PLAZA workshop

Orthologs, paralogs and phylogenetic trees

October 25-26, 2021

# Topics Covered

In his series of exercises you will learn how to efficiently search for the ortholog(s) of your gene of interest, in the (crop) species that you are working on. You will also learn how to deal with the presence of paralogs both in the model species and (crop) species. Further interpretation of the different relationships between the genes will be done by using the Multiple Sequence Alignments (MSA) and phylogenetic trees.

# Exercises

## Part 1: Integrative Orthology

1. Search for the *Zea mays* *(cv. B73)* gene **Lea14-A** in the **Monocot 4.5** PLAZA [Toolbar🡪Search]
   1. What are the orthologs in *Brachypodium distachyon* for this gene (using the Integrative Orthology approach) [Gene page 🡪Toolbox 🡪 Explore the orthologs using the Integrative Orthology Viewer]
      1. Is there a clear best ortholog?
      2. What are the possible paralogs of our *Zea mays* gene?
      3. Which *Brachypodium* gene do the paralogs of our *Zea mays* gene have as common ortholog? How many orthology types are there as support?
   2. What are the (integrative) orthologs in *Oryza sativa ssp. Japonica*? [Gene page 🡪Explore the orthologs using the Integrative Orthology Viewer]
      1. Which is the (putatively) best ortholog?
      2. Explore the associated phylogenetic tree [Integrative Orthology Viewer 🡪 Click on diamond representing the relationship]
         1. Search for our *Zea mays* gene AND search for our *Oryza sativa* gene. [PhyD3 🡪Search bar ] Are these two genes located closely together?
         2. Look within the tree for the *Triticum aestivum* genes that are most closely located to our *Zea mays* gene of interest [Hint: species code tae]
            1. How many *Triticum aestivum* genes do we find as putative tree-based ortholog(s) of our *Zea mays* gene?
            2. Repeat exercise 1.a. using *Triticum aestivum* instead of *Brachypodium distachyon*: does the Integrative Orthology approach correspond with what you saw in the phylogenetic tree?
         3. How many (out-paralogous) clades exist in the Poales for this gene family? [Lea14-A Gene family 🡪 Phylogenetic tree 🡪 tree with speciation/duplication information]
2. Search for the term **RABA2a** in *Zea mays (cv B73)* in **the Monocot 4.5** PLAZA [Toolbar🡪Search]
   1. How many results are there?
   2. Is there a difference between the Integrative Orthologs in *Arabidopsis thaliana* for each of these *Zea* genes [Gene page 🡪 Toolbox 🡪 Explore the orthologs using the Integrative Orthology Viewer]?
   3. Can you confirm these results using the phylogenetic tree of the sub-family?

## Part 2: Gene Family Tools

1. Search for the term **EGY1** in *Arabidopsis thaliana* in the **Monocot 4.5** PLAZA [Toolbar 🡪Search]
   1. Do the homologs of this gene exhibit large sequence similarity? [Gene family page Toolbox 🡪View the similarity heatmap]
   2. Given the gene prefixes, what species has the most outliers compared to the other species in this gene family? [Heatmap 🡪 Color coding]
2. Search for the term **fungal pathogens** in the **Dicots 4.5** PLAZA [Toolbar 🡪Search ]
   1. Which of these genes is located in the cytosol? [Search descriptions]
   2. Is the family that is associated with the cytosol gene, equally distributed over the species within PLAZA? [Family page 🡪Toolbox 🡪Explore the expansion/depletion of species in this gene family].
      1. For which species does this family not have any genes? Does this make biological sense?
   3. View the phylogenetic tree [Family page 🡪Toolbox 🡪Explore the phylogenetic tree]
      1. Are most genes single-exon or multi-exon in this family? [Tree page 🡪Toolbox]
      2. What is the length of MSA used to create the phylogenetic tree? [Tree page 🡪stripped length]