African Journal of Biotechnology Vol. 11(5), pp. 1106-1118, 16 January, 2012
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB11.773
ISSN 1684–5315 © 2012 Academic Journals

Full Length Research Paper

Genome-wide analysis of 1-amino-cyclopropane-1-carboxylate synthase gene family in *Arabidopsis*, rice, grapevine and poplar

Meihong Xu* and Ming-Hui Wang

School of Agriculture and Biology, Shanghai Jiao Tong University, Shanghai, 200240, People's Republic of China.

Accepted 21 October, 2011

Ethylene is an important hormone that is involved in a number of developmental processes and stress responses in higher plants. 1-Aminocyclopropane-1-carboxylate synthase (ACC) synthase as a key enzyme of the ethylene biosynthesis pathway, is in great need to be well studied. Though some efforts have been made to elucidate the structure, catalytic activity and evolutionary relationships of ACC synthase, there has not been a comprehensive study performed to statistically evaluate sequence conservation and functional divergence in this family. Our study was performed in four sequenced species, *Arabidopsis*, rice, grapevine and poplar, to determine the evolution mode in *ACS* gene family. Genome wide screening identified 12 *ACS* genes in *Arabidopsis*, six in rice, 10 in grapevine and 11 in poplar, while evolutionary pattern, site-specific dN/dS ratio tests, branch-site *dN/dS* ratio tests and diverge analysis were employed on their sequence. The results show that evolutionary pattern were quiet different in these four species, while strong purifying selection played the most important roles during evolution of the *ACS* gene family, and altered functional constraints may have taken place at some amino acid residues among main lineages (*A1, A2, A3 and B*). Besides, we mapped 11 residues that were probably involved in functional diverge of *ACS* family onto the sequence logo and three-dimensional structure of *Arabidopsis ACS1* to get more understanding of their functional diversity. In all, our results provide solid information for better understanding of the function and evolution of the *ACS* gene family in higher plants.

Key words: ACC synthase, phylogenetic, positive selection, functional divergence.

INTRODUCTION

Ethylene, as a plant hormone, is involved in the growth, development and stress-related processes of higher plants, including germination, ripening, auxin treating, wounding and anaerobiosis (Gray et al., 1992; Theologis, 1992; Yang and Hoffman., 1984; Kende, 1993). Ethylene biosynthesis depends on two main enzymes, 1-aminocyclopropane-1-carboxylate (ACC) synthase and 1-aminocyclopropane-1-carboxylate (ACC) oxidase. ACC synthase catalyzes the S-adenosylmethionine (SAM) to ACC; and ACC oxidase oxidizes ACC to ethylene (Kende, 1993, 1989). Both the enzymes belong to multigene families in various plants (Kende, 1993; Zarembinski and Theologis., 1994; Fluhr et al., 1996).

ACC synthase is a member of the PLP-dependent enzyme family, and its activity requires pyridoxal-5'-phosphate (PLP or vitamin B6) as the cofactor (Yip et al., 1990; Huai et al., 2001a). It was first identified from tomato (Boller et al., 1979), then a lot of ACC synthase genes were identified and cloned among various plant species, such as zucchini, *Arabidopsis*, rice, wheat, winter squash, apple, potato, pear, banana, carnation, mungbean, rose (Johnson et al., 1998; Ge et al., 2000) etc. Though ACC synthases are a divergent multigene family, their primary structure shares similar molecular size (441 to 496 amino acids) and contains seven highly conversed regions (Dong et al., 1991; Wong et al., 1999). Among them, the region 5 is the PLP-binding site which is necessary for attachment to the *ACS* family. The N-termini of ACC synthase has two highly conserved

*Corresponding author. E-mail: xmh-szh@sjtu.edu.cn. Tel : +86-21-34206936.

leucine and serine residues except subclade *A3*; while the C-termini is quite divergent and has a hyper-variable region consisting by 18 ~ 85 residues (Wong et al., 1999; Jakubowicz, 2002). ACC synthase is found to be active only in higher plants and fungus *Penicillium citrinum* (Kakuta et al., 2001). Normally, the microorganism generates ethylene from 2-oxoglutarate or from 2-oxo-4-methylthiobutyrate (Jia et al., 1999), but the ACC synthase was detected in fungus *P. citrinum* showing a low level of ethylene (Kakuta et al., 2001). Homologues of *ACS* were also identified in Japanese puffer fish (*Fugu rubripes*) and human, however these genes did not convert SAM to ACC (Peixoto et al., 2000; Koch et al., 2001). Moreover, the *ACS* amino acid identity of human with *F. rubripes* is higher than with plants (Jakubowicz and pacak., 2004).

Moreover, some efforts have been made to elucidate the structure, catalytic activity and evolutionary relationships of ACC synthase (Jakubowicz et al., 2002, 2004). However, there has been no comprehensive study performed to statistically evaluate sequence conservation and functional divergence in this family. Completion of the complete genome sequencing of *Arabidopsis* (Initiative 2000), rice (Matsumoto et al., 2005; Yu et al., 2002; Goff et al., 2002) grapevine (Jaillon et al., 2007) and poplar (Tuskan et al., 2006) would help us to better understand the *ACS* genes in these four plant species and better investigate the expansion patterns of this family.

In our study, 12, 6, 10 and 11 *ACS* genes were identified in *Arabidopsis*, rice, grapevine and poplar, respectively. A phylogenetic tree was constructed to evaluate the evolutionary relationships of *ACS* genes in the four plant species. Chromosomal distribution showed that *ACS* genes are dispersed throughout the respective genomes of the four species. Segmental duplication, tandem duplication, and transposition events are considered to be three main reasons that contribute to gene family evolution. As *ACS* gene family in the four species, segmental depilation played a leading role in *Arabidopsis* and poplar, and transposition events were more practical in rice and grapevine. To examine the driving force for duplicated genes, paralogous genes were used to calculate the non-synonymous (*dN*), synonymous (*dS*) and the ratio of non-synonymous (*dN*)/synonymous (*dS*). In particular, coefficient of functional divergence (θ) was introduced to predict which individual genes displayed difference in functional constraint. Finally, we predicted distinct amino acid residues that may be of importance in the functional divergence of the *ACS* paralogous. Our systematic analysis provides a solid foundation for further functional dissection of *ACS* genes in plants.

MATERIALS AND METHODS

Sequence database searches

The DNA and amino acids sequences of *ACS* genes were downloaded from Ensembl Database (http://plants.ensembl.org/index.html). Firstly, the generic names of *ACS*1 to *ACS*12 were first used as queries to get the *Arabidopsis ACS* genes. Then, all the orthologues of the *Arabidopsis ACS* genes in rice, grapevine and poplar were collected. A table with species names, abbreviations and accession numbers are provided in Supplementary Table 1. Since the new genes in rice, grapevine and poplar have no standard annotation, we assigned each of them an identity on the order of their locations on each chromosome. One additional sequence from the fungi *P. citrinum* (downloaded from http://www.ncbi.nlm.nih.gov) was used as an out-group to determine the earliest diverging in *ACS* gene family (Supplementary Table 1). The final data set included a total of 41 sequences from *Arabidopsis*, rice, grapevine and poplar. However, one *Arabidopsis* and one grapevine *ACS* gene, *AtACS3* and *VvACS1*, were excluded from the analysis since *AtACS3* may be a pseudogene and *VvACS1* is not a completed sequence. *AtACS2* owns splice variants, so the longer one, *AtACS*2B was chosen to perform the analysis.

Multiple sequence alignment and phylogenetic tree construction

The sequences of *ACS* proteins were aligned by MUSCLE (Edgar, 2004) with default setting. Incomplete sequences and highly divergent regions or gaps resulting in uncertain alignments were excluded from further analysis. The amino acid alignment was subsequently transformed into an aligned cds fasta file using PAL2NAL (Suyama et al., 2006). Sequence logo for the protein alignments (*ACS* dataset) was created online using WebLogo Version3 (http://weblogo.berkeley.edu/) (Crooks et al., 2004). The nucleotide alignment was then converted to nexus format with DnaSP (Rozas et al., 2003) version 4.10 for phylogenetic analysis. The full alignment of 38 amino acid sequences were used to perform phylogenetic tree reconstruction by neighbor-joining (NJ) based on p-distance (Saitou and Nei., 1987) in the molecular evolutionary genetics analysis (MEGA) program, version 4.0 (Tamura et al., 2007). Maximum likelihood (ML) implemented in Phyml and Bayesian inference (BI) performed in MrBayes were also used to construct phylogenetic tree. We analyzed four independent runs each, using the general time reversible (GTR) model plus gamma distribution plus invariant sites model of molecular evolution (GTR+G+I), selected by Model Test (Posada et al., 1998). The robustness of the trees was assessed by bootstrapping (500 pseudo- replicates) in Phyml. Bayesian analysis was conducted using the same model with MrBayes v3.1.2 (Huelsenbeck and Ronquist.. 2001; Taira et al., 2011).

Analysis of *ACS* gene expansion patterns

Segmental duplication, tandem duplication and transposition events are three main reasons for gene family expansion (Kong et al., 2007). Since transposition events were difficult to identify based only on the gene sequence analysis, we mainly focused on segmental and tandem duplication in this study. All the *ACS* genes in these four species were first checked for chromosomal location in order to study the expansion pattern. If a multiple members of gene family occurred within the same or near intergenic regions, it was characterized as tandem duplication. Segmental duplications were identified by a method similar to that of Maher et al. (2006). Firstly, the paralogs were identified at the terminus of the phylogenetic tree. Secondly, 10 protein-coding genes upstream and downstream of each pair of paralogs were respectively selected from *Arabidopsis*, rice, grapevine and poplar genomes. Lastly, the flanking 20 genes near one *ACS* gene were used as a database, and the genes flanking on the other *ACS* gene of one pair paralogs were used to

match this database. Finally, the paralogous pair was considered to originate from one duplication if they were both located in a region of conserved protein-coding genes.

Identifying selective forces

Several models have been reported to explain the preservation of duplicate genes, but they differ in the way of predicting the pattern of molecular evolution that follows gene duplications (Martinez-Castilla et al., 2003; Wagner, 2008). The selective pressures acting on coding regions were estimated by a phylogenetic-based maximum likelihood method, while the ratio of the non-synonymous (*dN*) to synonymous substitution rates (*dS*), ω = *dN/dS*, were estimated using codeml program implemented in the PAML package version 4.4 (Yang, 2007). The ratio of non-synonymous (*dN*) to synonymous (*dS*) substitution rate is always used to test the positive Darwinian selection, (ω = *dN/dS*). The ratio less than 1 means negative or purifying selective, while the ratio equal to 1 indicates neutral evolution, and positive selection is indicated if the estimate is significantly greater than 1 (Li et al., 1983). In this work, two codon models were used to explore the gene evolution, one is the site-specific model (Yang 2002; Wong et al., 2004), which intends to detect positive selection among codons. The other is branch-site model, which was used to detect positive selection in assigned lineages of phylogeny (Yang and Nielsen., 2002; Zhang et al., 2005).

Furthermore, three pairs of site models were introduced to detect positive selection: M0 (one ratio) vs. M3 (discrete), M1a (nearly neutral) vs. M2a (positive selection), and M7 (β) vs. M8 (β and ω), and likelihood ratio tests (LRTs) were conducted to verify whether ω was significantly different from 1 for each pair-wise comparison. When a signature of positive selection according to the LRTs showed, the Bayes empirical Bayes (BEB) method (Nielsen and Yang., 1998; Wong et al., 2004) was used to identify individual codons with dN/dS > 1. These models average over all sequences in the phylogeny, not accounting for molecular rate shifts or positive selection on specific lineages, may hide the evidence of positive selection.

In comparison, the branch-site model is more powerful to detect positive position, since the assigned branch is allowed to have a class of sites with ω > 1 (model A) and positive selection always happens on a small number of codon sites in a short period of evolution time. In branch-site model A, the branches were divided into two groups: one is of foreground branch, which was used to detect positive selections; the remaining branches are background branches. In our study, main lineage *A1, A2, A3* and *B* was assigned as foreground branch respectively, and each foreground branch was used to determine which one had evolved at a different rate, with sites under positive selection identified by posterior probability (*PP*) inference.

Functional divergence and structure analysis

To study the functional divergence after the gene duplications, we used the Diverge 2.0 software to estimate the coefficient of type I functional divergence (θ_I) (Gu 1999; Gu 2001). Type I sites represent amino acid residues conserved in one subfamily but highly variable in another, implying that these residues have been subjected to different functional constraint. Genes which have been predicted to be subjected to positive selection were used to search for homologous sequences in the PDB database of protein structures http://www.rcsb.org/pdb/home/home.do using Blastp (Altschul et al., 1990; Altschul et al., 1997). The Rasmol (http://rasmol.org/) was used for all structural manipulations and highlighting the relevant amino acid replacements identified in the evolutionary analysis.

RESULTS AND DISCUSSION

Collection of *ACS* genes in *Arabidopsis*, rice, grapevine and poplar

After carefully surveying the four plant genomes, 12 members of *ACS* gene family in *Arabidopsis*, six in rice, 10 in grapevine and 11 in poplar were identified. The CDS and protein sequences of *ACS* genes family were all downloaded from ensemble database (http://plants.ensembl.org/index.html). *Arabidopsis ACS* genes were dispersed on the five chromosomes: both two *ACS* genes were found on chromosomes 1 and 3, one on chromosome 2, four on chromosomes 4, and three on chromosome 5, respectively. The six *ACS* genes in rice were dispersed on five chromosomes except chromosome 2: one gene was located on chromosome 1, 3, 4 and 6, respectively, and two genes on chromosome 5. There were 11 *ACS* genes in poplar: each gene was located on chromosomes 1, 6, 7, 12, 14, 15 and 18, and two on chromosomes 2 and 3. Ten *ACS* genes were found in grapevine: three genes were located on chromosome 2, one was located on 11, 15, 16 and 18 respectively, and the positions of the other three were not clear.

Phylogenetic relationships of *ACS* gene family among the four plant species

To investigate the molecular evolution and phylogenetic relationships among *ACS*s in *Arabidopsis*, rice, grapevine and poplar, a phylogenetic tree based on 37 protein sequences with one *ACS* protein from *P. citrinum* was constructed using the neighbor-joining (NJ) method with p-distance and pair-wise deletion option (Figure 1). For statistical reliability, we conducted bootstrap analysis with 500 replicates. The NJ phylogenetic tree showed that all the *ACS* genes from the four higher plants were classified into two well-supported clades *A* and *B*. The sequences in clade *A* were further divided into four subclades: *A1* to *A4*. Among these, *A4* contained only one member (*VvACS3*), while *A3* was the largest clade containing 19 members. As shown in Figure 1, we can see that all subclades except for *A4* contain *Arabidopsis*, rice, grapevine and poplar *ACS* sequences, and the *ACS* genes from rice always make up one independent branch. For subclade *A2* for example, three genes from rice (*OsACS1*, *OsACS4* and *OsACS5*) formed one separated branch. These indicated that *ACS* genes have already duplicated before the monocot and dicot diverged.

More also, nine pairs of paralogous genes were identified from the phylogenetic tree (squares in Figure 1), three for *Arabidopsis*, one for rice, one for grapevine and four for poplar. This result meant that the *ACS* genes in *Arabidopsis*, rice, grapevine and poplar had expanded

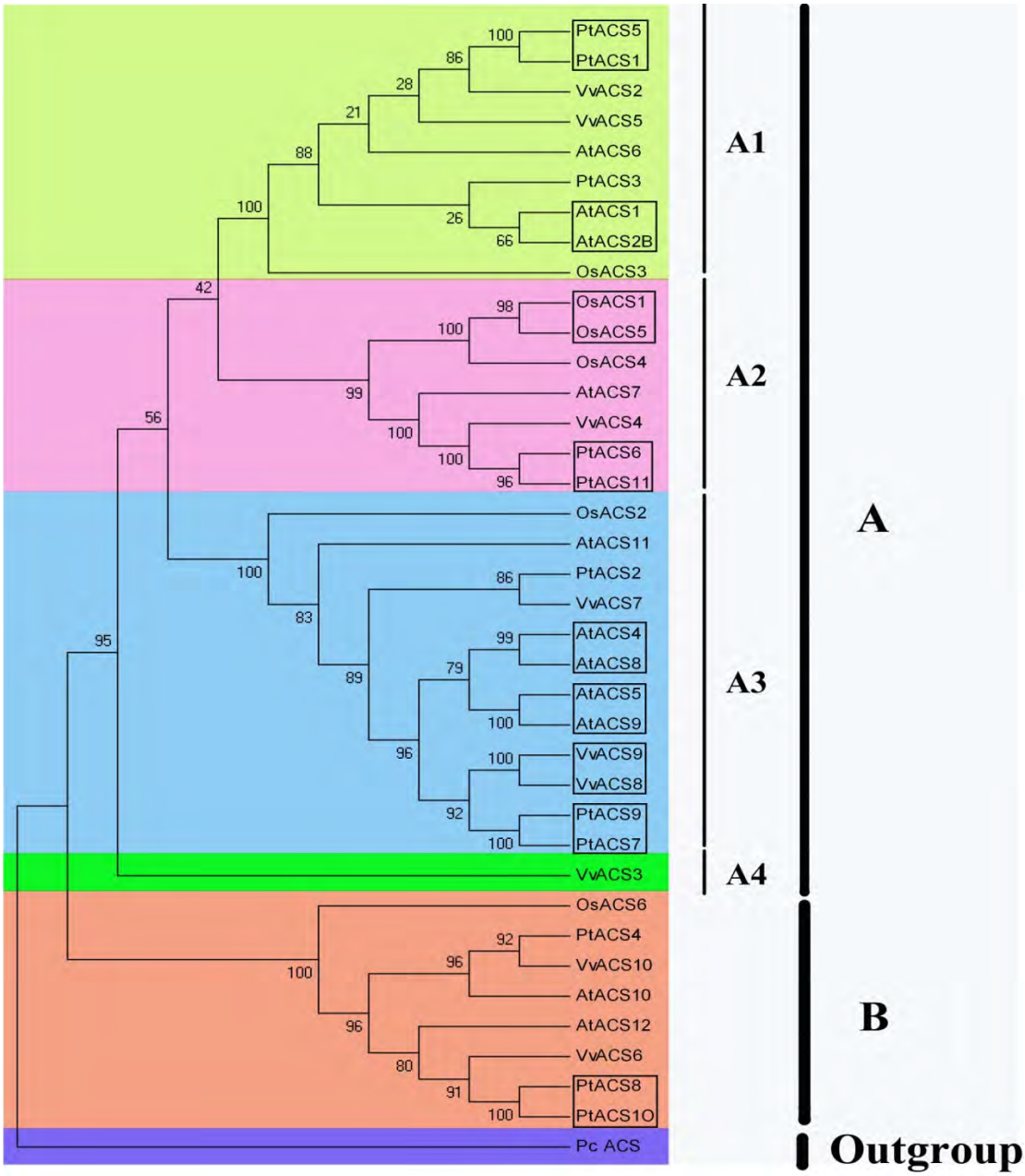


Figure 1. The phylogenetic tree of the *ACS* protein family in *Arabidopsis*, rice, grapevine and poplar. *P. citrinum* was used as outgroup. Species name abbreviations are shown in supplementary Table 1. Different shaded boxes denote the five lineages, and main lineage *A1, A2, A3* and *B* were used to set as foreground branches. The squares at the terminus of the tree mean the paralogous genes among *ACS* family.

in a species-specific manner, thus suggesting that most *ACS* gene members did not originate from the common ancestral genes that existed before the divergence of monocot and dicot. Moreover, our subclades classification was consistent with the report from Jakubowicz et al. (2004). Comparing the alignment, all amino acid residue makers of different types or groups reported by Jakubowicz et al. (2004), were also detected in our work.

Evolutionary patterns of *ACS* genes in *Arabidopsis*, rice, grapevine and poplar

Segmental duplication, tandem duplication and transposition events are three main reasons for gene family expansion (Kong et al., 2007). By searching both the sharing region and neighboring regions of *ACS* genes, no genes were found to be located in tandem repeats, suggesting that tandem duplication did not play

Table 1. Duplicated ACS genes and the number of conserved protein-coding genes flanking them in *Arabidopsis*, rice, grapevine and poplar

Duplicated ACS gene 1	Duplicated ACS gene 1	Number of conserved flanking protein-coding gene
AtACS4	*AtACS8*	3
AtACS5	*AtACS9*	4
PtACS1	*PtACS5*	12
PtACS6	*PtACS11*	1
PtACS7	*PtACS9*	3
PtACS8	*PtACS10*	11

the main role in expansion of this family. Thus, we assumed that segmental duplication and transposition events contributed more in evolution of the *ACS* gene family in *Arabidopsis*, rice, grapevine and poplar. To validate this hypothesis, nine pairs of paralogs were selected from the terminus of the phylogenetic tree (Figure 1), and then the similar protein-coding genes flanking each paralogous pair were searched. The number of conserved flanking protein-coding genes of each pair of paralogs is displayed in Table1. In *Arabidopsis*, three paralogs were found at the terminus of the phylogenetic tree. Except for *AtACS1* and *AtACS2*, the other two had conserved flanking protein-coding genes, which indicates that probably segmental duplication played a leading role in the expansion of *ACS* gene family.

In addition, there were four pairs of paralogs in poplar; only three *ACS* genes were not paralogs. Highly conserved genes were observed among the flanking regions for all pairs of paralogs genes, indicating that all of the paralogs *ACS* genes in poplar arose from segmental duplication events. There was, however, only one pair of paralogs in rice and poplar, respectively. Both these two paralogs did not find any conserved genes in flanking regions, which suggested that random insertion events rather than tandem duplications and segmental duplications play a leading role in the expansion of *ACS* gene family in these two species. In conclusion, segmental duplications were the main evolutionary pattern in *ACS* gene family of *Arabidopsis* and poplar; while random translocation and insertion events played a leading role in grapevine and rice.

Evidence for the action of selection in *ACS* gene evolution

In this work, two codon models were used to explore the selective force: one is site-specific model (Yang 2002; Wong et al., 2004), which intends to detect positive selection across all amino acid sites and the other is branch-site model, which was used to detect positive selection in pre-specified lineages of phylogeny (Yang et al., 2002; Zhang et al., 2005). In the site-specific model, the result from M1a against M2a had nearly identical log-likelihood scores (Table 2), thus suggesting that the amino acid changes were neutral or under purifying selection. The low ω estimation from M1a indicate the strong selective constraints, that is, 88.6% *ACS* genes displayed strong purifying selection. The result of M7 against M8 did not show any possible positive selection, either. These results provided consistently strong evidences that the tests did not find any positively selected sites across multiple sequences alignments.

Moreover, estimating ω ratios across all amino acid sites might hide the evidence of positive selection, since positive sites will likely appear at specific lineages on the phylogeny. Therefore, it is possible to find molecular rate shifts and positive selection after gene duplications when functional diversification processes affect one or both duplicates. In this study, we tried to elucidate whether or not positive selection had fixed certain amino acids at particular moments of *ACS* gene evolution, especially following the main duplication events. To achieve this, we used branch-site models that detect positive selection at specific coding sites and in particular branches of the phylogenetic tree. In branch-sites model, comparison of the model A to M1a was used to predict positive sites in particular assigned branches of the phylogenetic tree. As shown in Table 3, several positively selected sites (estimated ω greater than 1) were identified in previously assigned lineages by Bayes empirical Bayes (BEB) estimation, however, a class of sites in each was identified with ω2 = 1 (Table 3). Therefore, the relaxed functional constraint is most consistent with the results of molecular evolutionary analysis of the *ACS* data. In conclusion, the results from both site-specific and branch-sites indicate that strong purifying selection is the primary evolution model in *ACS* genes family.

Functional divergence analysis

Gene duplication-specific changes in the substitution rates (type I functional divergence) might reflect the difference in evolutionary rate at amino acid sites after gene duplication (Gu, 1999, 2001). In this study, Diverge program was applied to perform analysis in detecting evolutionary rate pattern shifts and to identify which amino acid sites may have contributed most to the

Table 2. Results of codon-based models of molecular evolution for *ACS* genes family.

Model	lnL[a]	Parameter[b]
M0 (single ω)	-19600.36	Ω = 0.101
M1(nearly neutral)	-19447.86	ω0 = 0.101 (P0 = 0.886) ω1= 1.000 (P1= 0.114)
M2a (selection)	-19447.86	ω0 = 0.101 (P0 = 0.886) ω1= 1.000 (P1= 0.040) ω2= 1.000 (P1= 0.074)
M3 (discrete)	-18999.54	ω0 = 0.026 (P0 = 0.503) ω1= 0.137 (P1= 0.325)
M7 (β)	-18990.88	p= 0.750 q= 5.493
M8 (β+ α)	-18990.88	p = 0.750 q = 5.493 P0 = 0.999 ω1= 42.863 (P1= 0.001)

[a]Log likelihood of the data; [b] Pi, the proportion of codons that fall in each category.

Table 3. Parameters estimation and likelihood ratio tests for the branch-site models for *ACS* gene family.

ACS gene family	Parameter estimated (aa)	lnL[b]	2Δlc	p value	Selective site
A1					
MA vs. M0	p0= 0.81495 p1 = 0.09279 (p2= 0.09227) ω0 = 0.09492 (ω1=1.00000) ω2 =1.00000	-19420.35	180.01	P < 0.01	**116*S** 173**C (P>0.95)
A2					
MA vs. M0	p0 = 0.85626 p1= 0.09717 (p2 = 0.04658) ω0 = 0.09746 (ω1=1.00000) ω2 =1.00000	-19432.44	167.92	P < 0.01	**116*s** 158*E 166*D (P>0.95)
A3					
MA vs. M0	p0 = 0.66027 p1= 0.09410 (p2= 0.24563) ω0= 0.08842 (ω1=1.00000) ω2 =1.00000	-19369.37	230.99	P < 0.01	**7*Q 24*C** 25*T **31*V 55**G** 56**R 78*M **122*Q 140**I** 165*N 181**V 201*Q **234**L** 248*H (P>0.95)
B					
MA vs. M0	p0 = 0.83908 p1= 0.07267 (p2 = 0.08825) ω0 = 0.09343 (ω1=1.00000) ω2 =1.00000	-19404.18	196.18	P < 0.01	39**K 40*Q 136*Q 158*E 210*E **260*R** 261*V (P>0.95)

[a]The number of free parameters; [b]likelihood of the MA model; [c] $2(l_1-l_0)$; *the amino acid residues indicated in bold were also found in the analysis of functional divergence.

functional divergence (Gu et al., 2002) for all pairs of *ACS* clusters. Six coefficient of functional divergence (θ) were displayed (Table 4), varying from 0.08 to 0.59, which meant that all clusters in *ACS* genes family exhibited significant evidence of type I functional divergence except *A2/A3*. This therefore implied that there were some amino acid sites with discrepancies in their evolutionary rates between these clusters except *A2/A3*. Proper cutoff value is normally used to predict the amino acids that are involved in the functional divergence after gene-duplication. In this paper, we chose the cut-off values: *Q(k)≥0.80*. As expected, most amino acids had very low posterior probability (*PP*) values, meaning that these sites would not be involved in the hypothetical functional

Table 4. Maximum likelihood estimates of the coefficient of functional divergence (θ) from pair-wise comparisons between *ACS* groups.

Comparison	θ[a]	SE[b](θ)	LRT[c] (θ)	Significance	Probability cut-off[d]
A1 vs B	0.59	0.06	78.66	p<0.01	0.80
A1 vs A2	0.22	0.06	14.12	p<0.01	0.80
A1 vs A3	0.17	0.05	12.69	p<0.01	0.80
B vs A2	0.49	0.08	35.81	p<0.01	0.80
B vs A3	0.49	0.05	76.53	p<0.01	0.80
A2 vs A3	0.08	0.06	2.02	P>0.05	0.80

[a]θ is the coefficient of functional divergence; [b]SE, standard error; [c]LRT (θ) is a likelihood ratio test; [d]probability cut-off is the minimal posterior probability for amino acids causing functional divergence.

divergence (Figure 2). Specifically, we detected 38, 3, 1, 16, 26 amino acid positions (with *PP* threshold values higher than 0.80) in the *A1/B, A1/A2, A1/A3, B/A2* and *B/A3* comparisons, respectively (Table 4). Besides, 11 of these functional divergence candidate positions were also detected by LRT in codeml. It may be related to relaxed selective constraint during *ACS* genes duplications (indicated in bold in Table 3).

It is known that gene duplication events can make gene functional diverge. Normally, after duplication, one daughter gene retains the original function, while the other accumulates changes, some of which survive leading to functional divergence. Once functional divergence has occurred between duplicated genes, there must have been changes within their coding regions. In this study, two ways were employed to detect possible sites involved in gene functional divergence. One is Diverge program (Gu, 1999, 2001), which was used to establish the posterior probability of type I divergence at each site in the alignment, and it predicted 84 candidate functional divergence-related sites with 0.80 as a cutoff value. However, all these sites were predicted in logic and we lacked a method to verify their contributions to the gene functional divergence. For comparative purpose, a LRT, like a Bayesian method was also used to analyze the same alignment and phylogeny, which provided a statistical framework to tell the positive residues. Combining the two ways, at last 11 statistically significant positions were deduced to be involved in functional diverge of *ACS* family.

Molecular modeling

We mapped 11 residues that were probably involved in functional diverge of *ACS* family onto the sequence logo (Figure 3) and three-dimensional structure (Figure 4) in order to get a better understanding of these sites on functional diversity. The balls in Figure 4 indicate the location of relaxed selected amino acid sites and detections of type I functional divergence (with high *PP* values) in *AtACS1*. It was interesting to note that most of these 11 sites were adjacent to the important residues for catalysis and structure of *ACS* synthase. For example, 173C was only one residue distance away from the active site D237, while 7Q and 31V were two and three residues away from Ala54 and Y92, respectively. The superscript number of active sites was consistent with tomato (Tarun et al., 1998a, 1998b), and the number before the amino acid is the position of sequence alignment.

Furthermore, some amino acid residues have been reported to play important roles in the *ACS* synthase activity in tomato and apple, and their functions were mainly involved in the following angles: 1) form the hydrogen bonds with O3 of the pyridine ring and N1 of PLP, respectively; 2) stabilize the pyridine ring of PLP and involve directly in catalysis; 3) form the correct orientation of PLP in the active sites; 4) work with SAM and AVG (White et al., 1994, Li et al., 1997, Huai et al., 2001b, McCarthy et al., 2001). The sequence variations in these regions are possibly involved in the functions described earlier. However, without further analysis, we cannot speculate on the role of these relaxed selected amino acid sites, although we provided evidence for some more possible sites playing an important functional role in *ACS* gene activity and also represented some sites for future mutagenesis experiment which could help our understanding of the structure-function relationships in *ACS* genes.

Conclusion

Genomic data afforded us an opportunity to gain a better understanding about the evolution of *ACS* gene using phylogenetic analyses. Our phylogenetic tree showed that duplication events triggered diversification of the *ACS* gene family into two groups in *Arabidopsis*, rice,

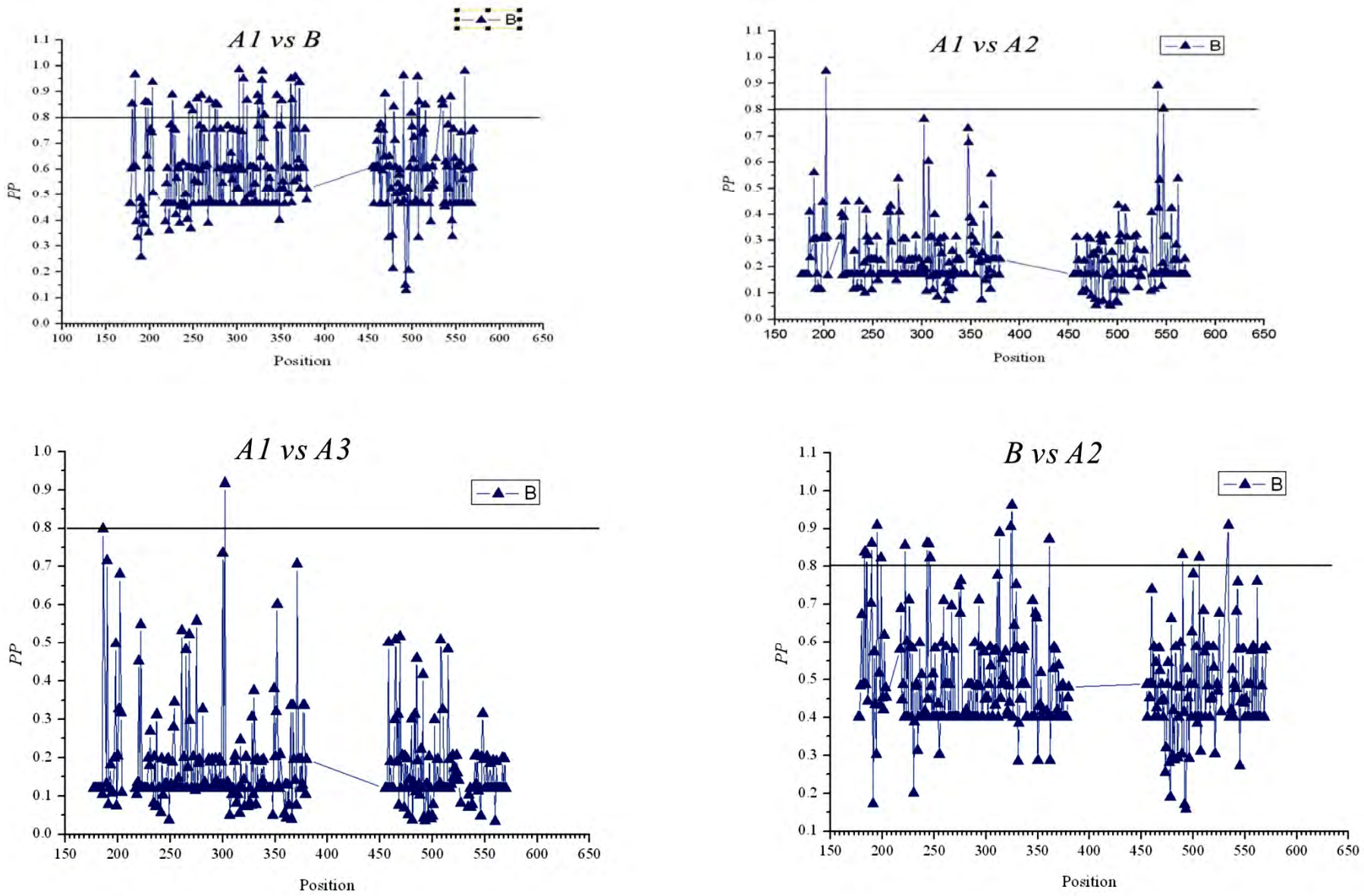
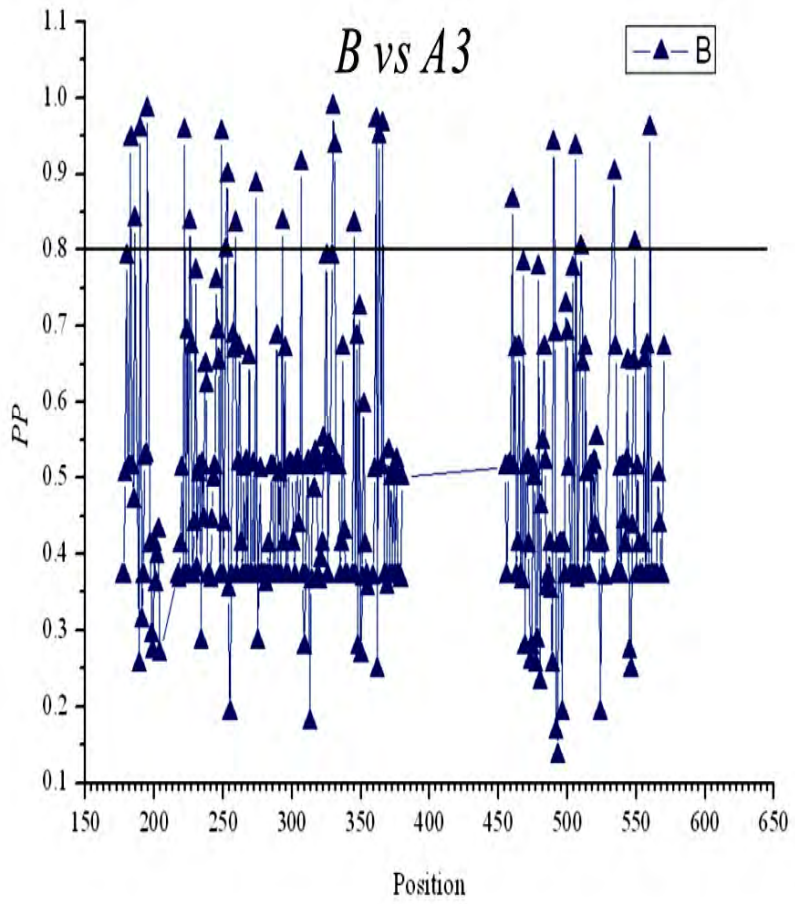


Figure 2. Site specific profiles for evolutionary rate change in the *ACS* protein family. The posterior probabilities of functional divergence for *ACS* protein family were obtained with Diverge. Individual cut-off values for each comparison are marked with black horizontal line.



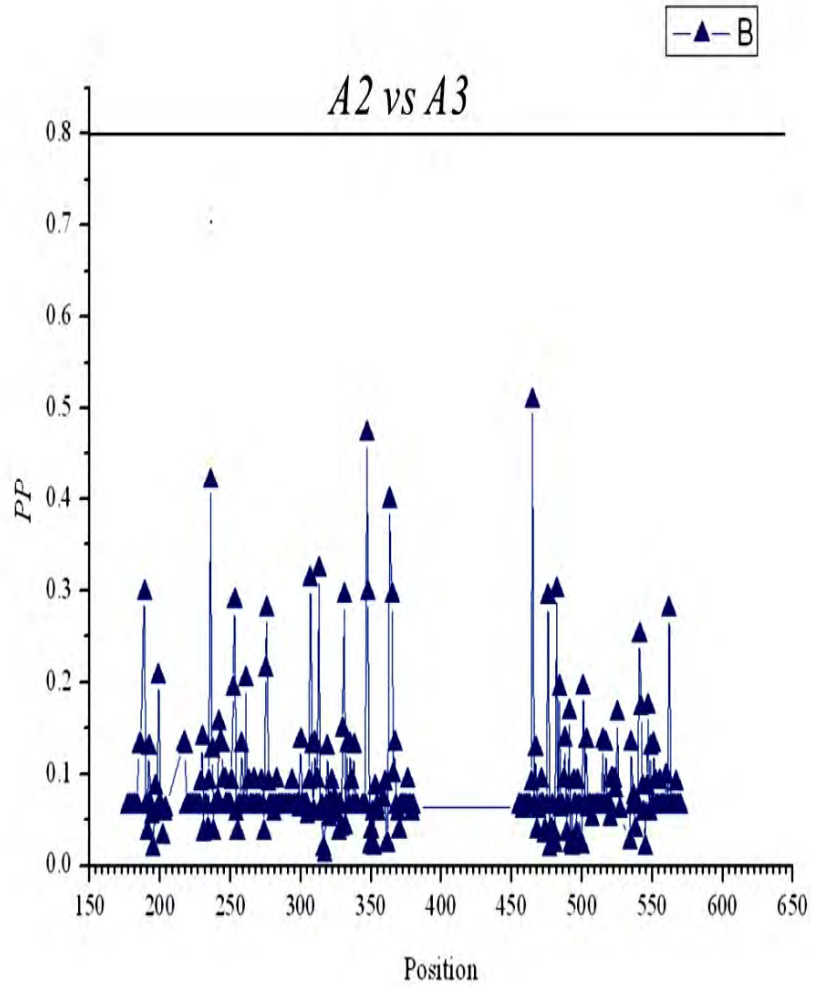


Figure 2. Contd.

grapevine and poplar. The evolution selection analysis indicated that relaxed selective constraint might play important roles on the evolutionary time and shape variation of some members of the family. Functional divergence analysis in ACS gene family suggests that some amino acids were involved in the hypothetical functional divergence. Finally, we mapped amino acids critical for functional divergence onto the sequence logo

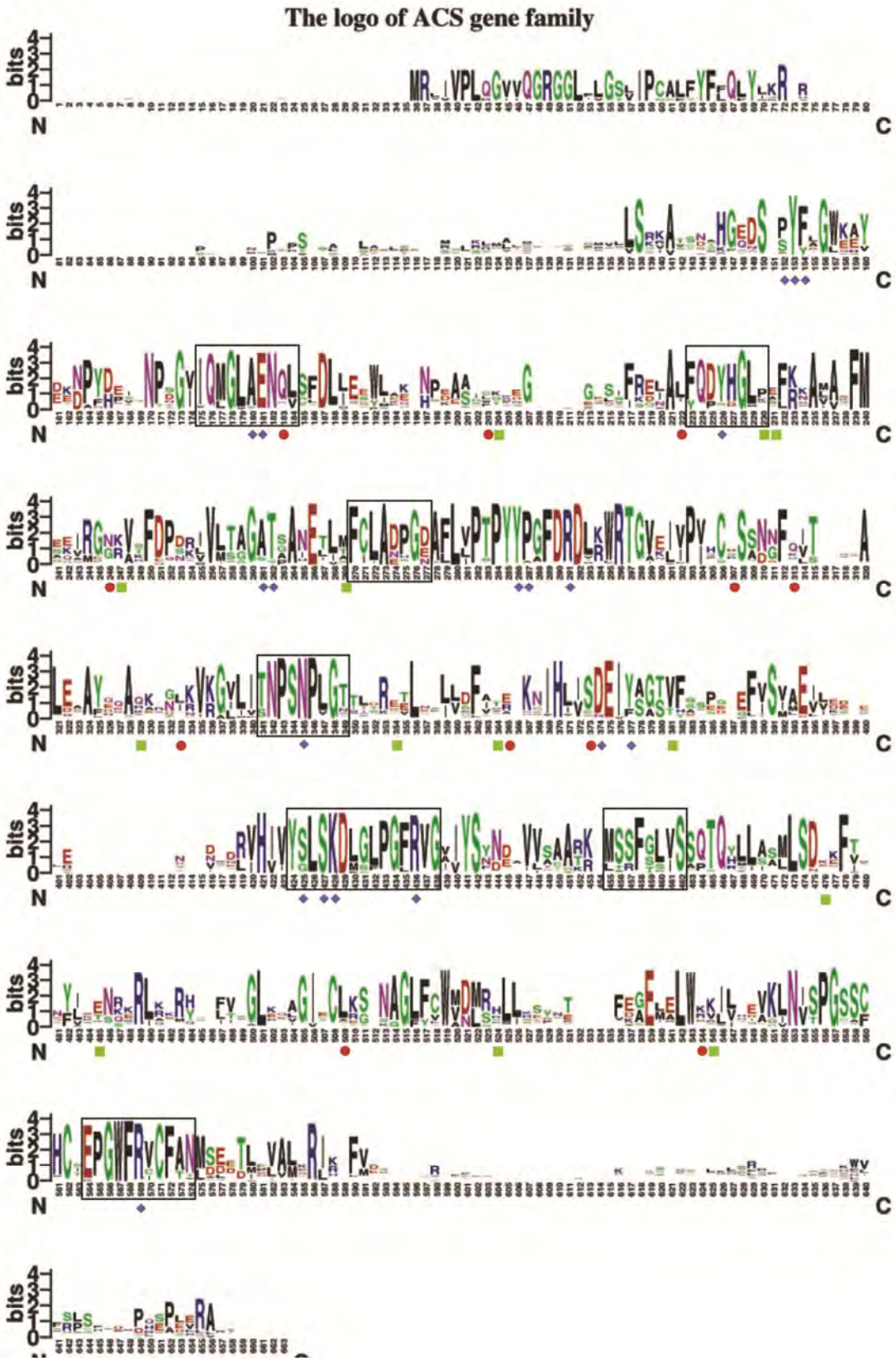


Figure 3. *ACS* sequence logo for the protein alignment. The boxes represent the conserve regions in *ACS* gene family. The pink balls stand for 11 statistically positions detected in both Diverge Program and LRT methods; the blue rhombus stand for the active sites reported in *ACS* family, and the green squares represents the positive sites detected by LRT method.

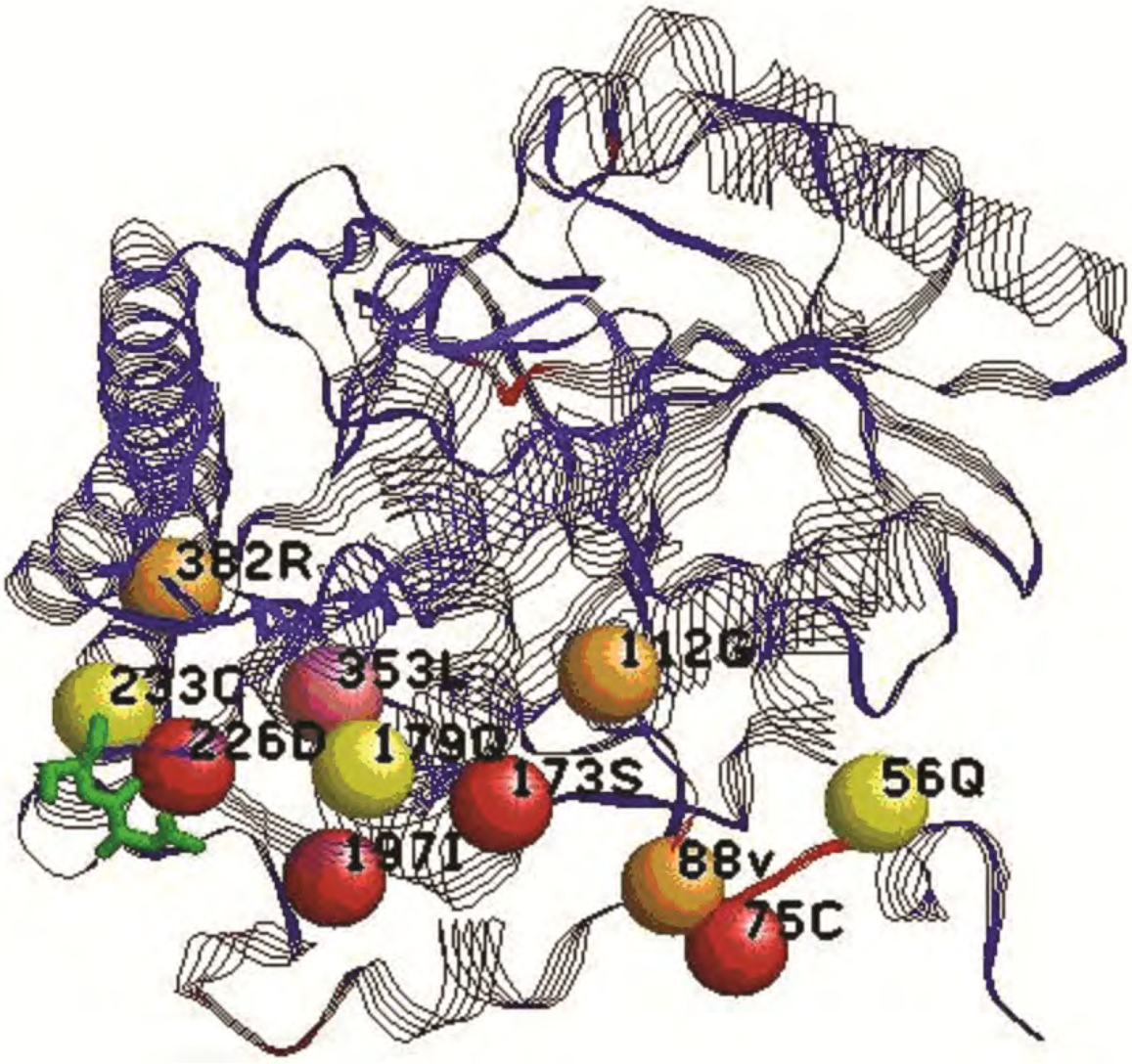


Figure 4. The modeled structure of *Arabidopsis ACS*1. Mapping the amino acid residues presumably submitted to altered functional constraints onto 3D structure, which shows in colorful balls. The green stick is the PLP binding site.

and three-dimensional structure, which were close to the reported active sites. All these studies will certainly contribute to a better understanding of the precise functional diversity of this family, and provide a starting point for further experimental verifications.

REFERENCES

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. J. Mol. Biol. 215(3): 403-410.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25(17): 3389-3402.

Boller T, Herner R, Kende H (1979). Assay for and enzymatic formation of an ethylene precursor, 1-aminocyclopropane-1-carboxylic acid. Planta, 145(3): 293-303.

Crooks G, Hon G, Chandonia J, Brenner S (2004). WebLogo: a sequence logo generator. Genome Res. 14(6): p. 1188.

Dong J, Kim W, Yip W, Thompson G, Li L, Bennett A (1991). Cloning of a cDNA encoding 1-aminocyclopropane-1-carboxylate synthase and expression of its mRNA in ripening apple fruit. Planta. 185(1): 38-45.

Edgar RC (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics, 5: p. 113.

Fluhr R, Mattoo A, Dilley D (1996). Ethylene biosynthesis and perception. Crit. Rev. Plant Sci. 15(5): 479-523.

Ge L, Liu J, Wong W, Hsiao W, Chong K, Xu Z, (2000). Identification of a novel multiple environmental factor-responsive 1-aminocyclopropane-1-carboxylate synthase gene, NT-ACS2, from tobacco. Plant Cell Environ. 23(11): 1169-1182.

Goff SA, Ricke D, Lan TH, Presting G, Wang R, Dunn M, (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. japonica). Science, 296 (5565): 92-100.

Gray J, Picton S, Shabbeer J, Schuch W ,Grierson D (1992). Molecular biology of fruit ripening and its manipulation with antisense genes. Plant Mol. Biol. 19(1): 69-87.

Gu X (1999). Statistical methods for testing functional divergence after gene duplication. Mol. Biol. Evol. 16(12): 1664-1674.

Gu X (2001). Maximum-likelihood approach for gene family evolution under functional divergence. Mol. Biol. Evol. 18(4): 453-464.
Gu X , Velden K (2002). DIVERGE: phylogeny-based analysis for functional-structural divergence of a protein family. Bioinformatics, 18(3): 500-501.
Huai Q, Xia Y, Chen Y, Callahan B, Li N and Ke H (2001a). Crystal structures of 1-aminocyclopropane-1-carboxylate (ACC) synthase in complex with aminoethoxyvinylglycine and pyridoxal-5'-phosphate provide new insight into catalytic mechanisms. J Biol. Chem. 276(41): 38210-38216.
Huai Q, Xia Y, Chen Y, Callahan B, Li N and Ke H (2001b). Crystal structures of 1-aminocyclopropane-1-carboxylate (ACC) synthase in complex with aminoethoxyvinylglycine and pyridoxal-5¡ä-phosphate provide new insight into catalytic mechanisms. J. Biol. Chem. 276(41): p. 38210.
Huelsenbeck JP, Ronquist F (2001). MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics. 17(8): 754-755.
Initiative A (2000). Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature, 408(6814): 796-815.
Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A (2007). The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature, 449(7161): 463-467.
Jakubowicz M (2002). Structure, catalytic activity and evolutionary relationships of 1-aminocyclopropane-1-carboxylate synthase, the key enzyme of ethylene synthesis in higher plants. Acta Biochim. Pol. 49(3): 757-774.
Jakubowicz M, Pacak A (2004). Relationships between ACC synthase isozymes calculated using phenetic data. The catalytical and evolutionary aspects. Acta Physiologiae Plantarum, 26(1): 5-27.
Jia YJ, Kakuta Y, Sugawara M, Igarashi T, Oki N, Kisaki M (1999). Synthesis and degradation of 1-aminocyclopropane-1-carboxylic acid by Penicillium citrinum. Biosci Biotechnol. Biochem. 63(3): 542-549.
Johnson PR, Ecker JR (1998). The ethylene gas signal transduction pathway: a molecular perspective. Annu Rev Genet. 32: 227-254.
Kakuta Y, Igarashi T, Murakami T, Ito H, Matsui H and Honma M (2001). 1-Aminocyclopropane-1-carboxylate synthase of *Penicillium citrinum*: primary structure and expression in *Escherichia coli* and *Saccharomyces cerevisiae*. Biosci. Biotechnol. Biochem. 65(7): 1511-1518.
Kende H (1989). Enzymes of ethylene biosynthesis. Plant Physiol. 91(1): 1-4.
Kende H (1993). Ethylene biosynthesis. Annu. Rev Plant Biol. 44(1): 283-307.
Koch KA, Capitani G, Gruetter MG, Kirsch JF (2001). The human cDNA for a homologue of the plant enzyme 1-aminocyclopropane-1-carboxylate synthase encodes a protein lacking that activity. Gene. 272(1-2): 75-84.
Kong H, Landherr LL, Frohlich MW, Leebens K, Mack J, Ma H, DePamphilis CW (2007). Patterns of gene duplication in the plant SKP1 gene family in angiosperms: evidence for multiple mechanisms of rapid gene birth. Plant J. 50(5): 873-885.
Li WH, Gojobori T (1983). Rapid evolution of goat and sheep globin genes following gene duplication. Mol. Biol. Evol. 1(1): 94-108.
Li Y, Feng L, Kirsch J (1997). Kinetic and Spectroscopic Investigations of Wild-Type and Mutant Forms of Apple 1-Aminocyclopropane-1-carboxylate Synthase. Biochemistry, 36(49): 15477-15488.
Maher C, Stein L, Ware D (2006). Evolution of Arabidopsis microRNA families through duplication events. Genome Res. 16(4): 510.
Martinez-Castilla LP, Alvarez-Buylla ER (2003). Adaptive evolution in the Arabidopsis MADS-box gene family inferred from its complete resolved phylogeny. Proc. Natl. Acad. Sci. USA. 100(23): 13407-13412.
Matsumoto T, Wu J, Kanamori H, Katayose Y, Fujisawa M, Namiki N (2005). The map-based sequence of the rice genome. Nature, 436(7052): 793-800.
McCarthy D, Capitani G, Feng L, Gruetter M, Kirsch J (2001). Glutamate 47 in 1-Aminocyclopropane-1-carboxylate Synthase Is a Major Specificity Determinant. Biochemistry, 40(41): 12276-12284.
Nielsen R, Yang Z (1998). Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics, 48(3): 929-936.
Peixoto B, Mikawa Y, Brenner S (2000). Characterization of the recombinase activating gene-1 and 2 locus in the Japanese pufferfish, Fugu rubripes. Gene, 246(1-2): 275-283.
Posada D ,Crandall KA (1998). MODELTEST: testing the model of DNA substitution. Bioinformatics. 14(9): 817-818.
Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics, 19(18): 2496-2497.
Saitou N, Nei M (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4(4): 406-425.
Suyama M, Torrents D, Bork P (2006). PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34(suppl 2): W609.
Taira C, Matsuda K, Kamijyo Y, Sakashita K, Ishida F, Kumagai T, (2011). Quantitative monitoring of single nucleotide mutations by allele-specific quantitative PCR can be used for the assessment of minimal residual disease in patients with hematological malignancies throughout their clinical course. Clin. Chim. Acta. 412(1-2): 53-58.
Tamura K, Dudley J, Nei M, Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24(8): 1596-1599.
Tarun AS, Lee JS, Theologis A (1998a). Random mutagenesis of 1-aminocyclopropane-1-carboxylate synthase: a key enzyme in ethylene biosynthesis. Proc. Natl. Acad. Sci. USA. 95(17): 9796-9801.
Tarun AS, Theologis A (1998b). Complementation analysis of mutants of 1-aminocyclopropane- 1-carboxylate synthase reveals the enzyme is a dimer with shared active sites. J. Biol. Chem. 273(20): 12509-12514.
Theologis A (1992). One rotten apple spoils the whole bushel: the role of ethylene in fruit ripening. Cell, 70(2): 181-184.
Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U (2006). The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science, 313(5793): 1596-604.
Wagner A (2008). Neutralism and selectionism: a network-based reconciliation. Nat. Rev. Genet. 9(12): 965-974.
White M, Vasquez J, Yang S, Kirsch J (1994). Expression of apple 1-aminocyclopropane-1-carboxylate synthase in *Escherichia coli*: kinetic characterization of wild-type and active-site mutant forms. Proc. Natl. Acad. Sci. USA. 91(26): 12428.
Wong W, Ning W, Xu P, Kung S, Yang S, Li N (1999). Identification of two chilling-regulated 1-aminocyclopropane-1-carboxylate synthase genes from citrus (Citrus sinensis Osbeck) fruit. Plant Mol. Biol. 41(5): 587-600.
Wong WS, Yang Z, Goldman N, Nielsen R (2004). Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics, 168(2): 1041-1051.
Yang S, Hoffman N (1984). Ethylene biosynthesis and its regulation in higher plants. Annu. Rev. Plant Physiol. 35(1):155-189.
Yang Z (2002). Likelihood and Bayes estimation of ancestral population sizes in hominoids using data from multiple loci. Genetics, 162(4): 1811-23.
Yang Z (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586-91.
Yang Z, Nielsen R (2002). Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol. Biol. Evol. 19(6): p. 908.
Yip WK, Dong JG, Kenny JW, Thompson GA ,Yang SF (1990). Characterization and sequencing of the active site of 1-aminocyclopropane-1-carboxylate synthase. Proc. Natl. Acad. Sci. USA. 87(20): 7930-7934.
Yu J, Hu S, Wang J, Wong GK, Li S, Liu B (2002). A draft sequence of the rice genome (*Oryza sativa* L. ssp. indica). Science, 296(5565): 79-92.
Zarembinski TI, Theologis A (1994). Ethylene biosynthesis and action: a case of conservation. Plant Mol. Biol. 26(5): 1579-1597.
Zhang J, Nielsen R, Yang Z (2005). Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol. Biol. Evol. 22(12): 2472-2479.

Supplementary Table 1. *ACS* homologues (n = 41) from *Arabidopsis*, rice, grapevine, poplar and *Penicillium citrinium.*

Arabidopsis thaliana	**Generic name**	**Genomic position**	**Protein length**
AtACS1	AT3G61510-TAIR	Chromosome 3	488
*AtACS2A**	AT1G01480.1-TAIR	Chromosome 1	390
AtACS2B	AT1G01480.2-TAIR		496
*AtACS3**	AT5G28360-TAIR	Chromosome 5	168
AtACS4	AT2G22810-TAIR	Chromosome 2	474
AtACS5	AT5G65800-TAIR	Chromosome 5	470
AtACS6	AT4G11280-TAIR	Chromosome 4	495
AtACS7	AT4G26200-TAIR	Chromosome 4	447
AtACS8	AT4G37770-TAIR	Chromosome 4	469
AtACS9	AT3G49700-TAIR	Chromosome 3	470
AtACS10	AT1G62960-TAIR	Chromosome 1	557
AtACS11	AT4G08040-TAIR	Chromosome 4	460
AtACS12	AT5G51690-TAIR	Chromosome 5	495
Oryza sativa			
OsACS1	LOC_Os01g09700	Chromosome 1	510
OsACS2	LOC_Os03g51740	Chromosome 3	487
OsACS3	LOC_Os04g48850	Chromosome 4	483
OsACS4	LOC_Os05g10780	Chromosome 5	437
OsACS5	LOC_Os05g25490	Chromosome 5	496
OsACS6	LOC_Os06g03990	Chromosome 6	542
Populus trichocarpa			
PtACS1	POPTR_0001s09850	Chromosome 1	481
PtACS2	POPTR_0002s11460	Chromosome 2	409
PtACS3	POPTR_0002s16530	Chromosome 2	474
PtACS4	POPTR_0003s11760	Chromosome 3	551
PtACS5	POPTR_0003s13220	Chromosome 3	480
PtACS6	POPTR_0006s15410	Chromosome 6	440
PtACS7	POPTR_0007s14640	Chromosome 7	467
PtACS8	POPTR_0012s14280	Chromosome 12	163
PtACS9	POPTR_0014s01300	Chromosome 14	468
PtACS10	POPTR_0015s14280	Chromosome 15	569
PtACS11	POPTR_0018s06960	Chromosome 18	446
Vitis vinifera			
*VvACS1**	GSVIVG00019920001	Chromosome 2	1808
VvACS2	GSVIVG00019414001	Chromosome 2	481
VvACS3	GSVIVG00004794001	Chromosome 2	506
VvACS4	GSVIVG00015239001	Chromosome 11	387
VvACS5	GSVIVG00026962001	Chromosome 15	438
VvACS6	GSVIVG00018494001	Chromosome 16	530
VvACS7	GSVIVG00009284001	Chromosome 18	452
VvACS8	GSVIVG00002708001	Chromosome Un	454
VvACS9	GSVIVG00004571001	Chromosome Un	445
VvACS10	GSVIVG00005455001	Chromosome Un	352
Penicillium citrinum			
	Genebank ID		
ACS	BAA92149.1		431

*These were excluded from the final analysis.