PLAZA workshop

Advanced orthology, synteny, and expression analysis in cereals

25-26 October 2021

# Topics Covered

In this series of exercises you will learn how to study the evolution of genes and genome organization between different plant species with varying levels of gene complexity, and combine this information with gene expression data available in public resources.

* ePlant maize - <http://bar.utoronto.ca/eplant_maize/> - will be used to browse maize gene expression data.
* Wheat Expression Browser - <http://www.wheat-expression.com/> - will be used to browse wheat gene expression data.

# Exercises

## Part 1: Exploration of gene family evolution and genome organization in maize

1. Starting from the gene GRMZM2G121308, encoding an expansin gene, identify how many copies exist in this gene family for the different maize genomes B73, B104 and PH207.

Note that for the maize B73 genome, gene identifiers changed between v3 and v4 of the gene annotation (search using GRMZM2G121308 = Zm00001d045861).

[PLAZA **Monocot 4.5** PLAZA: Search 🡪 Gene 🡪 Gene Family HOM04x5M000125]

1. Starting from Zm00001d045861, use the Integrative Orthology Viewer to identify orthologs in the other maize genomes. How many orthologs, per species, are supported by two or more supporting methods?

[Gene 🡪 Explore the orthologs using the Integrative Orthology Viewer]

1. Starting from Zm00001d045861, in how many species does this gene show conserved collinear genome organization? [Gene 🡪 View all collinear gene pairs]
2. Perform a whole-genome collinearity analysis using B73 and the other two maize genomes (B104 and PH207). How many anchorpoints (= colinear gene pairs) do you find per pairwise comparison?

Explain why for some maize genomes the observed genome organization is limited. [WGDotplot 🡪 Organism1: Zea mays B73, Organism2: Zea mays B104 ; repeat with Organis,2: Zea mays PH207]

1. Starting from the gene family HOM04x5M000125, use the Interactive Phylogenetics Module to build a phylogenetic tree selecting homologs from maize B73, *Oryza sativa ssp. Japonica* and add *Amborella trichopoda* as outgroup. Do you find for every Oryza gene exactly 1 ortholog in maize B73?

[Gene Family 🡪 Create a custom phylogenetic tree using this gene family as seed 🡪 Data settings: Species selection ; Program Settings: MUSCLE, FastTree, Editing: Positions only]

*This can take a while, continue with the next step in a new Browser Tab*

1. What is the prevalence of tandem gene duplication in this gene family / for these 3 maize species? Can you spot\* tandem duplicates in the phylogenetic tree generated in step 5?

[Gene Family 🡪 Follow link “Tandem/block/both gene duplicate percentage within this gene family”]

\* note that tandem genes are normally in close proximity and therefore typically have similar gene identifiers, only differing in the last digits.

## Part 2: Comparative gene expression analysis in maize and wheat

1. When/where is the maize gene Zm00001d002176 expressed?

[Search using the V3 gene identifier GRMZM2G175321]

[ePlant maize 🡪 Search Gene 🡪 Select Tissue and Experiment eFP viewers 🡪 select Sekhon Atlas]

1. Select orthologs in wheat (*T. aestivum*) using the Integrative Orthology Viewer (only keep putative orthologs supported by 2 methods) and check the expression levels of these genes in wheat. Are all orthologs markers for spikes?

[Wheat Expression Viewer 🡪 Enter genes 🡪 Select Developmental time-course of Chinese Spring + Developmental time-course of Azhurnaya]