Glossary

A

- Alignment editing : procedure removing non-homologous positions and partial/diverged sequences from a multiple sequence alignment prior to tree construction.
- Anchor point : colinear gene pair (~genes from the same gene family located in a colinear segment; see also Glossary - Colinearity)
- Angiosperms : The flowering plants or angiosperms (Angiospermae or Magnoliophyta) are the most diverse group of land plants. The flowering plants and the gymnosperms are the only extant groups of seed plants. The flowering plants are distinguished from other seed plants by a series of apomorphies, or derived characteristics.
- AnnoJ : a flexible genome browser designed for visualizing various features (eg. gene models, deep sequence data, etc.) onto a genome sequence.
- Arabidopsis: A genus in the family Brassicaceae. They are small flowering plants related to cabbage and mustard. This genus is of great interest since it contains (Arabidopsis thaliana), one of the model organisms used for studying plant biology and the first plant to have its entire genome sequenced.
- Archaeopteryx : java application based on FORESTER library to visualize trees.

B

- Bacillariophyta : Bacillariophyta is the division of the diatoms, a major group of algae and one of the most common types of phytoplankton. Most diatoms are unicellular, although they can exist as colonies in the shape of filaments or ribbons, fans, zigzags and other forms.
- BEPClade : A subfamily of the true grass family Poaceae.
- Bootstrap analysis : a type of statistical analysis used to test the reliability of specific branches in an evolutionary tree. The non-parametric bootstrap proceeds by re-sampling the original data, with replacement, to create a series of bootstrap samples of the same size as the original data. The bootstrap percentage of a node is the proportion of times that node is present in the set of trees that is constructed from the new data sets.
- Bonferroni correction : correction method for multiple testing that was used as part of the GO enrichment tool.
- Brassicales : The Brassicales are an order of flowering plants, belonging to the eurosids II group of dicotyledons under the APG II system.

C

- Chlamydomonadales : In taxonomy, the Volvocales, also known as Chlamydomonadales, are an order of flagellate or pseudociliate green algae, specifically of the Chlorophyceae. Volvocales can form planar or spherical colonies.
- *Cis*-Regulatory element : DNA sequences located within or next to transcribed genes and that either increase (enhancers) or decrease (repressor or silencer, depending on their mechanism of action) gene transcription. *Cis*-regulatory elements act by recruiting *trans*-acting transcriptional activator or repressor proteins.
- Chlorophyta : A division of green algae, includes about 7,000 species of mostly aquatic photosynthetic eukaryotic organisms. Like the land plants (bryophytes and tracheophytes), green algae contain chlorophylls a and b, and store food as starch in their plastids.
- Circle Plot : reports all colinear regions within a single species using a circular representation.
- Clustering : Edit in /app/scripts/data_files
- Colinearity : two genomic segments can be considered colinear if they share the same gene content (homologs) in the same order.
- Comparing : refers to the species included in an i-ADHoRe experiment to delineate genomic homology.

D

- Duplication consistency score : measure to find dubious duplication nodes in a reconciled tree that are artifacts from the tree construction procedure. Duplication nodes with a low consistency score can be considered speciation nodes.
- Duplication type : indicates Tandem or Block duplication event.

E

- Embryophyta : Embryophyta, or landplants, live primarily in terrestial habbits, in contrast to aquatic species such as the algae. The Embryophyta include trees, flowers, ferns, mosses, and various other green land plants.
- Eudicots : Eudicots and Eudicotyledons are terms introduced by Doyle & Hotton (1991) to refer to a group of flowering plants that had been called "tricolpates" or "non-Magnoliid dicots" by previous authors. The term means, literally, "true dicotyledons" as it contains the majority of plants that have been considered dicotyledons and have typical dicotyledonous characters.
- Eukaryotes : Eukaryotes are one of the main domains of life, the other domains being Bacteria and Archae, both prokaryotes. Eukaryotes are organisms with cells containing complex structures enclosed in membranes. The defining membrane-bound structure that sets eukaryotic cells apart from prokaryotic cells is the nucleus, or nuclear envelope, within which the genetic material is carried.
- Euphorbiaceae : The Spurge family (Euphorbiaceae) is a large family of flowering plants with 300 genera and around 7,500 species. Most are herbs, but some, especially in the tropics, are also shrubs or trees. Some are succulent and resemble cacti.
- Expansion Plot : explore the copy-number gene family variation between two groups of species.

F

- Fabids: Subgroup of the "true rosides". Also known as Fabidae or Eurosids I. Identified as being separate from Malvids through molecular analyses, although the relationships within eurosids I and II are not fully resolved.
- Functional clustering : physical clustering of functionally related genes.

G

- Galegoids : The Galegoids are mostly temperate and include clover, fava bean and the model legumes.
- Gene family : a set of homologous genes grouped by sequence similarity using Markov clustering.
- Gene family expansion: this tool reports the expansion/depletion of a species/lineage (through gene copy number variation compared to total proteome sizes of the species) within a gene family.
- Gene Family Finder : this tool enables to identify (expanded) gene families specific to one or more species.
- Gene type : refers to a locus encoding a protein-coding gene, RNA, pseudo gene or TE (Transposable Element).
- Genome Browser: tool that allows users to visually inspect the genomes, with various features mapped onto the raw sequence.
- GO : a controlled vocabulary to describe gene and gene product attributes in any organism.
- GO depth : indicates for a GO term the shortest distance (through parent-child relationships) to the root in the GO hierarchy.
- GO enrichment : the over-representation of a certain GO term in a gene set compared to the genome-wide background frequency. The statistical significance is determined using the hypergeometric distribution with Bonferroni correction.
- GO projection : methodology that uses orthology relations to transfer functional annotation between genes and/or species.
- GO projection source gene : refers to the orthologous source gene that was used to transfer the functional annotation. The tree icon links to the phylogenetic tree that was used to delineate the orthologous group.
- GO type : refers to the three organizing principles of GO being Cellular Component (CC), Biological Process (BP) and Molecular Function (MF).
- Green plants : Living organisms belonging to the kingdom Plantae. They include familiar organisms such as trees, herbs, bushes, grasses, vines, ferns, mosses, and green algae. The scientific study of plants, known as

botany, has identified about 350,000 extant species of plants, defined as seed plants, bryophytes, ferns and fern allies.

H

- Homologs: (or homologous genes) genes sharing similarity due to common ancestry

I

- i-ADHoRe : iterative Automated Detection of Homologous Regions, an algorithm to find genomic homology based on gene colinearity.
- i-ADHoRe experiments : refers to the species included in an i-ADHoRe experiment to delineate genomic homology.
- inparalogs : duplicated genes (or paralogs) that originated after a speciation event
- Integrative Orthology Viewer: tool to explore for a query gene the orthologous genes in other species using different evidences.
- Interactive Phylogenetics Module: tool to create custom phylogenetic trees based on PLAZA content.
- InterPro : a database of protein families, domains and functional sites in which identifiable features found in known proteins can be applied to new protein sequences.

J

- Jalview : Jalview is a multiple alignment editor written in Java. It is used widely in a variety of web pages but is available as a general purpose alignment editor.

K

- Keywords : Most frequent functional annotation terms associated with a gene family.
- Ks : the synonymous substitution rate reports the fraction of synonymous substitutions over all synonymous sites.
- Ks-dating tool: method to explore several Ks graphs of colinear gene pairs simultaneously.

L

- Land plants : The embryophytes are the most familiar group of plants. They include trees, flowers, ferns, mosses, and various other green land plants. All are complex multicellular eukaryotes with specialized reproductive organs. With very few exceptions, embryophytes obtain their energy through photosynthesis (that is, by absorbing light); and they synthesize their food from carbon dioxide.

M

- Magnoliophyta : Magnoliophyta, also known as angiospermae or flowering plants, are seed producing plants with a number of derived characteristics (such as flowers, endosperm within the seeds and the production of fruits) which distinguish them from other seed-producing plants such as the gymnosperms.
- Malpighiales : Malpighiales is one of the largest orders of flowering plants, containing about 16000 species. The order is very diverse and hard to recognize except with molecular phylogenetic evidence. It is not part of any of the classification systems that are based only on plant morphology.
- Malvids : subgroup of the "true rosides". Also known as Eurosids II. Identified as being separate from Fabids through molecular analyses, although the relationships within eurosids I and II are not fully resolved.
- Mamiellales : Mamiellales are an order of green algae, specifically primitive eukaryotic, marine green algae, with low cellular complexity.
- Markov clustering : a graph-based clustering method that delineates clusters in a protein-protein similarity graph in a process that is sensitive to the density and the strength of the connections.

- Micromonas : The Micromonas genus consists of algae which are distributed widely in both cold and warm waters.
- Monocots : Monocotyledons or monocots are one of two major groups of flowering plants (angiosperms) that are traditionally recognized, the other being dicotyledons or dicots. Monocot seedlings typically have one cotyledon (seed-leaf), in contrast to the two cotyledons typical of dicots.
- Multiple sequence alignment : or MSA, an alignment of two or more biological sequences, generally protein, DNA, or RNA. In general, the input set of query sequences are assumed to have an evolutionary relationship by which they share a lineage and are descended from a common ancestor.
- Multiplicon : a set of homologous genomic segments (detected with i-ADHoRe).
- Multiplicon level : the number of homologous genomic segments in a multiplicon.
- Multiplicon navigation : use the arrows to scroll through the multiplicon.
- Multiplicon segments : the genomic regions grouped within a multiplicon.
- Multiplicon View : displays the aligned gene strings of a set of homologous segments.

N

- N2FixingClade : The nitrogen-fixing clade contains a high number of actinorhizal plants (which have root nodules containing nitrogen fixing bacteria, helping the plant grow in poor soils). Not all plants in this clade are actinorhizal, however.
- Normalization : transformation of the original data by reporting relative frequencies (percentages) for the different series being displayed.

O

- Orthologs : (or orthologous genes) homologs created through a speciation event.
- Orthology type : indicates for a Source and Target species the number of orthologous genes per species (S-to-T).
- OrthoMCL : graph-based clustering algorithm to detect orthologous groups; here used to subdivide homologous gene families into sub-families (see also Markov clustering).
- Oryza : Oryza is a genus of seven to twenty species of grasses in the tribe Oryzeae, within the subfamily Bambusoideae, native to tropical and subtropical regions of Asia, Northern Australia and Africa. They are tall wetland grasses, growing to 1–2 m tall; the genus includes both annual and perennial species.
- Ostreococcus : Ostreococcus is a genus of unicellular coccoid or spherically shaped green alga belonging to the class Prasinophyceae. It includes prominent members of the global picoplankton community, which plays a central role in the oceanic carbon cycle.
- Outlier : a gene initially included in a gene family but only showing sequence similarity to a limited number of family members. Note that these genes are NOT included in MSAs and phylogenetic trees.

P

- PACCMADClade : Panicoideae is a subfamily of the true grass family. It probably constitutes two well-distinct lineages, the Andropogonodae and the Panicodae.
- Papilionoideae : Faboideae is a subfamily of the flowering plant family Fabaceae or Leguminosae. An acceptable alternative name for the subfamily is Papilionoideae. This subfamily is widely distributed and members are adapted to a wide variety of environments. Faboideae may be trees, shrubs or herbs. The flowers are classically pea shaped and root nodulation is very common.
- Paralogs : (or paralogous genes) homologs created through a duplication event
- Pathway : reactome pathway indicating biological processes and reactions.
- Phylogenetic profile : indicates the presence or absence of a gene family in a set of species.
- Phylogenetic tree : an evolutionary tree shows the relationships among various biological taxa that are believed to have a common ancestor.

R

- Reactome : reactome is a free, online, open-source, curated resource of core pathways and reactions in human biology (available for other species through projection).
- Reconciliation : reconciliation extracts information from the topological incongruence between gene and species trees to infer duplications and losses in the history of a gene family.
- Rosaceae : The Rosaceae or rose family is a large family of flowering plants, with about 2830 species in 95 genera. The name is derived from the type genus Rosa. Roses can be herbs, shrubs or trees. Most species are deciduous, but some are evergreen. They have a worldwide range, but are most diverse in the northern hemisphere.
- Rosids : The rosids are a large group of flowering plants, containing about 70,000 species, more than a quarter of all angiosperms. It is divided into 16 to 20 orders, depending upon circumscription and classification. These orders, in turn, together comprise about 140 families. The rosids and the asterids are by far the largest groups in the eudicots.

S

- Similarity heatmap : visualizes the similarities (based on BLAST bitscores) between all genes within a family.
- Similarity heatmap type : normalized values show sequence similarities relative to the highest score for the reference gene.
- Skyline plot : provides for a locus of interest an overview of the colinear regions within and between species.
- Stramenopiles : Stramenopiles, also known as heterokonts, are a major line of eukaryotes, most of them diatoms. The name heterokonts refers to the motile life cycle stage, in which the flagellate cells possess two differently-shaped flagella.
- Species : Edit in /app/scripts/data_files
- Strand orientation : Edit in /app/scripts/data_files
- Sub-family : subset of a gene family delineated using the OrthoMCL algorithm.
- Synteny plot : reports the local gene organization for homologous genes within a family.

T

- Tandem representative: all genes in a tandem cluster are remapped to a tandem representative by i-ADHoRe (as tandem genes would negatively influence the program's statistics to identify genomic homology).
- Trebouxiophyceae : a class of green algae, in the division Chlorophyta.
- Tree explorer: view the phylogenetic tree of a homologous gene family.
- Tribe-MCL: graph-based clustering algorithm that was used to group all homologous genes into gene families (see also Markov clustering).

V

- Vascular plants: Vascular plants (also known as tracheophytes or higher plants) are those plants that have lignified tissues for conducting water, minerals, and photosynthetic products through the plant. Vascular plants include the ferns, clubmosses, flowering plants, conifers and other gymnosperms.

W

- WGD : abbreviation for Whole Genome Duplication.
- WGDotplot : tool that reports in a pairwise manner all colinear regions between and/or within species.
- WGMapping : displays the organization of a set of genes on all chromosomes (for a selected species).
- Window size : the size of the region being analyzed (measured in genes).
- Workbench : toolkit included in PLAZA that allows researchers to perform analyses on user-defined gene sets.