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).