

A Comparative Genome Analysis of PME and PME1 Families Reveals the Evolution of Pectin Metabolism in Plant Cell Walls

Maojun Wang, Daojun Yuan*, Wenhui Gao, Yang Li, Jiafu Tan, Xianlong Zhang*

National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan, Hubei, China

Abstract

Pectins are fundamental polysaccharides in the plant primary cell wall. Pectins are synthesized and secreted to cell walls as highly methyl-esterified polymers and then demethyl-esterified by pectin methylesterases (PMEs), which are spatially regulated by pectin methylesterase inhibitors (PMEIs). Although PME and PME1 genes are pivotal in plant cell wall formation, few studies have focused on the evolutionary patterns of the PME and PME1 gene families. In this study, the gene origin, evolution, and expression diversity of these two families were systematically analyzed using 11 representative species, including algae, bryophytes, lycophytes and flowering land plants. The results show that 1) for the two subfamilies (PME and proPME) of PME, the origin of the PME subfamily is consistent with the appearance of pectins in early charophyte cell walls, 2) Whole genome duplication (WGD) and tandem duplication contribute to the expansion of proPME and PME1 families in land plants, 3) Evidence of selection pressure shows that the proPME and PME1 families have rapidly evolved, particularly the PME1 family in vascular plants, and 4) Comparative expression profile analysis of the two families indicates that the eudicot *Arabidopsis* and monocot rice have different expression patterns. In addition, the gene structure and sequence analyses show that the origin of the PME1 domain may be derived from the neofunctionalization of the pro domain after WGD. This study will advance the evolutionary understanding of the PME and PME1 families and plant cell wall development.

Citation: Wang M, Yuan D, Gao W, Li Y, Tan J, et al. (2013) A Comparative Genome Analysis of PME and PME1 Families Reveals the Evolution of Pectin Metabolism in Plant Cell Walls. PLoS ONE 8(8): e72082. doi:10.1371/journal.pone.0072082

Editor: Olga A. Zabolina, Iowa State University, United States of America

Received: April 7, 2013; **Accepted:** July 4, 2013; **Published:** August 12, 2013

Copyright: © 2013 Wang et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was financially supported by the National Natural Science Foundation of China (NO. 31230056) and National Basic Research Program (2010CB126001). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

* E-mail: xlzhang@mail.hzau.edu.cn (XZ); robert@mail.hzau.edu.cn (DY)

Introduction

Plant cell walls are complex and dynamic structures composed of diverse polysaccharides and proteins, mainly including cellulose, hemicellulose, and pectin [1], and are generally divided into two functional categories: primary walls surrounding the growing cell and secondary walls, which are thickened structures [2,3]. Pectins, a type of polysaccharides, appeared after the divergence of chlorophyta and charophyce [4,5], and are a major component of plant primary cell walls. Pectins are important for contributing to cellular structural integrity, cell adhesion and the mediation of defense responses [6]. These polysaccharides mainly accumulate in primary cell walls and occupy 35% in eudicots and noncommelinid monocots and only 2–10% in grass primary walls [7].

Pectins are assumed to be biosynthesized in *cis*-Golgi by a large number of glycosyltransferases, methyltransferases and acetyltransferases. The basic backbones of pectins are composed of homogalacturonan (HG), xylogalacturonan

(XGA), apiogalacturonan, rhamnogalacturonan I (RGI) and rhamnogalacturonan II (RGII) [8]. Homogalacturonan (HG) are methyl-esterified in *medial*-Golgi and transported to the primary cell wall in a highly methyl-esterified state, in which they are demethyl-esterified by pectin methylesterases (PMEs) [9], and spatially regulated by pectin methylesterase inhibitors (PMEIs) [10–12]. The demethyl-esterified pectins can bind Ca²⁺ to form a rigid gel, which plays a critical role in cell wall stiffening [13,14]. Therefore, the methyl-esterification status of HG crucially affects cell wall texture and mechanical properties. In addition, pectins can also be degraded by pectinases [15]. Research on the methyl-esterification of HG in primary walls helps clarify the mechanisms of cellular growth and cell shape modeling [16].

The mature and active region of PME genes mainly consists of the PME domain. In higher plants, the PME genes are classified into two types. Both of the two types of genes possess a PME domain. However, compared to type II PME (subsequently referred to as PME) gene, type I PME (subsequently referred to as proPME) gene possesses an

additional pro domain [17,18]. The pro domain is located at the N-terminus of proPME genes and shares similarities with the PME domain of the PMEI genes. The function of the pro domain has been reported, including involvement in maintaining unprocessed PME in the Golgi [19]. Because the sequence of the pro domain is similar to the PME domain, former studies inferred that it might play a role in auto-inhibitory activity of mature PME proteins (to prevent premature demethoxylation) [20]. However, these studies mainly focused on the function of a few genes using experimental methods, and there were hardly any comprehensive studies regarding how PME and PMEI genes originated. The recent availability of genome sequences for many plant species enables comparative genome analyses to make inferences regarding the origin and evolution of PME and PMEI genes.

The mechanisms underlying the origin of new genes are mainly *de novo* gene birth and the duplication-divergence hypothesis [21-23]. The modes of gene duplication can be summarized as six mechanisms (whole genome duplication (WGD), tandem duplication, proximal duplication, DNA based transposed duplication, retrotransposed duplication and dispersed duplication) [24]. In fact, WGD widely occurs in different species and plays a vital role in organism evolution [25], which often leads to the formation of new species or reflects adaptive evolution when organisms confront ecological stresses [26]. Genome evolution research has detected that the *Arabidopsis* and grape genomes shared an identical γ duplication event. Additionally, *Arabidopsis* experienced two species-specific WGDs (α and β). Both monocot rice and sorghum have experienced two WGDs (σ and ρ) [27,28]. Gene functional divergence, including subfunctionalization, neofunctionalization, pseudogenization and concerted evolution [29], is the direct consequence of gene duplication, which often results in gene family expansion. Study on gene family evolution provides important clues for explaining gene function divergence [30,31].

Currently, the phylogenetic appearance of pectin polymers in cell walls was discovered using chemical approaches, whereas there are no systematic reports of genetic evidence and functional research. To reveal the gene origination of the PME genes and the evolutionary relationship between pro and PMEI domains, we performed a comparative genome analysis of the PME and PMEI families across 11 species, including algae, bryophytes, lycophytes and flowering land plants. The results imply that the origin of PME genes coordinates with the appearance of pectin in cell walls. The PME domains may be derived from duplication and divergence of the pro domain and have rapidly evolved. The expression profile analyses of PME and PMEI families show different expression patterns in the model plants *Arabidopsis* and rice. In addition, a network analysis infers that the demethyl-esterification process of pectin conferred by PME and PMEI families is involved in a complex metabolic network. This study provides basic clues for further understanding the relationship between pectin metabolism and plant cell wall evolution.

Materials and Methods

Retrieval of PME and PMEI Gene Family Sequences

The gene models of *Arabidopsis thaliana* and *Oryza sativa* were downloaded from TAIR (The *Arabidopsis* Information Resource, <http://www.arabidopsis.org/>) and RGAP (Rice Genome Annotation Project, <http://rice.plantbiology.msu.edu/>). The gene files of *Carica papaya*, *Vitis vinifera*, *Populus trichocarpa*, *Solanum lycopersicum*, *Sorghum bicolor*, *Physcomitrella patens* and *Selaginella moellendorffii* were downloaded from Phytozome (<http://www.phytozome.net/>) [32]. The gene information of *Amborella trichopoda* was downloaded from <http://www.amborella.org/> [33]. The gene models of *Coleochaete orbicula* were downloaded from the National Center for Biotechnology Information (NCBI) GenBank (J0233843-J0252228) [34].

The Hidden Markov Model (HMM) profiles of PF01095 (PME domain) and PF04043 (PMEI domain) were downloaded from PFam database (<http://pfam.sanger.ac.uk/>), and the HMMER software package [35] was used to detect PME and PMEI genes with the best domain e-value cutoff of $1e-10$. These sequences were regarded as potential PME and PMEI genes. To validate the HMM search, these potential sequences were used as queries to search the NCBI non-redundant (nr) protein database with blastp program of GenBank and only the results with the best hits (an e-value less than $1e-5$) of “pectin methylesterases” and “pectin methylesterases inhibitor” were retained. Finally, partial genes were manually removed. An in-house Perl script was used to extract the domain sequences with the boundary site information from the HMM results. The domain sequences were further checked manually.

Orthomcl software was used to infer the orthologous genes among the species with the default settings [36], which initially required an all-vs-all blastp, and then the mcl clustering algorithm was used to deduce the relationship between genes. The orthologous genes were defined as genes in a cluster from at least three species.

Genome Synteny and Gene Duplication

The WGD information of *Arabidopsis*, grape, rice and sorghum was downloaded from former studies [27,28], and the PME and PMEI genes were detected. Tandem duplication genes were identified based on the physical location in the individual chromosome with no more than one intervening gene. To further analyze the genome synteny, the syntenic blocks among *Arabidopsis*, papaya, poplar, grape, tomato, rice and sorghum were downloaded from the Plant Genome Duplication Database (PGDD) [37]. All the PME and PMEI genes were mapped to the syntenic blocks for intra- and inter-genomic comparison. The Circos software was used to draw the syntenic diagram [38].

Motif Identification and the Exon-intron Structural Analysis

For all the PME, pro and PMEI domain sequences, the online MEME program (version 4.0.0) [39] (<http://meme.sdsc.edu/meme/cgi-bin/meme.cgi>) was employed to identify and analyze the conserved motifs among amino acid

sequences with the following parameters: number of repetitions, any; maximum number of motifs, 10; and optimum motif width set to >6 and <50.

The gene structure information of the PME and PMEI families were parsed from the General Feature Format (GFF) files of every species using an in-house Perl script. The diagrams of the exon-intron structures were drawn using the online program GSDS (<http://gsds.cbi.pku.edu.cn/>) [40].

Phylogenetic Analysis

For all the pro and PMEI domains, the sequences were aligned using Clustal X (version 2.0) [41], and the neighbor-joining (NJ) tree was constructed using PHYLIP software [42]. The maximum probability method was used to construct the consensus domain sequence for every species using the HMMER package. The MEGA5 program was then used to construct a maximum-likelihood (ML) tree of the pro and PMEI consensus domain sequences and the PME orthologous genes using the Whelan And Goldman (WAG) model based on the BIC scores (Bayesian Information Criterion) [43,44]. The molecular clock test was performed by comparing the ML value of the orthologous gene topology with and without the molecular clock constraints under the WAG model. The bootstrap value was 500 to construct the phylogenetic tree.

The Estimation of the Rates of Gene Evolution

The multiple alignment analysis of the protein domain sequence was performed using the Clustal X (version 2.0) [41], and the coding sequences were aligned and guided by alignments of protein sequences using the PAL2NAL software with the NOGAPS parameter [45]. The ratio of nonsynonymous substitutions per nonsynonymous site (Ka) to synonymous substitutions per synonymous site (Ks) (omega) homologous gene pairs was calculated with the yn00 procedure of the PAML package [46]. Based on the definition of Ka/Ks, a value less than 1 indicates negative or purifying selection acting on amino acid changes, whereas a value greater than 1 indicates positive selection, which may indicate adaptive evolution. The saturation effects were excluded by discarding the gene pairs in which Ks >2.5.

The Expression Analysis of the PME and PMEI families in *Arabidopsis* and Rice

The *Arabidopsis* microarray data were downloaded from the Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo/>) with the GSE series accession numbers GSE5629, GSE5630, GSE5631, GSE5632, GSE5633 and GSE5634. The expression profile data of rice PME and PMEI families were downloaded from the CREP database (<http://crep.ncpgr.cn>). Subsequent data processing was identical to former research [47].

Network Assembly and Functional Enrichment

In this study, the *Arabidopsis* PME and PMEI genes were submitted to the Arabidopsis Network Analysis Pipeline (ANAP) [48], which effectively integrated 11 publicly available *Arabidopsis* network databases. The functional enrichment

analysis of the genes involved in the PME and PMEI networks were conducted using Blast2GO software with the molecular function category of level three [49].

Results

Genome-wide Identification of PME and PMEI Genes

Through the genome-wide identification of PME and PMEI domains encoding genes, the number of the two families was summarized in Figure 1 and Table S1. The results show that only 15 PME genes exist in *C. orbicula*, a representative member of the charophytes that diverged after pectin appeared in the cell walls. There were 35 and 18 PME genes detected in the bryophyte *P. patens* and lycophyte *S. moellendorffii*, but only 12 and 5 proPME genes and one PMEI gene were detected in the two species, respectively. Furthermore, the results show that these two gene families widely appear in the basal angiosperms *A. trichopoda* and also the monocot and eudicot species. The PME family copy numbers in *Arabidopsis* are identical to former research, but 6 more in rice and 13 less in poplar [17], which may be due to the update of genome annotations.

Gene Family Expansion

The WGD and tandem duplication of the PME and PMEI genes were analyzed to study the gene family expansion. After an in-depth analysis of the WGD information [27,28], 5 PME genes of *Arabidopsis* were derived from the α duplication event, 21 and 7 proPME genes from the α and β events, respectively, 20 and 12 PMEI genes from the α and β events, respectively (Table 1). In grape, 3 PME and 3 proPME genes were derived from the γ duplication event. In the monocot rice, we found that 4 and 7 PME genes, 7 and 9 proPME genes, and 2 and 14 PMEI genes were derived separately from the σ and ρ duplication events, respectively. These results show that the number of original PME genes was relatively conserved, but the species-specific WGD events contributed to the expansion of the proPME subfamily and PMEI family. A gene tandem duplication analysis of the two families showed that 15 of the 71 PMEI genes in *Arabidopsis* were tandem duplication genes, among which we detected a cluster of 7 members in chromosome 5 (AT5G46930, AT5G46940, AT5G46950, AT5G46960, AT5G46970, AT5G46980, and AT5G46990). Although there are 4 more PMEI genes in sorghum than rice, they have a similar number of PME and proPME genes. Through the gene duplication analysis, we found that the proPME and PMEI families rapidly expanded, and possibly, such substantial gene expansion in the land plants was driven by their functional specialization of cell wall formation (Figure 1).

Genome synteny of the PME and PMEI families was then analyzed. The results show that there is good synteny through the inter-genomic comparison within the eudicots and monocots, but only a few genes are syntenic between eudicots and monocots (Figure 2), which might be reason of the different ancestral genome organization, because eudicot genomes might originate from 7 chromosomes and monocot genomes might originate from 5 chromosomes [29].

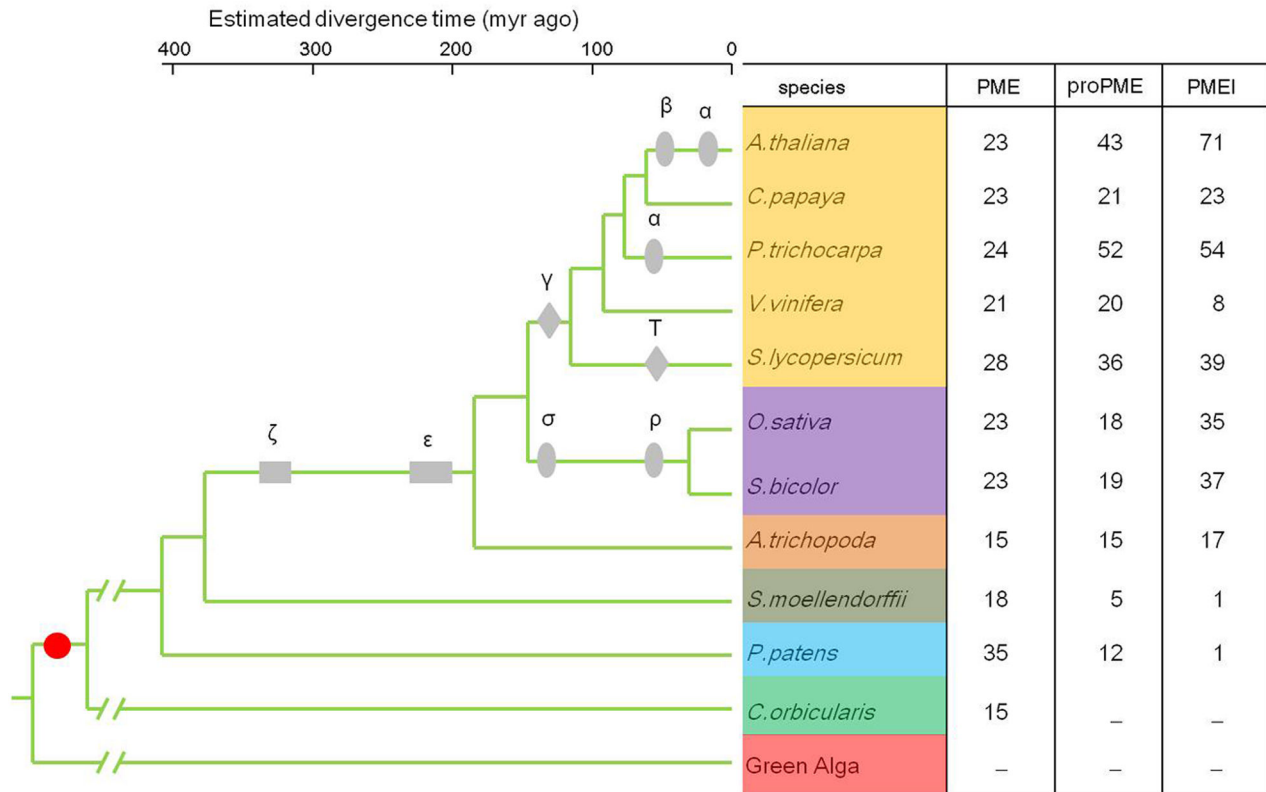


Figure 1. The copy number of PME and PME1 families in the collected species. The ultrametric tree was modified from Jiao et al. [23], Popper et al. [4] and Lee et al. [35]. The grey boxes (ζ , ϵ), diamonds (γ , T) and ellipses (α , β , σ , ρ) show the whole genome duplication events in the species, the red dot indicates when pectins appeared in the cell wall of organisms. -Indicates not detected.

doi: 10.1371/journal.pone.0072082.g001

Exon-intron Structure and Phylogenetic Analysis

Gene exon-intron structure analyses of the PME and PME1 families were conducted in 10 species (Table 2). Obviously, the average gene length of proPME is larger than PME. Compared to the PME and proPME genes, the PME1 genes are shortest and the average length is only several hundred base pairs. In addition, most PME genes possess more than 4 exons, except for monocots and the basal angiosperm *A. trichopoda*. The proPME genes contain only 2.55 exons on average (from 1.87 to 3.23), but the average exons length of proPME is at least twice that of PME. However, most genes in the PME1 family possess just one exon in all the species. With a graphical display, we depicted the PME gene structures of *Arabidopsis* (Figure 3). The results show that most PME domains of the PME genes consist of 3 exons. Conversely, most of the PME domains of the proPME genes possess only one intron. The pro and PME1 domains possess few introns (Figure S1).

Genome-wide gene identification indicates that angiosperm genomes have a relatively conserved number of PME genes (Figure 1). Domain sequence alignment of the PME genes shows that the PME domains are highly conserved in studied species (Figure S2). Then, Orthomcl software was used to detect the orthologous PME genes in the 11 genomes (Table S2) with the definition that genes in a cluster from at least 3

different species were orthologous. Compared to the species tree in Figure 1, the phylogenetic analysis of a well detected orthologous cluster shows good coordination (Figure S3). Based on the divergence time of the species tree, the divergence of the bryophyte mosses and charophyte *C. orbicula* could date back to 470 and 560 million years ago, which was slightly earlier than former results [50]. More accurate estimation should be performed with larger scale molecular data.

The phylogenetic analysis was conducted to study the evolutionary relationship of the pro and PME1 domains. Firstly, the neighbor-joining tree of the pro and PME1 domains of all the representative species shows that they are obviously clustered into two clades, except the 16 PME1 domains from 8 noncommelinid species are classified into the pro domain clade (Figure 4A). To verify this result, the HMMER package was then used to construct consensus sequences of the pro and PME1 domains in every species by selecting the maximum probability residue at each match state. The phylogenetic analysis reveals that the consensus domain tree is not completely in accord with the species tree (Figure S4). The main differences include the species divergence between *P. patens* and *S. moellendorffii* in the pro domain clade and the divergence between *A. trichopoda* and *V. vinifera* in the PME1

Table 1. A summary of the modes of PME, proPME and PME1 gene duplication in *A. thaliana*, *V. vinifera*, *O. sativa* and *S. bicolor*.

Species	Subfamily	Whole Genome		Other	Total
		Duplication	Tandem Duplication		
<i>A. thaliana</i>	PME	5 ^α	2	16	23
	proPME	7 ^β ; 21 ^α	6	9	43
	PME1	12 ^β ; 20 ^α	15	24	71
<i>V. vinifera</i>	PME	3 ^γ	2	16	21
	proPME	3 ^γ	4	13	20
	PME1	- ^{**}	3	5	8
<i>O. sativa</i>	PME	4 ^σ ; 7 ^ρ	2	10	23
	proPME	7 ^σ ; 9 ^ρ	2	0	18
	PME1	2 ^σ ; 14 ^ρ	6	13	35
<i>S. bicolor</i>	PME	4 ^σ ; 10 ^ρ	2	7	23
	proPME	6 ^σ ; 11 ^ρ	2	0	19
	PME1	4 ^σ ; 16 ^ρ	10	9	39

α and β indicate two recent *Arabidopsis* duplication events, and γ indicates the triplication event that all eudicots shared. σ and ρ indicate the duplications that rice and sorghum shared.

* Gene modes of proximal duplication, DNA-based transposed duplication, retrotransposed duplication and dispersed duplication are included.

** No genes detected in the relative mode.

domain clade. Although the two domains are attributed to each clade, further sequence alignment shows that they shared three conserved motifs, AL[KE] DCLEL[LY] [DS]D[AS] [VL] DELK, TW[LV] SAALT[DN] [QA] [DE] TC[LE] DG[FL] and LTSN[AS] LAL (Figure 4B), which indicates that pro and PME1 domains may have similar evolutionary origin.

Strong Selection of the pro and PME1 Domains

To assess the potential selective forces of the evolving domains, the PAML software was used to calculate the Ka and Ks values of each pair in the three domains. The result of Ka/Ks analysis reveals that the PME and pro domains experienced purifying selection, but the pro domain showed stronger selection than the original PME domain (Figure 5). However, a large portion of the PME1 domains in the eudicot species (tomato, *Arabidopsis* and poplar) experienced positive selection. We observed that most offspring of PME1 genes from the recent α and β WGD events experienced positive selection in *Arabidopsis*, implying that they were quickly expanding. In the monocot rice and sorghum, all three domains experienced purifying selection, thus indicating a relatively different evolutionary pattern between eudicots and monocots.

Comparative Expression Profile Analysis between *Arabidopsis* and Rice

Through expression data mining of the PME and PME1 families in the public databases, we presented the expression profiles of 63 tissue samples of *Arabidopsis* and 27 tissue

samples of rice. Based on hierarchical clustering, the expression patterns in *Arabidopsis* can be clustered into six groups. The genes in group A are mainly expressed in leaves, shoots and seeds, the genes in group B are mainly expressed in roots and seeds, the genes in group C are mainly expressed in roots, the majority of genes in group D and E are expressed in vegetable tissues, and the genes in group F are specifically expressed in flower-related organs [51] (Figure S5a). In addition, the results show that all six groups consist of the PME, proPME and PME1 genes, thus indicating that although there are different expression patterns in different tissues, the proportional distribution of the genes in the two families shows no obvious bias in any group. By contrast, the expression profile in rice shows that most genes of the three clusters are not expressed specifically in tissues (Figure S5b), which may reflect the differences of pectin composition in primary cell walls between the eudicots and monocots.

A Network Analysis of the *Arabidopsis* PME and PME1 Families

Because the genes involved in a biological pathway typically express cooperatively [52], the PME and PME1 gene networks were constructed to further analyze the pectin metabolic system based on the pipeline of ANAP. Eventually, 26 of 67 PME genes in the ANAP database involves of 257 unique genes exhibiting 514 interactions, including co-expression and protein-protein interactions (Figure 6, Table S3). A further functional enrichment analysis reveals that these genes are mainly involved in the molecular function of transferase activity, ion binding, and hydrolase activity (Figure S6). In detail, among the 514 interactions, ADF11 and expansion A7 and A18 are directly related to cell wall formation [53,54], which obviously correspond to the function of the PME gene of pectin metabolism. Fourteen galacturonosyltransferases involved in pectin biosynthesis, and 16 pectin lyase-like genes involved in the degradation of demethyl-esterified pectin, are also detected in this network, thus indicating that complex pectin synthesis-degradation system related genes may act together temporally. Plant pathogen resistance genes are vital to plant development, and the network analysis also reveals that PME genes may function together with pathogen resistant proteins through the detection of 4 CAP (Cysteine-rich secretory proteins, Antigen 5, and Pathogenesis-related 1 protein) genes, consistent with the expression and network analysis of the pectin lyase-like gene family in *Arabidopsis* [55]. However, only 10 of 71 PME1 genes are present in the database, which consist of 34 interactions of 33 unique genes (Figure S7).

Discussion

This study focused on the computational identification, gene duplication, evolution and functional analysis of the PME and PME1 families in fully or partially sequenced plant and algal genomes. We have found several important features of the relationship between the pectin demethyl-esterification process and the cell wall evolution.

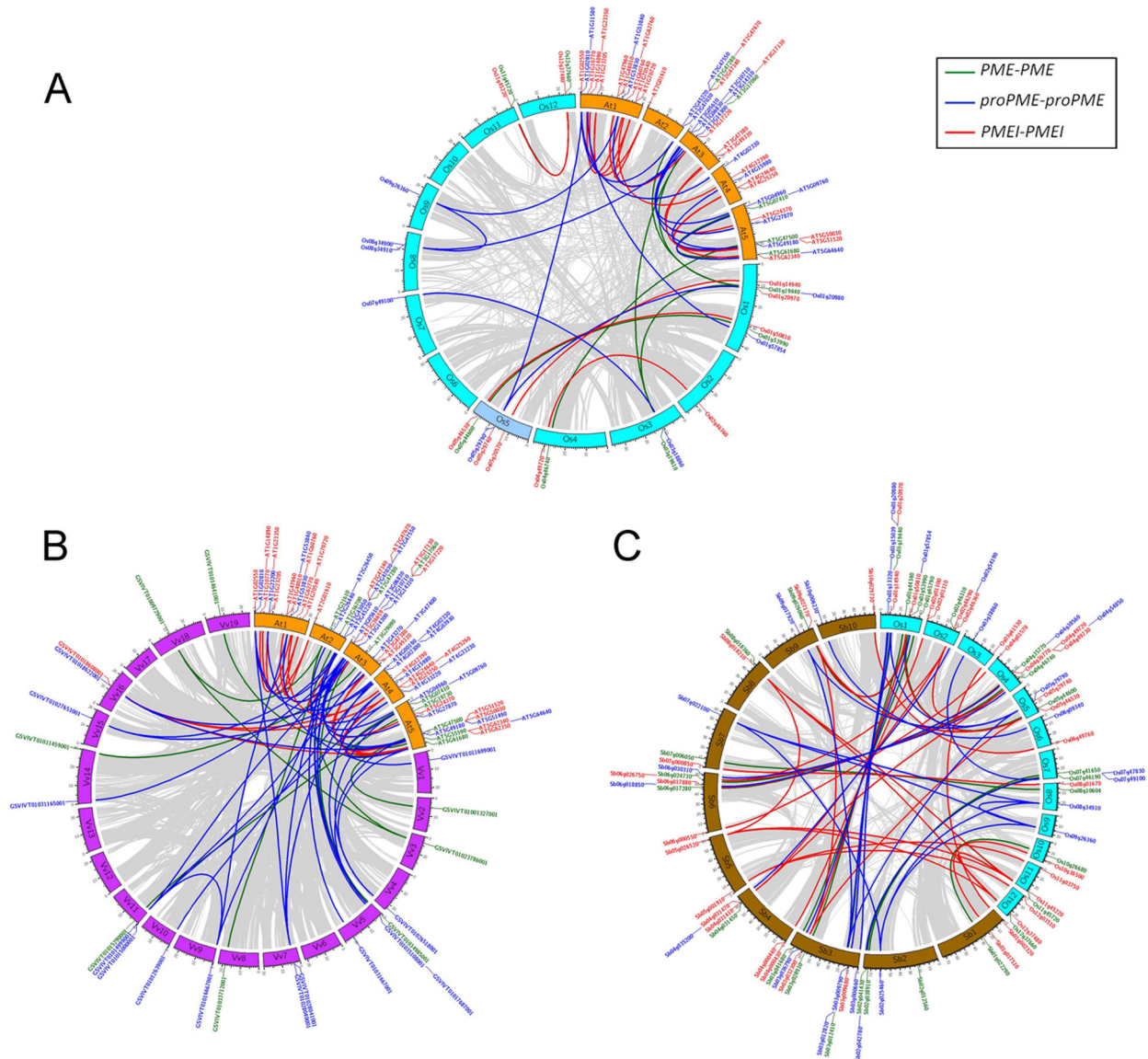


Figure 2. The intra- and inter-genomic comparison analyses showed gene synteny of the PME and PME I families in *A. thaliana* (At), *V. vinifera* (Vv), *O. sativa* (Os) and *S. bicolor* (Sb). The synteny gene pairs were parsed from the Plant Genome Duplication Database (PGDD). The gray lines indicate whole genome duplication blocks between species, the green lines indicate the synteny of PME genes, the purple lines indicate the synteny of proPME genes, and the red lines indicate the synteny of PME I genes. (A) gene synteny of PME and PME I families in *A. thaliana* (At) and *O. sativa* (Os), (B) gene synteny of PME and PME I families in *A. thaliana* (At) and *V. vinifera* (Vv), and (C) gene synteny of PME and PME I families in *O. sativa* (Os) and *S. bicolor* (Sb).

doi: 10.1371/journal.pone.0072082.g002

The Origin of the PME and PME I Gene Families

Genome-wide gene identification has shown that PME genes can be identified after the appearance of pectins in cell walls (Figure 1). To improve the resolution of the evolutionary detection of the PME domain origin, we attempted to identify PME and PME I gene families in chlorophyta (*C. reinhardtii* and *V. carteri*) [56,57], rhodophyta (*C. merolae*) [58], and diatoms (*P. tricornutum*) [59], but no such genes could be found.

Although former research has identified PME genes in bacteria, the evolutionary relationship between land plants and bacteria remains unclear [60]. Therefore, we consider that plant PME genes may have originated from charophytes. Notably, this origin coordinates with the appearance of pectin in cell walls.

The identification of 15 PME genes in *C. orbicula* indicates that they appeared first evolutionarily in comparison with proPME and PME I genes. The proPME appeared after the

Table 2. A summary of gene length, exon number, exon length and intron length of the PME, proPME, and PMEI in representative species.

Species	Average Gene Length			Average Exons per Gene			Average Exon Length			Average Intron Length ^a	
	PME	proPME	PMEI	PME	proPME	PMEI	PME	proPME	PMEI	PME	proPME
<i>P. patens</i>	2330	2692	-	4.28	3.23	-	273	546	-	666	688
<i>S. moellendorffii</i>	1284	2300	-	4.41	3	-	217	579	-	413	338
<i>A. trichopoda</i>	3159	2860	562	3.73	1.87	1.06	213	658	517	1027	2122
<i>S. bicolor</i>	1794	2917	878	3.35	2.05	1.16	327	872	561	667	2122
<i>O. sativa</i>	2623	3065	938	3.47	1.89	1.12	322	958	559	868	1709
<i>S. lycopersicum</i>	2683	3145	686	4	2.58	1.13	257	665	507	524	612
<i>V. vinifera</i>	2576	3052	960	4.94	2.93	1.33	248	434	562	558	1124
<i>P. trichocarpa</i>	2550	2654	885	4.63	2.25	1.09	228	747	556	454	1355
<i>C. papaya</i>	2600	2896	851	4.74	3.07	1.15	246	349	616	618	691
<i>A. thaliana</i>	1933	2476	817	4.67	2.67	1.25	227	648	470	429	942
Mean Value ^b	2353	2805	822	4.22	2.55	1.16	255	645	543	622	1170

^a The intron length of the PMEI genes were not shown because many of them do not have introns.

^b The mean values of the gene length, exon number, exon length and intron length in the 10 species.

Indicated only one PMEI gene was detected in *P. patens* and *S. moellendorffii*, and the structural information was not shown here.

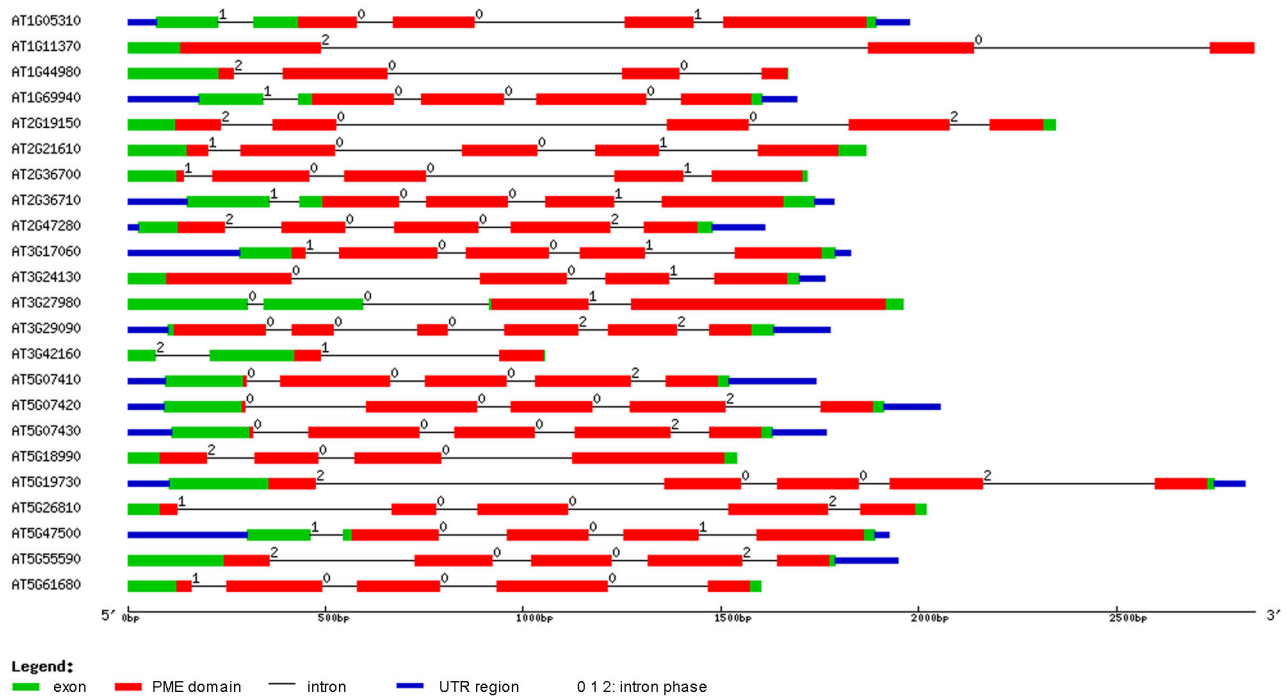


Figure 3. The exon-intron structural analysis of the 23 PME genes in *Arabidopsis*. The gene structures were drawn using the online tool GSDS. The legend shows that the blue boxes are UTR regions, the green boxes are exons, the black lines are introns, the red boxes are the PME domains, and numbers at the exon-intron joints are intron phases.

doi: 10.1371/journal.pone.0072082.g003

divergence of the bryophytes (Figure 1). These results suggest that PME genes occurred at an early stage of eukaryotes' evolutionary history, and the domain fusion event between PME and pro domains possibly occurred after the divergence of charophytes. Compared with the PME and proPME genes, the PMEI family appeared later, most likely because of the

gene neofunctionalization after the ζ and ε WGD events. In fact, the gene family synteny analysis shows that WGD and tandem duplication contributed to the expansion of the proPME and PMEI families in angiosperm species. Because both of PME and proPME genes possess the function of demethyl-esterification, further study should be carried to clarify the

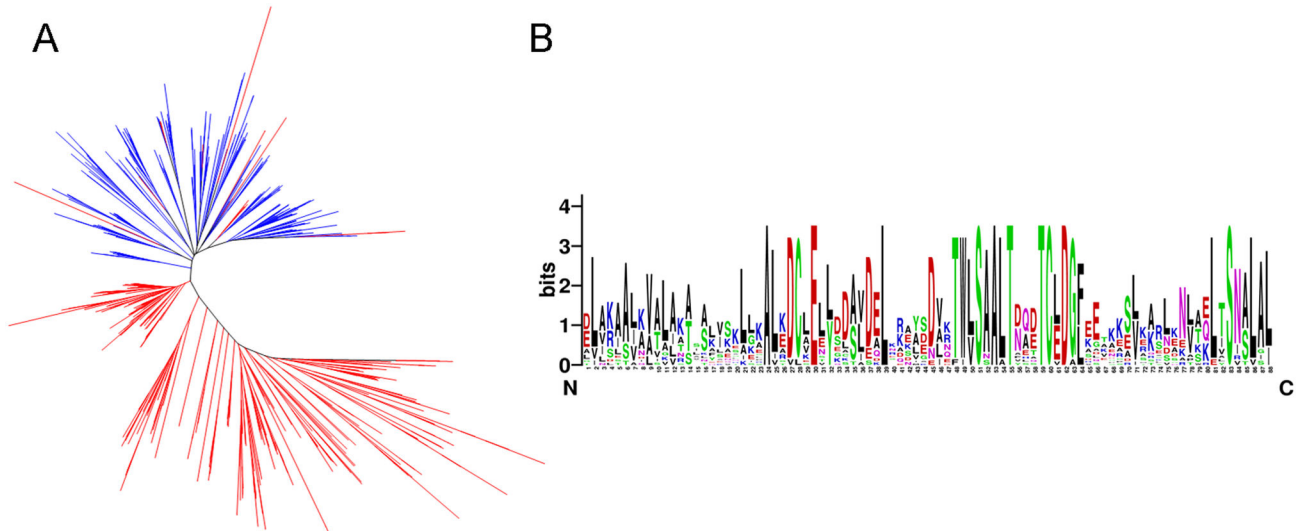


Figure 4. The molecular phylogenetic analysis of the orthologous PME genes, and pro and PMEI domains. (A) The phylogenetic analysis of the pro and PMEI domains in 10 species using the neighbor joining method. The blue lines represent the pro domain, and the red lines represent the PMEI domain. (B) The consensus sequence alignment shows the conserved motifs of the pro and PMEI domains in the representative species. The MEME program was then used to verify the conserved motifs.

doi: 10.1371/journal.pone.0072082.g004

relationship between gene family expansion and the neofunctionalization of proPME genes. Our results indicate that there are two turning points of the pectin methyl-esterification process: the initial point was the divergence of moss from charophytes represented by the appearance of proPME, and the second point was the appearance of the PMEI family in land plants, which contributed to the complexity of the methyl-esterification process.

A Putative Mechanism of PME, pro, and PMEI Domain Origin

Considering the sequence similarities between pro and PMEI domains, there must be some unclarifying evolutionary clues between them. Determining how the new pro domain emerged will be crucial to illustrate the evolutionary history of the PME and PMEI families.

Through the gene structure analysis, we observe that a large number of introns are lost after the domain fusion event by comparing the PME domain of proPME with the old PME genes (Figure S1). However, the origin of PME domain remains unknown. More comparative genomic studies should be performed with the genome sequencing of some important micro-algae. Notably, we find that the pro domains of two proPME genes in *Arabidopsis* (AT1G23200 and AT4G15980) show good sequence similarities with the introns of two original type II PME genes (AT1G44980 and AT3G27980). This observation may provide basic clues for an evolutionary interpretation of the pro domain origin (data not shown). The phylogenetic analysis and sequence alignment indicate that the PMEI domain may have originated from the pro domain, based on the sequence similarities, domain structures, and domain lengths and guided by genomic recombination or transposon-

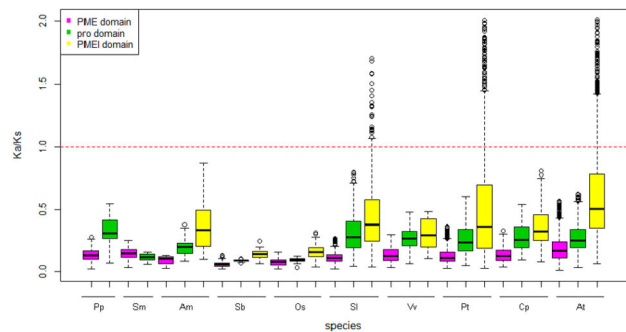


Figure 5. The Ka/Ks value distributions of the PME, pro and PMEI domains in 10 species. The red broken line indicates that genes are under positive selection (more than one) or negative selection (less than one). The short species names are *P. patens* (Pp), *S. moellendorffii* (Sm), *A. trichopoda* (Am), *S. bicolor* (Sb), *O. sativa* (Os), *S. lycopersicum* (Sl), *V. vinifera* (Vv), *P. trichocarpa* (Pt), *C. papaya* (Cp) and *A. thaliana* (At). The Ka and Ks values were computed using the PAML program.

doi: 10.1371/journal.pone.0072082.g005

based recombination (Figure 4) [61,62]. The selection pressure analysis of the pro and PMEI domains indicates that they experienced strong positive selection, namely, they have rapidly evolved, particularly in tomato, poplar and *Arabidopsis*, in which lineage-specific whole genome duplication events have largely contributed to the expansion of the proPME and PMEI families (Figure 5).

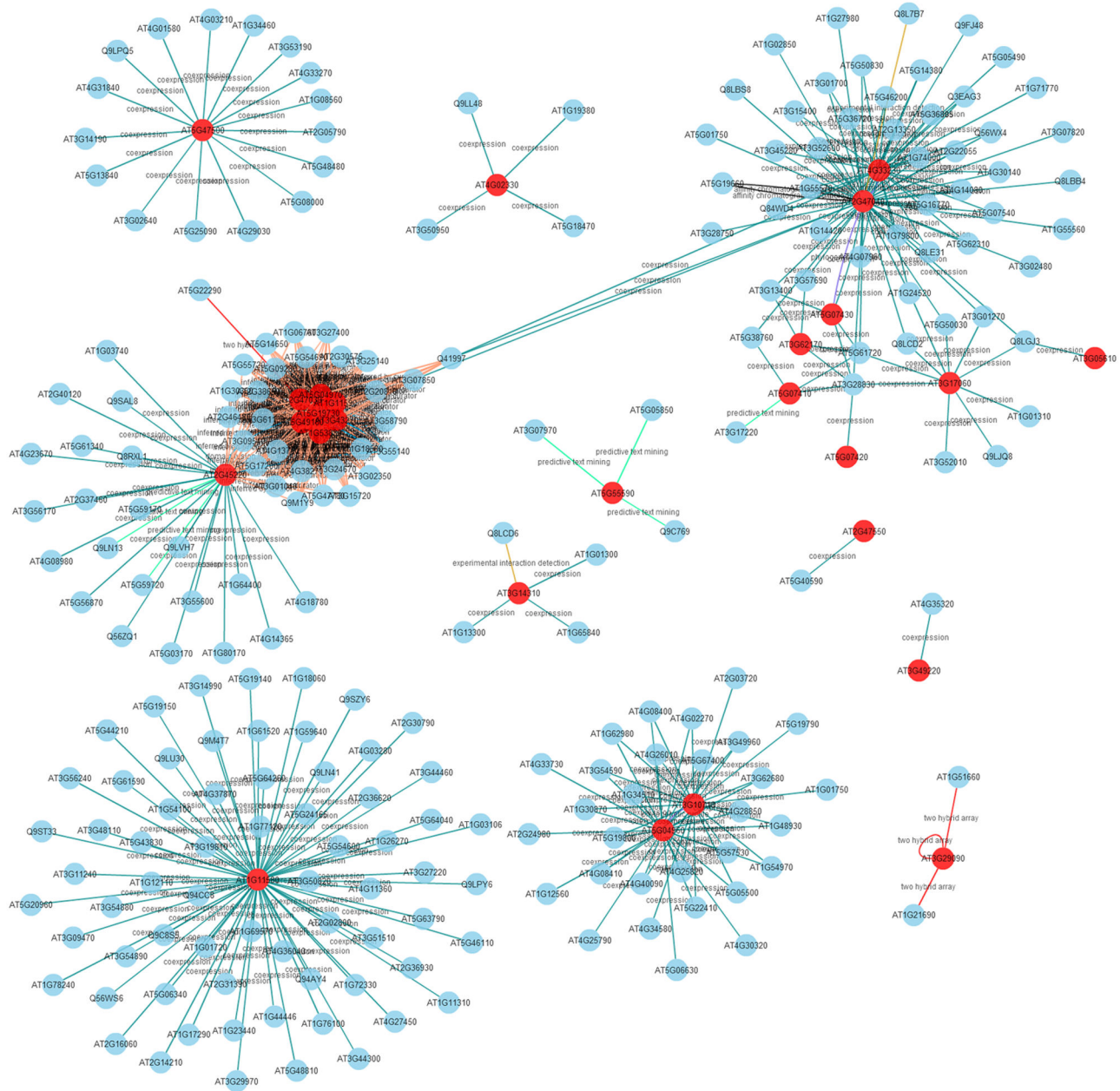


Figure 6. The Arabidopsis PME gene network. Twenty-six PME genes in *Arabidopsis* were mapped to the ANAP database. This network involves 257 unique genes exhibiting 514 interactions. The red nodes represent the PME genes.
doi: 10.1371/journal.pone.0072082.g006

Our model raised the possibility of domain origin through computer data mining, and additional research should be presented to further understand the evolution of the three domains. Meanwhile, an in-depth domain functional study can also clarify the evolutionary relationship of the pro and PME1 domains.

Pectin Metabolism and Cell Wall Evolution

The origin and early evolution of the land plants provide good opportunities for research on plant cell wall evolution, although there are no very good structural models suggested currently [63]. New biophysical and visualization methods are necessary to understand the wall organization of components in a single cell [2,64]. A component analysis has summarized the basic metabolite appearances using uni- and multicellular algae, mosses, basal angiosperms, eudicots and monocots. Cellulose

and cellulose synthases genes originated at the divergence of rhodophyta from glaucophyta [65]. Homogalacturonan (HG), a polymer of pectin, is widely present in the cell walls of uni- and multicellular species, which is frequently the case in embryophytes and land plants. Rhamnogalacturonan II (RGI), also a polymer, was initially observed in embryophytes and the content generally increased during the evolution of vascular plants, which is a trend that may satisfy the formation of lignified secondary walls but not in the monocot species. PME and PMEI genes spatially regulate the methyl-esterification of pectin polymers. In fact, pectin polysaccharides are complex in the cell wall and involved in large families of related synthesis and degradation genes. In addition to the PME and PMEI families, the *Arabidopsis* genome also encodes approximately 65 glycosyltransferases [6], 67 methyltransferases, and 67 pectin lyases [57]. The complex pectin metabolism may be related to functional diversity. Research has shown that demethyl-esterification of pectins in cell walls is related to organ initiation in *Arabidopsis* [66]. In this study, the functional network also shows that PME genes may be involved in plant-pathogen interactions and affect plant resistance to diseases (Figure 6). This result may be related to the methyl-esterification of pectins during plant-pathogen interactions [67]. Although the exact pathway and interaction network mediated by these genes remains unclear, we can speculate on and expect these putative mechanisms for further functional study.

Overall, these results generate a new insight into the genic evidence of pectin divergence and how these genes evolved in algae and land plants from the perspective of bioinformatics. Meanwhile, expression profiling and functional network analysis in model species may help us better understand the possible molecular mechanisms of the primary cell wall biosynthesis.

Supporting Information

Figure S1. The exon-intron structural analysis of the proPME and PMEI family in *Arabidopsis*.

The legend shows that the blue boxes are the UTR regions, the green boxes are exons, the black lines are introns, the red boxes are the PME, pro and PMEI domains, and the numbers at the exon-intron joints are the intron phases. (A) The exon-intron structural analysis showed the PME domain of the proPME. (B) The exon-intron structural analysis shows the pro domain of the proPME. (C) The exon-intron structural analysis shows the PMEI domain of the PMEI in *Arabidopsis*.

(TIF)

Figure S2. The consensus sequence alignment showed the conserved motifs of the PME domain in representative species.

HMMER package was used to trim the consensus sequence of the PME domain in the eleven species, and the NOGAPS sequence alignment was retained. MEME program was used to validate the conserved motifs.

(TIF)

Figure S3. The molecular phylogenetic analysis of the orthologous PME genes from eleven species.

The ortholog gene cluster was initially identified from the output of Orthomcl software and verified by a subsequent single cluster phylogeny analysis. The molecular clock test was performed by comparing the ML value for the given topology with and without the molecular clock constraints under the WAG model. The null hypothesis of an equal evolutionary rate throughout the tree was rejected at a 5% significance level. The evolutionary analysis was conducted in MEGA5.

(TIF)

Figure S4. The molecular phylogenetic analysis of the pro and PMEI domains using the maximum likelihood method.

The evolutionary history was inferred by using the Maximum Likelihood method based on the WAG model with 500 replications for the bootstrapping test. Each branch represents the consensus domain sequences trimmed using HMMER.

(TIF)

Figure S5. Expression profiling of the PME, proPME and PMEI families in *Arabidopsis* and rice.

The uppercase-lowercase 'At' and 'Os' are the PME genes, the lowercase-lowercase 'at' and 'os' are the proPME genes, and the uppercase-uppercase 'AT' and 'OS' are the PMEI genes. (a) The co-expression profile of *Arabidopsis* PME and PMEI families in 63 tissue samples. These genes were divided into 6 groups using the complete linkage clustering method. (b) Co-expression profiling of the rice PME and PMEI families in 27 tissue samples. The M indicates rice variety Minghui 36, and Z is variety Zhenshan 97.

(TIF)

Figure S6. Functional enrichment of the genes involved in the *Arabidopsis* PME gene network.

The GO enrichments were performed with Blast2GO software, and the molecular function category of level three was analyzed.

(TIF)

Figure S7. The *Arabidopsis* PMEI gene network.

This network involves 34 unique genes exhibiting 33 interactions. The red nodes represent the PMEI genes.

(TIF)

Table S1. The genome-wide identification of the PME, proPME and PMEI genes.

This table shows genes from 11 species, *C. orbicular* (Co), *P. patens* (Pp), *S. moellendorffii* (Sm), *A. trichopoda* (Am), *V. vinifera* (Vv), *C. papaya* (Cp), *P. trichocarpa* (Pt), *A. thaliana* (At), *S. lycopersicum* (Sl), *O. sativa* (Os) and *S. bicolor* (Sb).

(XLSX)

Table S2. The identified orthologous groups of the PME family genes in 11 representative species.

The orthologous genes were defined as genes in a cluster from at least three species. This analysis was conducted using Orthomcl software. (XLSX)

Table S3. The annotation summary of the genes involved in the PME network.

The annotation information was downloaded from TAIR (The *Arabidopsis* Information Resource, <http://www.arabidopsis.org/>). (XLSX)

Acknowledgements

We are grateful to Dr. Lingqiang Wang (National Key Laboratory of Crop Genetic Improvement, Biomass and

References

- Fry SC (2004) Primary cell wall metabolism: tracking the careers of wall polymers in living plant cells. *New Phytol* 161: 641-675. doi:10.1111/j.1469-8137.2004.00980.x.
- Keegstra K (2010) Plant cell walls. *Plant Physiol* 154: 483-486. doi:10.1104/pp.110.161240. PubMed: 20921169.
- Somerville C, Bauer S, Brininstool G, Facette M, Hamann T et al. (2004) Toward a systems approach to understanding plant cell walls. *Science* 306: 2206-2211. doi:10.1126/science.1102765. PubMed: 15618507.
- Sørensen I, Pettolino FA, Bacic A, Ralph J, Lu F et al. (2011) The charophycean green algae provide insights into the early origins of plant cell walls. *Plant J* 68: 201-211. doi:10.1111/j.1365-313X.2011.04686.x. PubMed: 21707800.
- Popper ZA, Michel G, Hervé C, Domozych DS, Willats WG et al. (2011) Evolution and diversity of plant cell walls: from algae to flowering plants. *Annu Rev Plant Biol* 62: 567-590. doi:10.1146/annurev-arplant-042110-103809. PubMed: 21351878.
- Caffall KH, Mohnen D (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res* 344: 1879-1900. doi:10.1016/j.carres.2009.05.021. PubMed: 19616198.
- Mohnen D (2008) Pectin structure and biosynthesis. *Curr Opin Plant Biol* 11: 266-277. doi:10.1016/j.pbi.2008.03.006. PubMed: 18486536.
- Harholt J, Suttangkakul A, Vibe Scheller H (2010) Biosynthesis of Pectin. *Plant Physiol* 153: 384-395. doi:10.1104/pp.110.156588. PubMed: 20427466.
- Sterling JD, Quigley HF, Orellana A, Mohnen D (2001) The catalytic site of the pectin biosynthetic enzyme alpha-1,4-galacturonosyltransferase is located in the lumen of the Golgi. *Plant Physiol* 127: 360-371. doi:10.1104/pp.127.1.360. PubMed: 11553763.
- Giovane A, Servillo L, Balestrieri C, Raiola A, D'Avino R et al. (2004) Pectin methylesterase inhibitor. *Biochim Biophys Acta* 1696: 245-252. doi:10.1016/j.bbapap.2003.08.011. PubMed: 14871665.
- Di Matteo A, Giovane A, Raiola A, Camardella L, Bonivento D et al. (2005) Structural basis for the interaction between pectin methylesterase and a specific inhibitor protein. *Plant Cell* 17: 849-858. doi:10.1105/tpc.104.028886. PubMed: 15722470.
- Juge N (2006) Plant protein inhibitors of cell wall degrading enzymes. *Trends Plant Sci* 11: 359-367. doi:10.1016/j.tplants.2006.05.006. PubMed: 16774842.
- de Freitas ST, Handa AK, Wu Q, Park S, Mitcham EJ (2012) Role of pectin methylesterases in cellular calcium distribution and blossom-end rot development in tomato fruit. *Plant J* 71: 824-835. doi:10.1111/j.1365-313X.2012.05034.x. PubMed: 22563738.
- Hongo S, Sato K, Yokoyama R, Nishitani K (2012) Demethylesterification of the primary wall by PECTIN METHYLESTERASE35 provides mechanical support to the *Arabidopsis* stem. *Plant Cell* 24: 2624-2634. doi:10.1105/tpc.112.099325. PubMed: 22693281.
- Yadav S, Yadav PK, Yadav D, Yadav KDS (2009) Pectin lyase: A review. *Proc Biochem* 44: 1-10. doi:10.1016/j.procbio.2008.09.012.
- Wolf S, Mouille G, Pelloux J (2009) Homogalacturonan methylesterification and plant development. *Mol Plant* 2: 851-860. doi:10.1093/mp/ssp066. PubMed: 19825662.
- Pelloux J, Rustérucci C, Mellerowicz EJ (2007) New insights into pectin methylesterase structure and function. *Trends Plant Sci* 12: 267-277. doi:10.1016/j.tplants.2007.04.001. PubMed: 17499007.
- Jolie RP, Duvetter T, Van Loey AM, Hendrickx ME (2010) Pectin methylesterase and its proteinaceous inhibitor: a review. *Carbohydr Res* 345: 2583-2595. doi:10.1016/j.carres.2010.10.002. PubMed: 21047623.
- Wolf S, Rausch T, Greiner S (2009) The N-terminal pro region mediates retention of unprocessed type-I PME in the Golgi apparatus. *Plant J* 58: 361-375. doi:10.1111/j.1365-313X.2009.03784.x. PubMed: 19144003.
- Bosch M, Cheung AY, Hepler PK (2005) Pectin methylesterase, a regulator of pollen tube growth. *Plant Physiol* 138: 1334-1346. doi:10.1104/pp.105.059865. PubMed: 15951488.
- Carvunis AR, Rolland T, Wapinski I, Calderwood MA, Yildirim MA et al. (2012) Proto-genes and de novo gene birth. *Nature* 487: 370-374. doi:10.1038/nature11184. PubMed: 22722833.
- Kaessmann H (2010) Origins, evolution, and phenotypic impact of new genes. *Genome Res* 20: 1313-1326. doi:10.1101/gr.101386.109. PubMed: 20651121.
- Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. *Nat Rev Genet* 12: 692-702. doi:10.1038/nrg3053. PubMed: 21878963.
- Wang Y, Wang X, Tang H, Tan X, Ficklin SP et al. (2011) Modes of gene duplication contribute differently to genetic novelty and redundancy, but show parallels across divergent angiosperms. *PLOS ONE* 6: e28150. doi:10.1371/journal.pone.0028150. PubMed: 22164235.
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L et al. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* 473: 97-100. doi:10.1038/nature09916. PubMed: 21478875.
- Fawcett JA, Maere S, Van de Peer Y (2009) Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. *Proc Natl Acad Sci U S A* 106: 5737-5742. doi:10.1073/pnas.0900906106. PubMed: 19325131.
- Bowers JE, Chapman BA, Rong J, Paterson AH (2003) Unravelling angiosperm genome evolution by phylogenetic analysis of chromosomal duplication events. *Nature* 422: 433-438. doi:10.1038/nature01521. PubMed: 12660784.
- Tang H, Bowers JE, Wang X, Paterson AH (2010) Angiosperm genome comparisons reveal early polyploidy in the monocot lineage. *Proc Natl Acad Sci U S A* 107: 472-477. doi:10.1073/pnas.0908007107. PubMed: 19966307.
- Abrouk M, Murat F, Pont C, Messing J, Jackson S et al. (2010) Palaeogenomics of plants: synteny-based modelling of extinct ancestors. *Trends Plant Sci* 15: 479-487. doi:10.1016/j.tplants.2010.06.001. PubMed: 20638891.
- Finet C, Berne-Dedieu A, Scutt CP, Marlétaz F (2013) Evolution of the ARF gene family in land plants: old domains, new tricks. *Mol Biol Evol* 30: 45-56. doi:10.1093/molbev/mss220. PubMed: 22977118.
- Li C, Li QG, Dunwell JM, Zhang YM (2012) Divergent evolutionary pattern of starch biosynthetic pathway genes in grasses and dicots. *Mol*

Bioenergy Research Centre, Huazhong Agricultural University) for providing the normalized *Arabidopsis* microarray data.

Author Contributions

Conceived and designed the experiments: DY XZ. Performed the experiments: MW YL JT. Analyzed the data: MW WG. Contributed reagents/materials/analysis tools: MW WG. Wrote the manuscript: MW. Proofread the manuscript and revised the manuscript: DY XZ.

- Biol Evol 29: 3227-3236. doi:10.1093/molbev/mss131. PubMed: 22586327.
32. Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD et al. (2012) Phytosome: a comparative platform for green plant genomics. *Nucleic Acids Res* 40: D1178-D1186. doi:10.1093/nar/gkr944. PubMed: 22110026.
 33. Zuccolo A, Bowers JE, Estill JC, Xiong Z, Luo M et al. (2011) A physical map for the *Amborella trichopoda* genome sheds light on the evolution of angiosperm genome structure. *Genome Biol* 12: R48. doi: 10.1186/1465-6906-12-S1-P48. PubMed: 21619600.
 34. Timme RE, Delwiche CF (2010) Uncovering the evolutionary origin of plant molecular processes: comparison of *Coleochaete* (Coleochaetales) and *Spirogyra* (Zygnematales) transcriptomes. *BMC Plant Biol* 10: 96. doi:10.1186/1471-2229-10-96. PubMed: 20500869.
 35. Eddy SR (2011) Accelerated Profile HMM Searches. *PLOS Comput Biol* 7: e1002195. PubMed: 22039361.
 36. Li L, Stoekert CJ Jr, Roos DS (2003) OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res* 13: 47-56. PubMed: 12952885.
 37. Lee TH, Tang H, Wang X, Paterson AH (2013) PGDD: a database of gene and genome duplication in plants. *Nucleic Acids Res* 41: D1152-D1158. doi:10.1093/nar/gks1104. PubMed: 23180799.
 38. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R et al. (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19: 1639-1645. doi:10.1101/gr.092759.109. PubMed: 19541911.
 39. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res* 34: W369-W373. doi:10.1093/nar/gkl198. PubMed: 16845028.
 40. Guo AY, Zhu QH, Chen X, Luo JC (2007) [GSDS: A gene structure display server]. *Yi Chuan* 29(8): 1023-1026. doi:10.1360/yc-007-1023. PubMed: 17681935.
 41. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA et al. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23: 2947-2948. doi:10.1093/bioinformatics/btm404. PubMed: 17846036.
 42. Felsenstein J (1989) PHYLIP -- Phylogeny Inference Package (Version 3.2). *Cladistics* 5: 164-166.
 43. Tamura K, Peterson D, Peterson N, Stecher G, Nei M et al. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731-2739. doi:10.1093/molbev/msr121. PubMed: 21546353.
 44. Whelan S, Goldman N (2001) A general empirical model of protein evolution derived from multiple protein families using a maximum-likelihood approach. *Mol Biol Evol* 18: 691-699. doi:10.1093/oxfordjournals.molbev.a003851. PubMed: 11319253.
 45. Suyama M, Torrents D, Bork P (2006) PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 34: W609-W612. doi:10.1093/nar/gkl315. PubMed: 16845082.
 46. Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* 24: 1586-1591. doi:10.1093/molbev/msm088. PubMed: 17483113.
 47. Wang L, Guo K, Li Y, Tu Y, Hu H et al. (2010) Expression profiling and integrative analysis of the *CESA/CSL* superfamily in rice. *BMC Plant Biol* 10: 282. doi:10.1186/1471-2229-10-282. PubMed: 21167079.
 48. Wang C, Marshall A, Zhang D, Wilson ZA (2012) ANAP: An Integrated Knowledge Base for Arabidopsis Protein Interaction Network Analysis. *Plant Physiol* 158: 1523-1533. doi:10.1104/pp.111.192203. PubMed: 22345505.
 49. Götz S, García-Gómez JM, Terol J, Williams TD, Nagaraj SH et al. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res* 36: 3420-3435. doi:10.1093/nar/gkn176. PubMed: 18445632.
 50. Zimmer A, Lang D, Richardt S, Frank W, Reski R et al. (2007) Dating the early evolution of plants: detection and molecular clock analyses of orthologs. *Mol Genet Genomics* 278: 393-402. doi:10.1007/s00438-007-0257-6. PubMed: 17593393.
 51. Jetz W, McPherson JM, Guralnick RP (2012) Integrating biodiversity distribution knowledge: toward a global map of life. *Trends Ecol Evol* 27: 151-159. doi:10.1016/j.tree.2011.09.007. PubMed: 22019413.
 52. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* 95: 14863-14868. doi:10.1073/pnas.95.25.14863. PubMed: 9843981.
 53. Ruzicka DR, Kandasamy MK, McKinney EC, Burgos-Rivera B, Meagher RB (2007) The ancient subclasses of Arabidopsis ACTIN DEPOLYMERIZING FACTOR genes exhibit novel and differential expression. *Plant J* 52: 460-472. doi:10.1111/j.1365-313X.2007.03257.x. PubMed: 17877706.
 54. Li Y, Darley CP, Ongaro V, Fleming A, Schipper O et al. (2002) Plant expansins are a complex multigene family with an ancient evolutionary origin. *Plant Physiol* 128: 854-864. doi:10.1104/pp.010658. PubMed: 11891242.
 55. Cao J (2012) The Pectin Lyases in *Arabidopsis thaliana*: Evolution, Selection and Expression Profiles. *PLOS ONE* 7: e46944. doi:10.1371/journal.pone.0046944. PubMed: 23056537.
 56. Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ et al. (2007) The Chlamydomonas genome reveals the evolution of key animal and plant functions. *Science* 318: 245-250. doi:10.1126/science.1143609. PubMed: 17932292.
 57. Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM et al. (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. *Science* 329: 223-226. doi:10.1126/science.1188800. PubMed: 20616280.
 58. Matsuzaki M, Misumi O, Shin-i T, Maruyama S, Takahara M, et al (2004) Genome sequence of the ultrasmall unicellular red alga *Cyanidioschyzon merolae* 10D. *Nature* 428: 653-657. doi:10.1038/nature02398. PubMed: 15071595.
 59. Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K et al. (2008) The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* 456: 239-244. doi:10.1038/nature07410. PubMed: 18923393.
 60. Markovic O, Janecek S (2004) Pectin methylesterases: sequence-structural features and phylogenetic relationships. *Carbohydr Res* 339: 2281-2295. doi:10.1016/j.carres.2004.06.023. PubMed: 15337457.
 61. Zhu Z, Zhang Y, Long M (2009) Extensive structural renovation of retrogenes in the evolution of the *Populus* genome. *Plant Physiol* 151: 1943-1951. doi:10.1104/pp.109.142984. PubMed: 19789289.
 62. Kaessmann H, Vinckenbosch N, Long M (2009) RNA-based gene duplication: mechanistic and evolutionary insights. *Nat Rev Genet* 10: 19-31. doi:10.1038/ni0109-19. PubMed: 19030023.
 63. Sarkar P, Bosneaga E, Auer M (2009) Plant cell walls throughout evolution: towards a molecular understanding of their design principles. *J Exp Bot* 60: 3615-3635. doi:10.1093/jxb/erp245. PubMed: 19687127.
 64. Haigler CH, Betancur L, Stiff MR, Tuttle JR (2012) Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Front Plant Sci* 3: 104. PubMed: 22661979.
 65. Fangel JU, Ulvskov P, Knox JP, Mikkelsen MD, Harholt J et al. (2012) Cell wall evolution and diversity. *Front Plant Sci* 3: 152. PubMed: 22783271.
 66. Peaucelle A, Braybrook SA, Le Guillou L, Bron E, Kuhlemeier C et al. (2011) Pectin-induced changes in cell wall mechanics underlie organ initiation in *Arabidopsis*. *Curr Biol* 21: 1720-1726. doi:10.1016/j.cub.2011.08.057. PubMed: 21982593.
 67. Lionetti V, Raiola A, Camardella L, Giovane A, Obel N et al. (2007) Overexpression of pectin methylesterase inhibitors in Arabidopsis restricts fungal infection by *Botrytis cinerea*. *Plant Physiol* 143: 1871-1880. doi:10.1104/pp.106.090803. PubMed: 17277091.