Identification and Bioinformatics Analysis of the Expansin Gene Family in *Physcomitrium patens* Postprint

Authors: Lan Yuchun, Huang Bin, Wei Jiao, Jiang Shan

Date: 2019-08-27T00:00:00+00:00

Abstract

Expansins (EXP) constitute a gene family that participates in nearly all aspects of plant development, from seed germination to fruit ripening. In this study, bioinformatics approaches were employed to identify members of the Expansin gene family in Physcomitrella patens, and to analyze their gene structure, chromosomal localization, and phylogenetic relationships. The results revealed that the Physcomitrella patens genome contains 32 Expansin A (EXPA) and 6 Expansin-like A (EXLA) members, while no Expansin-like B (EXLB) or Expansin B (EXPB) were detected. The amino acid sequence lengths of expansins range from 228–290 residues, with the encoded proteins harboring two conserved domains, Pollen_allerg_1 and DPBB_1. Protein subcellular localization predictions indicated that approximately four-fifths of EXP family genes in Physcomitrella patens were predicted to localize extracellularly by the CELLO online tool, whereas Euk-mPLoc predicted all EXP gene family members to be extracellularly localized. Gene structure analysis demonstrated that approximately 68% of Expansin genes in Physcomitrella patens contain 1–3 introns. This study provides a comprehensive analysis of the basic characteristics of the expansin gene family in Physcomitrella patens, laying a foundation for further investigation into the molecular evolution and biological functions of its expansin genes.

Full Text

Preamble

Identification and Bioinformatic Analysis of the Expansin Gene Family in *Physcomitrella patens*

Lan Yuchun[1], Huang Bing[1], Wei Jiao[1], Jiang Shan[2]*
(1. School of Life Sciences, Guizhou Normal University, Guiyang 550001, China;

2. School of International Education, Guizhou Normal University, Guiyang 550001, China)

Abstract

Expansins (EXP) constitute a gene family that participates in nearly all aspects of plant development, from seed germination to fruit ripening. Using bioinformatic approaches, we identified members of the expansin gene family in the moss *Physcomitrella patens* and analyzed their gene structures, chromosomal locations, and phylogenetic relationships. The results revealed that the *P. patens* genome contains 32 Expansin A (EXPA) genes and 6 Expansin-like A (EXLA) genes, while no Expansin-like B (EXLB) or Expansin B (EXPB) members were detected. The encoded expansin proteins ranged from 228 to 290 amino acids in length and possessed two conserved domains: Pollen_allerg_1 and DPBB_1. Subcellular localization predictions indicated that approximately four-fifths of EXP family genes were localized extracellularly according to the CELLO online tool, while Euk-mPLoc predicted all EXP family members to be extracellular. Gene structure analysis showed that about 68% of *P. patens* expansin genes contain 1–3 introns. This study provides fundamental information on the expansin gene family in *P. patens*, laying the groundwork for further investigation into the molecular evolution and biological functions of these genes.

Keywords: *Physcomitrella patens*, expansin, gene family, bioinformatics

Introduction

The plant cell wall is a crucial structure that determines cell shape, provides mechanical support and rigidity, and serves as the final barrier against pathogens. Expansins are cell wall proteins that mediate cell wall loosening by disrupting non-covalent bonds between cell wall components, thereby increasing wall extensibility and participating in numerous plant developmental processes. Expansins were first discovered in studies of acid-induced extension in cucumber (*Cucumis sativus*) hypocotyl cell walls, where they were shown to restore extensibility to heat-inactivated cell walls without exhibiting lysozyme activity, suggesting a role in regulating cell wall expansion. Subsequent research has linked expansins to various developmental processes, including fruit softening, pollen tube elongation, leaf abscission, and plant stress responses. Early studies proposed that expansins function through an enzymatic mechanism to loosen cell wall components, facilitate cell expansion, and enhance cellular flexibility under adverse environmental conditions.

With the advent of genome sequencing technologies, expansin genes have been identified in an increasing number of species. Based on evolutionary relationships, the expansin superfamily is classified into four subfamilies: , , , and , now designated as EXPA, EXPB, EXLA, and EXLB, respectively. EXPA and EXPB subfamily proteins exhibit cell wall extension activity and participate in cell elongation and other developmental processes, whereas EXLA and EXLB

subfamilies are only known from their gene sequences, with no experimental evidence demonstrating cell wall-loosening functions.

Physcomitrella patens, a model organism for non-vascular plants, occupies a unique evolutionary position and has a fully sequenced genome, making it an ideal system for functional genomics, developmental biology, plant physiology, and evolutionary studies. The *P. patens* genome is 511 Mb and contains 27 chromosomes. Its advantages include high efficiency of homologous recombination with exogenous genes, ease of cultivation, short life cycle, and readily observable phenotypes. Based on the completed genome sequence of *P. patens* (http://www.cosmoss.org/), this study employed bioinformatic methods to identify and characterize the expansin gene family, including phylogenetic analysis, chromosomal mapping, and gene structure analysis, providing a reference framework for future functional studies.

Materials and Methods

1.1 Experimental Materials

This study utilized the model plant *Physcomitrella patens* as the research subject. The complete genome data were downloaded from the Ensembl Plants database (http://plants.ensembl.org/index.html) using the most recent release available.

1.2 Identification of the *P. patens* Expansin Gene Family

Genomic data for *P. patens* were obtained from the Ensembl database (http://asia.ensembl.org/index.html). Hidden Markov Model (HMM) profiles for conserved protein domains were downloaded from the Pfam database (http://pfam.xfam.org/). The HMMER software (Finn et al., 2011) was used to search the *P. patens* predicted proteome for sequences containing the Pollen_allerg_1 (PF01357) and DPBB_1 (PF03330) conserved domains (Li et al., 2014). Sequences with E-values $< 1 \times 10^{-2}$ were selected as preliminary candidates. These candidate sequences were then used to construct a *P. patens*-specific HMM profile, which was employed for a second round of screening (E-value < 0.01). To ensure accuracy, the resulting candidate proteins were manually verified using the SMART website (http://smart.embl-heidelberg.de/) (Letunic et al., 2012) to confirm the presence of both Pollen_allerg_1 and DPBB_1 domains, with sequences lacking either domain being eliminated.

1.3 Protein Characterization and Subcellular Localization Prediction

Molecular weight (Mw) and isoelectric point (pI) data for the identified *P. patens* expansin proteins were obtained from the NCBI database (https://www.ncbi.nlm.nih.gov/). Subcellular localization was predicted using the CELLO online tool (http://cello.life.nctu.edu.tw/) and the Euk-mPLoc website (http://www.csbio.sjtu.edu.cn/bioinf/euk-multi-2/).

1.4 Phylogenetic Analysis of the *P. patens* Expansin Family

Protein sequences of *P. patens* expansins were aligned with expansin sequences from *Arabidopsis thaliana* retrieved from the TAIR (https://www.arabidopsis.org/) and NCBI databases. Multiple sequence alignment was performed using ClustalX, and phylogenetic trees were constructed using MEGA 7.0 with the Maximum Likelihood method and bootstrap validation (1,000 replicates).

1.5 Conserved Motif and Gene Structure Analysis

The MEME suite was used to identify conserved motifs in *P. patens* expansin gene sequences. Intron and exon distribution patterns were extracted from genome annotation files and visualized using TBtools software.

1.6 Chromosomal Mapping of *P. patens* Expansin Genes

Chromosomal location information was obtained from the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html) and mapped using MapInspect software (Yi et al., 2015).

Results

2.1 Members of the *P. patens* Expansin Gene Family

HMMER screening of the *P. patens* predicted proteome yielded 91 candidate proteins. After manual verification using SMART, 38 expansin proteins were ultimately identified (Table 1), a number comparable to that found in the dicot model *Arabidopsis thaliana* (Seader et al., 2016). Protein characterization revealed that the 38 *P. patens* expansin genes encode proteins ranging from 25.14 to 73.95 kDa in molecular weight, with approximately 87% falling within the 20–30 kDa range. The isoelectric points ranged from 4.14 to higher values, and the proteins comprised 228–290 amino acids. Subcellular localization predictions indicated that about one-fifth of EXP family genes were predicted to be periplasmic by CELLO, while Euk-mPLoc predicted all members to be extracellular. Pfam database verification confirmed that all 38 expansin proteins contained both Pollen_allerg_1 and DPBB_1 characteristic domains.

2.2 Conserved Domains and Gene Structure Analysis

MEME analysis identified 10 highly conserved motifs in *P. patens* expansin gene sequences (Figure 1 [Figure 1: see original paper]A). The arrangement of these motifs followed distinct patterns: the EXPA subfamily exhibited a stable architecture of motif 2–(motif 8)–motif 4–motif 1–motif 3, except for PpEXPA32 which lacked motif 1. The EXLA subfamily displayed a consistent structure of motif 2–motif 9–motif 7–motif 10–motif 6, with no instances of motif loss, gain, or substitution.

Analysis of intron-exon structure among the 38 *P. patens* EXP genes (Figure 1B) revealed that approximately 49% of EXPA genes contained two introns, about one-fifth contained a single intron, 9% contained three introns, and approximately 22% were intronless. In contrast, the EXLA subfamily almost exclusively contained one intron, except for PpEXLA1 and PpEXLA4. Notably, PpEXLA1 contained five introns, the highest number observed in the expansin gene family.

Note: A. Distributions of conserved motifs in Expansin genes. Ten putative motifs are indicated in different colored boxes (see Table 2 for details). B. Exon/intron organization of Expansin genes. Green boxes represent exons and black lines represent introns. The upstream/downstream regions of Expansin genes are indicated in yellow boxes. Exon lengths can be inferred from the scale at the bottom.

Figure 1. Gene structure and conserved motif analysis of *Physcomitrella patens* Expansin genes.

Table 2. List of putative motifs of Expansin proteins

Motif	E-value	Width
motif1	2.2e-911	-
motif2	2.8e-1148	-
motif3	6.3e-610	-
motif4	1.1e-521	-
motif5	6.30E-298	-
motif6	2.9e-399	-
motif7	8.40E-146	-
motif8	2.60E-133	-
motif9	1.10E-121	-
motif10	9.20E-107	-

2.3 Phylogenetic Analysis of the *P. patens* Expansin Gene Family

Phylogenetic analysis based on amino acid sequence alignment revealed three distinct clades in the *P. patens* expansin gene family (Figure 2 [Figure 2: see original paper]), with both subfamilies forming independent evolutionary branches. The EXPA subfamily comprised two separate evolutionary branches, indicating divergent evolutionary paths among its members during long-term evolution. Some expansin genes exhibited long branch lengths, suggesting early divergence and substantial sequence differentiation, though they maintain clear evolutionary relationships.

To investigate the evolutionary relationship between *P. patens* and *Arabidopsis thaliana* EXP genes, a phylogenetic tree was constructed using 38 expansin amino acid sequences from each species (Figure 3 [Figure 3: see original paper]). The tree clearly divided the 38 *P. patens* EXP genes into two subfamilies. Twelve

pairs of putative orthologous proteins were identified, with four pairs showing bootstrap values of 99, while only one paralogous pair was detected within the species.

Figure 2. Phylogenetic tree of Expansin genes in *Physcomitrella patens*.

Figure 3. Phylogenetic tree of Expansin genes in *Physcomitrella patens* and *Arabidopsis thaliana*.

2.4 Chromosomal Distribution of *P. patens* EXP Genes

Chromosomal location information obtained from Phytozome revealed that the 38 *P. patens* expansin genes are distributed across 15 chromosomes. Chromosomes 8 and 14 each harbor six EXP genes (Figure 4 [Figure 4: see original paper]). A small gene cluster was identified on chromosome 14 based on the definition of gene clusters. The remaining genes were unevenly distributed, with seven chromosomes containing a single gene and three chromosomes containing two genes each.

Figure 4. Chromosome distribution of the Expansin gene family in *Physcomitrella patens*.

Discussion

Expansins are essential components of the plant cell wall that participate in cell expansion and various developmental processes involving cell wall modification. This study identified 38 expansin genes in *P. patens*, which can be classified into EXPA and EXLA subfamilies. The two subfamilies exhibit conserved features in intron-exon distribution and motif architecture that represent their distinctive characteristics. Approximately 70% of EXPA subfamily genes contain 1–2 introns, while about 22% are intronless. In contrast, the EXLA subfamily almost exclusively contains a single intron. Motif analysis revealed that both subfamilies share motif 2, while EXLA genes consistently exhibit the motif 2–motif 9–motif 7–motif 10–motif 6 structure, showing greater conservation than the EXPA subfamily. The conserved intron-exon structures and motif features within each subfamily, along with motif diversification between subfamilies, suggest that *P. patens* EXP genes participate in multiple metabolic pathways. Chromosomal mapping revealed random distribution of expansin genes across 15 chromosomes, with minimal gene clustering except on chromosome 14.

EXP genes play crucial physiological roles in plant growth, development, and stress responses, making them a focal point in plant functional genomics. However, their specific functions in *P. patens* remain largely uncharacterized. This study leveraged existing database resources and bioinformatic tools to identify 38 expansin gene family members in *P. patens* and analyze their protein characteristics, evolutionary relationships, gene structures, and chromosomal locations. These findings provide a theoretical foundation for future investigations

into expansin gene function at the molecular level and illuminate species-specific features of *P. patens*.

References

BAI J, PENNILLA LA, NING JC, et al., 2002. Diversity in nucleotide binding site-leucine-rich repeat genes in cereals [J]. *Cytogenet Genome Res*, 12(12): 1871-84.

BOLSER DM, RUQURTI B, ROBERTS J, et al., 2017. Ensembl plants: Integrating tools for visualizing, mining, and analyzing plant genomic data [J]. *Methods Mol Biol*, 1533: 1-31.

BEDINGER P, DURACHKO DM, 1997. Group I allergens of grass pollen as cell wall-loosening agents [J]. *Proc Natl Acad Sci USA*, 94(12): 6559-6564.

COSGROVE DJ, 2000. Loosening of plant cell walls by expansins [J]. *Nature*, 407(6802): 321-6.

COSGROVE DJ, 2005. Growth of the plant cell wall [J]. *Nat Rev Mol Cell Biol*, 6(11): 850-61.

COSGROVE DJ, LI LC, CHO HT, et al., 2002. The growing world of expansins [J]. *Plant Cell Physiol*, 43(12): 1435-1444.

ERIC JB, BENEDETTO R, JEREMY AR, et al., 2005. Changes in expansin activity and gene expression during ethylene-promoted leaflet abscission in *Sambucus nigra* [J]. *J Exp Bot*, 56(413): 817-23.

HAN YY, LI AX, LI F, WANG W, 2012. Characterization of a wheat (*Triticum aestivum* L.) expansin gene, *TaEXPB23*, involved in the abiotic stress response and phytohormone regulation [J]. *Plant Physiol Biochem*, 54: 49-58.

HANS K, KENT JB, DAVID AB, HYUNG-TAEG C, et al., 2004. Nomenclature for members of the expansin superfamily of genes and proteins [J]. *Plant Mol Biol*, 55(3): 311-314.

IVICA L, TOBIAS D, PEER B, 2012. SMART 7: Recent updates to the protein domain annotation resource [J]. *Nucl Acids Res*, 40(Database issue): D302-5.

JAVIER S, ROBERT EC, DANIEL JC, 1992. Two endogenous proteins that induce cell wall extension in plants [J]. *The Plant Cell*, 4(undefined): 1425-33.

JIN HQ, CHEN YH, JIN YF, 2006. The development of expansin (cell wall expansion protein) [J]. *Chin Bull Life Sci*, 18(2): 168-174. [金慧清, 陈英豪, 金勇丰, 2006. Expansin(细胞壁松弛蛋白) 的发展 [J]. 生命科学, 18(2): 168-174.]

LI HY, SHI Y, DING YN, et al., 2014. Bioinformatics analysis of expansin gene family in poplar genome[J]. *J Beijing For Univ*, 36(2):59-67. [李昊阳, 施杨, 丁亚娜, 等, 2014. 杨树扩展蛋白基因家族的生物信息学分析 [J]. 北京林业大学学报, 36(2):59-67.]

ROBERT DF, JODY C, SEAN RE, 2011. HMMER web server: interactive sequence similarity searching [J]. *Nucl Acids Res*, 39(Web Server issue): W29-37.

RENSING SA, ROMBAUTS S, VAN DEP, et al., 2002. Moss transcriptome and beyond [J]. *Trends Plant Sci*, 7(12): 535-8.

STEFAN AR, DANIEL L, ANDREAS Z, et al., 2008. The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants [J]. *Science*, 319(5859): 64-9.

SAMPEDRO J, CAREY RE, COSGROVE DJ, 2006. Genome histories clarify evolution of the expansin superfamily: New insights from the poplar genome and pine ESTs [J]. *J Plant Res*, 119(1): 11-21.

VICTORIA HS, JENNIFER MT, ROBERT EC, 2016. Utility of the *Amborella trichopoda* expansin superfamily in elucidating the history of angiosperm expansins [J]. *J Plant Res*, 129(2): 199-207.

YI JM, HUANG T, HUANG Y, et al., 2015. Phylogenetic analysis of MADS-box gene family in *Physcomitrella patens*[J]. *Plant Physiol J*, 51(2): 197-206. [易吉明, 黄婷, 黄勇, 等, 2015. 小立碗藓 MADS-box 基因家族的系统进化分析 [J]. 植物生理学报, 51(2): 197-206.]

Note: Figure translations are in progress. See original paper for figures.