

# ***Brassica* Informatics in the UK**

## **A vision for the future**

Sarah Ayling, Dan Bolser & Pierre Carion

EMBL-EBI



ROTHAMSTED  
RESEARCH

**TGAC**   
The Genome Analysis Centre™

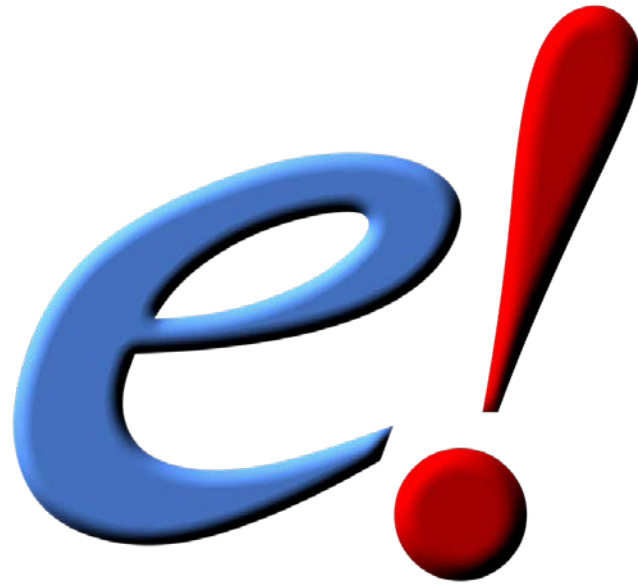


# The Aim

- Simplify the interactions between the various databases, stripping out redundancies.
- Develop user input interfaces
  - Plant Lines, Linkage Maps, Phenotypes
- Provide more information in the context of the Genome
- Develop Application Interfaces (API)

# InterStore<sup>DB</sup>

FOR *Brassica*

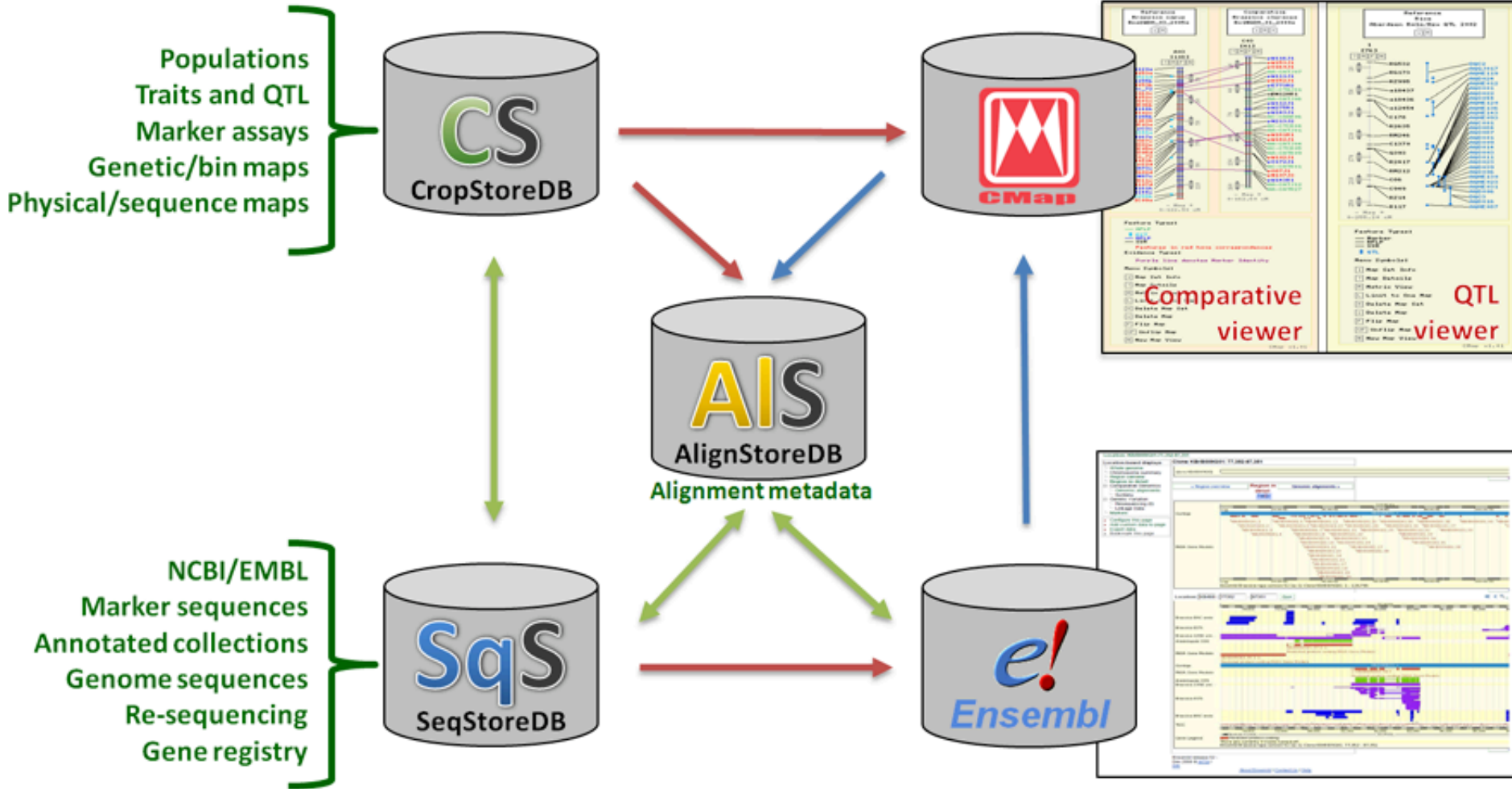


# InterStore<sub>DB</sub>

data framework

Community Input

Graphical Output





# Brassica rapa

*Brassica rapa*

Taxonomy ID [51351](#)

<http://plants.ensembl.org>

e.g. [Bra022475](#) or [A02:9566690-9570878](#) or [Carboxypeptidase](#)

## About *Brassica rapa*

Ensembl databases for *Brassica rapa* are made possible through a joint effort by the Ensembl Genomes group and [Rothamsted Research](#). The data are also browsable at [BrassEnsembl](#).



*Brassica rapa* is a widely cultivated leaf and root vegetable. The genome was sequenced as a contribution to the [Multinational Brassica Genome Sequencing Project](#) and was published in August 2011 ([Wang X, et. al. Nature 2011](#)).

This [Ensembl Genomes](#) site was originally developed by [Rothamsted Research](#) (RRes), allowing people to access the Brassica genome through the Ensembl user interface. From release 13 of Ensembl Genomes, the EBI will be maintaining the genome browser for *B. rapa* in the context of [Ensembl Plants](#).

[More information and statistics](#)

## Genome assembly: IVFCAASv1

[More information and statistics](#)

[Download DNA sequence \(FASTA\)](#)



View karyotype



Example region

## Gene annotation

**What can I find?** Protein-coding and non-coding genes, splice variants, cDNA and protein sequences, non-coding RNAs.

[More about this genebuild](#)

[Download genes, cDNAs, ncRNA, proteins \(FASTA\)](#)

[Update your old Ensembl IDs](#)



Example gene



Example transcript

## Comparative genomics

**What can I find?** Homologues, gene trees, and whole genome alignments across multiple species.

[More about comparative analysis](#)

[Download alignments \(EMF\)](#)



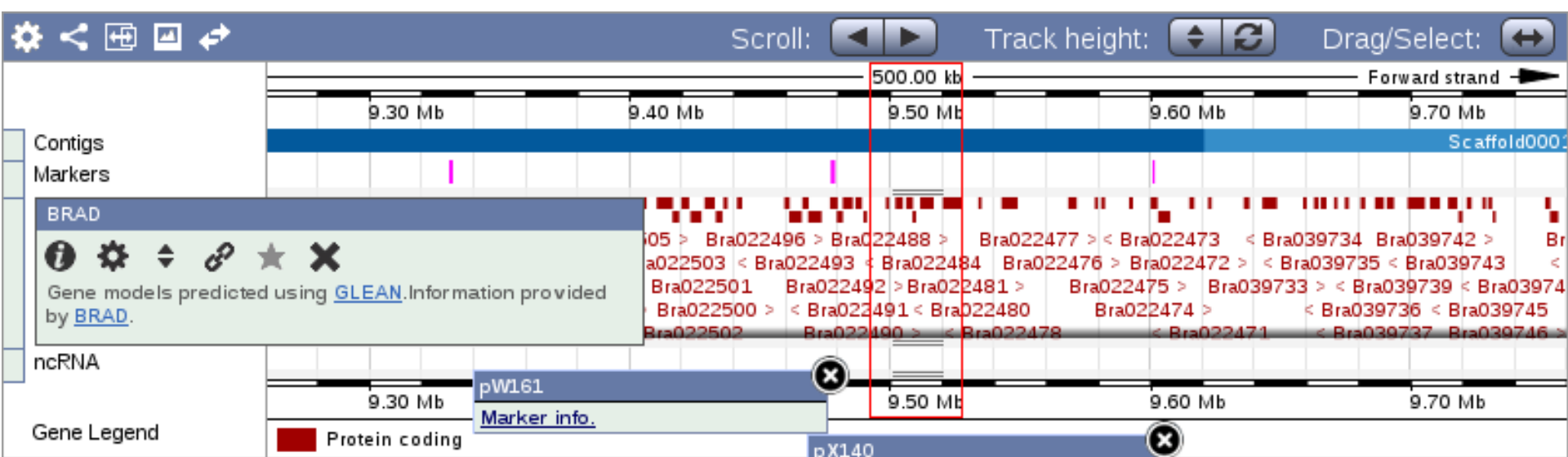
Example gene tree

## Variation

This species currently has no variation database. However you can process your own variants using the Variant Effect Predictor:

[Variant Effect Predictor](#)





Location:

Gene:

Navigation: [Left Arrow] [Left Arrow] [Zoom In] [Slider] [Zoom Out] [Right Arrow] [Right Arrow]

Novel transcript

Transcript [Bra022482.1](#)

Gene [Bra022482](#)

Protein product [Bra022482.1-P](#)

Gene GC 47.32

Location [Chromosome A02: 9,502,212-9,504,356](#)

Gene type Predicted protein coding

Transcript type Predicted protein coding

Strand Forward

Base pairs 2,145

Amino acids 714

Analysis GLEAN prediction

Prediction method Gene models predicted using [GLEAN](#). Information provided by [BRAD](#).

Contigs

BRAD

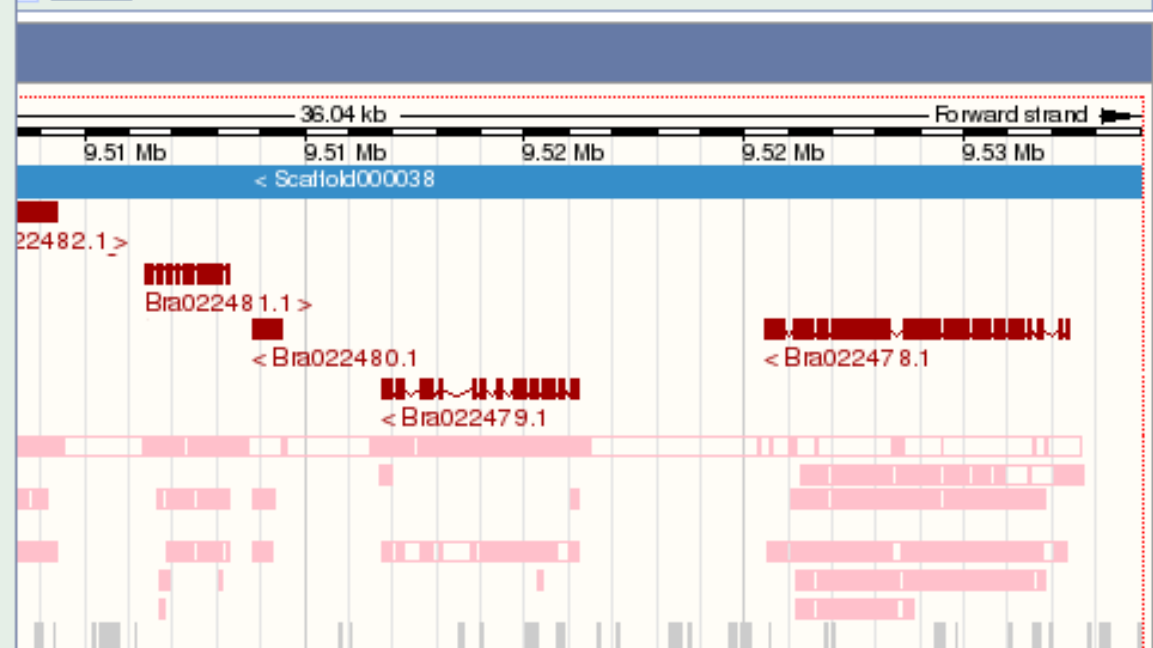
BRAD

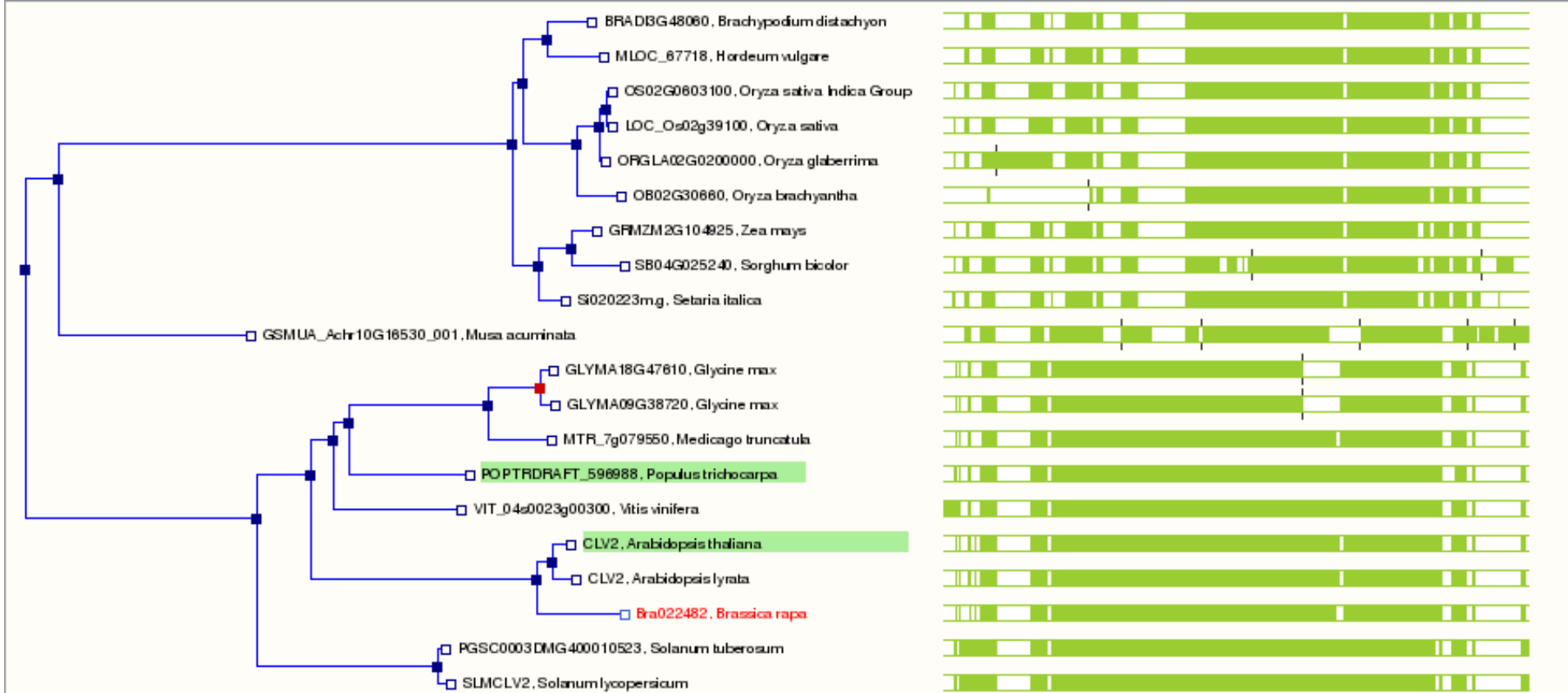
Arabidopsis thaliana

Oryza sativa

Vitis vinifera

All repeats





**LEGEND**

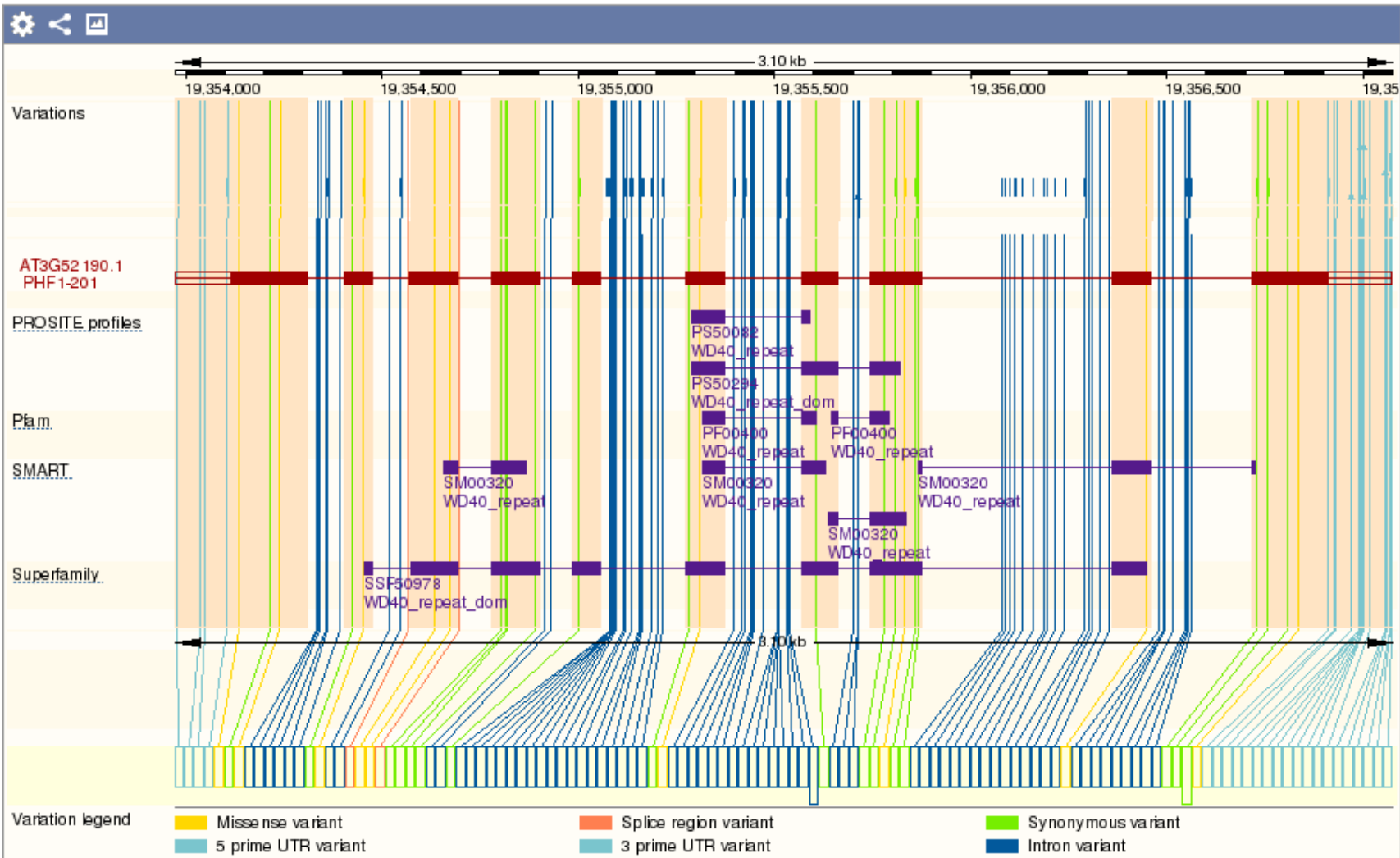
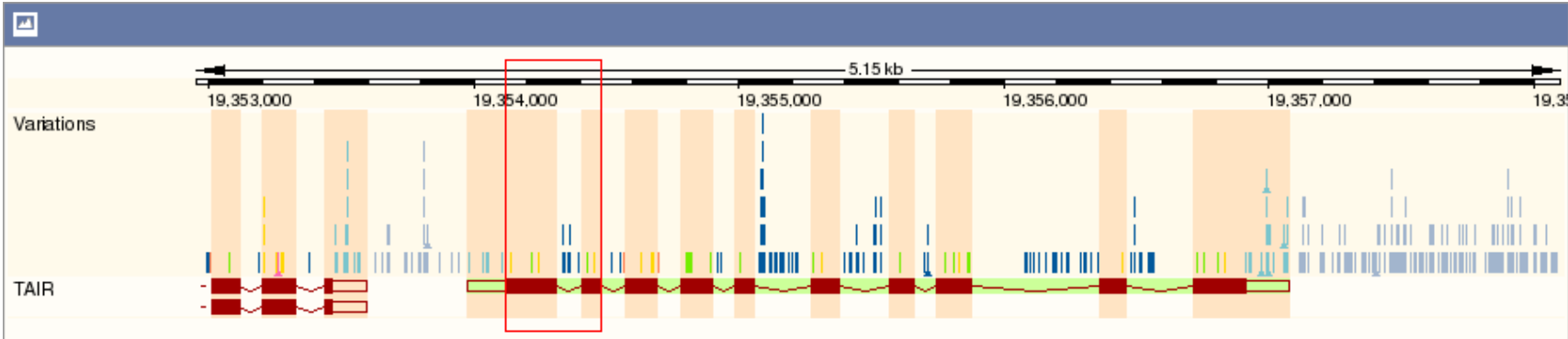
|                          |                                    |                    |                                |                              |
|--------------------------|------------------------------------|--------------------|--------------------------------|------------------------------|
| <b>Branch Length</b>     | <b>Gene s</b>                      | <b>Nodes</b>       | <b>Collapsed Node s</b>        | <b>Collapsed Alignme nts</b> |
| — x1 branch length       | Gene ID gene of interest           | □ gene node        | ◀ collapsed sub-tree           | □ 0 - 33% Aligned AA         |
| - - - x10 branch length  | Gene ID within-sp. paralog         | ■ speciation node  | ▶ collapsed (paralog)          | ■ 33 - 66% Aligned AA        |
| - - - x100 branch length | Gene ID an notated with GO:0009909 | ■ duplication node | ▶ collapsed (gene of interest) | ■ 66 - 100% Aligned AA       |
|                          |                                    | ■ ambiguous node   |                                |                              |
|                          |                                    | ■ gene split event |                                |                              |

**Expanded Alignments**

□ Gap

■ Aligned AA





# *Brassica* Informatics in the UK

## Questions and Discussion Time

