PLAZA workshop

Genes, Gene Families, and Genome Organization

October 25-26, 2021

# Topics Covered

In this series of exercises you will learn how to efficiently search the PLAZA platform for genes of interest, as well as browse the associated linked pages for additional information. By using the tools available on the gene and gene family pages you will be able to further explore and deduce the homologous and orthologous relationships between genes. Furthermore, you will learn how the interpret genome colinearity both within and between species.

# Exercises

## Part 1: Searching and browsing the PLAZA platform

1. Search for the *Zea mays (cv. B73)* gene **Lea14-A** in the **Monocot 4.5** PLAZA [Toolbar🡪Search]
   1. Identifiers
      1. What is the gene identifier? Where does this identifier come from?
      2. What is the old v3 Maize identifier? Search the PLAZA platform with this v3 identifier. Is this correct?
   2. Structural Annotation
      1. On which chromosome and which strand of the genome is the gene located?
   3. Gene Family
      1. How many homologs does the gene have? [Gene page 🡪 Family ]
      2. What is the clade that certainly does NOT contain Lea14-A homologs? [Gene family page 🡪 Smallest encompassing clade + Species Tree ]
      3. What is the percentage of species with a Lea14-A homolog? Does this match with the indicated smallest encompassing clade?
2. Search for the genes with the description **E2F3** in the **Dicots 4.0** PLAZA
   1. How many results do we get?
   2. Search comparison
      1. Search for the *Arabidopsis thaliana* gene **E2F** in the **Dicots 4.5** PLAZA.
      2. How many results do we get?
      3. What is the difference compared to searching for **E2F3**?
   3. Structural Annotation
      1. View the sequence of the gene (**E2F3 – Arabidopsis thaliana**). Does the structure of the sequence match the strand of the gene and the displayed gene model? [Toolbox 🡪View sequences]
   4. Description
      1. In what pathway is this gene (**E2F3 – Arabidopsis thaliana**) involved? [Gene page 🡪 GO table]
   5. Gene Family
      1. Explore the synteny of the sub family of the gene (**E2F3 – Arabidopsis thaliana**) [Gene family page 🡪Explore the local gene organization]
         1. Is the genome organization conserved between *Arabidopsis thaliana* and *Arabidopsis lyrata* for all orthologous copies? [Synteny Plot 🡪 Species selection]
         2. Substitute *Arabidopsis lyrata* for *Populus trichocarpa* and *Gossypium raimondii* in above question. Can we infer any knowledge about conserved gene positions?
3. Within the **Monocots 4.5** PLAZA, search for gene families that are *only* found in the BOP-clade [Analyze 🡪 Gene Family Finder ]
   1. How many families are there?
   2. Which species clearly has the largest number homologs in the found families? Can you explain why?

## Part 2: Genome organization and colinearity

1. Explore the genome colinearity within *Ananas comosus* in the **Monocots 4.5** PLAZA. [Analyze 🡪 WGDotplot 🡪 JavaScript WGDotplot]
   1. How many conserved duplicated regions have been detected?
   2. Approximately, what is the mean Ks-value of one of the larger colinear regions? [JavaScript WGDotplot 🡪Click region OR JavaScript WGDotplot 🡪hover over color legend]
2. In the **Monocots 4.5** PLAZA, find the colinear regions that are conserved in all 3 Oryza and all 3 Zea genomes. [Analyze 🡪 Colinear Regions Finder]
   1. How many colinear regions are found?
   2. Explore the colinear region with the most segments.
      1. Is there a gene position that is conserved across ALL species that contain this conserved region? [Group species together]