

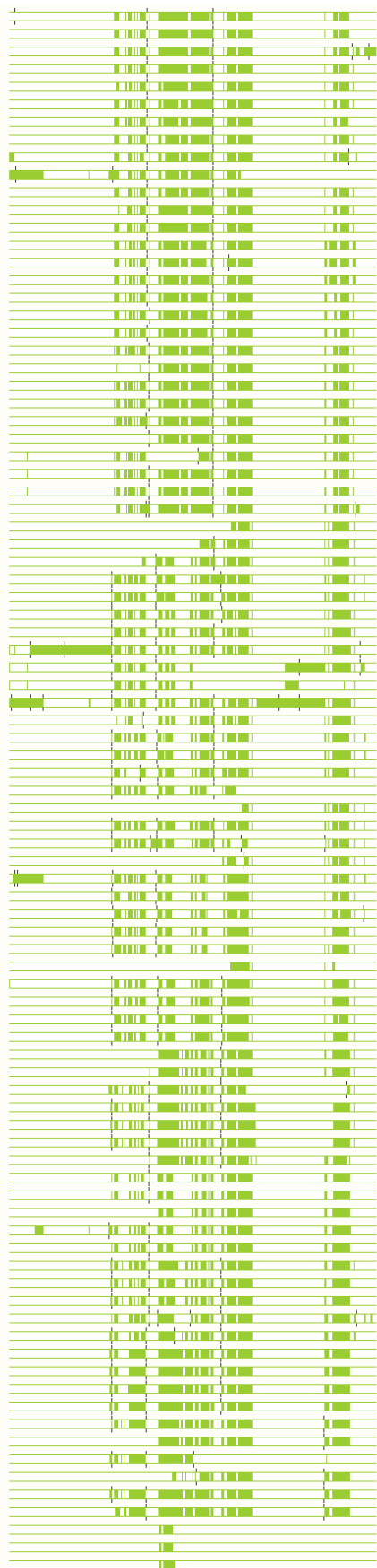
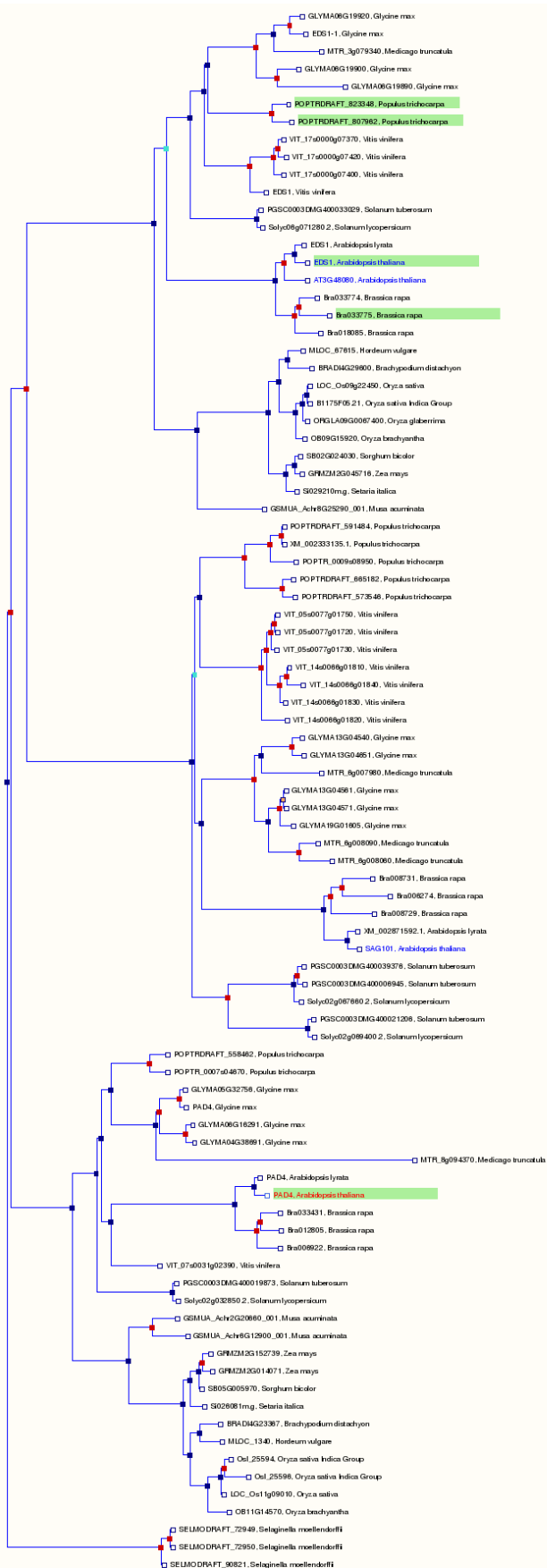
Figure 2 Synteny analyses available through Ensembl Plants

<i>Oryza sativa japonica</i>	N/A													
<i>Arabidopsis thaliana</i>	-	N/A												
<i>Brachypodium distachyon</i>	YES	-	N/A											
<i>Oryza sativa indica</i>	-	-	-	N/A										
<i>Populus trichocarpa</i>	-	YES	-	-	N/A									
<i>Sorghum bicolor</i>	YES	-	YES	-	-	N/A								
<i>Vitis vinifera</i>	-	-	-	-	YES	-	N/A							
<i>Arabidopsis lyrata</i>	-	YES	-	-	YES	-	YES	N/A						
<i>Zea mays</i>	YES	-	-	-	-	YES	-	-	N/A					
<i>Solanum lycopersicum</i>	-	-	-	-	-	-	-	-	-	N/A				
<i>Solanum tuberosum</i>	-	-	-	-	-	-	-	-	-	YES	N/A			
<i>Hordeum vulgare</i>	YES	-	YES	-	-	-	-	-	-	-	-	N/A		
	O.sat	A.tha	B.dis	O.ind	P.tri	S.bic	V.vin	A.lyr	Z.may	S.lyc	S.tub	H.vul		

The bread wheat genome *Triticum aestivum* is not yet sufficiently well assembled to be presented as a species in its own right in Ensembl, but ESTs, DNA contigs and polymorphic loci have been aligned to the syntenic locations in barley and *Brachypodium* to enable the wheat sequences to be seen in an appropriate context. Bread wheat will be promoted to a first class status within Ensembl Plants when a more contiguous assembly becomes available.

The protein-centric analysis has (as of July 2013) placed 8,919,135 proteins from 23 plant genomes and selected outlying eukaryotic species (*Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Ciona intestinalis* and *Saccharomyces cerevisiae*) into 43,771 clusters. The bread wheat precursor genomes, which are the only genomes in Ensembl Plants presently missing from these analyses, will be included in the protein-centric analysis with the release due September 2013. For each cluster, an alignment has been performed and an evolutionary history has been inferred, with putative speciation and gene duplication events inferred. **Figure 3** shows a graphical representation of a typical tree, one of the representations of the data that can be visualised through the Ensembl Plants site.

Figure 3: Visualisation of a gene tree containing the *Arabidopsis thaliana* PAD4 gene. The graphic illustrates a new feature: users can select certain annotations (from the Gene Ontology or InterPro), and highlight the genes within the tree that have been annotated with such terms (in this diagram, the highlighted genes have been annotated with the GO term “regulation of hydrogen peroxide metabolic process”).



LEGEND

Branch Length	Genes	Nodes	Collapsed Node s	Collapsed Alignme nts	Expanded Alignments
— x1 branch length	Gene ID gene of interest	□ gene node	▶ collapsed sub-tree	0 - 33% Aligned AA	□ Gap
- - - x10 branch length	Gene ID within-sp. paralog	■ speciation node	▶ collapsed (paralog)	33 - 66% Aligned AA	■ Aligned AA
- - - x100 branch length	Gene ID annotated with GC0010310	■ duplication node	▶ collapsed (gene of interest)	66 - 100% Aligned AA	
		■ annotated with GC0010310			
		■ ambiguous node			
		■ gene split event			