

BrassEnsembl as a window on the genome
UK-BRC
7th May, 2010

Chris Love
Rothamsted Research



ROTHAMSTED
RESEARCH

Outline

- ▶ Introduction
- ▶ BrassEnsembl overview
- ▶ BlastView
- ▶ Future developments
- ▶ Conclusions





ROTHAMSTED
RESEARCH

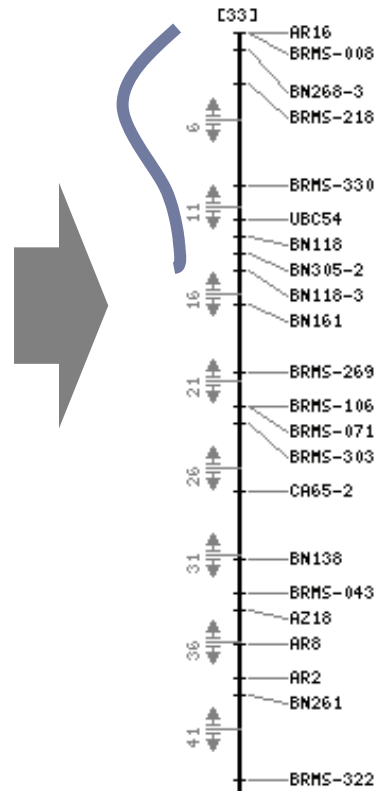
Introduction

Phenotypic
measurements



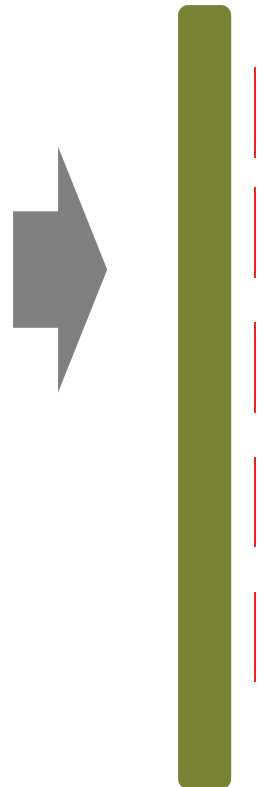
QTL

Linkage
maps



Genomic
sequence

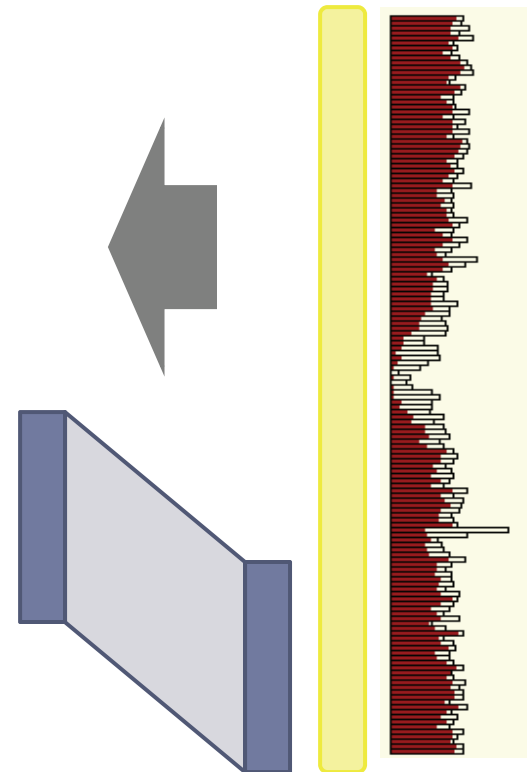
B. rapa chromosome A3



Comparative
genomics

Arabidopsis chromosome 4

Known genes



BrassEnsembl

- ▶ *Brassica rapa* centric Ensembl genome browser
- ▶ Ensembl version 52
- ▶ Current build includes:
 - ▶ 974 BACs only (no scaffolds included yet)
 - ▶ ~118 Mb sequence data
 - ▶ BAC sequence retrieved from GenBank



Current features presented in BrassEnsembl

36,066	Total Genes (17,512 PASA, 18,554 SNAP identified)
13,642 (77.9%)	Of PASA identified genes have functional annotation provided by homology to <i>Arabidopsis</i> genes
17,775	<i>Arabidopsis</i> coding sequences aligned using BLAT with links to <i>Arabidopsis</i> Ensembl (AtEnsembl)
6,895	95K <i>Brassica</i> unigene alignments using BLAT
681,689	<i>Brassica</i> ESTs alignments using BLAST (Evalue<1E-20)
174,907	<i>B. rapa</i> BAC end sequence alignments (Evalue<1E-5)





Location-based displays

- Whole genome
- Chromosome summary
- Region overview
- Region in detail**
- Comparative Genomics
 - Genomic alignments
 - Synteny
- Genetic Variation
 - Resequencing (0)
 - Linkage Data
- Markers

- Configure this page
- Add custom data to page
- Export data
- Bookmark this page

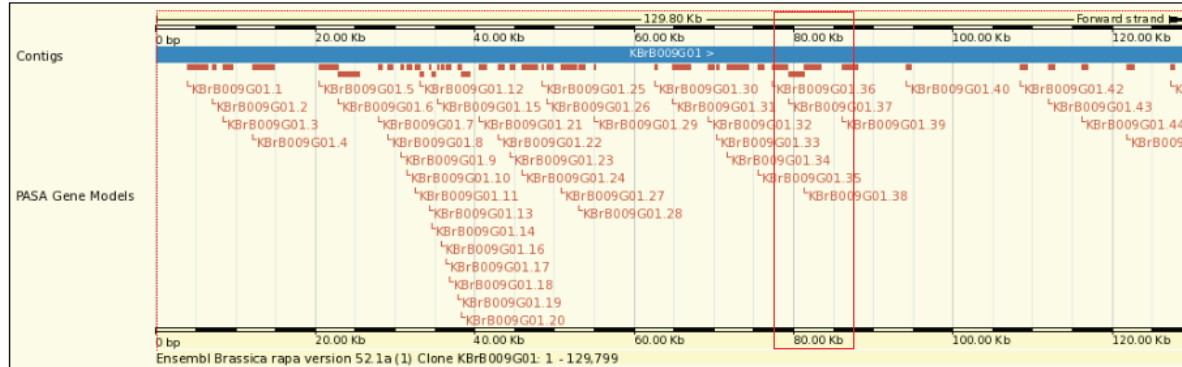
Clone KBrB009G01: 77,352-87,351

clone KBrB009G01 Export image

« Region overview

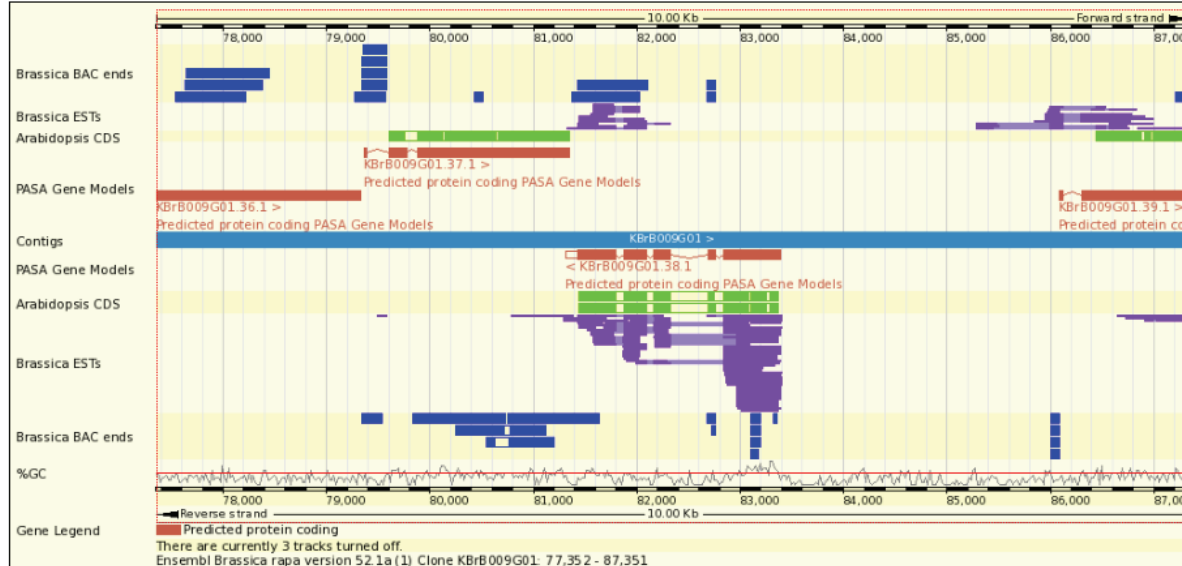
Region in detail [help](#)

Genomic alignments »



Export image

Location: KBrB0 : - « < > >>



Export image

Configuring the display

You currently have 0 tracks in the overview panel and 3 tracks in the main panel turned off. To change the tracks you are displaying, use the "Configure this page" link on the left.



BlastView- The Ensembl BLAST interface

- ▶ User submission of query sequence
- ▶ Available BLASTable databases:
 - ▶ Current genomic sequence
 - ▶ CDS of predicted gene models
 - ▶ Peptide translations of gene models
 - ▶ *B. rapa* R-o-18 Illumina 76bp read dataset
- ▶ Aligning gene features or regions
link to region within BrassEnsembl

brassica.info

BLAST/BLAT | Docs | FAQs

refresh Online Help

new SETUP CONFIG RESULTS DISPLAY

Important Notice
We use Blast as our default DNA search.

Enter the Query Sequence

Either Paste sequences (max 30 sequences) in FASTA or plain text:

Or Upload a file containing one or more FASTA sequences

Or Enter a sequence ID or accession (EMBL, UniProt, RefSeq)

Or Enter an existing ticket ID:

dna queries
 peptide queries

Select the databases to search against

Select species:
Use 'ctrl' key to select multiple species

dna database
 peptide database



Select the Search Tool

Search sensitivity:
Optimise search parameters to find the following alignments

Summary

- ▶ setup
○ Not yet initialised
- ▶ configure
○ Not yet initialised
- ▶ results
○ Not yet initialised
- ▶ display
○ Not yet initialised

BlastView

[Home](#)
[BLAST/BLAT](#) | [Docs & FAQs](#)

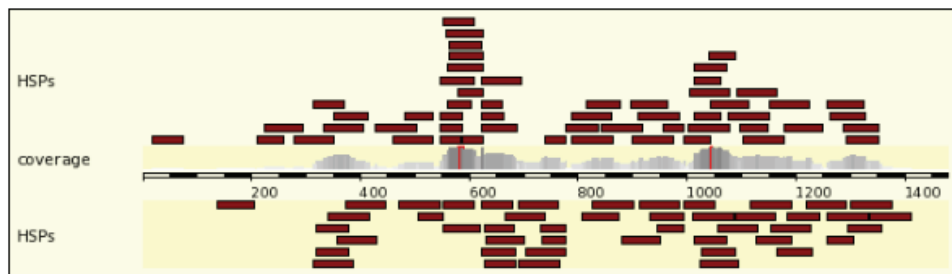
new
SETUP
← CONFIG
← RESULTS
← DISPLAY

Displaying **KBrB009G01.38:KBrB009G01.38.1** sequence alignments vs **Brassica_rapa RO18** database

Showing top alignments of 112, sorted by [refresh](#)

Alignment Locations vs. Karyotype (click arrow to hide)

Alignment Locations vs. Query (click arrow to hide)



Alignment Summary (click arrow to hide)

Select rows to include in table, and type of sort
(Use the 'ctrl' key to select multiples) [refresh](#)

Query	Subject	Chromosome	Clone	Stats	Sort By
<input type="checkbox"/> off	<input type="checkbox"/> off	<input type="checkbox"/> off	<input type="checkbox"/> _off_	<input type="checkbox"/> off	>Clone
<input type="checkbox"/> Name	<input type="checkbox"/> Name	<input type="checkbox"/> Name	<input type="checkbox"/> Name	<input type="checkbox"/> Score	<Score
<input type="checkbox"/> Start	<input type="checkbox"/> Start	<input type="checkbox"/> Start	<input type="checkbox"/> Start	<input type="checkbox"/> E-val	>Score

Links	Query Name	Subject Name	Stats
			Score E-val %ID Length
[A] [S] [G]	KBrB009G01.38:KBrB009G01.38.1	1	76 5e-34 100.00 76
[A] [S] [G]	KBrB009G01.38:KBrB009G01.38.1	1	76 5e-34 100.00 76
[A] [S] [G]	KBrB009G01.38:KBrB009G01.38.1	1	76 5e-34 100.00 76
[A] [S] [G]	KBrB009G01.38:KBrB009G01.38.1	1	76 5e-34 100.00 76
[A] [S] [G]	KBrB009G01.38:KBrB009G01.38.1	1	76 5e-34 100.00 76
[A] [S] [G]	KBrB009G01.38:KBrB009G01.38.1	1	76 5e-34 100.00 76
[A] [S] [G]	KBrB009G01.38:KBrB009G01.38.1	1	76 5e-34 100.00 76

[refresh](#)
[Online Help](#)

Summary

- [▶ setup](#)
 - [• Brassica_rapa](#)
 - [• R-o-18 Illumina](#)
 - [• BLASTN](#)
 - [• Custom sensitivity](#)
- [▶ configure](#)
 - [• -e: 0.0001](#)
 - [• -b: 250](#)
 - [• RepeatMasker](#)
- [▶ results](#)
- [▶ display](#)



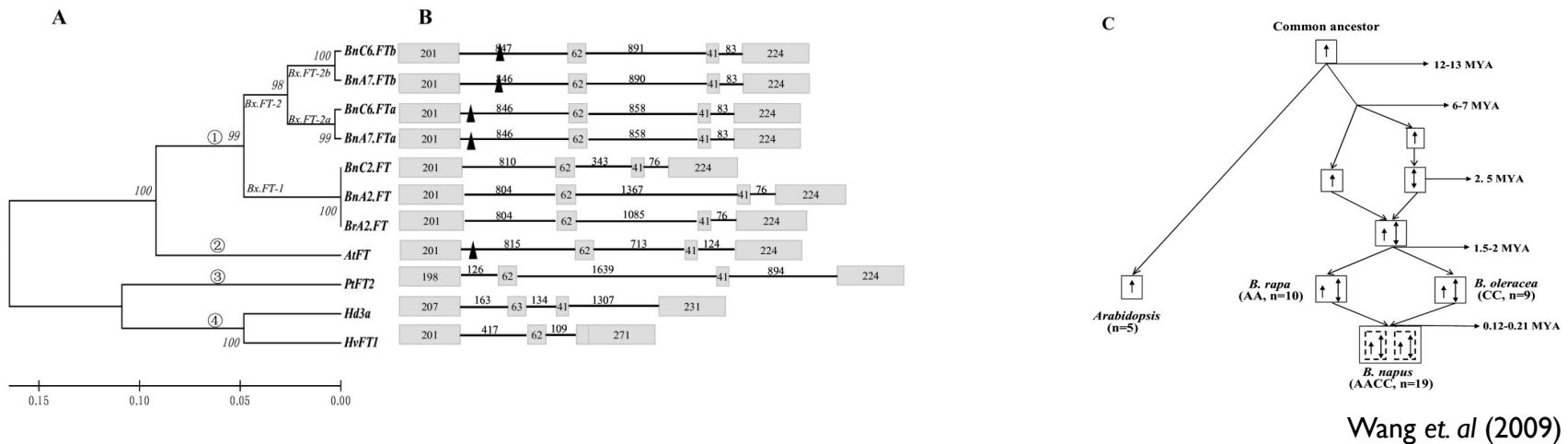
Future Developments:

- ▶ Update Ensembl to version 56
- ▶ Addition of new datasets:
 - ▶ *B. rapa* R-o-18 alignments
 - ▶ 135K Brassica unigene alignments
 - ▶ annotated *B. rapa* sequence scaffolds (when available)
- ▶ Application of JIC/JCVI Annotate Pipeline
- ▶ Links to AlignStore_{DB} and SeqStore_{DB}
- ▶ Closer association with CropStore_{DB} via sequence based markers



Future Developments: Ensembl ComparaView

- ▶ Determination of phylogeny
- ▶ Utilises Ensembl Compara pipeline to identify orthologs and paralogs
- ▶ Example using FLOWERING LOCUST in *B. napus* (FT)



- ▶ Results from Ensembl Compara pipeline on FT in *Arabidopsis* EnsemblGenomes



- x1 branch length
- - - x10 branch length
- - - x100 branch length
- Gene ID current gene
- Gene ID within-sp. paralog
- speciation node
- duplication node
- ambiguous node
- AA alignment match/mismatch
- AA consensus > 66% (mis)match
- AA consensus > 33% (mis)match
- AA alignment gap

Conclusions

- ▶ Combination of associated tools simplify browsing between genomic sequence, genetic and phenotypic data
- ▶ Close collaboration with:
 - ▶ EBI Ensembl Genomes team
 - ▶ NASC AtEnsembl
 - ▶ Brassica Gbrowse
- ▶ Future funding necessary for further developments



Acknowledgements

- ▶ Graham King (RRes)
- ▶ Pierre Carion (RRes)
- ▶ Ambrose Andongabo (RRes)
- ▶ Martin Trick (JIC)
- ▶ Nick James, Sean May (NASC)
- ▶ Ensembl team (EBI)

<http://www.brassica.info/BrassEnsembl>

