**transPLANT milestone report**

**MS18 (work package 7): 10 reference genomes incorporated in transPLANT hub and submitted to comparative analysis**

EMBL-EBI are providing access to many plant genomes described in the registry available for interactive and programmatic analysis through the Ensembl Plants (<http://plants.ensmebl.org>) site. These data are shared with other partners through the use of DAS and other web services, and will be available for search through the transPLANT website (<http://www.transplantdb.org>) in the near future. Over the course of the project, we will work to develop further tools to promote interoperability among all the resources in development by project partners.

Ensembl, originally developed in the course of the Human Genome Project but subsequently applied to other domains, is a powerful tool suite for the analysis and display of genome scale data, and Ensembl Plants is the EBI’s primary user interface for accessing plant data. We have used transPLANT funding to increase our capacity to include additional reference genomes incorporated in Ensembl Plants. In the first year of the grant, we have made 5 releases of Ensembl Plants, and incorporated the following additional genomes: *Brassica rapa* (turnip)*,*  [*Chlamydomonas reinhardtii* (a red alga), *Glycine max* (soybean)*, Cyanidioschyzon merolae*](http://ensemblgenomes.org/oryza_brachyantha)(a green alga), *Oryza glabberina* (African rice), *Physcomitrella patens* (a moss), *Selaginella moellendorffii* (spikemosse), *Setaria italic* (foxtail millet), *Solanum lycopersicum* (tomato)*,* and *Zea mays* (maize); taking the total number of genomes in the system to 19. The inclusion of plants *senua latae* – i.e. including both the green and the red algae – within Ensembl Plants has been undertaken specifically to support the comparative aspects.

Genome, and proteins coding, sequences have been analysed comparatively using the Ensembl Compara functional genomics pipeline, which has 2 elements: a protein-based analysis, which infers evolutionary relationships after clustering and alignment (and which are performed over the domain of all plants) and a parwise DNA-based analysis, performed using the alignment tools blastZ and lastZ. The currently available analyses are given in the table 1.

**Table 1**: Pairwise DNA-based comparative analyses available through Ensembl Plants

|  |  |  |
| --- | --- | --- |
| **Genome 1** | **Genome 2** | **Method** |
| *Arabidopsis thaliana* | *Arabidopsis lyrata* | blastz |
| *Arabidopsis thaliana* | *Brachypodium distachyon* | blastz |
| *Arabidopsis thaliana* | *Brassica rapa* | lastz |
| *Arabidopsis thaliana* | *Chlamydomonas reinhardtii* | blastz |
| *Arabidopsis thaliana* | *Cyanidioschyzon merolae* | lastz |
| *Arabidopsis thaliana* | *Glycine max* | blastz |
| *Arabidopsis thaliana* | *Oryza brachyantha* | lastz |
| *Arabidopsis thaliana* | *Oryza glaberrima* | blastz |
| *Arabidopsis thaliana* | *Oryza indica* | blastz |
| *Arabidopsis thaliana* | *Oryza sativa* | blastz |
| *Arabidopsis thaliana* | *Physcomitrella patens* | blastz |
| *Arabidopsis thaliana* | *Populus trichocarpa* | blastz |
| *Arabidopsis thaliana* | *Selaginella moellendorffii* | blastz |
| *Arabidopsis thaliana* | *Setaria italica* | lastz |
| *Arabidopsis thaliana* | *Solanum lycopersicum* | lastz |
| *Arabidopsis thaliana* | *Sorghum bicolor* | blastz |
| *Arabidopsis thaliana* | *Vitis vinifera* | blastz |
| *Oryza sativa* | *Arabidopsis lyrata* | blastz |
| *Oryza sativa* | *Arabidopsis thaliana* | blastz |
| *Oryza sativa* | *Brachypodium distachyon* | blastz |
| *Oryza sativa* | *Brassica rapa* | lastz |
| *Oryza sativa* | *Chlamydomonas reinhardtii* | blastz |
| *Oryza sativa* | *Cyanidioschyzon merolae* | lastz |
| *Oryza sativa* | *Glycine max* | blastz |
| *Oryza sativa* | *Oryza brachyantha* | lastz |
| *Oryza sativa* | *Oryza glaberrima* | blastz |
| *Oryza sativa* | *Oryza indica* | blastz |
| *Oryza sativa* | *Physcomitrella patens* | blastz |
| *Oryza sativa* | *Populus trichocarpa* | blastz |
| *Oryza sativa* | *Selaginella moellendorffii* | blastz |
| *Oryza sativa* | *Setaria italica* | lastz |
| *Oryza sativa* | *Solanum lycopersicum* | lastz |
| *Oryza sativa* | *Sorghum bicolor* | blastz |
| *Oryza sativa* | *Vitis vinifera* | blastz |
| Oryza sativa | *Zea mays* | blastz |
| Oryza indica | *Brachypodium distachyon* | blastz |
| Zea mays | *Sorghum bicolor* | blastz |

The protein-centric analysis has placed 646,588 proteins from the 19 plant genomes and selected outlying eukaryotic species. (*Homo sapiens*, *Drosophila melangoster*, *Caenorhabditis elegans, Ciona intestinalis* and *Saccharomyces cerevisiae*) into 38,154 clusters. For each cluster, an alignment has been performed and an evolutionary history has been inferred, with putative speciation and gene duplication events inferred. **Figure 1** 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 1:** Visualisation of a gene tree for a WD40 domain containing protein.

