Phenylalanine ammonia-lyase through evolution: A bioinformatic approach

Shiva Hemmati[1,2*]

[1]Department of Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, I.R. Iran.
[2]Pharmaceutical Research Center, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, I.R. Iran.

Abstract

Phenylalanine ammonia-lyase (PAL) is the first entry enzyme of the phenylpropanoid pathway that converts phenylalanine to cinnamic acid which is the precursor of various secondary metabolites. PAL is recently formulated for phenylketonuric patients in pegylated forms; therefore, screening a PAL with the highest affinity to the substrate is of a great importance. PAL exists in all higher plants and some fungi and few bacteria. Ancestors of land plants have been adopted by evolving metabolic pathways. A multi-gene family encodes PAL by gene duplication events in most plants. In this study, the taxonomic distribution and phylogeny of *pal* gene found in land plants, fungi and bacteria have been analyzed. It seems that the ancestor of plants acquired a *pal* gene via horizontal gene transfer in symbioses with bacteria and fungi. Gymnosperms have kept a diverse set of *pal* genes that arose from gene duplication events. In angiosperms, after the divergence of dicotyledons from monocots, *pal* genes were duplicated many times. The close paralogues of *pal* genes in some species indicate expansion of gene families after the divergence in plant *pal* gene evolution. Interestingly, some of the plant *pal*s clustered by species in a way that *pal*s within one species are more closely related to each other than to homologs in the other species which indicates this duplication event occurred more recently.

Keywords: Phenylalanine ammonia lyase, Phylogenetic tree, Bioinformatic, Evolution, Phenylketonuria

1. Introduction

A vast number of secondary metabolites in plants, such as flavonoids, lignins, phytoalexins, and hormones are produced by the phenylpropanoid pathway (1). The first step in the biosynthesis of phenylpropanoid metabolism is controlled by the enzyme phenylalanine ammonia lyase (PAL). The enzyme catalyses the nonoxidative deamination of phenylalanine to trans-cinnamic acid and ammonia (2). The important issue about the enzyme is its role in connecting plant primary metabolism to the phenylpropanoid pathway (3). On the other hand, PAL is recently formulated as and oral PEGylated form for the treatment of phenylketonuric (PKU) patients (4). In this way, phenylalanine hydroxylase (which needs the cofactor BH_4 in excess) is substituted by recombinant PAL which provides another biochemical mean to catabolize phenylalanine and consequently decrease blood phenylalanine levels. It is a more convenient way of treatment for PKU patients than reduced-phenylalanine diet (5); hence screening for a *pal* gene which catalyses a PAL enzyme with the highest affinity to the substrate is of great importance for this purpose. The first step is to survey species containing the gene and how the corresponding gene is appeared and evolved.

The appearance of land plants from green algal ancestors was about 500 million years ago (6). Since land plants were under harsh environments and important stresses as UV radiation and attack by microbial soil communities, adaptations such as specialized secondary metabolic pathways have been developed (6,7). The phenylpropanoid pathway provides vital compounds such as lignin for vascularization, flavonoids to color and to protect against UV light and phytoalexins to deteriorate microbes (8).

In most plant species, a small gene family of 3~5 genes encodes PAL. In potato, about 40-50 genes make up *pal* gene family (9). In contrast, loblolly pine *pal* gene family consists of a single gene (10). The PAL protein has 595 to 750 amino acids except *Botrytis cinerea*, with 1,131 amino acids (11).

A few copy of the *pal* gene is reported in gymnosperms, whereas numerous *pal* gene families have been reported in angiosperms (12). During the

Corresponding Author: Shiva Hemmati
Department of Biotechnology, School of Pharmacy, Shiraz University of Medical Sciences, Shiraz, Iran
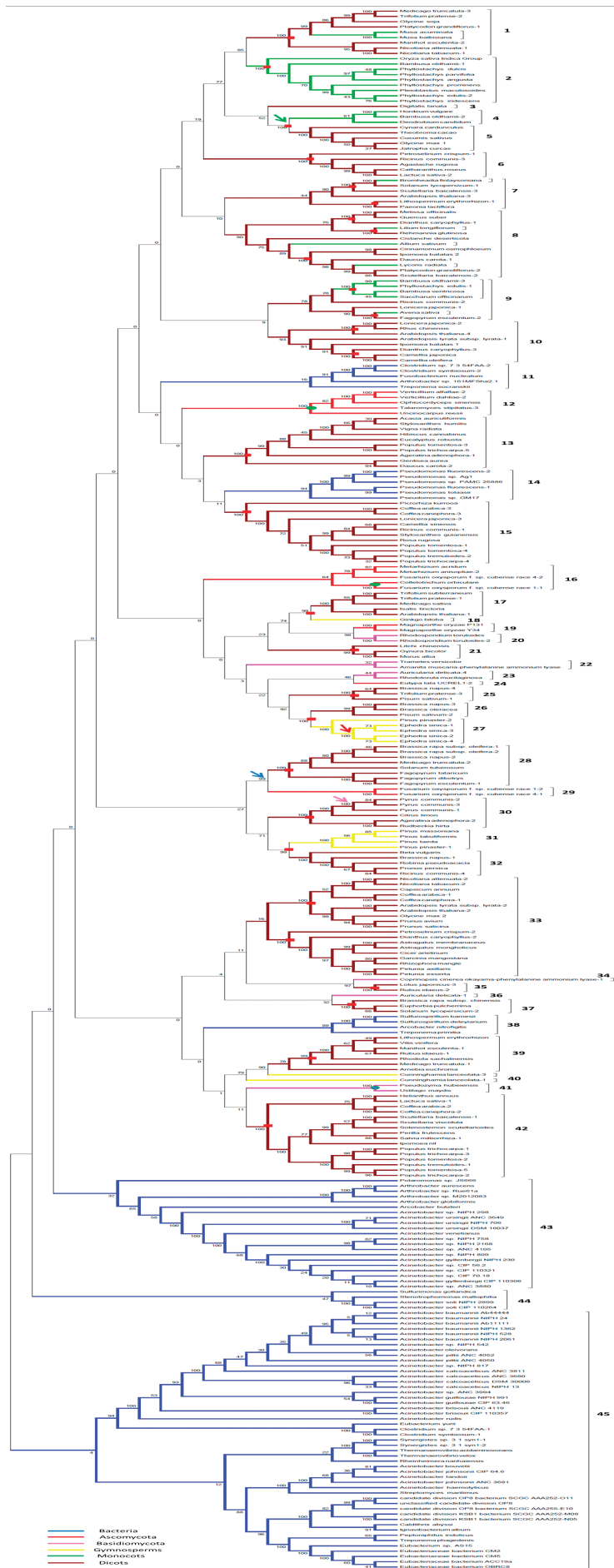P.O.Box: 71345-1583
E-mail: hemmatish@sums.ac.ir

Figure 1. Unrooted phylogenetic tree generated for 369 aligned amino acid sequences of PAL by the UPGMA method. The percentages of replications in which the taxa are related in the bootstarp test (2000 replicates) are shown next to the branches.

evolution of gymnosperms, as a phylum of seed plants, they have experienced large environmental and distributional changes, back to the Mesozoic era (13). Some gene families are substantially larger in gymnosperms than in angiosperms (14), indicating the importance of gene duplication as an important mechanism for genome expansion in conifers. It seems that large multi-gene families are correlated with conifer genome size (15).

The *pal* gene family in plants is likely to be expanded by gene duplication, including tandem duplication, segmental duplication, and whole-genome duplication (16). For example, seven *pal* genes are arranged in tandem in two duplication blocks in cucumber, indicating that the *pal* gene family may have expanded in the cucumber genome via tandem duplications (17). In contrast, there are more than 20 clustered *pal* genes in the tomato genome, indicating an unusually active duplication event (18). There are a few members of *pal* gene family in many species, such as *Arabidopsis* (19), and there are a dozen or more copies in other species, such as tomato (18) or potato (20). The members of *pal* genes express in response to different environmental conditions or in different growing tissues and organs.

Microorganisms also use some of the pathways for metabolism of phenylalanine in a similar way as animals and plants (21). Although *pal* is mentioned as a higher plant enzyme, it is also found in a few bacteria, such as *Streptomyces maritimus* and *Sorangium cellulosum* where PAL is involved in benzoyl-CoA biosynthesis (22,23) and *Streptomyces verticillatus* for biosynthesis of cinnamamide (24). On the other hand, PAL enzyme only exists in certain fungi (25).

Accurate relation between genes in the gene family in evolutionary genomics are defined as orthologs or paralogs (26). The genes that have been diverged as the result of speciation events are orthologs, while those that have been diverged following duplication events are paralogs (27). As a result, ortholog genes which are originated from a single gene in the last common ancestor of a series of present species, have often retained identical biological functions.

In this study, we present a phylogenetic analysis of the *pal* protein family in prokaryotes, fungi and plants and a comparison of the *pal* families between these kingdoms and, in details, between different groups of plant.

2. Materials and methods

In this study, screening for nucleotide translation products was performed using the sequence database at NCBI. Based on exhaustive preliminary phylogenetic analysis, 295 representative taxa were chosen for final tree construction and 369 aligned amino acids were selected for analysis. Alignments of multiple amino acid sequences were carried out using the Clustal W tool in the MEGA 5.2 program. An unrooted phylogenetic tree of the nonredundant *pal* proteins was

then produced with 2,000 bootstrap trials and UPGMA statistical method using MEGA 5.2.

3. Results

Based on preliminary exhaustive phylogenetic analyses, 295 representative taxa were chosen for final tree construction. These sequences are well conserved and allowed the selection of 369 aligned amino acid positions for analysis. The resulting unrooted phylogenetic tree is shown in Figure 1. The prokaryotic part of the tree is not congruent with species phylogeny, indicating extensive gene duplications, losses, and horizontal gene transfer (HGT) within bacteria. The bacterial *pal* are not monophyletic between clades 11, 14, 38, 43, 44 and 45 (Figure 1), and it would be interesting to characterize them.

The eukaryotic *pal* cluster contains exclusively orthologues from plants and fungi but no other eukaryotic lineage and these form some monophyletic sister groups (Figure 1). The plant *pal* clusters include only members from land plants. We found no orthologues in available *pal* sequences data from the red and green algae lineages which branch prior to the divergence of land plants within the phylum Plantae. The monophyly of plants *pal* orthologues in some species of clades and subclades 1-10, 13, 15, 17, 21, 25-28, 30-33, 35, 37, 39 and 42, N (bootstrap value) >99 (Figure 1, shown as red square) indicates that in each clade or subclade, members have a single origin and derive from a gene that was already present in their ancestors. The monophyly of the fungal *pal* orthologues in few species of clades 12 and 16, N>99 (Figure 1, shown as green circle) indicates that in each clade or subclade, members have a single origin and derive from a gene that was present in their common ancestor, and possibly earlier.

Ascomycota and basidiomycota PALs in Figure 1 (clades and subclades 12, 16, 19, 20, 22-24, 29, 34, 36 and 41) don't show meaningful relationship (N>99) and reveal the variation in the *pal* protein sequences among fungi.

In Figure 1, gymnosperm PALs are distributed among distinct clades or subclades 18, 27, 31 and 40. Subclades 18, 27 and 31 are monophyletic with angiosperms in the clades and subclade 17, 26 and 32. The high bootstrap values (N>99) provide support for the organisation of these gymnosperm genes into these three distinct clusters. It seems that the four distinct *pal* genes detected in Ephedra sinica N>99 (Figure 1, clade 26, shown as red arrow) are derived from a recent tandem duplication event. Interestingly, similar to Ephedra sinica PALs, most of the plant PALs clustered by species so that PALs within one species are more closely related to each other than to homologs in the other species. The diversification of the PALs occurred separately in each species.

The phylogenetic relationships of the angiosperms

(monocots and dicots), N>99 (Figure 1, shown as red square) has been mentioned above. PALs of Monocots, in subclade 4, show monophylogenecity (N>99) to dicots in subclade 5 (Figure 1, shown as green arrow).

As indicated in Figure 1, PALs of some monocot species such as *Bromheadia finlaysoniana* in clade 7, *Lilium longiflorum*, *Allium sativum* and *Lycoris radiata* in clade 8, *Avena sativa* in clade 9 is scattered between dicots and are monophyletic (N>99) with some species in these clades. In monocots, *Musa cuminata* and *Musa balbisiana* PALs are situated in clade 1 between dicots, N.99 (Figure 1). In dicots, in clade 30, the *Pyrus communis* PALs are clustered with each other and formed a distinct subgroup, N>99 (Figure 1, purple arrow). It seems that these are derived from a recent tandem duplication event.

4. Discussion

During colonization of terrestrial environments by pioneer land plant ancestors, it was crucial that they can associate with fungi and soil bacteria. For example, N2 fixing cyanobacteria started symbiosis with early fungal lineages (28) and land plants. Fungi (Glomeromycota) started arbuscular-mycorrhizal (AM) symbioses with the first land plants (28-31). To assess the origin of *pal* it's reported that *pal* is emerged in bacteria with an antimicrobial role then a member of a pioneer fungal lineage which was existed before the divergence of Ascomycota and Basidiomycota obtained a *pal* via HGT from a soil bacterium through an early symbiosis (28); fungi transferred *pal* to an ancestor of land plants via an ancient AM symbiosis. This was the starting point of the phenylpropanoid pathway development, and the distribution of plants on terrestrial environments. According to our findings the evolutionary relatedness of *pal* orthologues from land plants from clade 28 and fungi from clade 29 indicates a common origin N>99 (Figure 1, shown as blue arrow). However, the phylum Plantae does not share an exclusive ancestor with Fungi (32).

Other directions of gene transfer cannot be excluded. For example gene transfer from a soil bacterium to an ancient land plants occurred through an ancient symbiosis, then from this to an ancestor of Ascomycota and Basidiomycota fungi (or an earlier branching lineage) through an ancient AM symbiosis (30-31). If an endosymbiotic gene transfer was the route of *pal* gene transfere from cyanobacteria to the host nucleus in the ancestor of the phylum Plantae during symbiosis, it is not obvious why that it has been lost multiple times independently in some algal lineages. Another hypothesis indicates that those bacteria that transferred *pal* to the ancestor of land plants were different from the bacteria that transferred *pal* to the ancestor of fungi. Consequently, if a *pal* orthologue was present in the ancestor of all eukaryotes, it would have been subsequently lost in all eukaryotic lineages

to the exception of land plants and fungi (32). Its likely that *pal* is originated in the ancestor of land plants or in the ancestor of fungi, then it had been transferred via HGT between these two phyla (32).

It's reported that gymnosperm *pal* genes have been clustered into three clades (12). Bagal's study shows that the origin of the most ancient clade is estimated to predate the origin of vascular plants and the origin of the other two clades is via gene duplication within a seed-plant ancestor before the divergence of angiosperms and conifers. The phylogeny of the *pal* gene family identified in this study showed four distinctive branching patterns for the gymnosperm clades. The members of each gymnosperm *pal* clade (Figure 1) may include genes encoding *pal* isoforms that have similar functions or are regulated by similar developmental control mechanisms (33). In contrast to angiosperms, gymnosperms have retained a diverse set of *pal* genes distributed among four major clades that arose from gene duplication events.

Duplication events are an important issue in the evolution of the *pal* gene family. Different protein isoforms may express from duplicate copies of genes, or each duplicate copy may have a distinct expression pattern for response to different physiological conditions, such as tissue development or resistance to environmental stresses (34). Duplication events in *Lonicera japonica* 1 and *Lonicera japonica* 2 in clades 9 and 10, *Populus tomentosa* 3 and *Populus stomentosa* 1 & 4 in clade 13 and 15, also in *Trifolium pratense* 1 and *Trifolium pratense* 3 in subclade 17 and 25 (Figure 1) can be identified in the *pal* cladogram.

Tandem duplication events have resulted from a recent duplication occurrence and they may still have overlapping functions (12). Tandem duplication events in *Ephedra sinica* in clade 27 and *Pyrus communis* in clade 30 (Figure 1) can be identified in the *pal* tree. *Pal* enzymes that do not cluster together (Figure 1) are probably encoding *pal* isozymes with unique functional gene.

In conclusion plants had acquired a *pal* gene via HGT from bacteria and fungi. Duplication events and divergence resulted to a diverse set of *pal* genes through evolution.

Conflict of Interest:
None declared.

6. References
1. Jorrin J, Dixon RA. Stress Responses in Alfalfa (Medicago sativa L.): II. Purification, characterization, and induction of phenylalanine ammonia-lyase isoforms from elicitor-treated cell suspension cultures. Plant physiol. 1990; 92: 447-55.

2. Lois R, Dietrich A, Hahlbrock K, Schulz W. A phenylalanine ammonia lyase gene from parsley: structure, regulation and identification of elicitor and light responsive cis-acting elements. EMBO J. 1989; 8: 1641-8.

3. Chaman ME, Copaja SV, Argandona VH. Relationships between salicylic acid content, phenylalanine ammonia-lyase (PAL) activity, and resistance of barley to aphid infestation. J Agric Food Chem. 2003; 51: 2227-31.

4. Veronese FM, Harris J. M. Introduction and overview of peptide and protein pegylation. Adv Drug Deliv Rev. 2002; 54: 453-6.

5. Wang L, Gamez A, Sarkissian CN, Straub M, Patch MG, Han GW, Striepeke S, Fitzpatrick P, Scriver CR, Stevens RC. Structure-based chemical modification strategy for enzyme replacement treatment of phenylketonuria. Mol Genet Metab. 2005; 86: 134-40.

6. Kenrick P, Crane PR. The origin and early evolution of plants on land Nature. 1997; 389: 33-9.

7. Waters ER. Molecular adaptation and the origin of land plants. Mol Phylogenet Evol. 2003; 29: 456-63.

8. Dixon RA, Achnine L, Kota P, Liu C-J, Reddy MSS, Liangjiang Wang L. The phenylpropanoid pathway and plant defence -a genomics perspective. Mol Plant Pathol. 2002; 3:371-90.

9. Joos HJ, Halhbrock K. Phenylalanine ammonia-lyase in potato (SolanumtuberosumL.) genomic complexity, structural comparison of two selected genes, and modes of expression. Eur J Biochem. 1992; 204: 621-9.

10. Whetten RW, Sederoff RR. Phenylalanine ammonia-lyase from loblolly pine: purification of the enzyme and isolation of complementary DNA clones. Plant Physiol. 1992; 98: 380-6.

11. Ferrer JL, Austin MB, Stewart C Jr, Noel JP. Structure and function of enzymes involved in the biosynthesis of phenyl-propanoids. Plant Physiol Biochem. 2008; 46: 356-70.

12. Bagal UR, Leebens-Mack JH, Lorenz WW, Dean J. The phenylalanine ammonia lyase (PAL) gene family shows a gymnosperm-specific lineage. BMC Genomics. 2012; 13 (Suppl 3): S1.

13. Eckert AJ, Hall BD. Phylogeny, historical biogeography, and patterns of diversification for Pinus (Pinaceae): phylogenetic tests of fossil-based hypotheses. Mol Phylogenet Evol. 2006; 40: 166-82.

14. Perry DJ, Furnier GR. Pinus banksiana has at least seven expressed alcohol dehydrogenase genes in two linked groups. Proc Natl Acad Sci USA. 1996; 93: 13020-3

15. Ahuja MR, Neale DB. Evolution of genome size in conifers. Silvae Genet. 2005; 54: 126-37.

16. Cannon SB, Mitra A, Baumgarten A, Young ND, May. The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana. BMC Plant Biol. 2004; 4:10.

17. Shang QM, Li L, Dong CJ. Multiple tandem duplication of the phenylalanine ammonia-lyase genes in Cucumis sativus L. Planta. 2012; 236: 1093-105.

18. Chang A, Lim MH, Lee SW, Robb EJ, Nazar RN. Tomato phenylalanine ammonia-lyase gene family, highly redundant but strongly underutilized. J Biol Chem. 2008; 283:33591–601.

19. Raes J, Rohde A, Christensen JH, Van de Peer Y, Boerjan W. Genome-wide characterization of the lignification toolbox in Arabidopsis. Plant Physiol. 2003; 133: 1051-71.

20. Joos HJ, Hahlbrock K. Phenylalanine ammonia-lyase in potato (Solanum tuberosum L.). Eur J Biochem. 1992; 204: 621-9.

21. Wat CK, Towers GH. Metabolism of the aromatic amino acids by fungi. Recent Adv Phytochem. 1977; 12: 371-432.

22. Bezanson, G.S, Desaty, D, Emes, A.V, Vining, L.C. Biosynthesis of cinnamamide and detection of phenylalanine ammonia-lyase in Streptomyces verticillatus. Can J Microbiol. 1970; 16: 147-51.

23. Hill, A.M, Thompson, B.L, Harris, J. P, Segret, R. Investigation of the early stages in soraphen A biosynthesis. Chem Commun. 2003; 21: 1358-9.

24. Xiang L, Moore BS. Inactivation, complementation and heterologous expression of encP, a novel bacterial phenylalanine ammonia-lyase gene. J Biol Chem. 2002; 277: 32505-9.

25. Moore K, SubbaRao PV, Towers GH. Degradation of phenylalanine and tyrosine by Basidiomycetes. Life Sci. 1967; 62: 29-33.

26. Koonin EV. Orthologs, paralogs, and evolutionary genomics. Annu Rev Genet. 2005; 39: 309-38.

27. Fitch WM. Distinguishing homologous from analogous Proteins. Syst Zool. 1970; 19: 99-113.

28. Parniske M. Arbuscular mycorrhiza: the mother of plant root endosymbioses. Nat Rev Microbiol. 2008; 6: 763-71.

29. Brundeztt MC. Coevolution of roots and mycorrhizas of landplants. New Phytol. 2002; 154: 275-304.

30. Raven JA. The evolution of cyanobacterial symbioses. Proc Royal Irish Academy. 2002; 1: 3-6.

31. Yuan X, Xiao S, Taylor TN. Lichen-Like Symbiosis 600 MillionYears Ago. Science. 2005; 308: 1017-20.

32. Emiliani M, Fondi M, Fani R, Simonetta G. A horizontal gene transfer at the origin of phenylpropanoid metabolism: a key adaptation of plants to land. Biol Direct. 2009; 4:7

33. Pina A, Errea P. Differential induction of phenylalanine ammonia-lyasegene expression in response to in vitro callus unions of Prunus spp. J Plant Physiol. 2008; 165: 705-14.

34. Logemann E, Parniske M, Hahlbrock K. Modes of expression and common structural features of the complete phenylalanine ammonia-lyase gene family in parsley. Proc Natl Acad Sci USA. 1995; 92: 5905-9.