Molecular Phylogenetics and Evolution 126 (2018) 266–278

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

New insights into the phylogeny of the TMBIM superfamily across the three of life: Comparative genomics and synteny networks reveal independent evolution of the BI and LFG families in plants

Samuel D. Gamboa-Tuz[a,1], Alejandro Pereira-Santana[b,1], Tao Zhao[b], M. Eric Schranz[b], Enrique Castano[c], Luis C. Rodriguez-Zapata[a,*]

[a] Biotechnology Unit, Centro de Investigacion Cientifica de Yucatan, 97205 Yucatan, Mexico
[b] Biosystematics Group, Wageningen University and Research, 6708 PB Wageningen, The Netherlands
[c] Biochemistry and Molecular Biology Unit, Centro de Investigacion Cientifica de Yucatan, 97205 Yucatan, Mexico

ARTICLE INFO

Keywords:
Bax inhibitor 1
Gene family evolution
Lifeguard
Programmed cell death
Synteny network
TMBIM

ABSTRACT

The Transmembrane BAX Inhibitor Motif containing (TMBIM) superfamily, divided into BAX Inhibitor (BI) and Lifeguard (LFG) families, comprises a group of cytoprotective cell death regulators conserved in prokaryotes and eukaryotes. However, no research has focused on the evolution of this superfamily in plants. We identified 685 TMBIM proteins in 171 organisms from Archaea, Bacteria, and Eukarya, and provided a phylogenetic overview of the whole TMBIM superfamily. Then, we used orthology and synteny network analyses to further investigate the evolution and expansion of the BI and LFG families in 48 plants from diverse taxa. Plant BI family forms a single monophyletic group; however, monocot BI sequences transposed to another genomic context during evolution. Plant LFG family, which expanded trough whole genome and tandem duplications, is subdivided in LFG I, LFG IIA, and LFG IIB major phylogenetic groups, and retains synteny in angiosperms. Moreover, two orthologous groups (OGs) are shared between bryophytes and seed plants. Other several lineage-specific OGs are present in plants. This work clarifies the phylogenetic classification of the TMBIM superfamily across the three domains of life. Furthermore, it sheds new light on the evolution of the BI and LFG families in plants providing a benchmark for future research.

1. Introduction

Programmed Cell Death (PCD) is essential for cellular homeostasis, development, and environmental responses of multicellular organisms (Lord and Gunawardena, 2012); its importance in unicellular organisms is also recognized (Jiménez-Ruiz et al., 2010). Most research regarding PCD has focused on apoptosis and the animal B-Cell Lymphoma 2 (BCL-2) gene family, which contains both pro- and anti-apoptotic members (Chipuk et al., 2010). No BCL-2 homologues have been identified outside the animal kingdom. However, biological functions of the BCL-2 family remain conserved in fungi and plants. For example, heterologous expression of the human pro-apoptotic BCL-2 Associated X (BAX) gene/protein activates cell death in budding yeast (*Saccharomyces cerevisiae*) and Arabidopsis (*Arabidopsis thaliana*) (Baek et al., 2004; Priault et al., 2003). Another group of PCD regulators, linked to BCL-2, is the Transmembrane BAX Inhibitor Motif containing (TMBIM) superfamily, also referred to as BAX Inhibitor-1 (BI-1). The name of this superfamily derives from the first characterized human BI-1 gene (HsBI-1/TMBIM6), which inhibits BAX-induced PCD (Xu and Reed, 1998). This HsBI-1 protein is 237 amino acids (aa) long, contains a domain composed of 6–7 transmembrane regions, and locates in the endoplasmic reticulum (ER) (Xu and Reed, 1998). In mammals, this protein participates in the regulation of cytosolic calcium concentrations, protection against ER stress, and cancer development, among other functions (Henke et al., 2011; Robinson et al., 2011). Other five TMBIM proteins (some with similar anti-PCD functions) are coded in the human genome: Responsive to Centrifugal Force and Shear Stress 1 (RECS1)/TMBIM1, Lifeguard (LFG)/TMBIM2, Glutamate Receptor Ionotropic N-

Abbreviations: ER, endoplasmic reticulum; FtsH, filamentous temperature-sensitive; GA, Golgi apparatus; GAAP, Golgi anti-apoptotic protein; GHITM, growth-hormone inducible transmembrane; GRINA, glutamate receptor ionotropic N-methyl-D-aspartate associated; HGT, horizontal gene transfer; HMM, hidden Markov model; LFG, lifeguard; Ma, million years ago; MeJA, methyl jasmonate; MSA, multiple sequence alignment; NCBI, National Center of Biotechnology Information; OG, orthologous groups; PCD, programmed cell death; RECS1, responsive to centrifugal force and shear stress 1; SMSA, structural multiple sequence alignment; TMBIM, transmembrane BAX inhibitor motif; UPR, unfolded protein response

* Corresponding author.
E-mail address: lcrz@cicy.mx (L.C. Rodriguez-Zapata).
[1] Contributed equally.

https://doi.org/10.1016/j.ympev.2018.04.032
Received 1 December 2017; Received in revised form 19 April 2018; Accepted 20 April 2018
Available online 25 April 2018

Methyl-D-Aspartate Associated (GRINA)/TMBIM3, Golgi Anti-Apoptotic Protein (GAAP)/TMBIM4, and Growth-Hormone Inducible Transmembrane (GHITM)/TMBIM5 (Rojas-Rivera and Hetz, 2014). In contrast to the BCL-2 family, the TMBIM superfamily is widespread across both evolutionarily-distant and closely-related organisms and is highly conserved in function and structure (Henke et al., 2011; Rojas-Rivera and Hetz, 2014). Similar to HsBI-1, BXI1p protein from budding yeast provides protection against the heterologous expression of BAX and other stressful stimuli such as ER-stress, heat shock, and ethanol- and glucose-induced PCD (Cebulski et al., 2011; Chae et al., 2003). In prokaryotes, BsYetJ protein from the bacteria *Bacillus subtilis* works as a pH-regulated calcium channel (Chang et al., 2014). In plants, BI-1 homologues are key regulators of PCD, cytosolic calcium concentrations (Ihara-Ohori et al., 2006), sphingolipid metabolism (Nagano et al., 2012), autophagy (Xu et al., 2017), and Methyl Jasmonate (MeJA)-induced senescence (Yue et al., 2012). Furthermore, plant BI-1 proteins determine the outcome of plant-pathogen interactions with biotrophic and necrotrophic fungi (Babaeizad et al., 2009); and provide protection against several types of abiotic stresses such as heat, drought, oxidative stress, and salt stress (Duan et al., 2010; Isbat et al., 2009; Ishikawa et al., 2010; Kawai-Yamada et al., 2004; Watanabe and Lam, 2006).

In eukaryotes, the TMBIM superfamily is further divided into the BI and Lifeguard (LFG) families (Henke et al., 2011; Hu et al., 2009). The BI family is composed of homologues of the GHITM/TMBIM5 and BI-1/TMBIM6 proteins. TMBIM5 orthologues are present in the choanoflagellate *Monosiga brevicollis*, the hemichordate *Saccoglossus kowleskii*, the nematode *Caenorhabditis elegans*, and vertebrates (Henke et al., 2011). No TMBIM5 orthologues are present in plants. TMBIM6 orthologues are present in protists, algae, animals, and plants. Moreover, the LFG family is composed of homologues of the human TMBIM1-4 proteins plus the Tmbim1b protein from cow (*Bos taurus*) (Zhou et al., 2008). These proteins have been renamed as LFG1-5 (Hu et al., 2009). The LFG family expanded from a single LFG4-like ancestor before the divergence of major eukaryotic groups 2,000 million years ago (Ma) (Hu et al., 2009). In animals, this ancestor was duplicated and gave rise to LFG4 (GAAP/TMBIM4) and the precursor of LFG1 (GRINA/TMBIM3). Subsequently, this LFG1 precursor underwent additional duplications leading to the vertebrate proteins LFG2 (LFG/TMBIM2) and LFG3 (RECS1/TMBIM1). LFG5 (Tmbim1b) protein possibly derived from an LFG2- or LFG3-like precursor (Hu et al., 2009). In plants, LFG proteins have been reported in bryophytes, gymnosperms, and a few angiosperms (Hu et al., 2009; Weis et al., 2013). Apparently, LFG proteins in plants have undergone several rounds of duplications similarly as seen in animals.

Gene and genome duplications have been major players in the acquisition of novel traits during plant evolution (Flagel and Wendel, 2009). Gene duplication occurs by means of small duplication events (tandem, segmental, transposon-mediated duplications), and whole genome duplications (WGDs) and triplications (WGTs) (Panchy et al., 2016). The latter two, also known as polyploidizations, are drastic events that lead to the abrupt increment of both genome size and gene content followed by gene loss (fractionation) (Fawcett et al., 2013), and are considered a common mode of speciation (Van de Peer et al., 2017). Several ancient polyploidization events (paleopolyploidy) occurred during plant evolution. A WGD (ζ) occurred in the common ancestor of seed plants 319 Ma, and another (ε WGD) occurred in the common ancestor of angiosperms 192 Ma (Jiao et al., 2011). A WGT (At-γ) occurred in the common ancestor of most eudicots, and two additional WGD (At-α and At-β) occurred in the common ancestor of Brassicaceae (Jaillon, 2007). In monocots, several rounds of WGDs (τ, σ, ρ) occurred in the common ancestor of grass species, such as rice (*Oryza sativa*) and wheat (*Triticum aestivum*) (Jiao et al., 2014; Tang et al., 2010). Other WGDs have been detected in other economically and ecologically important plant species. The TMBIM superfamily in plants was also expected to take place according to these different duplications events.

Despite efforts to understand the molecular and biological functions of the TMBIM proteins in plants, little attention has been paid to their evolution considering paleopolyploidy events. As more sequenced genomes become available, comparative genomics allows us to get a deeper understanding about the evolution and duplication of the TMBIM superfamily in plants and other organisms. In the present study, we provide a phylogenetic overview and general motif analysis of the TMBIM superfamily of proteins across a wide range of genomes from Archaea, Bacteria, and Eukarya. Then, to get new insights about the evolution of the TMBIM superfamily in plants, we deepened in the individual analysis of the BI and LFG families in 48 plants through the integration of the available WGD/T information, phylogenetic analysis, detection of orthologous groups, and synteny network analysis to propose distinct models of evolution. Additionally, we briefly discuss new findings regarding the classification of the TMBIM proteins in prokaryotes, fungi, and animals. Our main goal was to determine the evolutionary history of the TMBIM superfamily in plants, information needed for future functional studies.

2. Materials and methods

2.1. HMM search and retrieval of TMBIM protein sequences

We conducted BLAST and PSI-BLAST (Altschul et al., 1997, 1990) searches against the non-redundant protein database of the National Center of Biotechnological Information (NCBI) (https://www.ncbi.nlm.nih.gov/) using as query the protein sequences of TMBIM1-6 (human), AtBI-1 and AtLFG1-5 (Arabidopsis), BIX1p (budding yeast), EYccA (*Escherichia coli*), and BsYetJ (*B. subtilis*). Then, we used the BLAST-Hits results of these searches (954 sequences) to construct a custom TMBIM-HMM profile and searched on the proteomes of 256 selected species from Archaea, Bacteria, and Eukarya (Table S1). HMM profile construction (hmmbuild) and searches (hmmbuild) were done with HMMER 3.1b2 (http://hmmer.org/; (Eddy, 1998)), and identified sequences were retrieved with Seqret from EMBOSS 3.0 (Rice et al., 2000).

2.2. Phylogenetic analysis pipeline

Since no large-scale analysis on the evolution of the TMBIM superfamily had been performed, we first conducted a phylogenetic overview across Archaea, Bacteria, and Eukarya (Fig. S1a). The TMBIM sequences (see Section 2.1) were aligned in blocks using the [-profile] option of MUSCLE v3.8.31 (Edgar, 2004). The resulting Multiple Sequence Alignments (MSA) were edited in UGENE v1.25.0 (Okonechnikov et al., 2012) as follows: N- and C-terminal regions were trimmed, and positions with more than 20% of gaps were removed, leaving only the TMBIM domain. Evolutionary Model testing was performed in ProtTest v3.4 (Darriba et al., 2011). All phylogenetic trees were based on the LG substitution model (the best that fitted our data) and were inferred through the Maximum Likelihood method in RaxML 8.2.9 (Stamatakis, 2014). Whenever indicated in figure captions, the number of rapid bootstrap replicates was determined by the bootstopping criterion using the [-autoMRE] option (Pattengale et al., 2009). Since we analyzed numerous sequences from evolutionary distant organisms we had to remove several positions, which may contain evolutionary relevant information, from the MSA in order to diminish the amount of gaps. Hence, we conducted specific phylogenetic analyses of the BI and LFG families in plants (Fig. S1e). In addition, we also performed specific phylogenetic analyses of the TMBIM proteins of prokaryotes, fungi, and animals in order to discuss particular findings about their classification (Fig. S1b–d). Phylogenetic trees were visualized with FigTree 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/) and iTOL v3.4 (Letunic and Bork, 2016).

2.3. Motif discovery and HMMLogo analyses

As part of our overview of the TMBIM superfamily (Fig. S1a), the complete set of retrieved TMBIM sequences (Section 2.1) from all organisms (Prokaryotes and Eukaryotes) was subjected to a motif discovery analysis using a stand-alone version of MEME v4.10.2 (Bailey et al., 2009, 1994). Parameters were set to find 60 motifs with a minimum and maximum width of 10 to 25 amino acids, and minimum E-value threshold with no limit. Motif architecture of every sequence is depicted in the phylogenetic tree of Fig. S2. For easier interpretation, the motifs were clustered based on their frequency among the main phylogenetic groups of TMBIM sequences. To do so, we divided the TMBIM sequences in 13 groups according to our phylogenetic analyses: Archaea BI, Bacteria BI, Fungi GHITM, Metazoa GHITM, Fungi BI, Metazoa BI, Plant BI, Fungi LFG, Metazoa LFG4, Metazoa LFG1-like, Plant LFG I, Plant LFG IIA, and Plant LFG IIB. The number of sequences in each group that contained any of the 60 motifs were counted, and a percentage matrix was generated. Based on this matrix, the 60 motifs were hierarchically clustered using the euclidean distance (dist()) and complete clustering (hclust()) methods, and visualized through ComplexHeatmap v3.5 (Gu et al., 2016) in R v3.4.1 (R Development Core Team, 2008). The 60 motifs are briefly described in Text S1 and Fig. S6. HMMlogos were generated by means of Skyling using default settings (http://skylign.org/; (Wheeler et al., 2014))

2.4. Structural multiple sequence alignment

To highlight the most conserved positions, regions, and motifs of the entire TMBIM superfamily, a Structural Multiple Sequence Alignment (SMSA) of representative TMBIM sequences was performed with PROMALS3D (Pei et al., 2008) and colored with BOXSHADE 3.21 (Hofmann and Baron, 1996); protein coordinates of the 3D model of BsYetJ were included (Fig S1a).

2.5. Determination of orthologous groups (OGs) in plants

We detected TMBIM OGs in 48 plants with ProteinOrtho v5.15 (Lechner et al., 2011) using default options. ProteinOrtho implements an extended version of the reciprocal best alignment heuristic (based on BLAST searches) approach to detect co-orthologous proteins/genes among multiple species, and employs spectral partitioning to cluster them in OGs (Lechner et al., 2011). Methods based on constructing OGs from heuristic pairwise comparisons may be too inclusive and create mixed groups that not accurately represent evolutionary relationships (Kristensen et al., 2011); hence, we mapped the detected (co-)orthologous TMBIM proteins of each OG to specific phylogenetic trees of the plant BI and LFG families. We only retained those OGs whose sequences were contained within single clades (monophyletic) in the phylogenetic trees (Fig. S1f–g). On the other hand, the OGs were also mapped to the species tree in Fig. 2 to easily point shared species and infer the possible moments of origin of each OG during plant evolution.

2.6. Tandem gene analysis in plants

To detect tandem duplicates of the TMBIM sequences in the 48 analyzed plants, each genome was compared against itself in SynMap (Lyons et al., 2008) in CoGe (Lyons and Freeling, 2008). SynMap implements the algorithm LAST (Kiełbasa et al., 2011) to find homologous genes or regions between two genomes, and identify syntenic pairs by finding collinear series of putative homologous sequences with DAGChainer (Haas et al., 2004). SynMap also implements the blast2raw algorithm to detect tandem duplicates. This analysis was performed with the online tool of SynMap (https://genomevolution.org/coge/SynMap.pl) with DAGChainer option settings as follows: relative gene order, maximum distance (-D) 30, minimum number of aligned pairs (-A) 5, tandem duplication distance of 15, and C-score 0.1 (Fig. S1h).

2.7. Synteny network approach in plants

We performed a synteny analysis of the TMBIM sequences of 47 plant species; tobacco (*Nicotiana tabacum*) was not included because of its fragmented genome assembly. Parallel coordinated plots and pairwise dot plots are commonly used to analyze synteny between two or among a few genomes; however, as the number of genes and genomes to be analyzed increases, such approaches become impractical. Hence, we opted to follow the recently developed Synteny Network approach described by Zhao and Schranz (2017) and Zhao et al. (2017) that allows the simultaneous analysis and easy visualization of syntenic blocks among a greater number of genes and genomes. In this approach all synteny relationships are depicted as a network, where “nodes” represent sequences and “edges” or “connections” represent pairwise synteny relationships. This syntenic network approach can lead to the detection of transposed genes to another genomic region (separate subnetworks) to discover specific evolutionary patterns on gene families. Briefly, we performed a reciprocal all-against-all pairwise comparison between the entire proteomes of 47 plants by means of the protein similarity search tool RAPSearch2 (Zhao et al., 2012). Similarity results and gene position information (GFF files) were used to calculate synteny blocks between all pairwise compared genomes with MCScanX (Wang et al., 2012) using default parameters, creating a score matrix of all synteny relationships among all gene families (i.e. a network). Synteny information of the TMBIM superfamily were retrieved from the matrix, and used to construct dense connected network clusters or communities (k-clique = 4) by using the Clique percolation method implemented in CFinder (Derényi et al., 2005; Palla et al., 2005). The resulting network communities were visualized with Cytoscapev3.5.1 (Shannon et al., 2003) and Gephi v0.9.1 (Bastian et al., 2009). In addition, to know the relationship between the detected synteny networks and the phylogeny of the TMBIM proteins of plants, we performed a phylogenetic profiling of the network communities by mapping them to a phylogenetic tree with all of the TMBIM sequences detected in the 48 plant species (including those missing from the network). For this specific analysis, a phylogenetic tree was built from the alignment of all of the TMBIM sequences from plants made with hmmalign [-trim] in HMMER 3.1b2 (Eddy, 1998) and computed in RaxML8.2.9 (Stamatakis, 2014) by using the LG substitution model. For detailed information about the synteny network approach please see Zhao and Schranz (2017) and Zhao et al. (2017) (Fig. S1h).

2.8. Generation of species trees

Species trees were generated with PhyloT (Letunic, 2015) and visualized with iTOL v3.4 (Letunic and Bork, 2016). Information about WGD events was taken from CoGepedia (https://genomevolution.org/wiki/index.php/Plant_paleopolyploidy, last accessed on 1/12/2017) and Renny-Byfield and Wendel (2014). Information about ploidy level, genome size, and chromosome number was obtained from the Plant DNA C-values database (http://data.kew.org/cvalues/CvalServlet?querytype=1) and available published genomes (Table S1).

3. Results

3.1. Phylogeny overview of the TMBIM superfamily across the three domains of life

We identified a total of 685 TMBIM proteins in 171 species distributed among archaea, bacteria, protists, fungi, animals, algae, and plants (Tables S1 and S2). Prokaryotic TMBIM proteins were divided into the Archaea BI and Bacteria BI clades (Fig. 1a and b). Archaea BI clade contained proteins from six methanogenic archaeal species within Euryarchaeota, and seven bacterial species distributed among Acidobacteria, Spirochaetes, Chlamydie, Planctomycetes, γ proteobacteria, and Formicutes (Clostridia) (Fig. S3a and b). Bacteria BI clade

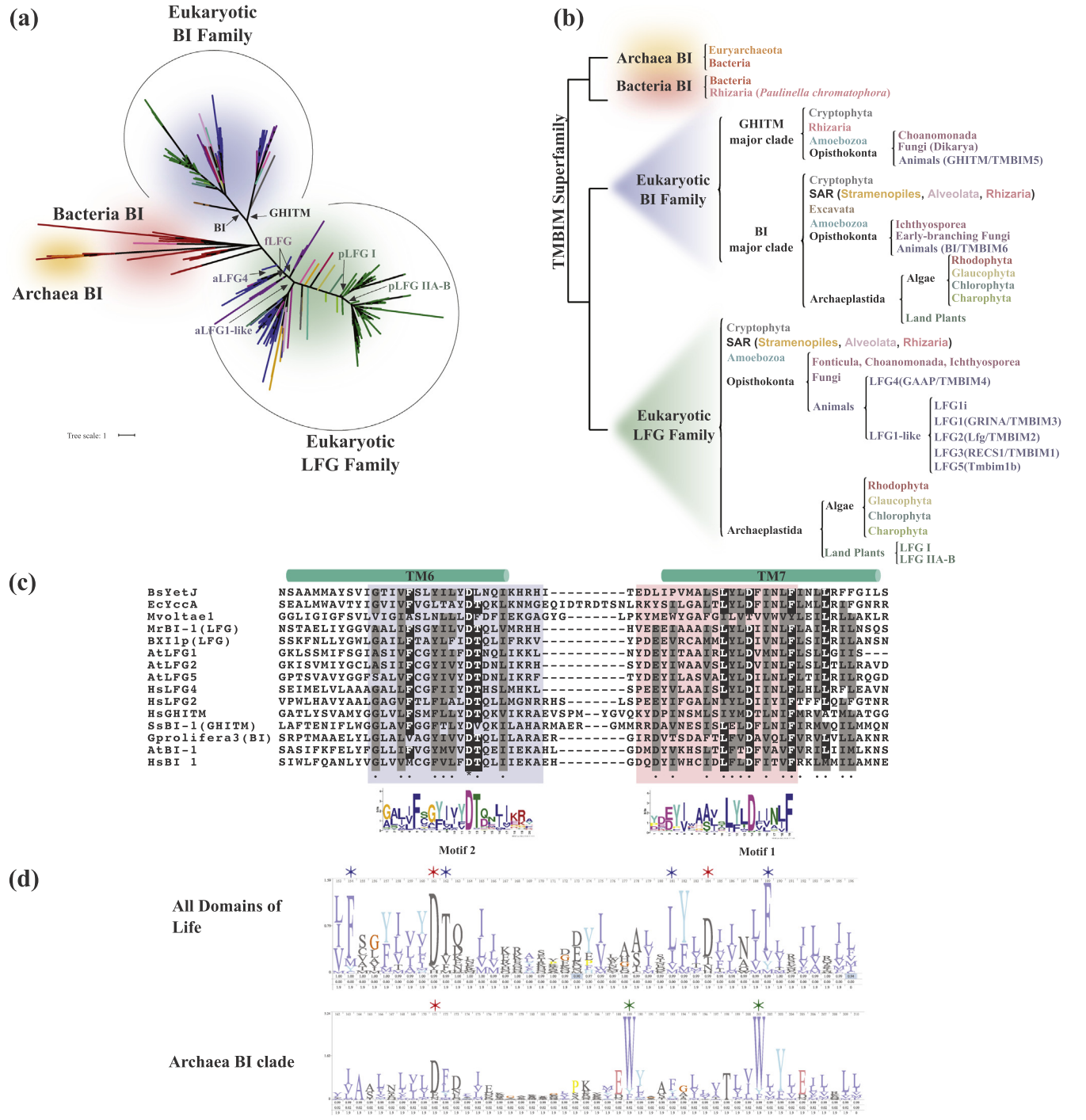


Fig. 1. Classification and conserved sequence regions of the TMBIM superfamily of proteins. (a) Maximum likelihood unrooted tree inferred from the aligned domains of 685 TMBIM proteins from 171 species. Branch colors represent different groups of organisms as indicated in (b). Branches leading to BI and GHITM major clades, and branches leading to animal (aLFG4 and aLFG1-like), fungi (fLFG) and plant (pLFG I and Plant IIA-B) LFG groups are indicated with arrows. The complete tree is shown in Fig. S2. (b) Diagram depicting the proposed general classification of the TMBIM proteins on the studied groups of organisms. (c) Multiple sequence alignment of the TM6 and TM7 transmembrane regions of representative TMBIM proteins. TM6 and TM7 of the BsYetJ protein from *Bacillus subtilis* are indicated with bars. Motifs M1 and M2 are blue and red shaded, respectively, and their sequence LOGOs are shown below each one. Full alignment is shown in Fig. S4. (d) HMMLogo of the C terminal region of the aligned domains of the total 685 identified TMBIM proteins, and the Archaea BI group. Full HMMLogos are depicted in Fig. S7. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

contained proteins from 33 bacterial species distributed in Fusobacteria, Spirochaetes, Bacteroidetes, Chlamydie, Planctomycetes, Proteobacteria (α, β, γ, δ, and ε), and Formicutes (Bacilli); and the sequence from the protist *Paulinella chromatophora* (Rhizaria) (FIg. S3a and b). This *P. chromatophora* sequence (Pchromatophora1) clustered together with the TMBIM proteins from the two cyanobacteria *Prochlorococcus marinus* (Promarinus1) and *Trichodesmium erytrhaeum* (Terytrhaeum1) (Fig. S3b). *P. chromatophora* contains chromatophores, photosynthetic entities of α- cyanobacterial origin (Marin et al., 2007). Hence, it is possible that the TMBIM sequence from *P. chromatophora* was acquired by Horizontal Gene Transfer (HGT) from a cyanobacterial symbiont. Furthermore, the phylogenetic placement of prokaryotic

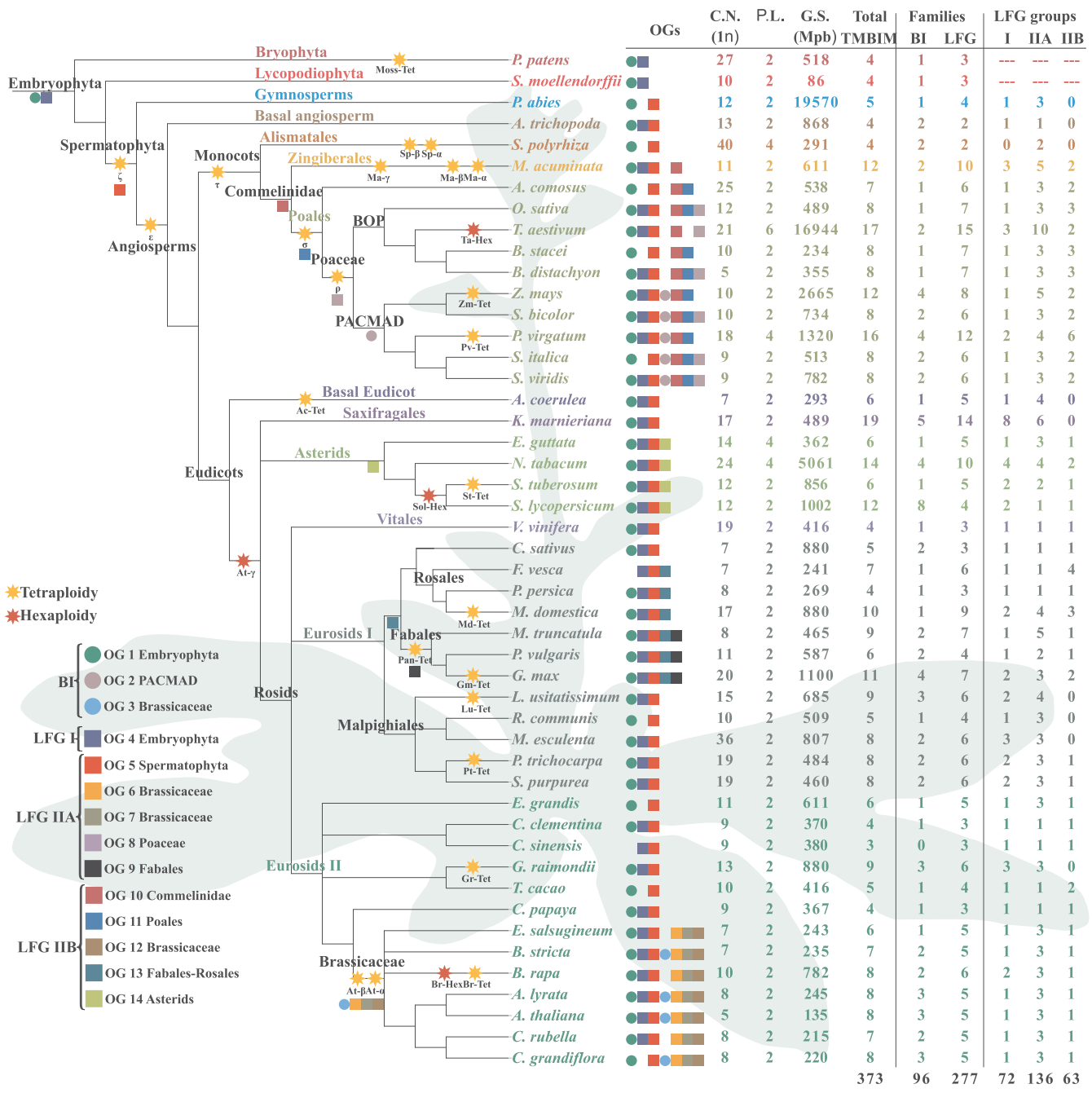


Fig. 2. Phylogenetic relationships of the 48 analyzed plant species. Phylogenetic relationships among plant species were based on the NCBI taxonomy by using PhyloT program (Letunic, 2015) and modified according to the APG taxonomy (http://www.mobot.org/MOBOT/research/APweb/). Suggested positions of orthologous groups (OGs; colored circles and squares) and occurrence of ancient whole genome duplications (WGDs; yellow stars) and triplications (WGT; red stars) are indicted in the branches of the tree. Names of WGDs and WGTs are indicated. On the table: OGs = Presence/absence of OGs in each species, C.N. = 1n chromosome number, P.L. = ploidy level, G.S. = 1C genome size in megabase pairs (Mbp). Total number of identified TMBIM proteins per species, family, and groups are also indicated in the table. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

TMBIM proteins was rather scattered among different phyla also suggesting HGT, a widely recognized mechanism in prokaryotes (Soucy et al., 2015). YccA (Ecoli1) and BsYetJ (Bsubtilis1) proteins from *E. coli* and *B. subtilis*, respectively, were also included in the Bacteria BI clade (Fig. S3b). YccA inhibits the activity of the Filamentous Temperature-Sensitive (FtsH) protease in *E. coli* and acts as an inhibitor of PCD in budding yeast (van Stelten et al., 2009); BsYetJ is a pH-regulated calcium channel in *B. subtilis* (Chang et al., 2014). On the other hand, Eukaryotic TMBIM proteins were unmistakably divided into the BI and LFG families, which is in agreement with literature (Hu et al., 2009) (Fig. 1a and b). The eukaryotic BI family was further divided into the GHITM and BI major clades (Fig. 1a and b). The GHITM major clade included proteins from animals (Metazoa GHITM), fungal species from Dykaria (Fungi GHITM) (Figs. 1b and S2), and species from the protist lineages Cryptophyta, Rhizaria, Amoebozoa, and Choanomonada (Ophistokonta) (Table S3). The BI major clade included proteins from all analyzed protist lineages (Table S3), animals (Metazoa BI), early-branching fungi (Fungi BI), red and green algae, and land plants (Plant BI) (Figs. 1b and S2). The eukaryotic LFG family was mainly divided in a lineage specific manner (Fig. 1b). This family contained proteins from all the analyzed protist lineages (except for Excavata) (Table S3), as well as fungi (Fungi LFG), animals (Metazoa LFG4 and Metazoa LFG1-

like), algae, and plants (Plant LFGs) (Figs. 1b and S2).

3.2. Sequence and motif conservation indicate a pH-regulated calcium channel activity in TMBIM proteins in eukaryotes and bacteria, but not in Archaea

Transmembrane proteins evolve to a faster rate than water-soluble proteins due to their exposure to stronger adaptive forces in the boundaries of organelles and cells, but their inner transmembrane regions evolve at a slower rate than their aqueous portions (Sojo et al., 2016). Accordingly, our SMSA indicates little overall sequence conservation of the TMBIM superfamily of proteins (as a whole) except for some highly conserved residues within the transmembrane regions, particularly TM4, TM6, and TM7, which contained M6, M2 and M1 motifs, respectively (Fig. S4). Motifs M1 and M2 were the most abundant in the whole family being present in both the BI and LFG families, and in the Bacteria BI clade, but not in Archaea BI clade (Figs. S5, S6, Table S4, and Text S1). TM6 (M2) contains a Phenylalanine (F), an Aspartate (D), and a Threonine (T) highly conserved residues; and TM7 (M1) contains a Leucine (L), an D, and a F highly conserved residues (Fig. 1c and d). These M1 and M2 motifs together are similar to the motifs 7, 8, and 10 reported by (Hu et al., 2009) in LFG proteins of animals and plants. However, we found that M1 and M2 motifs were also present in LFG proteins from yeast, several protists, and bacteria (Figs. S2 and S6). Moreover, the conserved D residues in Motifs M1 and M2 (together) conform a Di-aspartyl pH sensor that regulates the calcium channel properties of the BsYetJ protein in *B. subtilis* (Chang et al., 2014). Our data highlights the evolutionary conservation of this region in most proteins of the TMBIM superfamily suggesting a possible conserved function of the pH-regulated calcium channel in Bacteria and Eukarya. On the other hand, Archaea BI clade presents a distinct sequence composition in TM6 and TM7 containing motifs M51 and M37 instead of M2 and M1, respectively (Fig. S2). Motif 37 (in TM7) in the Archaea BI clade contains two well conserved Tryptophan (W) residues (instead of an D), suggesting a distinct function or mechanism of the TMBIM proteins in this Archaea BI clade (Figs. 1d and S7); further sampling and functional analyses of these archaeal proteins are necessary.

3.3. Two major phylogenetic groups are found in the LFG family in plants

We analyzed the proteomes of 57 organisms from Archaeplastida covering 9 algae and 48 plants. We found a total of 19 TMBIM proteins in seven algal species (Table S3). No TMBIM proteins were found in the Chlorophytes *Micromonas pusilla* and *Ostreococcus lucimarinus* possibly due to independent gene losses. Algal TMBIM proteins were present in both the BI (only within the BI major clade) and LFG families (Fig. 1b and Table S3). Interestingly, and like the results of (Hu et al., 2009), LFG proteins from Chlorophytes were more closely related to the Metazoa LFG1-like clade. The exception was *C. subellipsoidea* that possessed LFG proteins close to the Metazoa LFG1-like clade and LFG clade from plants (Fig. S2 and Table S3). This is not surprising since many homologous sequences of Chlamydomonas were present in the common ancestor of both animals and plants, and some genes shared by Chlamydomonas and animals have been lost in plants (Merchant et al., 2007).

In plants we found a total of 373 TMBIM proteins in 48 species; 96 and 277 sequences belonged to the BI and LFG families, respectively (Fig. 2). Regarding the BI family, plants possess an average of two sequences per species, with a minimum of one in several species, and a maximum of eight in tomato (*Solanum lycopersicum*) (Fig. 2). This plant BI family formed a single monophyletic group that was divided in a lineage-specific manner, mainly between monocots and eudicots (Figs. 3a and S8). However, seven sequences from tomato (*Solanum lycopersicum*) belonged to a specific long-branching clade (Slyopersicum-only clade) formed out of the Slycopersucum7 sequence and the tandem duplicated array Slycopersicum6, 8–12 (Figs. 3a and S8). The single remaining BI sequence from tomato (Slycopersicum4) was clustered inside an asterid specific clade next to tobacco, potato (*Solanum tuberosum*), and seep monkeyflower (*Erythranthe* guttata; synonym: *Mimulus guttatus*) (Fig. S8).

Regarding the LFG family, plants possess an average of 5.7 sequences per species (about 2.8-fold more than in plant BI family), with a minimum of two in Amborella (*Amborella trichopoda*) and common duckweed (*Spirodela polyrhiza*), and a maximum of 19 sequences in *Kalanchoe marnieriana* (Fig. 2). In monocots (Commelinidae), the number of LFG proteins seems to be related to particular paleopolyploidy events: banana (*Musa acuminata*), wheat, maize (*Zea mays*) and switchgrass (*P. virgatum*) contain more LFG proteins in their genomes, and they have also experienced particular polyploidy events apart of the τ, σ, and ρ WGDs shared by most monocots (Fig. 2). Our phylogenetic analysis revealed that the LFG family in plants was firstly separated into Bryophytes-Lycophytes (the moss *Physcomitrella patens* and the clubmoss *Selaginella moellendorffii*) and seed plants (gymnosperms and angiosperms) (Fig. 3b). LFG proteins in seed plants were divided in two major groups supported by high bootstrap values: > 60 for LFG I group, and > 80 for LFG II group (Fig. S9). This result is similar to that found by (Weis et al., 2013) when they examined the LFG proteins from Arabidopsis and barley (*Hordeum vulgare*). LFG I group contained 72 sequences from 45 species (average of 1.6 sequences per species) (Fig. 2); common duckweed was the only absent. LFG II group was subsequently divided in two subgroups: IIA and IIB (Fig. 3b). LFG IIA subgroup contained 136 sequences from 46 species (average of 2.9 sequences per species) (Fig. 2); and LFG IIB subgroup contained 63 sequences from 37 species (average number of 1.7 sequences per species) (Fig. 2).

3.4. Two OGs of the TMBIM superfamily are shared by most plants

Detection of OGs is interesting to comparative genomics and functional analysis, since orthologues commonly perform equivalent biological functions in different species. We found 14 OGs of TMBIM proteins in plants (Table S5) and these OGs were mapped to the plant species tree (Fig. 2) and specific phylogenetic trees of the BI and LFG families in plants (Figs. 3, S8 and S9). Two of them (OGs 1 and 4) are shared by mosses, lycophytes, gymnosperms (OG1 only) and angiosperms; so, these OGs are probable shared by most land plants taxa (Embryophyte) (Fig. 2). OG1 corresponded to the BI family (Fig. 3a) and was shared by 46 species; only missing strawberry (*Fragaria vesca*) and sweet orange (*Citrus sinensis*) (Fig. 2). OG4 corresponded to the LFG I group (Fig. 3b) and was shared by 39 species; missing Norway spruce, common duckweed, pineapple, *Brachypodium stacei*, *Setaria italica*, castor bean (*Ricinus communis*), rose gum (*Eucalyptus grandis*), cacao (*Threoboma cacao*), and *Capsella grandiflora* (Fig. 2). Interestingly, Arabidopsis AtB-1 and AtLFG5 from OGs 1 and 4, respectively, are highly expressed in most tissues (Fig. S10). OG5, which corresponded to the LFG IIA subgroup (Fig. 3b), was shared between the gymnosperm Norway spruce and all sampled angiosperms; thus this OG5 might have appeared due to the duplication of an ancestral protein of OG4 in the common ancestor of seed plants prior their diversification after ζ-WGD (Fig. 2). Arabidopsis representative of OG5, AtLFG4, is highly expressed in the flower stamen and pollen (Fig. S10) probably having some particular function in reproduction. Other two (2–3), four (6–9), and five (10–14) lineage-specific OGs are present in the BI family, the LFG IIA subgroup, and the LFG IIB subgroup, respectively (Fig. 3a and b). OGs 3, 6, 7, and 12 were shared by Brassicaceae, each OG containing one sequence from Arabidopsis: AtBI-2, AtLFG3, AtLFG2, and AtLFG1, respectively (Figs. S8 and S9). Brassicaceae has experimented two additional WGDs (α and β) in addition to the WGT shared by all eudicots (γ) (Fig. 2); so, one or both of these WGDs might have contributed to the appearance of the ancestor sequences of these lineage-specific OGs during Brassicaceae speciation. AtLFG2 (OG7, LFG IIA) presents higher

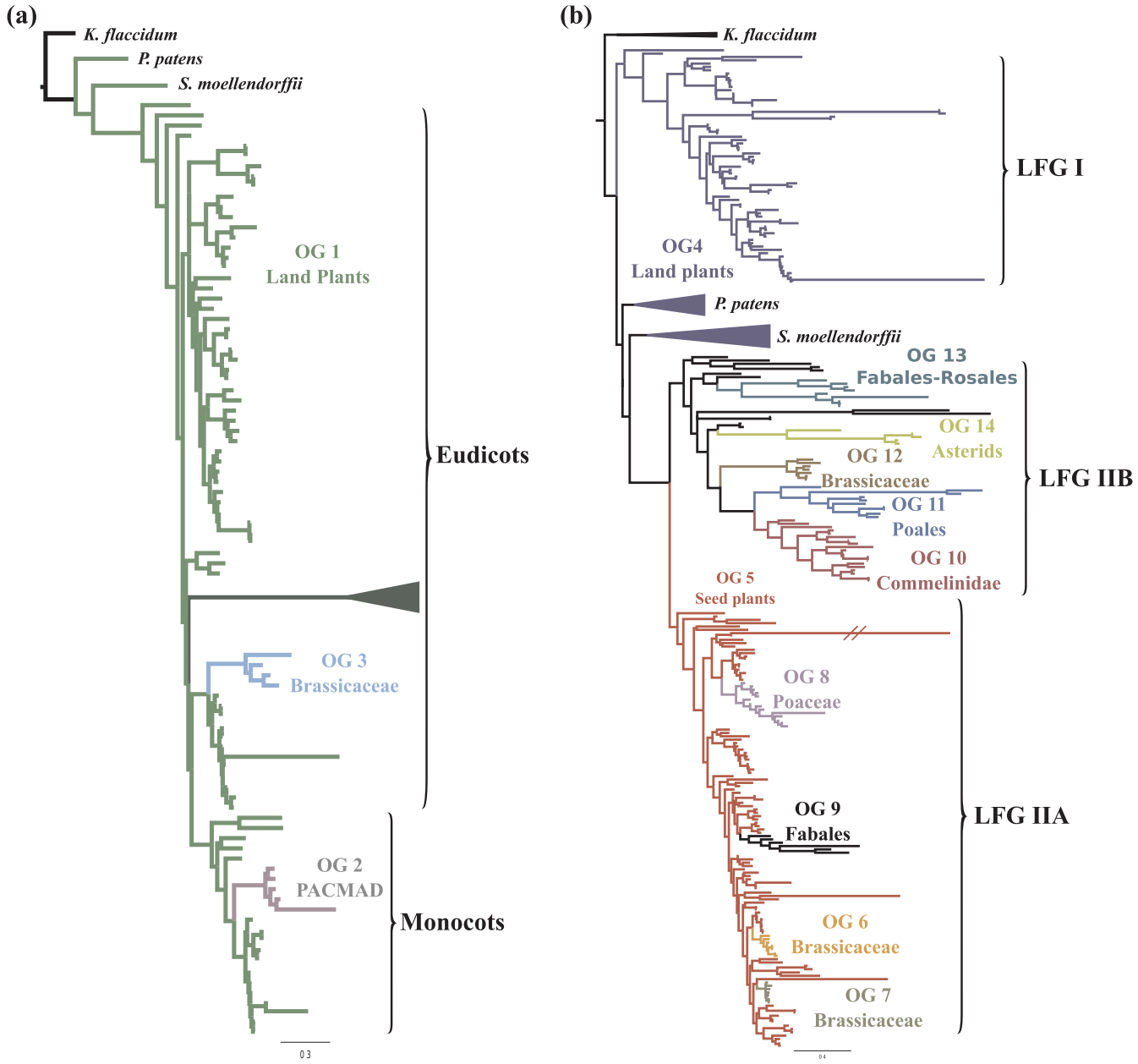
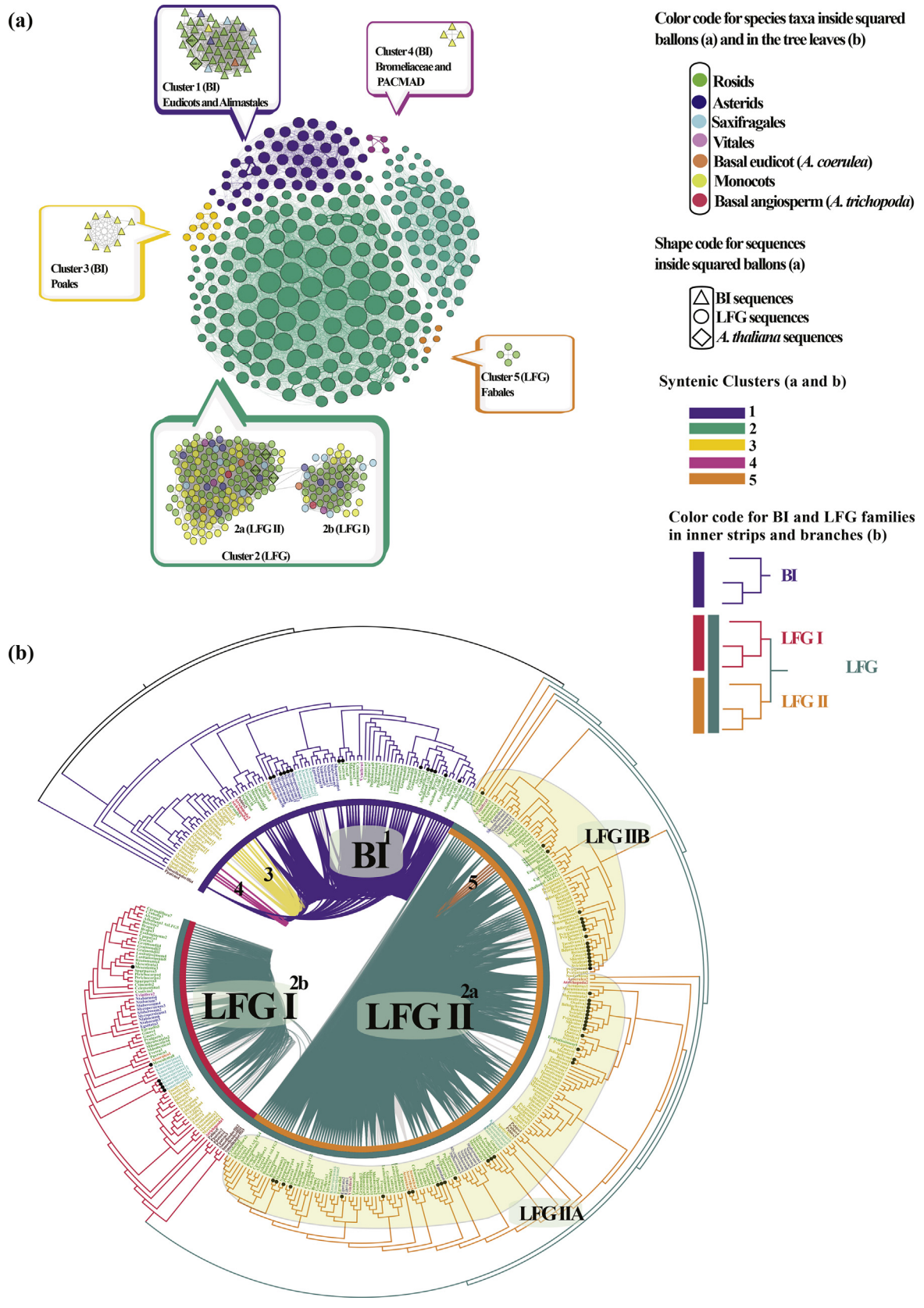


Fig. 3. Phylogenetic relationships of BI and LFG families in plants. Maximum likelihood trees of 96 and 373 proteins of BI (a) and LFG (b) families, respectively, from 48 land plants. Sequences of the filamentous terrestrial alga *Klebsormidium flaccidum* were included as outgroups. In (a) the "Slycopersicum-only" branch of the BI family is shown in dark-green color. Name and number of OGs are indicated in the trees branches with different colors. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

expression in roots and seeds, while AtLFG3 (OG6, LFG IIA) is highly expressed in seeds only (Fig. S10). On the contrary, AtBI-2 (OG3, BI) is expressed at low levels in all tissues (Fig. S10). In monocots, OGs 10, 11, 8, and 2 are shared by Commelinidae, Poales, Poaceae, and PACMAD grasses, respectively (Fig. 2); σ and ρ WGDs might have contributed to the appearance of OGs 11 and 8 during Poales and Poaceae speciation, respectively (Fig. 2). OG13 (LFG IIB; Fig. 3a) might have appeared during the speciation of nitrogen-fixing fabids and was lost in Cucurbitales, but retained in Fabales and Rosales (Fig. 2). Lastly, OG 14 was shared by Asterids (Fig. 2). OGs are convenient when describing evolutionary relationships across species (Gabaldón and Koonin, 2013) since orthologous genes typically conserve similar biological functions in different organisms (Koonin, 2005). Ancestor genes of OG1 (BI) and OG4 (LFG I) were probably present since the last common ancestor of all plants, probably playing a conserved function in cell maintenance and survival, which might be the reason why they are highly expressed in all tissues of Arabidopsis (Fig. S10). Duplication events gave rise to linage-specific OGs, which may represent specific functions and expression patterns in Brassicaceae, legumes, asterids, and monocots. Further functional and expression analyses are required to confirm this.

3.5. Syntenic network and tandem analyses of the TMBIM superfamily in plants

The plant TMBIM synteny network contained 270 nodes (sequences) linked by 5,029 edges (pairwise synteny relationship) (Table S6). No sequences from the moss *P. patens*, the clubmoss *S. moellendorffii*, and the gymnosperm Norway spruce (*Picea abies*) were found to be connected to the network due to their phylogenetic distance from



(caption on next page)

Fig. 4. Synteny Network Analysis of the TMBIM superfamily in plants. (a) Complete synteny network of the TMBIM gene superfamily based on clique percolation method at k = 4. Node size indicates the number of connections with other nodes. The complete cluster network is shown next to each community inside squared balloons. (b) Phylogenetic reconstruction of the TMBIM gene superfamily of land plants and syntenic relationship of the five clusters found in the network. The BI family is marked with purple inner stripe and also purple branches. The LFG family is marked with dark turquoise inner stripe and then split into LFG I group (red inner stripe and branches) and the LFG II group (orange inner stripe and branches). Branches for subgroups in LFG II (LFG IIA and LFG IIB) are enclosed on the tree with light-orange background color. Tandem duplicated genes are marked with black dots at the end of the branches. Clique size k = 4 was used to depict the five communities. Clique connections below 4 (k < 4) are show in gray colored lines. Communities are coded from 1 to 5 for easy tracking (Table S7). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

angiosperms and the large genome size of Norway spruce. We used clique percolation clustering at size k = 4 (implemented in CFinder) to define more dense and connected clusters or communities numbered from 1 to 5 (Figs. 4a, S11a, and Table S7). These five communities were used for phylogenetic profiling (Fig. 4b). Syntenic clusters 1, 3, and 4 were specific for the BI family, while clusters 2 and 5 corresponded to the LFG family (Fig. 4b). No syntenic connections were found between BI and LFG families (Fig. 4b), confirming that these families splitted apart before the evolutionary emergence of plants from a more ancestral organism, probably a prokaryote.

Cluster 1 comprises 48 nodes connected by 634 edges from 29 species (Tables S7 and S8), mostly eudicots; 22.5% of these sequences (only eudicots) came from tandem duplication events (Fig. 4b). Only one monocot sequence from common duckweed (Spolyrhiza4), which is one of the most ancient monocotyledonous clade (Alimastales), was part of this syntenic cluster and it was well connected with eudicots (36 connections; Fig. S11a). Nevertheless, Cluster 1 lacked syntenic sequences from grapevine (*Vitis vinifera*) and Amborella, but contained a sequence (Acoerulea6) from the basal eudicot columbine (*Aquilegia caerulea)*. Regarding asterids, only one representative from seep monkeyflower, tomato, and potato were found in this BI syntenic cluster (Table S8). AtBI-1 and AtBI-2 sequences from Arabidopsis were clustered inside this group and were syntenic to each other (but not AtBI-3, a tandem of AtBI-1) (diamonds in Fig. S11); both sequences belong to a syntenic block emerged from the At-α WGD (Wang et al. 2016). BI sequences from monocots were clustered apart in the linage-specific syntenic clusters 3 and 4, showing evidences of ancient transposition events (Figs. 4a, b and S11). Cluster 3 was formed by 10 nodes connected by 38 edges from eight species within Poales (Fig. S11a), and Cluster 4 was formed by 4 nodes from three species within PACMAD group and pineapple (*Ananas comosus*), a Bromeliaceae (Fig. S11a). No tandem duplicated sequences were found in clusters 3 and 4 (Fig. 4b). The early-diverging monocot common duckweed and banana were absent in these clusters (Fig. S11a and Table S8). Nevertheless, as mentioned above, a sequence from common duckweed (Alimastales) was syntenic to the rest of eudicot plants and grouped in BI Cluster 1 (Fig. S11a). Since cluster 3 contained sequences from both PACMAD and BOP (OG1) but not the Bromeliaceae pineapple (Fig. S11a and b), and cluster 4 contained sequences from Bromelieaceae (OG1) and the PACMAD (OG2) but not the BOP group of grasses (Fig. S11a and b); we infer that the common ancestor of Poales (most recent common ancestor of Bromeliaceae and Poaceae) experienced a replicative transposition (Monocot transposition 2 in Fig. S11a and b) and cluster 3 was only retained by PACMAD and BOP, and cluster 4 was only retained in Bromeliaceae and PACMAD. These ancient transpositions could be the reason why we found that BI sequences from monocots and eudicots were clustered apart in our phylogenetic analysis (Fig. 3a).

The LFG family was formed by two syntenic clusters: Cluster 2 contained all considered angiosperm species (44), and Cluster 5 contained the three sampled Fabales species (Tables S7 and S8). Cluster 2 was formed of 198 nodes connected by 4,342 edges (Table S7). This cluster was divided in two communities corresponding to the LFG I (cluster 2b) and LFG II (cluster 2a) gene groups (Fig. 4a and b). Even when this two communities are poorly connected to each other, they shared syntenic relationship since both contain sequences from the pivotal species grapevine, columbine, and Amborella (Fig. S11 and Table S8). LFG I community (cluster 2b; red inner stripe in the tree in Fig. 4b) was formed by 57 nodes connected by 771 edges from 39 species (Table S8); only 10.5% of its sequences came from tandem duplication events. This LFG I (cluster 2b) community contained sequences from both monocots and eudicots, including AtLFG5 from Arabidopsis (diamond in Fig. S11), a sequence from Amborella (Atrichopoda1) and one from banana (clique k = 4). The LFG II community (cluster 2b; orange inner stripe in the tree of Fig. 4b) was formed by 141 nodes connected by 3,567 edges from all of the 44 analyzed angiosperm species (Table S8) suggesting that this community has retained its synteny since the last common ancestor of angiosperms. LFG II community contains up to 32.6% of tandem duplicate genes, most of them in monocots, suggesting that it has also expanded through small-scale duplication events (black dots at the end of branches in Fig. 4b and see (Table S2). Only one sequence from Amborella was found inside this community (Atrichopoda2). AtLFG1-4 from Arabidopsis were grouped in this LFG II community, but only AtLFG3 and AtLFG4 were directly syntenic to each other (diamonds if Fig. S11). These sequences came from a syntenic block that emerged from At-α WGD (Wang et al., 2016). AtLFG1 and AtLFG2 (but not AtLFG3 and AtLFG4) were syntenic to each of the two sequences from columbine and grapevine, suggesting that these genes derived from an ancient WGD event (At-γ). AtLFG2 is one of the few sequences that links the LFG II community to the LFG I community, and according to (Wang et al., 2016) this sequence is a relocated γ gene, i.e. a gene that relocated or transposed to another genomic context after the At-γ WGT. The lineage-specific syntenic cluster 5 was formed by 4 nodes connected by 6 edges from three Fabales species belonging to OG13 (Fig. S11b). These data indicate a probable transposition of some OG13 members of the Fabales ancestor to another genomic context after the legume-specific WGD, the Panilionoid-Tetraploidy (Pan-Tet).

4. Discussion

4.1. Eukaryotic BI and LFG families diverged early in prokaryotes

Previous studies on the evolution of the TMBIM superfamily covered a limited number of sequences and/or organisms, mainly animals (Henke et al., 2011; Hu et al., 2009). Hence, detailed knowledge about the evolution of this superfamily in other lineages like prokaryotes, protists, fungi, algae, and plants is still lacking. Here we examined the evolution of the TMBIM superfamily across a wide range of species from the three domains of life, with greater detail in plants. We report for the first time the presence of TMBIM proteins in Archaea and several bacterial species (Fig. S3). Previous work suggested that the BI and LFG proteins make two different families in eukaryotes with a probable very distant ancestor (Hu et al., 2009). Our results support this hypothesis since the BI and LFG families were splitted apart from each other and from the prokaryotic TMBIM proteins (Fig. 1a). Furthermore, within the Bacteria BI clade we found a smaller cluster of four bacterial sequences comprising the proteins from *Gimesia maris, Lentisphaera araneosa, Capnocytophaga gingivalis* and *Riemerella anatipestifer* that seemed to be closer to the Eukaryotic LFG family (green square in Fig. S2); these proteins might be closer representatives of the common ancestor of Eukaryotic LFG family and Prokaryotic TMBIM proteins. Sampling of more bacterial genomes and metagenomes are still necessary to clarify

this. On the other hand, similarly to our results (Fig. S3), a scarce and scattered occurrence pattern have also been reported for other PCD-related proteins in bacteria and archaea. For instance, (Asplund-Samuelsson et al., 2012) found that only 262 of 1463 prokaryotic genomes contained metacaspases (caspase homologs); and only five of those species belonged to Euryarchaeota. Some hypotheses have suggested that eukaryotic PCD molecules and pathways had their origins in prokaryotes as several homologs of eukaryotic PCD-related genes, such as Caspases, Apoptosis-Inducing Factor, and Cytochrome *C* have been found in archaeal and bacterial species (Koonin and Aravind, 2002; Taylor-Brown and Hurd, 2013). Our findings also add up to the hypothesis that the eukaryotic PCD core machinery comes from a prokaryotic origin.

4.2. New insights into the TMBIM superfamily in fungi, and a possible archetype BI sequence

We found no TMBIM proteins in Microsporidia and Neocallimastigomycota, similarly to the results of Chen et al. (2015). Microsporidia are obligated intracellular parasites that have experienced many gene losses and genome reductions during evolution (Corradi, 2015), possibly explaining the lack of TMBIM proteins in these fungi. Previous studies have identified a single BI-l like protein in fungi: BXIp in budding yeast (Cebulski et al., 2011), MrBI-1 in *Metarhizium robertsii* (Chen et al., 2015), and SsBI-1 in *Sclerotinia sclerotiorum* (Yu et al., 2015). Our phylogenetic analysis indicates that both BI and LFG families are present in most fungi (Fig. S12). Moreover in the BI family, we found that early-diverging fungal species and Dykaria only retained homologues of the BI-1/TMBIM6 and GHITM/TMBIM5 proteins, respectively (FIg. S12). Considering that GHITM/TMBIM5 homologues are absent in algae, plants, and many protists but present in animals (Fig. 1b), and that BI-1/TMBIM6 was present in almost all eukaryotes, we infer that an archetype BI-1/TMBIM6 protein ancestor was present in the common ancestor of Eukaryotes. Then, a GHITM/TMBIM5 homologue appeared as a duplicate of BI-1/TMBIM6 during Eukaryote evolution, which was retained in Dykaria and Animals but lost in early-diverging fungi, algae, plants, and many protists; early-diverging fungi retained this archetype BI-1/TMBIM6 protein. Therefore, MrBI-1 and BXI1p are LFG homologues, and SsBI-1 is a GHITM/TMBIM5 homolog. These insights shed new light on the classification of the TMBIM proteins in Fungi, and will help guiding and interpreting future functional analyses.

4.3. Two ancient transposition events of the BI family occurred in monocots

The Plant BI family experienced few duplication events and remained as a single monophyletic group containing a single major OG(1) (Figs. 3a and S11b). In monocots, after the split of Alimastales from the rest of monocot lineages, the BI genes from OG1 transposed to another genomic context in the common ancestor of Commelinidae (Fig. 5a). In duckweed this BI protein from OG1 seems to have been duplicated, and transposed to another genomic context (Fig. 5a). This explains why a BI-1 sequence (not from OG1) from common duckweed still conserved the same synteny as eudicot plants (Fig. S11a and b). A subsequent replicative transposition occurred in the common ancestor of Poales during σ WGD (Figs. 5a and S11b). This transposition was lost in the common ancestor of the BOP clade but retained in the common ancestors of Bromeliaceae and PACMAD grasses, ultimately giving rise to the appearance of OG2 in the latter. This second transposition could be the reason why a sequence from pineapple remains syntenic to the sequences of the OG2 in the PACMAD group (cluster 4 in Fig. 11b). Further sampling of more monocot genomes is necessary to clarify this. A duplication during the At-α/β WGD in eudicots (Fig. 5a) possibly gave rise to OG 3 in Brassicaceae.

4.4. Expansion of the LFG family in plants and vertebrates

An LFG I protein, ancestor of OG4, was present in the common ancestor of plants. During ζ WGD this protein was duplicated in the common ancestor of seed plants (but not mosses and clubmosses) giving rise to the ancestor of OG5 and the LFG IIA group (Fig. 5b). LFG I and LFG II groups belong to two separated syntenic communities, therefore evolved under different genomic contexts; however, evidence sustains that both groups were syntenic in the common ancestor of monocots and dicots (Fig. 4a). LFG I (OG4) group remained mostly unexpanded during angiosperm diversification, similar to the case of Metazoa LFG4 clade in animals (Fig. S13). On the contrary, the LFG IIA group (OG5) expanded during angiosperm evolution through WGDs and tandem duplications giving rise to LFG IIB subgroup and several lineage-specific OGs (Fig. 5b), similar to the expansion of Metazoa LFG1-like clade in animals. Metazoa LFG1-like clade experienced several gene duplications (Hu et al., 2009), giving place to 4 small subclades corresponding to LFG1, LFG2, LFG3, and LFG5 proteins in vertebrates (Fig. S13). Even when WGDs are rare in animals (in contrast to plants) is worth noting their important influence on the expansion of gene families. Vertebrate genomes have experienced two rounds of ancient WGD (1R-WGD and 2R-WGD) and a linage-specific duplication event in teleost fishes (3R-WGD) (Fig. S13). These WGDs have shaped the diversity of organisms through preferential retention of duplicated genes (Nakatani et al., 2007). For example, in the teleost zebrafish (*Danio rerio*), which experienced the 3R-WGD (Glasauer and Neuhauss, 2014; Inoue et al., 2015), TMBIM genes appear to be preferentially expanded only in the LFG1-like clade. On the other hand, in plants, OGs 10–13 (within LFG IIB) probably originated from the duplication of a putative LFG IIB ancestor (which originated from OG5). The ancestor of OG13 in Fabales species (but not Rosales) was transposed to another genomic context, possibly during the Pan-Tet (Fig. 5b). The possible functional consequences of this transposition needs to be addressed through experimental studies. OGs 6–9 (within LFG IIA) originated from duplications of OG5 proteins.

In humans, the duplicate members of Metazoa LFG1-like clade diverged to perform both overlapping and specific functions (Carrara et al., 2017; Hu et al., 2009; Rojas-Rivera and Hetz, 2014); a similar scenario could take place in plants. Overexpression of AtLFG1-2 genes in Arabidopsis supports the infection by the biotrophic fungus *Erysiphe cruciferarum*; conversely, mutant plants of these genes are less susceptible. Similar phenotypes were found in epidermal cells of barley overexpressing or silencing the HvLFGa gene, respectively (Weis et al., 2013). AtLFG2 (BIL4) mediates cell elongation under the Brassinosteroid signaling pathway (Yamagami et al. 2009). Similar to our results, (Weis et al., 2013) found that LFG proteins from Arabidopsis and Barley were divided in two main groups: AtLFG5 and HvLFGe were more closely related to each other, and HvLFGa-d and AtLFG1-4 were clustered together in a different group. By expanding the sampling of analyzed plant genomes, we found that AtLFG1-3 (and its orthologues) are specific to Brassicaceae. However, as shown by (Weis et al., 2013) these proteins present overlapping functions with barley. Other lineage-specific functions for AtLFG1-2 probably exist. According to our data, functional analysis of AtLFG4 and AtLFG5 would exemplify more representative functions of the LFG family in plants.

5. Conclusions

We identified TMBIM proteins in previously unexplored organisms from all the three domains of life. Our findings indicate that the Eukaryotic BI and LFG families emerged from independent distant relatives in bacteria. However, in order to elucidate the origin of each family in eukaryotes, further search across sequenced bacterial genomes and metagenomes is necessary. We also provide a wider scope of the phylogenetic relationships of these proteins and clarify their relationships among some taxa such as fungi. The major contribution of

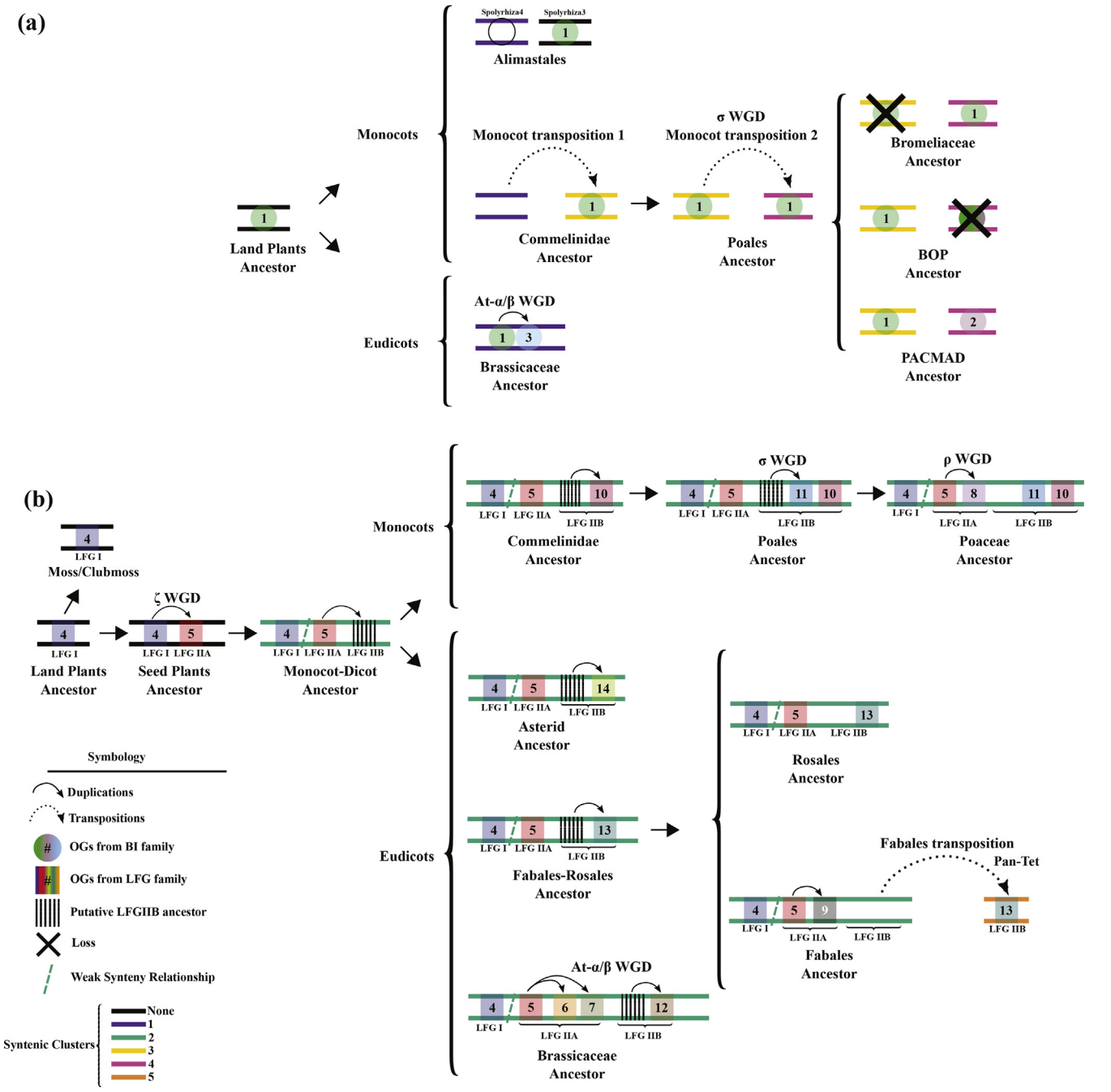


Fig. 5. Proposed model of evolution and expansion of BI and LFG families in plants. (a) Evolution and expansion of the BI family in plants. A BI gene was present in the last common ancestor of land plants, which represents the ancestor of OG1 (green circle). In monocots, after the split of Alimastales, the BI genes transposed to another genomic context in the common ancestor of Commelinidae. In duckweed this BI protein from OG1 seems to have been duplicated, and transposed to another genomic context. Therefore a BI-1 squence (not from OG1) from common duckweed still conserves the same synteny as eudicot plants. Then, this transposed OG 1 (in monocots) experienced a subsequent replicative transposition to another genomic context in the ancestor Poales. This duplication was lost in the common ancestor of the BOP clade, but was retained in the common ancestor of Bromeliaceae and the common ancestor of the PACMAD group of grasses, leading to the appearance of OG2 in the latter (purple circle). This second transposition could be the reason why a sequence from pineapple remains syntenic to the sequences within the PACMAD group. On the other hand, in Eudicots a duplication during the At-α/β WGD possibly gave rise to OG 3 in Brassicaceae (blue circle). (b) Evolution and expansion of the LFG family in plants. An LFG gene was present in the last common ancestor of land plants, which represents the ancestor of OG4 (blue square). This gene represents the ancestor of the LFG I group and the entire LFG family in land Plants. Possibly, during ζ WGD a duplication gave rise to the LFG IIA group in seed plants (OG 5, red square). A subsequent duplication event in the common ancestor of monocots and dicots gave rise to the putative ancestor gene of LFG IIB (stripped box). Subsequent duplication events of LFG IIA (OG5) and IIB, but not LFG I (OG4), gave rise to several lineage-specific OGs in angiosperms, some of which coincide with WGDs in Brassicaceae, Poales and Poaceae. OG13 (LFG IIB) in the ancestor of Fabales was transposed to another genomic context, possibly during the Panilionoid tetraploidy (Pan-Tet). Symbology: Parallel lines represent the plant genomes and are colored according to the syntenic clusters in Fig. 4. Dashed "/" line represents the poor syntenic relationship between clusters 2a and 2b. Number of the OGs are indicated inside squares and circles. Solid and doted arrows represent duplications and transpositions, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

this work is that we provide a deep analysis about the evolutionarily history of the TMBIM superfamily in land plants, and that we detected OGs that arose across plant evolution due to specific WGD events. Brassicaceae and monocots, with four specific OG each, represent a good example of the crucial role of WGD on gene family expansions, creating new material for gene functional diversification. Further functionality and expression data from several species remain necessary to understand the TMBIM superfamily. Our results provide a benchmark to carry out research in this regard and comprehend the role of these proteins, not only in the PCD regulation but in other functions as well.

6. Declarations of interest

None.

Acknowledgments

This work was supported by the "Consejo Nacional de Ciencia y Tecnología" (CONACyT), Mexico (grant number 215098 and 247355, 438058 to S.D.G.T, and 240187 to A.P.S.). We thank the Centro de Investigacion Cientifica de Yucatan (CICY) for the support in the realization of this study. The sequence data of *Aquilegia Coerulea*, *Cucumis sativus*, *Salix purpurea*, *Brassica rapa*, *Brachypodium stace*i, *Panicum virgatum*, and *Kalanchoe marneriana* were produced by the US Department of Energy Joint Genome (DOE-JGI) Institute http://www.jgi.doe.gov/ in collaboration with the user community. We thank the US DOE-JGI for prepublication access to such data.

Author contributions

SDGT, APS, and LCRZ conceived and designed the research.
SDGT performed the phylogenetic and orthology analyses.
APS and TZ performed the syntenic and tandem analysis.
SDGT and APS wrote the manuscript. MES and EC contributed to the analysis and interpretation of the manuscript and provided feedback.
LCRZ supervised the research and secured funding. All authors have read and approved the final manuscript. SDGT and APS contributed equally.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.04.032.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402. http://dx.doi.org/10.1093/nar/25.17.3389.

Asplund-Samuelsson, J., Bergman, B., Larsson, J., 2012. Prokaryotic caspase homologs: phylogenetic patterns and functional characteristics reveal considerable diversity. PLoS One 7, e49888.

Babaeizad, V., Imani, J., Kogel, K.H., Eichmann, R., Hückelhoven, R., 2009. Over-expression of the cell death regulator BAX inhibitor-1 in barley confers reduced or enhanced susceptibility to distinct fungal pathogens. Theor. Appl. Genet. 118, 455–463. http://dx.doi.org/10.1007/s00122-008-0912-2.

Baek, D., Nam, J., Koo, Y.D., Kim, D.H., Lee, J., Jeong, J.C., Kwak, S., Chung, W.S., Lim, C.O., Bahk, J.D., Hong, J.C., Lee, S.Y., Kawai-yamada, M., Uchimiya, H., Yun, D., 2004. Bax-induced cell death of Arabidopsisis meditated through reactive oxygen-dependent and -independent processes. Plant Mol. Biol. 56, 15–27. http://dx.doi.org/10.1007/s11103-004-3096-4.

Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W., Noble, W.S., 2009. MEME SUITE: tools for motif discovery and searching. Nucleic Acids Res gkp335.

Bailey, T.L., Elkan, C., et al., 1994. Fitting a mixture model by expectation maximization to discover motifs in bipolymers.

Bastian, M., Heymann, S., Jacomy, M., et al., 2009. Gephi: an open source software for exploring and manipulating networks. Icwsm 8, 361–362.

Carrara, G., Parsons, M., Saraiva, N., Smith, G.L., 2017. Golgi anti-apoptotic protein: a tale of camels, calcium, channels and cancer. Open Biol. 7, 170045. http://dx.doi.org/10.1098/rsob.170045.

Cebulski, J., Malouin, J., Pinches, N., Cascio, V., Austriaco, N., 2011. Yeast bax inhibitor, Bxi1p, is an ER-localized protein that links the unfolded protein response and programmed cell death in Saccharomyces cerevisiae. PLoS One 6, 1–7. http://dx.doi.org/10.1371/journal.pone.0020882.

Chae, H.J., Ke, N., Kim, H.R., Chen, S., Godzik, A., Dickman, M., Reed, J.C., 2003. Evolutionarily conserved cytoprotection provided by Bax Inhibitor-1 homologs from animals, plants, and yeast. Gene 323, 101–113. http://dx.doi.org/10.1016/j.gene.2003.09.011.

Chang, Y., Bruni, R., Kloss, B., Assur, Z., Kloppmann, E., Rost, B., Hendrickson, W.A., Liu, Q., 2014. Structural basis for a pH-sensitive calcium leak across membranes. Science 344, 1131–1135. http://dx.doi.org/10.1126/science.1252043.

Chen, Y., Duan, Z., Chen, P., Shang, Y., Wang, C., 2015. The Bax inhibitor MrBI-1 regulates heat tolerance, apoptotic-like cell death, and virulence in Metarhizium robertsii. Sci. Rep. 5, 10625. http://dx.doi.org/10.1038/srep10625.

Chipuk, J.E., Moldoveanu, T., Llambi, F., Parsons, M.J., Green, D.R., 2010. The BCL-2 Family Reunion. Mol. Cell 37, 299–310. http://dx.doi.org/10.1016/j.molcel.2010.01.025.

Corradi, N., 2015. Microsporidia: eukaryotic intracellular parasites shaped by gene loss and horizontal gene transfers. Annu. Rev. Microbiol. 69, 167–183.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27, 1164–1165.

Derényi, I., Palla, G., Vicsek, T., 2005. Clique percolation in random networks. Phys. Rev. Lett. 94, 160202.

Duan, Y., Zhang, W., Li, B., Wang, Y., Li, K., Sodmergen, Han, C., Zhang, Y., Li, X., 2010. An endoplasmic reticulum response pathway mediates programmed cell death of root tip induced by water stress in Arabidopsis. https://doi.org/NPH3207[pii]r10.1111/j.1469-8137.2010.03207.x. New Phytol. 186, 681–695.

Eddy, S.R., 1998. Profile hidden Markov models. Bioinformatics 14, 755. http://dx.doi.org/10.1093/bioinformatics/14.9.755.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. http://dx.doi.org/10.1093/nar/gkh340.

Fawcett, J.A., de Peer, Y., Maere, S., 2013. Significance and biological consequences of polyploidization in land plant evolution. In: Greilhuber, J., Dolezel, J., Wendel, J.F. (Eds.), Plant Genome Diversity Volume 2: Physical Structure, Behaviour and Evolution of Plant Genomes. Springer Vienna, Vienna, pp. 277–293. http://dx.doi.org/10.1007/978-3-7091-1160-4_17.

Flagel, L.E., Wendel, J.F., 2009. Gene duplication and evolutionary novelty in plants. New Phytol. 183, 557–564. http://dx.doi.org/10.1111/j.1469-8137.2009.02923.x.

Gabaldón, T., Koonin, E.V., 2013. Functional and evolutionary implications of gene orthology. Nat. Rev. Genet. 14, 360–366.

Glasauer, S.M.K., Neuhauss, S.C.F., 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. Genom. 289, 1045–1060. http://dx.doi.org/10.1007/s00438-014-0889-2.

Gu, Z., Eils, R., Schlesner, M., 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32, 2847–2849. http://dx.doi.org/10.1093/bioinformatics/btw313.

Haas, B.J., Delcher, A.L., Wortman, J.R., Salzberg, S.L., 2004. DAGchainer: a tool for mining segmental genome duplications and synteny. Bioinformatics 20, 3643–3646. http://dx.doi.org/10.1093/bioinformatics/bth397.

Henke, N., Lisak, D.A., Schneider, L., Habicht, J., Pergande, M., Methner, A., 2011. The ancient cell death suppressor BAX inhibitor-1. Cell Calcium 50, 251–260. http://dx.doi.org/10.1016/j.ceca.2011.05.005.

Hofmann, K., Baron, M.D., 1996. Boxshade 3.21. Pretty Print. shading Mult. files. Kay Hofmann ISREC Bioinforma. Group, Lausanne, Switz.

Hu, L., Smith, T.F., Goldberger, G., 2009. LFG: a candidate apoptosis regulatory gene family. Apoptosis 14, 1255–1265. http://dx.doi.org/10.1007/s10495-009-0402-2.

Ihara-Ohori, Y., Nagano, M., Muto, S., Uchimiya, H., Kawai-Yamada, M., 2006. Cell death suppressor Arabidopsis bax inhibitor-1 is associated with calmodulin binding and ion homeostasis. Plant Physiol. 143, 650–660. http://dx.doi.org/10.1104/pp.106.090878.

Inoue, J., Sato, Y., Sinclair, R., Tsukamoto, K., Nishida, M., 2015. Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. Proc. Natl. Acad. Sci. 112, 14918–14923. http://dx.doi.org/10.1073/pnas.1507669112.

Isbat, M., Zeba, N., Kim, S.R., Hong, C.B., 2009. A BAX inhibitor-1 gene in Capsicum annuum is induced under various abiotic stresses and endows multi-tolerance in transgenic tobacco. J. Plant Physiol. 166, 1685–1693. http://dx.doi.org/10.1016/j.jplph.2009.04.017.

Ishikawa, T., Takahara, K., Hirabayashi, T., Matsumura, H., Fujisawa, S., Terauchi, R., Uchimiya, H., Kawai-Yamada, M., 2010. Metabolome analysis of response to oxidative stress in rice suspension cells overexpressing cell death suppressor bax inhibitor-1. Plant Cell Physiol. 51, 9–20. http://dx.doi.org/10.1093/pcp/pcp162.

Jaillon, et al., 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449, 463–467.

Jiao, Y., Li, J., Tang, H., Paterson, A.H., 2014. Integrated syntenic and phylogenomic analyses reveal an ancient genome duplication in monocots. Plant Cell 26, 2792–2802. http://dx.doi.org/10.1105/tpc.114.127597.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J., dePamphilis, C.W., 2011. Ancestral polyploidy in seed plants and angiosperms. Nature 473, 97–100.

Jiménez-Ruiz, A., Alzate, J.F., MacLeod, E.T., Lüder, C.G.K., Fasel, N., Hurd, H., 2010. Apoptotic markers in protozoan parasites. Parasites Vectors 3, 104. http://dx.doi.org/10.1186/1756-3305-3-104.

Kawai-Yamada, M., Ohori, Y., Uchimiya, H., 2004. Dissection of Arabidopsis Bax inhibitor-1 suppressing Bax-, hydrogen peroxide-, and salicylic acid-induced cell death. Plant Cell 16, 21–32. http://dx.doi.org/10.1105/tpc.014613.

Kiełbasa, S.M., Wan, R., Sato, K., Horton, P., Frith, M.C., 2011. Adaptive seeds tame genomic sequence comparison. Genome Res. 21, 487–493. http://dx.doi.org/10.1101/gr.113985.110.

Koonin, E.V., 2005. Orthologs, paralogs, and evolutionary genomics. Annu. Rev. Genet. 39. http://dx.doi.org/10.1146/annurev.genet.39.073003.114725.

Koonin, E.V., Aravind, L., 2002. Origin and evolution of eukaryotic apoptosis: the bacterial connection. Cell Death Differ. 9, 394–404.

Kristensen, D.M., Wolf, Y.I., Mushegian, A.R., Koonin, E.V., 2011. Computational methods for Gene Orthology inference. Brief. Bioinform. 12, 379–391.

Lechner, M., Findeiß, S., Steiner, L., Marz, M., Stadler, P.F., Prohaska, S.J., 2011. Proteinortho: detection of (Co-)orthologs in large-scale analysis. BMC Bioinf. 12, 124. http://dx.doi.org/10.1186/1471-2105-12-124.

Letunic, I., 2015. phyloT: phylogenetic Tree Generator. Phylot.biobyte.de. Nucl. Acids Res 10, 725. http://dx.doi.org/10.1093/nar/gkw290.

Letunic, I., Bork, P., 2016. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. Nucl. Acids Res gkw290.

Lord, C.E.N., Gunawardena, A.H.L.A.N., 2012. Programmed cell death in C. elegans, mammals and plants. Eur. J. Cell Biol. 91, 603–613. http://dx.doi.org/10.1016/j.ejcb.2012.02.002.

Lyons, E., Freeling, M., 2008. How to usefully compare homologous plant genes and chromosomes as DNA sequences. Plant J. 53, 661–673.

Lyons, E., Pedersen, B., Kane, J., Freeling, M., 2008. The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. Trop. Plant Biol. 1, 181–190. http://dx.doi.org/10.1007/s12042-008-9017-y.

Marin, B., Nowack, E.C.M., Glöckner, G., Melkonian, M., 2007. The ancestor of the Paulinella chromatophore obtained a carboxysomal operon by horizontal gene transfer from a Nitrococcus-like γ-proteobacterium. BMC Evol. Biol. 7, 85. http://dx.doi.org/10.1186/1471-2148-7-85.

Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Maréchal-Drouard, L., Marshall, W.F., Qu, L.-H., Nelson, D.R., Sanderfoot, A.A., Spalding, M.H., Kapitonov, V.V., Ren, Q., Ferris, P., Lindquist, E., Shapiro, H., Lucas, S.M., Grimwood, J., Schmutz, J., Cardol, P., Cerutti, H., Chanfreau, G., Chen, C.-L., Cognat, V., Croft, M.T., Dent, R., Dutcher, S., Fernández, E., Fukuzawa, H., González-Ballester, D., González-Halphen, D., Hallmann, A., Hanikenne, M., Hippler, M., Inwood, W., Jabbari, K., Kalanon, M., Kuras, R., Lefebvre, P.A., Lemaire, S.D., Lobanov, A.V., Lohr, M., Manuell, A., Meier, I., Mets, L., Mittag, M., Mittelmeier, T., Moroney, J.V., Moseley, J., Napoli, C., Nedelcu, A.M., Niyogi, K., Novoselov, S.V., Paulsen, I.T., Pazour, G., Purton, S., Ral, J.-P., Riaño-Pachón, D.M., Riekhof, W., Rymarquis, L., Schroda, M., Stern, D., Umen, J., Willows, R., Wilson, N., Zimmer, S.L., Allmer, J., Balk, J., Bisova, K., Chen, C.-J., Elias, M., Gendler, K., Hauser, C., Lamb, M.R., Ledford, H., Long, J.C., Minagawa, J., Page, M.D., Pan, J., Pootakham, W., Roje, S., Rose, A., Stahlberg, E., Terauchi, A.M., Yang, P., Ball, S., Bowler, C., Dieckmann, C.L., Gladyshev, V.N., Green, P., Jorgensen, R., Mayfield, S., Mueller-Roeber, B., Rajamani, S., Sayre, R.T., Brokstein, P., Dubchak, I., Goodstein, D., Hornick, L., Huang, Y.W., Jhaveri, J., Luo, Y., Mart'inez, D., Ngau, W.C.A., Otillar, B., Poliakov, A., Porter, A., Szajkowski, L., Werner, G., Zhou, K., Grigoriev, I.V., Rokhsar, D.S., Grossman, A.R., 2007. The chlamydomonas genome reveals the evolution of key animal and plant functions. Science 318, 245–250. http://dx.doi.org/10.1126/science.1143609.

Nagano, M., Takahara, K., Fujimoto, M., Tsutsumi, N., Uchimiya, H., Kawai-Yamada, M., 2012. Arabidopsis sphingolipid fatty acid 2-hydroxylases (AtFAH1 and AtFAH2) are functionally differentiated in fatty acid 2-hydroxylation and stress responses. Plant Physiol. 159, 1138–1148. http://dx.doi.org/10.1104/pp.112.199547.

Nakatani, Y., Takeda, H., Kohara, Y., Morishita, S., 2007. Reconstruction of the vertebrate ancestral genome reveals dynamic genome reorganization in early vertebrates. Genome Res. 17, 1254–1265. http://dx.doi.org/10.1101/gr.6316407.

Okonechnikov, K., Golosova, O., Fursov, M., et al., 2012. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28, 1166–1167.

Palla, G., Derenyi, I., Farkas, I., Vicsek, T., 2005. Uncovering the overlapping community structure of complex networks in nature and society. Nature 435, 814–818.

Panchy, N., Lehti-Shiu, M., Shiu, S.-H., 2016. Evolution of gene duplication in plants. Plant Physiol. 171, 2294–2316. http://dx.doi.org/10.1104/pp.16.00523.

Pattengale, N.D., Alipour, M., Bininda-Emonds, O.R.P., Moret, B.M.E., Stamatakis, A., 2009. How many bootstrap replicates are necessary? In: Batzoglou, S. (Ed.), Proceedings of the 13th Annual International Conference, Research in Computational Molecular Biology, RECOMB 2009, Tucson, AZ, USA, May 18–21, 2009. Springer Berlin Heidelberg, Berlin, pp. 184–200. http://dx.doi.org/10.1007/978-3-642-02008-7_13.

Pei, J., Kim, B.-H., Grishin, N.V., 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. Nucl. Acids Res. 36, 2295. http://dx.doi.org/10.1093/nar/gkn072.

Priault, M., Camougrand, N., Kinnally, K.W., Vallette, F.M., Manon, S., 2003. Yeast as a tool to study Bax/mitochondrial interactions in cell death. FEMS Yeast Res. 4, 15 LP-27.

R Development Core Team, 2008. R: A Language and Environment for Statistical Computing.

Renny-Byfield, S., Wendel, J.F., 2014. Doubling down on genomes: polyploidy and crop plants. Am. J. Bot. 101, 1711–1725. http://dx.doi.org/10.3732/ajb.1400119.

Rice, P., Longden, I., Bleasby, A., 2000. EMBOSS: the European molecular biology open software suite. Trends Genet. 16, 276–277.

Robinson, K.S., Clements, A., Williams, A.C., Berger, C.N., Frankel, G., 2011. Bax inhibitor 1 in apoptosis and disease. Oncogene 30, 2391–2400. http://dx.doi.org/10.1038/onc.2010.636.

Rojas-Rivera, D., Hetz, C., 2014. TMBIM protein family: ancestral regulators of cell death. Oncogene 34, 1–12. http://dx.doi.org/10.1038/onc.2014.6.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504. http://dx.doi.org/10.1101/gr.1239303.

Sojo, V., Dessimoz, C., Pomiankowski, A., Lane, N., 2016. Membrane proteins are dramatically less conserved than water-soluble proteins across the tree of life. Mol. Biol. Evol. 33, 2874–2884.

Soucy, S.M., Huang, J., Gogarten, J.P., 2015. Horizontal gene transfer: building the web of life. Nat. Rev. Genet. 16, 472–482.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Tang, H., Bowers, J.E., Wang, X., Paterson, A.H., 2010. Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. Proc. Natl. Acad. Sci. 107, 472–477. http://dx.doi.org/10.1073/pnas.0908007107.

Taylor-Brown, E., Hurd, H., 2013. The first suicides: a legacy inherited by parasitic protozoans from prokaryote ancestors. Parasites Vectors 6, 108. http://dx.doi.org/10.1186/1756-3305-6-108.

Van de Peer, Y., Mizrachi, E., Marchal, K., 2017. The evolutionary significance of polyploidy. Nat. Rev. Genet. 18, 411–424.

van Stelten, J., Silva, F., Belin, D., Silhavy, T.J., 2009. Effects of antibiotics and a proto-oncogene homolog on destruction of protein translocator SecY. Science 325, 753–756. http://dx.doi.org/10.1126/science.1172221.

Wang, Y., Ficklin, S.P., Wang, X., Feltus, F.A., Paterson, A.H., 2016. Large-scale gene relocations following an ancient genome triplication associated with the diversification of core eudicots. PLoS One 11, e0155637.

Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T., Jin, H., Marler, B., Guo, H., Kissinger, J.C., Paterson, A.H., 2012. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucl. Acids Res. 40, e49.

Watanabe, N., Lam, E., 2006. Arabidopsis Bax inhibitor-1 functions as an attenuator of biotic and abiotic types of cell death. Plant J. 45, 884–894. http://dx.doi.org/10.1111/j.1365-313X.2006.02654.x.

Weis, C., Hückelhoven, R., Eichmann, R., 2013. LIFEGUARD proteins support plant colonization by biotrophic powdery mildew fungi. J. Exp. Bot. 64, 3855–3867. http://dx.doi.org/10.1093/jxb/ert217.

Wheeler, T.J., Clements, J., Finn, R.D., 2014. Skylign: a tool for creating informative, interactive logos representing sequence alignments and profile hidden Markov models. BMC Bioinf. 15, 7. http://dx.doi.org/10.1186/1471-2105-15-7.

Xu, G., Wang, S., Han, S., Xie, K., Wang, Y., Li, J., Liu, Y., 2017. Plant Bax Inhibitor-1 interacts with ATG6 to regulate autophagy and programmed cell death. Autophagy 13, 1161–1175. http://dx.doi.org/10.1080/15548627.2017.1320633.

Xu, Q., Reed, J.C., 1998. Bax inhibitor-1, a mammalian apoptosis suppressor identified by functional screening in yeast. Mol. Cell 1, 337–346. http://dx.doi.org/10.1016/S1097-2765(00)80034-9.

Yamagami, A., Nakazawa, M., Matsui, M., Tujimoto, M., Sakuta, M., Asami, T., Nakano, T., 2009. Chemical genetics reveal the novel transmembrane protein BIL4, which mediates plant cell elongation in brassinosteroid signaling. Biosci. Biotechnol. Biochem. 73, 415–421. http://dx.doi.org/10.1271/bbb.80752.

Yu, Y., Xiao, J., Yang, Y., Bi, C., Qing, L., Tan, W., 2015. Ss-Bi1 encodes a putative BAX inhibitor-1 protein that is required for full virulence of Sclerotinia sclerotiorum. Physiol. Mol. Plant Pathol. 90, 115–122. http://dx.doi.org/10.1016/j.pmpp.2015.04.005.

Yue, H., Nie, S., Xing, D., 2012. Over-expression of Arabidopsis Bax inhibitor-1 delays methyl jasmonate-induced leaf senescence by suppressing the activation of MAP kinase 6. J. Exp. Bot. 63, 4463–4474. http://dx.doi.org/10.1093/jxb/ers122.

Zhao, T., Holmer, R., de Bruijn, S., Angenent, G.C., van den Burg, H.A., Schranz, M.E., 2017. Phylogenomic synteny network analysis of MADS-Box transcription factor genes reveals lineage-specific transpositions, ancient tandem duplications, and deep positional conservation. Plant Cell.

Zhao, T., Schranz, M.E., 2017. Network approaches for plant phylogenomic synteny analysis. Curr. Opin. Plant Biol. 36, 129–134. http://dx.doi.org/10.1016/j.pbi.2017.03.001.

Zhao, Y., Tang, H., Ye, Y., 2012. RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. Bioinformatics 28, 125–126.

Zhou, J., Zhu, T., Hu, C., Li, H., Chen, G., Xu, G., Wang, S., Zhou, J., Ma, D., 2008. Comparative genomics and function analysis on BI1 family. Comput. Biol. Chem. 32, 159–162. http://dx.doi.org/10.1016/j.compbiolchem.2008.01.002.