

Jay Moore, Graham King, James Lynn

Data integration for
Brassica comparative genomics

- How best to bring together diverse data about *Brassica* genome organisation?
- How best to make data accessible and useful to people?
- How informative is the *Arabidopsis* genome sequence in determining *Brassica* genome organisation?

Data sources for *Brassica* comparative genomics

➤ Arabidopsis genome and genome annotations (NASC, 2005)

➤ *Brassica* sequence data

- genomic DNA sequences
- EST sequences
- BAC-end and BAC sequences

(Genbank, 2006)

➤ *Brassica* chromosome length estimates

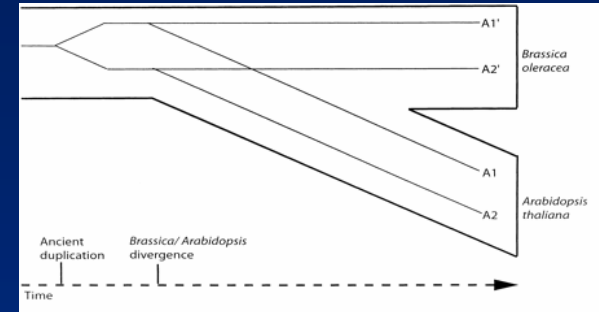
(Arumuganathan and Earle, 1991; Johnston *et al.*, 2005; Bennett and Leitch, 1995; cited in *Brassica.info*, 2005)

➤ *Brassica* genetic linkage maps

(Cavell *et al.*, 1998; Lombard and Delourme, 2001; Lowe *et al.*, 2004; Yang *et al.*, 2005; Parkin *et al.*, 2005; Piquemal *et al.*, 2005; Kowalski *et al.*, 1994; Lan *et al.*, 2000; Sebastian *et al.*, 2000; Lukens *et al.*, 2003)

➤ *Brassica*/Arabidopsis synteny maps

(Kowalski *et al.*, 1994; Ryder *et al.*, 2001; Parkin *et al.*, 2002; Lukens *et al.*, 2003; Parkin *et al.*, 2005)



Lukens *et al.*, 2003 <http://www.genetics.org/cgi/content/full/164/1/359>

<http://ensembl.warwick.ac.uk>



The screenshot shows the Ensembl Genome Browser interface. The browser window title is "Ensembl Genome Browser - Microsoft Internet Explorer". The address bar shows "http://bid62.warwick.ac.uk/index.html". The page header includes the "e!WHRI Ensembl" logo and a search bar with "Anything" entered. Below the header, there are navigation tabs for "Docs and downloads" and "browse a genome".

Docs and downloads

- Information
- What's New
- About Ensembl
- Ensembl data
- Software

Other links

- Home
- Sitemap
- Ensembl

Brassica

- Brassica A genome** [BINFO 1] (*B. rapa*)
browse | what's new
- Brassica B genome** [BINFO 1] (*B. nigra*)
browse | what's new
- Brassica C genome** [BINFO 1] (*B. oleracea*)
browse | what's new

Other Plants

- Arabidopsis thaliana** [TIGR 5]
browse | what's new

Warwick HRI Ensembl 2

This site is a joint project between [Warwick HRI](#), [Graham King \(RRes\)](#), and [NASC](#) to develop a *Brassica*-oriented viewer for comparative *Brassica* genomic and genetic information.

This work is part of an ongoing collaboration involved in the integration of *Brassica* and *Arabidopsis* information for the benefit of the [Brassica Research Community](#), and as a contribution to the Multinational *Brassica* Genome Project.

This project is currently supported through [BBSRC](#) CSG funding to Warwick HRI, and a [BBSRC](#) studentship under the supervision of [Graham King](#) and [James Lynn](#).

The *Brassica* sequence data within this viewer are derived from data downloaded from [Genbank](#) at 11/10/2005 13:00 GMT. We aim to carry out updates quarterly, unless we are advised of a significant input of new data from any source.

For all enquiries about this Warwick HRI Ensembl service, or to advise of new data sources, please contact [Jay Moore](#) or [Giles Weaver](#).

A joint project between [AAFC, Saskatoon](#) and Warwick HRI was set up to generate *B. oleracea* ESTs. Alignments to these ESTs are currently displayed as a separate track, but will be incorporated into the *B. oleracea* mRNA alignments track when Genbank submission has been completed. The *B. rapa* end-sequences incorporated have been generated as the first stage of the [Multinational Brassica rapa Genome Sequencing Project](#).

The current map of synteny between *Brassica* and *Arabidopsis* is based on data derived from [Isobel A.P. Parkin and colleagues](#). [Jay Moore](#) is currently working on likelihood estimates of possible arrangements of *Brassica* sequences, which will be incorporated in future versions.

For all enquiries about the Ensembl software and website, please [contact the Ensembl HelpDesk](#) (helpdesk@ensembl.org).

What's New in Warwick HRI Ensembl

- New download of sequence data, new BLAST analyses, new *Brassica*-centric organisation of tracks
- New map of synteny between *Brassica* A and C genomes and *Arabidopsis thaliana*
- New *Brassica* A genome assembly and genebuild estimates under construction
- New *Brassica* C genome assembly and genebuild estimates under construction
- Comparative *Brassica* genomics browser
- New look'n'feel for the Ensembl website

Have you tried...?

Brassica C genome (*B. oleracea*)

Likelihood-based assembly and genebuild under construction

© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.

<http://ensembl.warwick.ac.uk>

WHRI Ensembl v32: Arabidopsis thaliana Features on Chromosome 1 23316470-23366469 - Microsoft Internet Explorer

Address: http://bid62.warwick.ac.uk/Arabidopsis_thaliana/contigview?c=1:2331469.5;w=50000;h=1

Use Ensembl to...
 Export data
 Does and downloads
 Information
 What's New
 About Ensembl
 Using Ensembl data
 Software
 Other links
 Home
 Sitemap
 sanger
 EBI

Detailed view

Gene legend
 ■ ENSEMBL PREDICTED GENES (KNOWN) ■ ENSEMBL PREDICTED GENES (PSEUDOGENE)

Features ▾ DAS Sources ▾ Repeats ▾ Decorations ▾ Export ▾ Image size ▾ Help ▾

Jump to region 1 : 23316470 : 23366469 Refresh

<< 5MB < 2MB < 1MB Zoom Window > 1MB > 2MB > 5MB >>

Chr. 1
 Length 23.32 Mb 23.33 Mb 23.34 Mb 23.35 Mb 23.36 Mb
 Forward strand

B.napus mRNA
 B.napus DNA
 B.napa mRNA
 B.napa DNA
 B.oleracea mRNA (WHRI)
 B.oleracea DNA
 Arabidopsis CDNs
 Ensembl trans.
 DNA(contigs)
 Ensembl trans.
 Arabidopsis CDNs
 Length
 Gene legend
 ■ ENSEMBL PREDICTED GENES (KNOWN) ■ ENSEMBL PREDICTED GENES (NOVEL) ■ ENSEMBL PSEUDOGENES

There are currently 14 tracks switched off; use the menus above the image to turn these on.
 Ensembl Arabidopsis_thaliana 1:23316470-23366469 Tue Feb 28 11:59:49 2006

Basepair view

© 2006 WTSI / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.
 This Warwick HRI Service is *Under Development*. Please verify any results independently.

<http://ensembl.warwick.ac.uk>

WHRI Ensembl v32: Arabidopsis thaliana Synteny - Microsoft Internet Explorer

Address: http://bid62.warwick.ac.uk/Arabidopsis_thaliana/synteniview?otherspecies=Brassica_C_genome;chr=1;loc=23341469.5

Search *e! A.thaliana*: e.g. 1.4

WHRI Ensembl v32 - Jul 2005

Chromosome 1

- View Chromosome 1
- View Chr 1 Synteny
 - vs Brassica A genome (*Brassica A genome*)
 - vs Brassica B genome (*Brassica B genome*)
 - vs Brassica C genome (*Brassica C genome*)

Use Ensembl to...

- Export data

Docs and downloads

- Information
- What's New
- About Ensembl
- Using Ensembl data
- Software

Other links

- Home
- Sitemap

Change Chromosome

Chromosome

Fields marked with * are required

© 2006 WTSJ / EBI. Ensembl is available to [download for public use](#) - please see the [code licence](#) for details.
This Warwick HRI Service is *Under Development*. Please verify any results independently.

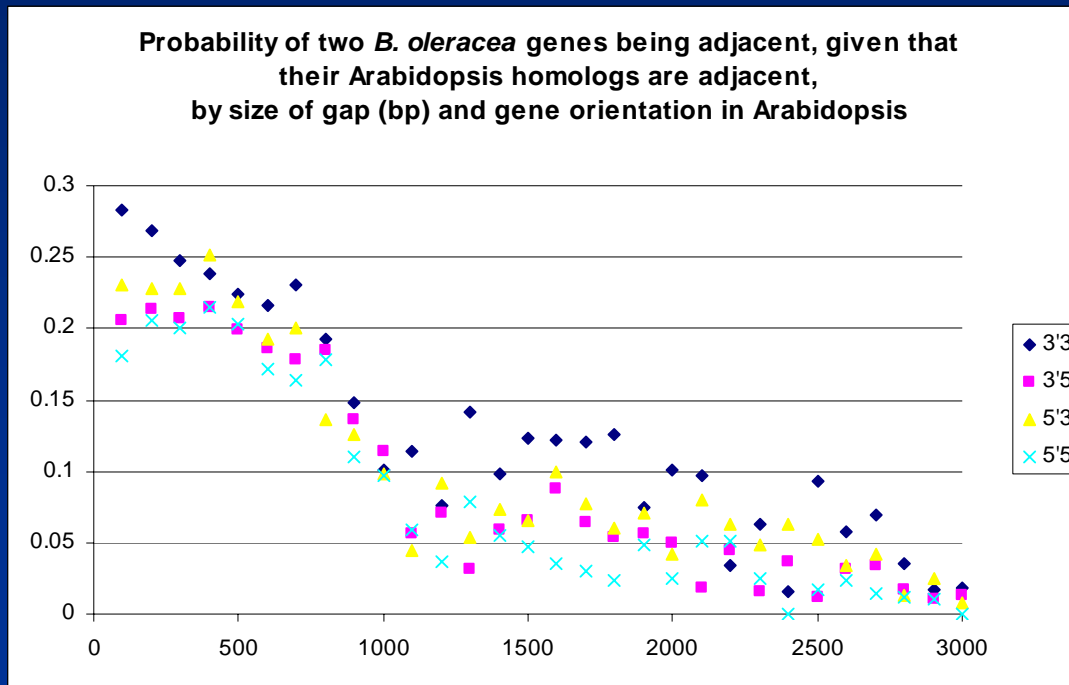
Start | Untitled Docu... | http://www2... | WHRI Ensem... | IT Services del... | Inbox - Micros... | RE: Presentati... | Lab_book_014... | Data Sources | MySQL Admini... | Microsoft Pow... | 12:01

Generating likely hypotheses of Brassica genome organisation

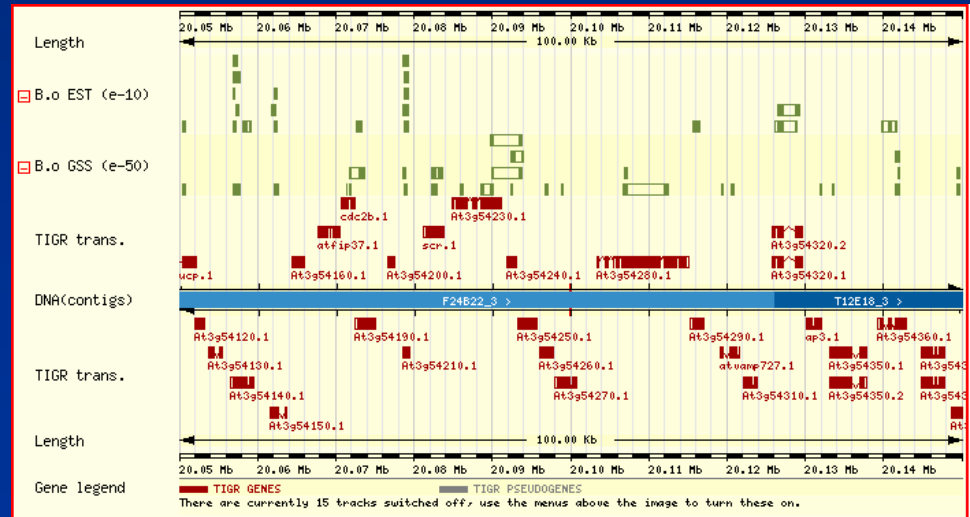
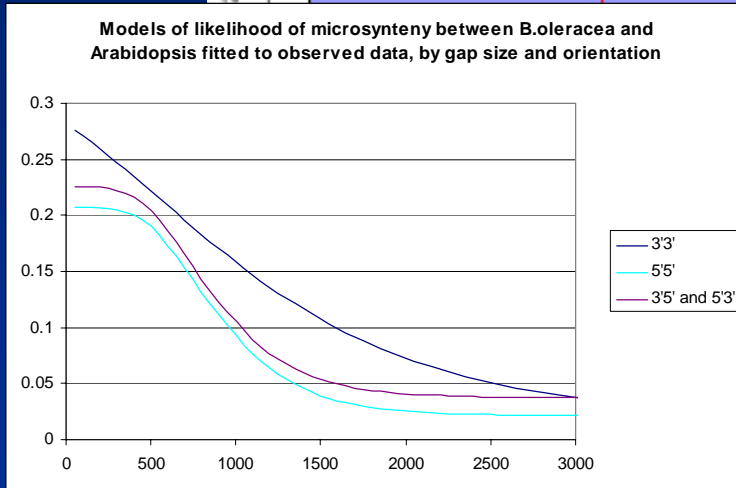
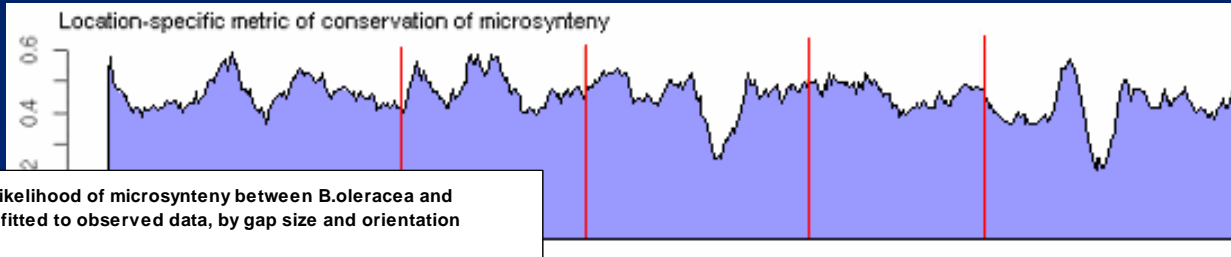
- Analysing the relationship between Arabidopsis genome sequence and known information about *Brassica* genomes
 - Generating alignments between the Arabidopsis genome sequence and shorter *Brassica* genomic, EST, and marker sequences (BLAST, S-search, WABA)
 - Analysing the alignments to develop likelihood-based models of microsynteny between *Brassica* and Arabidopsis (Perl, SQL)
 - Analysing the relationships between the Arabidopsis genome sequence and *Brassica* genetic and synteny maps to develop likelihood-based models of large-scale synteny (Perl, SQL)
 - Turning the results of the analyses into Ensembl objects, dynamic web pages (SQL, Bioperl, Ensembl API)

Characterising microsynteny in *B. oleracea*

- If two *Brassica* genes have homologs in *Arabidopsis*, which are adjacent, how likely are the genes to be adjacent in *Brassica*?

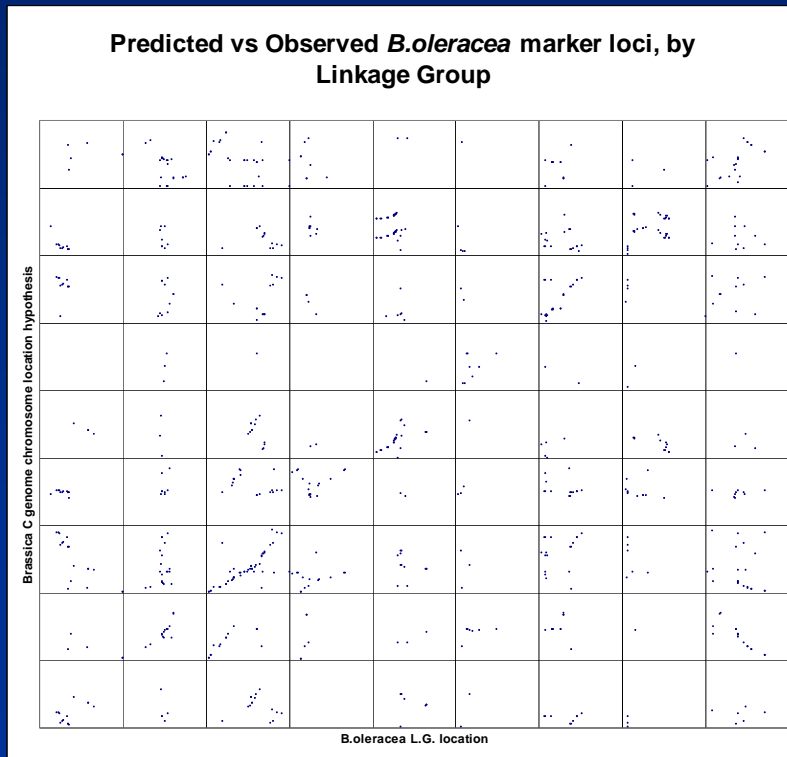


Using models of microsynteny to generate likelihood-based predictions of micro-scale Brassica genome organisation



Characterising genetic-scale synteny between *Arabidopsis* and Brassica A and C genomes

- Used the Parkin *et al* (2005) synteny map to predict the loci of sequence-tagged *B.oleracea* genetic markers:



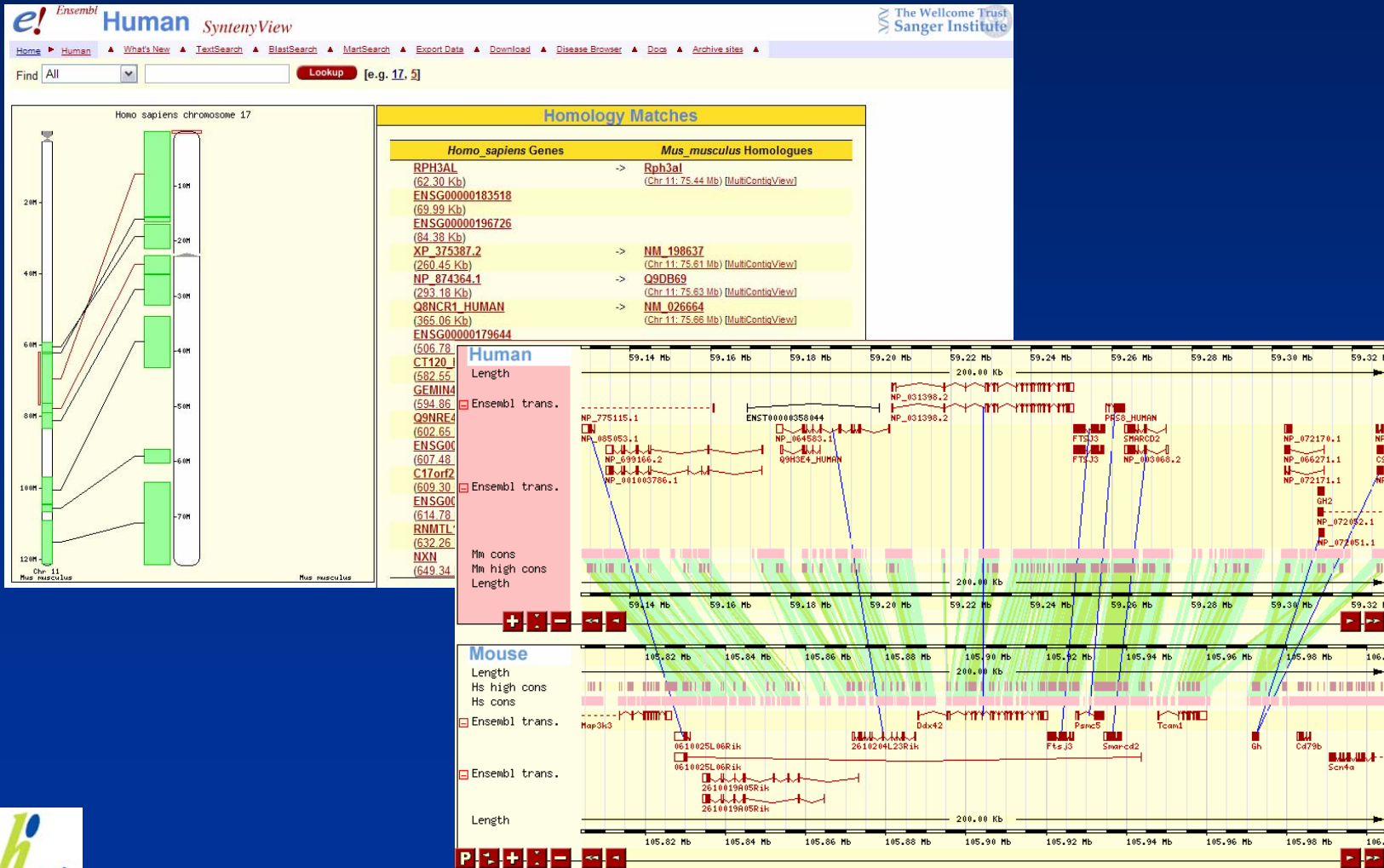
- Towards a likelihood function for the specificity and accuracy of predictions derived from the Parkin *et al* map.

Applications

- 17 *B.oleracea* ESTs were identified as candidates for investigation as disease-resistance candidate genes - high hit rate.
- 670 *Peronospora* ESTs were screened for contamination with *B.oleracea* sequence, 82 were annotated with putative housekeeping functions by homology.
- Composition of *B.oleracea* retrotransposons characterised, compared with *Arabidopsis*, and results published.
- Integration of Polish *B.oleracea* genetic map with the Warwick HRI integrated *B.oleracea* genetic map is ongoing.

Future Work

Towards a community comparative *Brassica*/Arabidopsis genome browser



Questions...