of the clones with the *O. sativa* genome, as summarised by the top track.

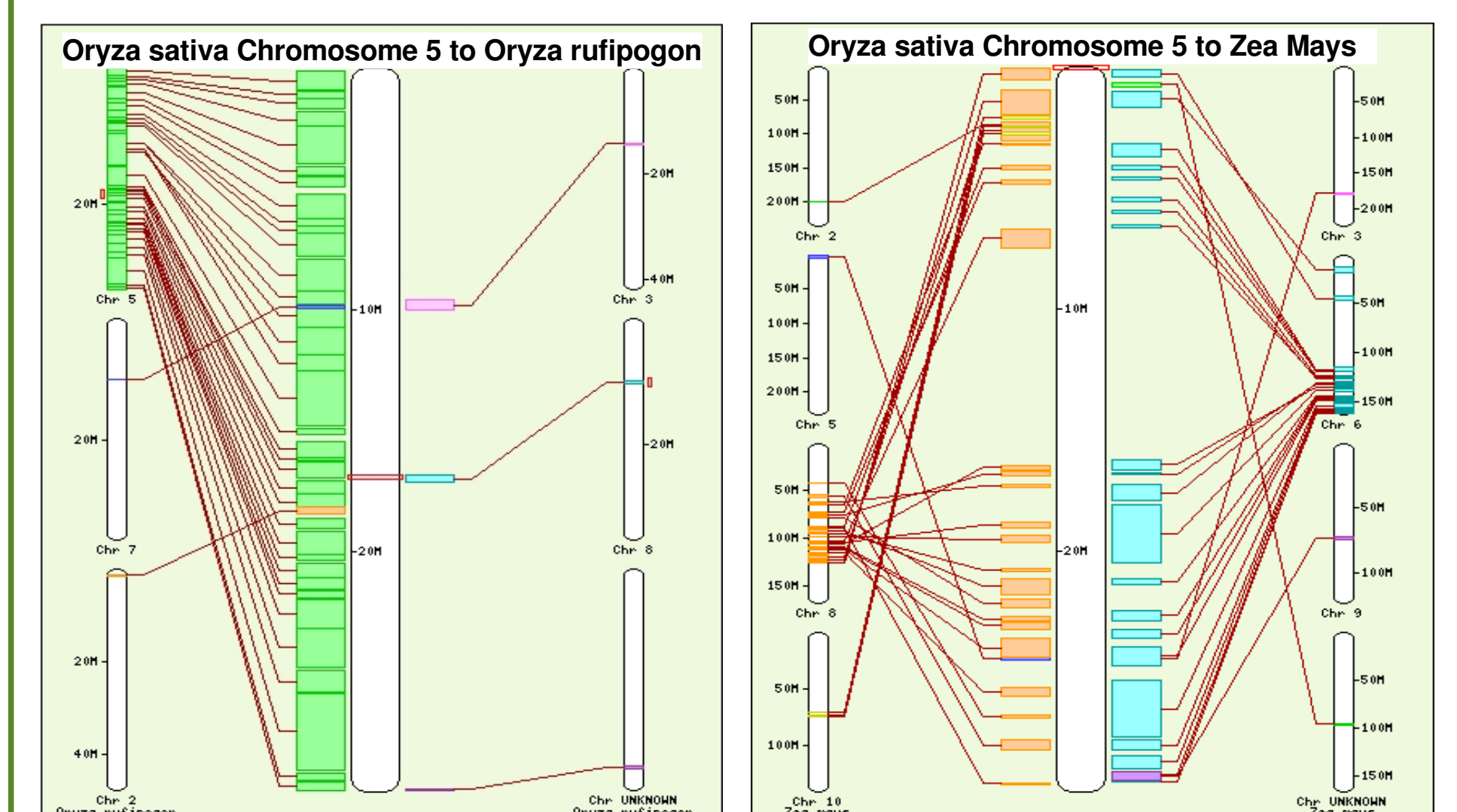
Alignments of clone end sequences from FPC map to sequenced genome are used to project the tiling path onto rice.



A 5Mbase region of the genome assembly for *O. sativa* Chr 5 showing a projection of the the BAC clone tiling paths from *O. rufipogon*. Colour coding shows the syntenic relationship of the clones with the *O. rufipogon* FPC map, as summarised by the top track.

### Cross-genome Synteny

Syntenic blocks between sequenced genomes and FPContig maps are determined from conserved order/orientation of aligned sequences.

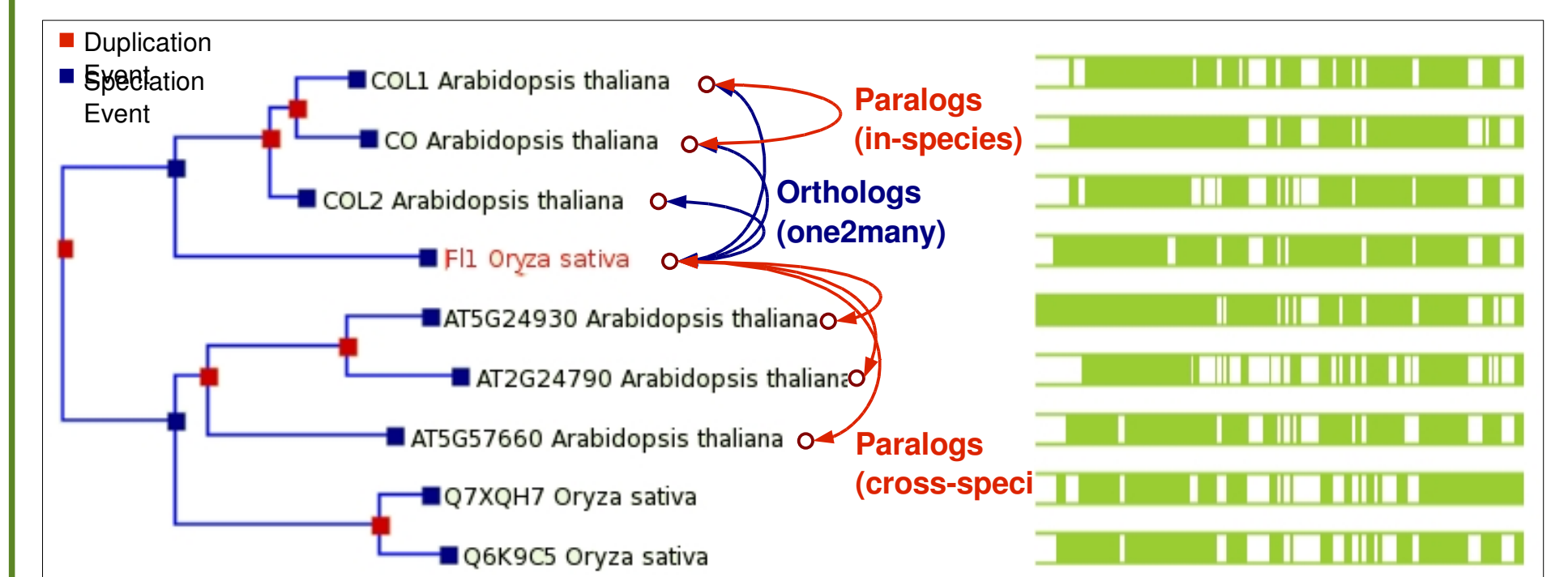


Syntenic blocks between rice chromosome 5 and a closely related *Oryza rufipogon*, and a more distant Poaceae relative, *Zea mays*.

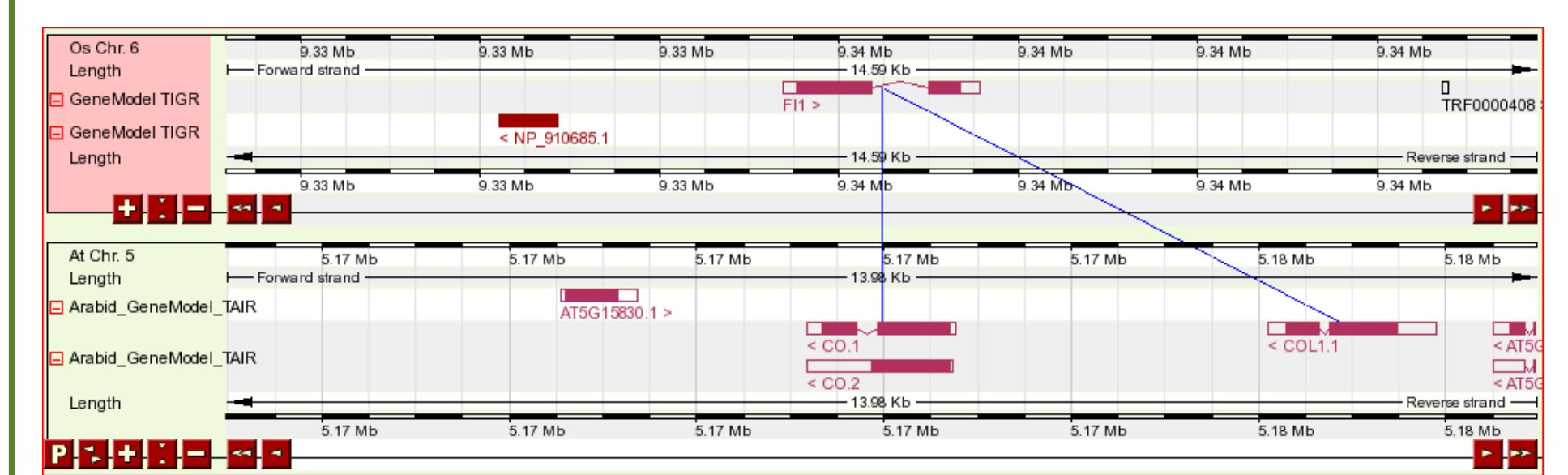
Syntenic blocks between the sequenced genomes of rice and Arabidopsis will be released shortly.

### Gene Trees

Gene orthologs and paralogs are determined from maximum likelihood phylogenetic trees based on multiple alignments of homologous gene clusters.



Annotated gene tree centred on rice gene *F11* (LOC\_Os06g16370).

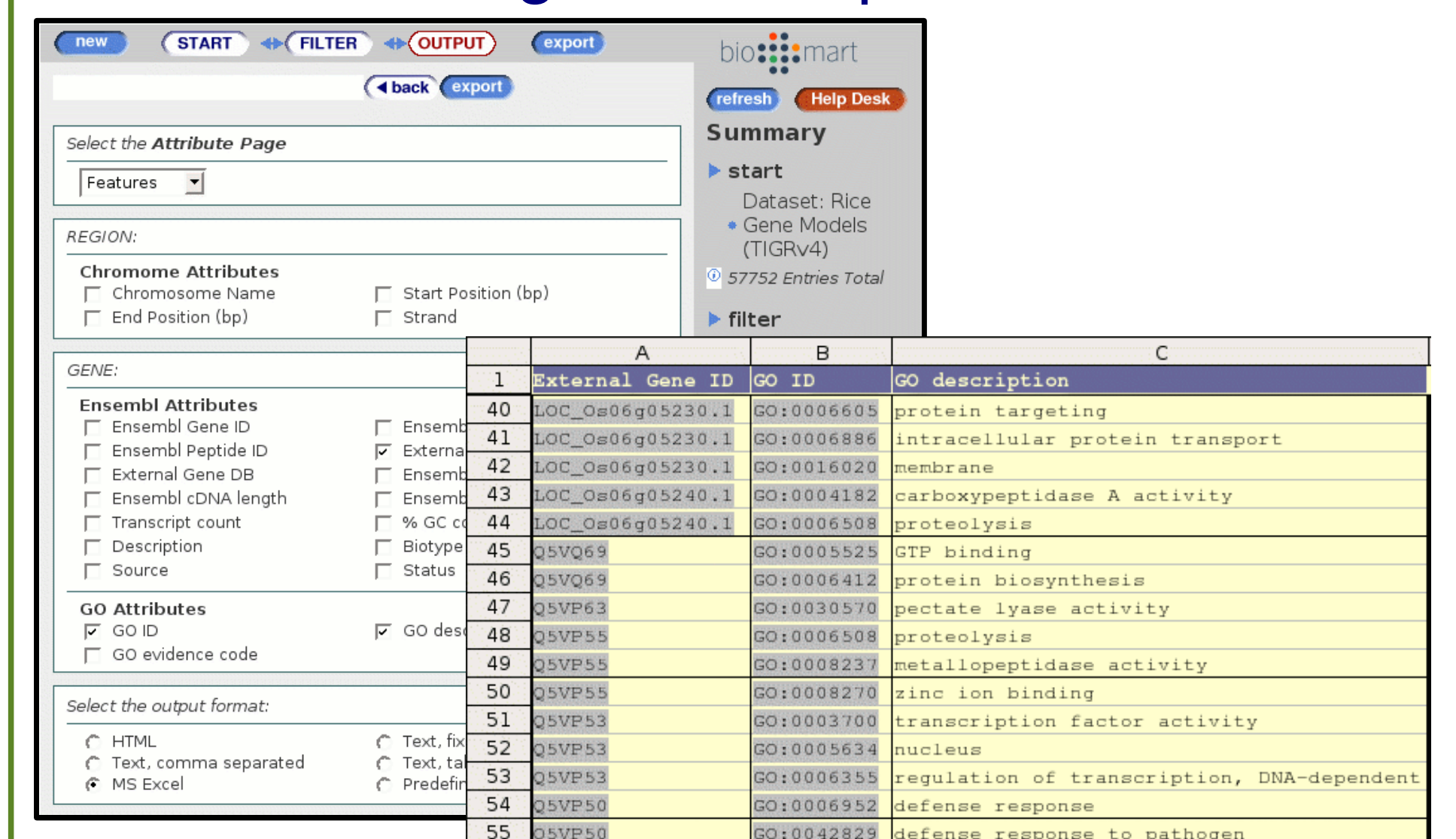


Genomic context of rice gene *F11*, and two Arabidopsis orthologs, *CO* and *COL1*.

### Data Download/upload

All databases are available via FTP.

The BioMart-based GrameneMart provides an interface for flexible downloading of user-specified datasets.



The genome browser can also be used as a DAS server or client

Gramene (<http://www.gramene.org>) is a curated, open-source, web-accessible data resource for comparative genome analysis in the grasses. Many techniques developed by Gramene for the annotation and visualisation of sequenced genomes have built on the established Ensembl software framework (<http://www.ensembl.org>). The Gramene project is supported by National Science Foundation grant No. 0321685 and USDA ARS.