Molecular Phylogenetics and Evolution 49 (2008) 691–701

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Divergent evolution of the thiolase superfamily and chalcone synthase family

Chenguang Jiang, Sun Young Kim, Dae-Yeon Suh *

Department of Chemistry and Biochemistry, University of Regina, 3737 Wascana Parkway, Regina, SK, Canada S4S 0A2

ARTICLE INFO

Article history:
Received 5 July 2007
Revised 14 August 2008
Accepted 1 September 2008
Available online 12 September 2008

Keywords:
Enzyme evolution
Thiolase superfamily
Chalcone synthase family
Active site architecture
Divergent evolution
Gene birth-and-death
Most recent common ancestor

ABSTRACT

Enzymes of the thiolase superfamily catalyze the formation of carbon–carbon bond via the Claisen condensation reaction. Thiolases catalyze the reversible non-decarboxylative condensation of acetoacetyl-CoA from two molecules of acetyl-CoA, and possess a conserved Cys–His catalytic diad. Elongation enzymes (β-ketoacyl-acyl carrier protein synthase (KAS) I and KAS II and the condensing domain of polyketide synthase) have invariant Cys and two His residues (CHH triad), while a Cys–His–Asn (CHN) triad is found in initiation enzymes (KAS III, 3-ketoacyl-CoA synthase (KCS) and the chalcone synthase (CHS) family). These enzymes all catalyze decarboxylative condensation reactions. 3-Hydroxyl-3-methylglutaryl-CoA synthase (HMGS) also contains the CHN triad, although it catalyzes a non-decarboxylative condensation. That the enzymes of the thiolase superfamily share overall similarity in protein structure and function suggested a common evolutionary origin. All thiolases were found to have, in addition to the Cys–His diad, either Asn or His (thus C(N/H)H) at a position corresponding to the His in the C*H*H and C*H*N triads. In our phylogenetic analyses, the thiolase superfamily was divided into four main clusters according to active site architecture. During the functional divergence of the superfamily, the active architecture was suggested to evolve from the C(H)H in archaeal thiolases to the C(N/H)H in non-archaeal thiolases, and subsequently to the CHH in the elongation enzymes and the CHN in the initiation enzymes. Based on these observations and available biochemical and structural evidences, a plausible evolutionary history for the thiolase superfamily is proposed that includes the emergence of decarboxylative condensing enzymes accompanied by a recruitment of the His in the C*H*H and C*H*N triads for a catalytic role during decarboxylative condensation. In addition, phylogenetic analysis of the plant CHS family showed separate clustering of CHS and non-CHS members of the family with a few exceptions, suggesting repeated gene birth-and-death and re-invention of non-CHS functions throughout the evolution of angiosperms. Based on these observations, predictions on the enzymatic functions are made for several members of the CHS family whose functions are yet to be characterized. Further, a moss CHS-like enzyme that is functionally similar to a cyanobacterial enzyme was identified as the most recent common ancestor to the plant CHS family.

1. Introduction

In biological systems, a carbon–carbon bond is formed predominantly by aldol reaction or related Claisen condensation (Walsh, 1979). Nonenzymatically, the aldol reaction occurs between the α-carbon of an aldehyde or ketone and the carbonyl carbon of another aldehyde or ketone, while the Claisen condensation involves the α-carbon of an ester and the carbonyl carbon of another ester, ketone, or aldehyde molecule. In both reactions, the nucleophilic α-anion attacks the electrophilic carbonyl carbon of another molecule, forming a new carbon–carbon bond. In enzymatic reactions, the nucleophilic component in Claisen condensation is the α-anion of acylthioesters. Depending on how to generate the nucleophilic α-anion, there are two types of reactions: non-decarboxylative or decarboxylative. A group of enzymes belonging to the thiolase superfamily (Table 1) catalyze Claisen condensation reactions and share overall similarity in protein structure and function (Dawe et al., 2003; Haapalainen et al., 2006). Thiolases and 3-hydroxyl-3-methylglutaryl-CoA synthase (HMGS, E.C. 2.3.3.10) catalyze non-decarboxylative Claisen condensation reactions and use acetyl-CoA as the nucleophilic extender in carbon–carbon bond formation (Fig. 1). There are two types of thiolases. While thiolase I (3-ketoacyl-CoA thiolase, E.C. 2.3.1.16) catalyzes the removal of an acetyl group from an acyl-CoA in fatty acid β-oxidation, thiolase II (acetoacetyl-CoA synthase, E.C. 2.3.1.9) catalyzes the reverse reaction, the Claisen condensation of two molecules of acetyl-CoA to acetoacetyl-CoA (Kursula et al., 2005). HMGS, meanwhile, catalyzes the condensation of acetyl-CoA with acetoacetyl-CoA to form HMG-CoA in the mevalonate pathway of cholesterol and isoprenoid biosynthesis (Campobasso et al., 2004).

Several members of the thiolase superfamily that are involved in fatty acid and polyketide syntheses catalyze decarboxylative Claisen reactions and use malonyl thioesters as the nucleophilic

* Corresponding author. Fax: +1 306 337 2409.
E-mail address: suhdaey@uregina.ca (D.-Y. Suh).

1055-7903/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2008.09.002

Table 1
Enzymes of the thiolase superfamily

Enzyme	Condensation mechanism	Active site architecture	Distribution	Remarks
Archaeal thiolases	Non-decarboxylative	C(H)H	Archaea	
Thiolase I (3-ketoacyl-CoA thiolase)	Non-decarboxylative	C(N/H)HC	Bacteria, eukaryota	Reversible, degradative
Thiolase II (acetoacetyl-CoA synthase)	Non-decarboxylative	C(N/H)HC	Bacteria, eukaryota	Reversible, biosynthetic
HMGS (3-hydroxy-3-methylglutaryl-CoA synthase)	Non-decarboxylative	*E*CHN	Archaea, bacteria, eukaryota	Irreversible
KAS I (β-ketoacyl-acyl-carrier-protein (ACP) synthase I)	Decarboxylative	CHH	Bacteria, eukaryota, some archaea	
KAS II (β-ketoacyl-acyl-carrier-protein (ACP) synthase II)	Decarboxylative	CHH	Bacteria, eukaryota	
KS domain of PKS (β-ketosynthase domain of polyketide synthase)	Decarboxylative	CHH	Bacteria, fungi	
KAS III (β-ketoacyl-acyl-carrier-protein (ACP) synthase III)	Decarboxylative	CHN	Bacteria, eukaryota	
KCS (β-ketoacyl-CoA synthase)	Decarboxylative	CHN	Plantae	
CHS family (chalcone synthase family)	Decarboxylative	CHN	Bacteria, fungi, Plantae	

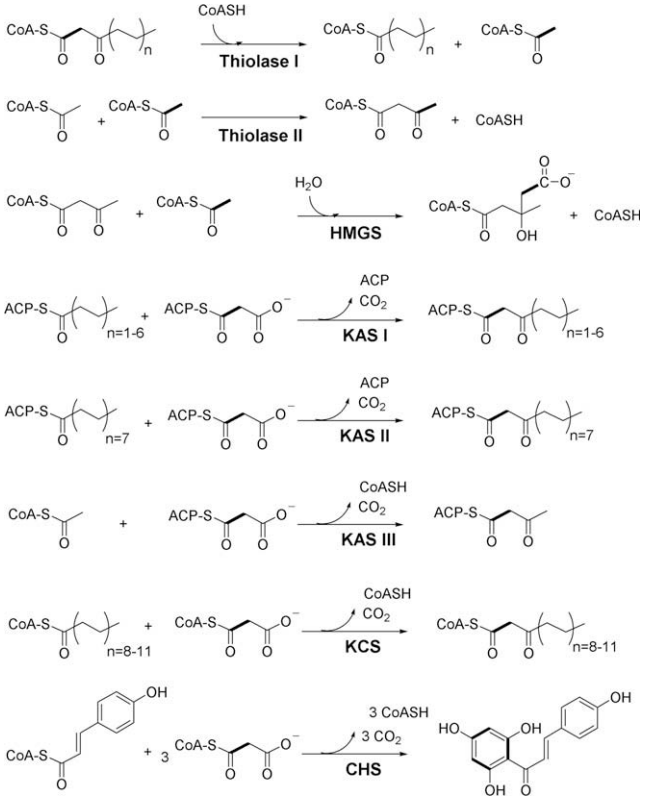

Fig. 1. Reactions catalyzed by enzymes of the thiolase superfamily.

extender (Fig. 1) (Heath and Rock, 2002). Thus, β-ketoacyl-acyl-carrier-protein (ACP) synthase I (KAS I, E.C. 2.3.1.41) and KAS II (E.C. 2.3.1.179) catalyze the chain elongation step of dissociated (type II) fatty acid synthesis in plants and bacteria, and add a C_2 unit derived from malonyl moiety to the growing fatty acid chain (von Wettstein-Knowles et al., 2000). These elongation enzymes use acyl thioesters of ACP as substrates, but they differ in substrate specificity in that palmitoleoyl-ACP is the preferred substrate for KAS II. Closely related to these enzymes are the β-ketosynthase (KS) domains of polyketide synthases (PKS) that are responsible for the condensation of a C_2 unit onto the growing polyketide chain in bacteria and fungi (Hopwood and Sherman, 1990).

Another group of enzymes, often called initiation enzymes, also catalyze decarboxylative condensation reactions. KAS III (E.C. 2.3.1.180) initiates fatty acid biosynthesis in plants and bacteria by catalyzing the first elongation step from C_2 to C_4 (Tsay et al., 1992), whereas 3-ketoacyl-CoA synthase (KCS) of fatty acid elongase system converts C_{18} fatty acids to C_{20} and C_{22} fatty acids in the biosyntheses of wax and seed storage lipids in plants (Blacklock and Jaworski, 2006). Chalcone synthase (CHS, E.C. 2.3.1.74) and related enzymes of the CHS family also catalyze decarboxylative condensation. Members of the CHS family differ in the choice of starter substrate, to which a C_2 unit from malonyl-CoA is condensed at a time, the number of condensation reactions catalyzed, and the cyclization mechanism by which the intermediate oligoketide is cyclized. As a result, the CHS family collectively synthesizes a variety of natural products including chalcones, stilbenes, and phloroglucinols (Austin and Noel, 2003) (Fig. 1). Well studied plant enzymes belonging to this family (also referred to as type III polyketide synthases) include stilbene synthase (STS) (Austin et al., 2004), benzalacetone synthase (Abe et al., 2003), benzophenone synthase (Liu et al., 2003), phlorisovalerophenone synthase (VPS) (Paniego et al., 1999), acridone synthase (Lukacin et al., 2001), and biphenyl synthase (Liu et al., 2007). Although the CHS family was once thought to be plant specific, a growing number of related enzymes has been found in bacteria and fungi (Moore et al., 2002; Gross et al., 2006).

All thiolase superfamily enzymes have an active site Cys that acts as nucleophile and anchors the growing chain during condensation reactions. Thiolases also possess a conserved His that activates the Cys nucleophile (thus, CH diad) and another Cys that acts as a base and deprotonates the α-carbon of acetyl-CoA (Mathieu et al., 1997; Modis and Wierenga, 2000). The decarboxylative condensing enzymes can be divided into two groups according to active site architecture. KAS I (Olsen et al., 2001), KAS II (Huang et al., 1998) and the KS domain of PKS all have invariant Cys and two His residues (CHH triad), while a Cys–His–Asn (CHN) triad is found in KAS III (Davies et al., 2000), KCS (Blacklock and Jaworski, 2006) and the CHS family (Ferrer et al., 1999). HMGS also contains the CHN triad in addition to a Glu that acts as a base during non-decarboxylative condensation (Campobasso et al., 2004). Crystal structures of enzymes of the thiolase superfamily have been determined to reveal that, in spite of very low overall sequence homology, these enzymes exhibit a common fold with a five-layered α-β-α-β-α core structure, where α is made of two α-helices and each β is a five-stranded, mixed β-strands (Fig. 2). The structural similarity is not limited in overall fold, but the active site residues are located in similar or identical spatial configurations to an extent that the catalytic triads can be closely superposed at corresponding residues (Fig. 2) (Dawe et al., 2003; Haapalainen et al., 2006). Thus, the enzymes of the thiolase superfamily are mechanistically related and share conserved structural features, suggesting a common evolutionary root. Several phylogenetic studies have been reported on the phylogeny of eukaryotic thiolases (Peretó et al., 2005), bacterial PKS (Jenke-Kodama et al.,

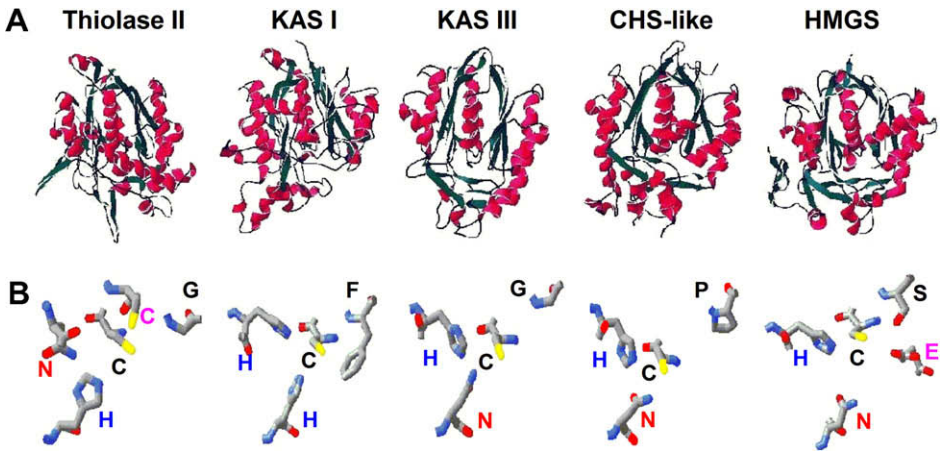

Fig. 2. The comparison of the tertiary structures and the active site architectures of the five enzymes of the thiolase superfamily. (A) The monomer structure is shown in ribbon drawing with α-helices in red and β-strands in green. (B) Active site residues of the enzymes. The active site nucleophile Cys is positioned at the center. It can be seen that the His residues of the thiolase C*H* diad and the KAS I CH*H* triad are equivalent to the Asn residues of the CH*N* triads of KAS III, CHS and HMGS. The location of the base in non-decarboxylative condensing enzyme (Cys in thiolase and Glu in HMGS, both in pink) differ in thiolase and HMGS. The Pro in the CHS GFGPG loop and the corresponding residues in other enzymes are also shown. The structures shown are those of *Zoogloea ramigera* thiolase II (1dlu), *E. coli* KAS I (2buh), *E. coli* KAS III (1hn9), *Mycobacterium tuberculosis* CHS-like enzyme, pks18 (1ted), and *Enterococcus faecalis* HMGS (1x9e).

2005), the KS domains of type I PKS (Kroken et al., 2003), non-reducing KS domains from fungi (Schmitt et al., 2005), and the CHS family in plants (Tropf et al., 1994) and in bacteria and fungi (Gross et al., 2006). However, these studies were limited to individual members of the thiolase superfamily, and studies on evolutionary relationships among different groups of the enzymes that employ different condensation mechanisms (non-decarboxylating vs decarboxylating) and possess different active site architectures (CHH vs CHN triads) have not been reported.

Gross et al. (2006) conducted an extensive phylogenetic analysis of more than 50 bacterial and fungal enzymes of the CHS family. Their results suggested that the plant CHS family might have evolved from bacterial enzymes. Recent phylogenetic studies on the plant CHS family showed that CHSs and non-CHS enzymes form separate clusters (Brand et al., 2006; Liu et al., 2007; Wanibuchi et al., 2007). This separate grouping of CHSs and non-CHSs was explained by an evolutionary scenario involving gene duplication and gene loss (Huang et al., 2004) and raised an interesting possibility that putative function of a new member of the family can be predicted on the basis of the enzyme's location in phylogenetic trees. Despite recent progress, many aspects of the evolution of the plant CHS family, for example, the ancestry of the family remain to be clarified. This study was carried out to provide a holistic view on the evolutionary history of the thiolase superfamily and further insights into the evolution of the plant CHS family.

2. Materials and methods

2.1. Sequence retrieval and alignment

Amino acid sequences were retrieved from NCBI (http://www.ncbi.nih.gov), TIGR Rice (http://www.tigr.org/tdb/e2k1/osa1/), and JGI Physcomitrella (http://www.shake.jgi-psf.org/Phypa1/Phypa1.home.html) databases. For the thiolase superfamily, enzyme sequences of representative organisms from different taxa were selected for phylogenetic analysis (Table S1 in Supplementary data). Whenever possible, different enzymes from a representative species of each taxonomic category, for example, *Escherichia coli* for bacteria and *Arabidopsis thaliana* for Plantae, were selected. For the CHS family, BlastP searches were conducted using the protein sequences of *Physcomitrella patens* CHS (ABB84527) and *Synechococcus* sp. CHS-like enzyme (CAE07508) as the query sequences. Over 900 plant sequences were examined to remove redundancies and partial sequences, and the remaining sequences were grouped according to taxa. In order to eliminate similar sequences from closely related species, preliminary phylogenetic trees were generated with the sequences belonging to the same class (e.g., Liliopsida) or subclass (e.g., eurosids I) using Bayesian inference methods (Ronquist and Huelsenbeck, 2003). Subsequently selected for detailed analysis were 105 plant sequences including all non-CHS plant enzymes and representative CHSs from the same species from which a non-CHS was found or, if not available, CHSs from most closely related species (Table S1 in Supplementary data).

Amino acid sequence alignments for the thiolase superfamily were created both with the entire sequences and with the active site sequences using ClustalX (Thompson et al., 1997) with the BLOSUM series of alignment matrices, and the alignments were manually refined. For the CHS family, the entire sequences were aligned with ClustalX (1.83) after the protruding N- and C-ends were truncated to be flush with the sequence of *Medicago sativa* CHS2 (P30074). Sequence alignments are provided as Supplementary data (Figs. S1 and S2).

2.2. Phylogenetic analysis

Phylogenetic analyses of the thiolase superfamily were conducted both with the entire sequences and with the active site sequences using Bayesian inference and bootstrap Neighbor-joining (NJ) methods. The Bayesian trees were constructed using the MrBayes (v. 3.1.2) program (Ronquist and Huelsenbeck, 2003). The "mixed" amino acid analysis model was used, allowing model jumping between various fixed-rate amino acid models including Poisson and BLOSUM62 during the analysis process. Markov chain Monte Carlo analysis (MCMC) was performed for 1 million generations with four independent chains, and the Markov chain was sampled every 100 generations. At the end of the run, average standard deviation of split frequencies and potential scale reduction factor (PSRF) were used as evidence for the state of convergence. All trees that were sampled before reaching the convergence state (typically 250,000 generations) were discarded, while the remaining trees were used to construct a consensus tree and to calculate posterior clade probabilities. The NJ trees were built with Poisson correction or Jones–Taylor–Thornton (JTT) model as implemented in the MEGA3.1 program (Kumar et al., 2004). The confidence of the tree topology was assessed by a bootstrap set of 1000 replicates. Similarly, phylogenetic analysis of the plant

CHS family was performed both with the Bayesian inference (MrBayes) and with the bootstrap NJ method implemented in ClustalX (seed = 111, $n = 1000$) or MEGA3.1 (Poisson or JTT model). Trees were displayed using MEGA3.1.

Ancestral active site sequences at interior nodes in the phylogenetic tree of the thiolase superfamily were inferred using the hierarchical Bayesian approach as implemented in the MrBayes program (Hall, 2006).

3. Results and discussion

3.1. Sequence alignments

Overall sequence similarity among the enzymes of the thiolase superfamily from a single organism and across taxa is very low. For example, enzymes from *A. thaliana* show sequence identity of lower than 30%; 29% between thiolase I and thiolase II, 29% between KAS I and mitochondrial KAS II, and 19% between KAS III and CHS. Despite low similarity in sequence, it has been noted that the members of the thiolase superfamily share similarity in overall structure and spatial arrangements of the active site residues (Fig. 2) (Dawe et al., 2003; Haapalainen et al., 2006). This is also evident in structure-based alignment of active site sequences (Fig. 3). In the thiolase superfamily, the nucleophile Cys and the His and Asn (or His) residues that play catalytic roles (see below) are located in the loops that converge at the active site pocket. When these residues are aligned, a similar pattern of secondary structures emerges surrounding these residues, although there is little sequence similarity. Also found at the active site is the Gly-rich loop located close to C-end. In KAS I and KAS II, this loop has a consensus sequence of SFGFG of which the second Phe serves as a gate keeper (Huang et al., 1998). The corresponding sequence in the CHS family, however, is GFGPG of which the Pro was shown to constitute an important part of cyclization scaffold (Suh et al., 2000a). The Cys in thiolases and the Glu in HMGS that serve as the base during non-decarboxylative condensation reactions are conserved only in respective enzymes (Fig. 3).

When active site sequences of KCS were similarly aligned, their predicted secondary structures around the active site residues compare favorably with those of other enzymes (Fig. 3). KCS is predicted to be a membrane protein with two N-terminal, membrane-spanning domains (Todd et al., 1999). Although crystal structure of KCS has not been reported, our structure-based sequence alignment presented in Fig. 3 suggests that active site architecture (and possibly overall fold of the non-membrane domain) of KCS be similar with those of other enzymes of the superfamily.

3.2. Divergent evolution of the thiolase superfamily

For phylogenetic analysis of the thiolase superfamily, representative sequences from all levels of the Tree of Life were collected, which include 8 thiolases, 12 KAS I, KAS II, and KS/PKS, 6 HMGS and 3 putative HMGS from archaea, 7 KAS III, 6 KCS, and 15 enzymes of the CHS family. In addition, 3 archaeal sequences annotated as thiolases and 1 archaeal thiolase homolog were included. With a total of 61 sequences thus selected, phylogenetic trees were constructed with the entire sequences and with the partial sequences surrounding the actives site residues. When the entire sequences were used, the Bayesian method and the bootstrap NJ method with Poisson and JTT models yielded phylogenetic trees virtually identical in tree topology. The Bayesian tree is shown in Fig. 4.

The phylogenetic tree of the thiolase superfamily diverges into four major clusters when the archaeal thiolase homolog from *Aeropyrum pernix* is placed as an outgroup. The *A. pernix* thiolase homolog of unknown function shows 25–30% sequence identity to other archaeal thiolases examined, but apparently does not contain the Cys–His catalytic diad characteristic of thiolases. The four main clusters in the tree are formed by archaeal thiolases, thiolases from bacteria and eukaryotes, the elongation enzymes (KAS I, KAS II, and KS/PKS), and the initiation enzymes (HMGS, KAS III, KCS, and the CHS family). The archaeal thiolase cluster appears ancestral to other enzymes of the superfamily, and non-thiolase members of the superfamily are clustered according to their active site architecture. Thus, the elongation enzymes possessing the CHH triad form a single cluster, while the initiation enzymes with the CHN triad form another cluster. The initiation enzyme cluster is further divided into two groups, one of HMGS, the other containing KAS III, KCS, and the CHS family. Therefore, in this cluster, the four enzymes are divided in accordance with condensation mechanism. HMGS is present in archaea, eukaryotes and a number of Gram-positive bacteria. In the phylogenetic tree, archaeal HMGSs are found in the same cluster with HMGSs from other organisms. One possible explanation is that HMGS evolved in ancestral cells before the divergence of archaea and the rest of the life forms. This would agree with the fact that the archaea possess unique membrane

Fig. 3. Structure-based sequence alignment of representative enzymes of the thiolase superfamily. Sequences surrounding the active site residues of a bacterial thiolase II from *Zoogloea ramigera*, a bacterial KAS I from *E. coli*, a bacterial HMGS from *Enterococcus faecalis*, a bacterial KAS III from *E. coli*, a plant KCS (KCS FAE1) from *Arabidopsis thaliana*, a bacterial CHS-like enzyme (pks18) from *Mycobacterium tuberculosis*, and a plant CHS from *Medicago sativa* are aligned to show similar secondary structures in the active site regions. No crystal structure of KCS has been reported; thus, secondary structure of *A. thaliana* KCS was predicted based on multiple sequence alignment as implemented in PredictProtein program (Rost et al., 2004). Included in this analysis were the KCS sequences from *A. thaliana* (AAA70154), *Brassica napus* (AAA96054), *Sinapis arvensis* (AAX58617), *Marchantia polymorpha* (AAO48425), *Zea mays* (CAC01441), and *Dunaliella salina* (AAK11266). Underlined are conserved active site residues as well as the proline in the CHS-signature GFGPG loop and the corresponding residues. α-Helices are in red, and β-strands are in blue.

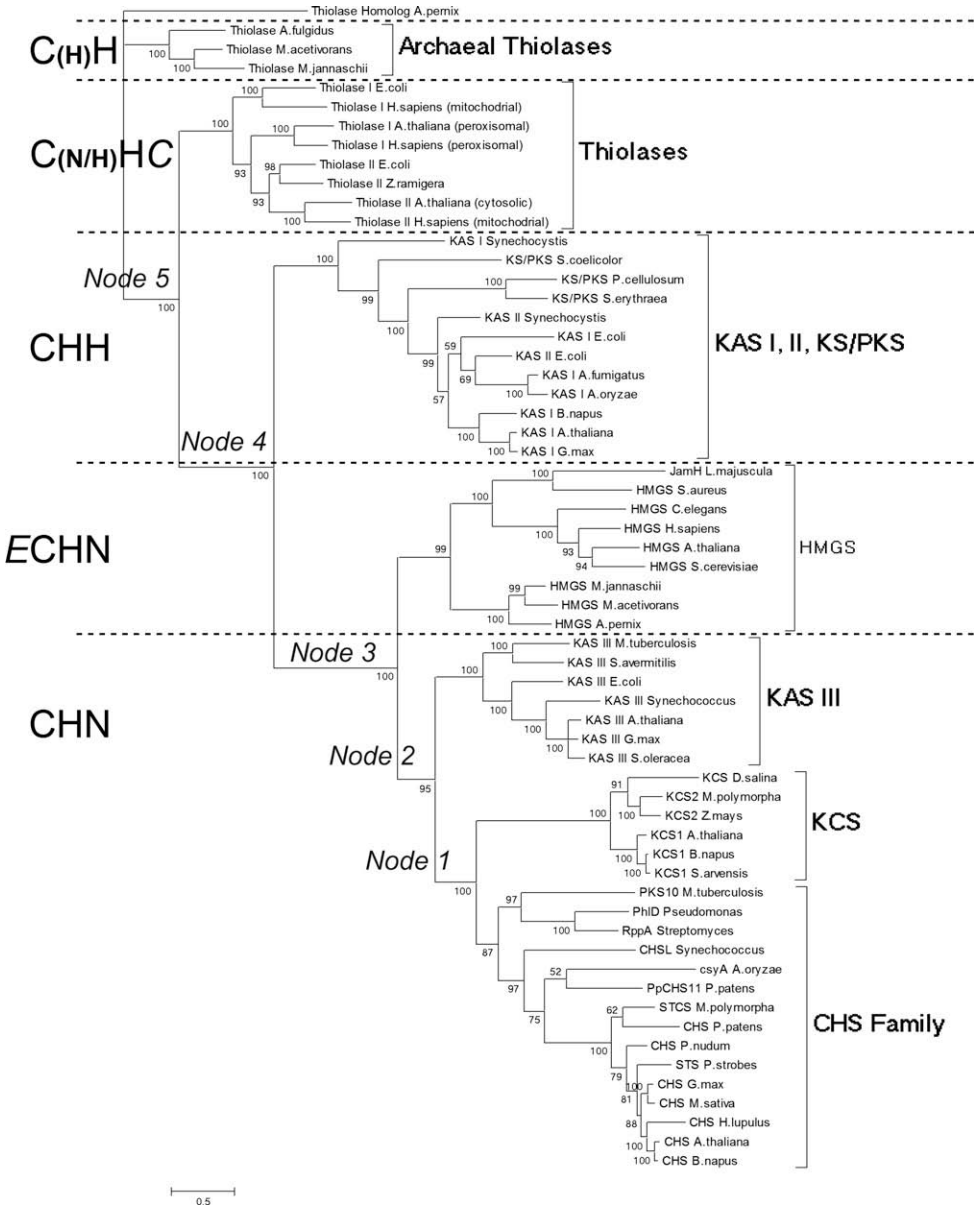

Fig. 4. Amino acid sequence-based phylogeny obtained for selected enzymes of the thiolase superfamily, inferred by Bayesian analysis. Numbers above branches indicate the probability of each clade in the tree, and branch length indicates number of expected substitutions per site. At the end of 1 million generations, average standard deviation of split frequencies reached close to 0.0063, and PSRF approached 1.000 for all the run parameters, indicating that the state of convergence was obtained. While the "mixed" amino acid model was employed, the BLOSUM62 model (Henikoff and Henikoff, 1992) solely contributed to the consensus tree (posterior probability of 1.000). An archaeal thiolase analog from *Aeropyrum pernix* served as an outgroup during tree construction. The active site architectures of different enzymes are indicated. The bases responsible for non-decarboxylative condensations (Cys in thiolases and Glu in HMGS) are in italic. Accession Nos. of the enzymes included in the analysis are available as Supplementary data (Table S1).

phospholipids of which hydrophobic side chains are generally isoprenoid derivatives instead of long-chain fatty acids. HMGS catalyzes the second reaction in the mevalonate pathway for isoprenoid biosynthesis and, thus, was likely present in early archaeal cells. More significantly, HMGSs do not cluster with thiolases that share a common mechanism (non-decarboxylative condensation) but cluster with other initiation enzymes that share the same type of catalytic triad (Fig. 4). The close relationship of HMGS and KAS III has been implied from a striking structural semblance between these two enzymes (Campobasso et al., 2004). It is noted that the Bayesian and the bootstrap NJ (Poisson) trees constructed with amino acid sequences surrounding the active site residues (the base Glu of HMGS, the CH(N/H) triad, and the base Cys of thiolases) are virtually identical with the trees built with the entire sequences except a few minor difference in branch patterns. These active site sequence-based trees are provided along with sequence alignment in Supplementary data (Fig. S2).

It is generally accepted that enzymes of secondary metabolism have been derived from pre-existing enzymes, ultimately from those in primary metabolism *via* gene duplication and mutations. For example, synthases, hydroxylases, and reductases of the flavonoid pathways may have evolved from the primary metabolism (Kubitzki, 1987; Firn and Jones, 2000). Our data also lend support to this proposition in that the CHS family appeared to have diverged from the KAS III lineage of primary metabolism. This evolutionary relationship of KAS III and the CHS family becomes more evident when phylogenetic trees of the thiolase superfamily from a single organism (e.g., *A. thaliana*) are constructed. Regardless of the phylogenetic method used, the CHS family invariably is linked to KAS III (Fig. S3 in Supplementary data).

3.3. Functional divergence and active site architecture of the thiolase superfamily

The phylogenetic tree (Fig. 4) illustrates how the active site architecture may have evolved during the functional divergence of the superfamily. As mentioned above, all members of the superfamily contains a nucleophilic Cys to which a starter substrate is attached via a thioester bond prior to condensation. Thiolases also contain an invariable His residue at active sites (His348 in *Zoogloea ramigera* thiolase II, Fig. 3). In yeast thiolase I, the His residue is within hydrogen bonding distance from the nucleophilic Cys sulfur atom and was thus proposed to act as a base to activate the nucleophile Cys (Mathieu et al., 1997). A similar role for the His was also proposed for *Z. ramigera* thiolase II (Modis and Wierenga, 2000). Thiolase also contains an additional Cys at the active site (Cys378 in *Z. ramigera* thiolase II, Fig. 3) that serves as a base to deprotonate the methyl group of the second acetyl-CoA molecule for non-decarboxylative condensation. In archaeal thiolase sequences the Cys base is not found, suggesting either that another unidentified residue acts as the base or that these archaeal enzymes employ different mechanisms. Although archaeal thiolases have been scarcely studied, a kinetic analysis of a purified halophilic thiolase I suggested that the archaeal thiolase may have a different catalytic mechanism (Liu et al., 2002).

Thiolases were thought to differ from the rest of the superfamily in that they possess a catalytic diad of CH. However, our close inspection revealed that all known archaeal thiolases contain a His and non-archaeal thiolases contain either an Asn or a His at a position corresponding to the His in the C*H*H or C*H*N triads of the non-thiolase enzymes. For example, *E. coli* thiolase I and human mitochondrial thiolase I contain a His, while other thiolases in Fig. 4 possess an Asn. When a total of 84 non-archaeal thiolases that are annotated as either thiolase I or II were examined, Asn was found in 14 thiolase II and 10 thiolase I sequences, whereas His was found in 50 thiolase I and 10 thiolase II sequences. Thirty-four thiolase sequences were selected based on annotated functions and taxonomical distribution and subjected to phylogenetic analysis. In the Bayesian inferred phylogeny, the Asn-possessing thiolases appear basal to the His-possessing thiolases (Fig. S4, Supplementary data). In the NJ tree, the Asn-possessing and His-possessing thiolase groups tend to form separate clusters (Fig. S4). All the His-possessing thiolases form a separate cluster with some of the Asn-possessing thiolases, while the remaining Asn-possessing thiolases form another cluster. In both trees, the archaeal thiolases serve as a basal cluster. This suggests that the Asn and His residues have been evolutionarily conserved.

As most of these enzymes are annotated without biochemical characterization, whether this choice of Asn or His at the active site is of biochemical significance is currently unknown. In the literature, this Asn or His residue has been proposed to play mainly a structural role in thiolases. Asn316 of *Z. ramigera* thiolase II was found to form a part of active site hydrogen bonding network, and similar hydrogen bonds are also suggested for thiolase I (Modis and Wierenga, 1999). Kursula et al. (2005) also found that in the crystal structure of human cytosolic thiolase II a water molecule is connected to the side chain of Asn321 and the nucleophilic Cys via hydrogen bonds. The authors proposed that the Asn321-a water molecule unit plays a catalytic role as a part of oxyanion hole during condensation. Although the functional role(s) of the Asn or His residue remains to be further investigated, it is clear that the active site architecture of thiolases can be described as C(H)H for archaeal thiolases and C(N/H)H*C* (the base C in italic) for non-archaeal thiolases to better reflect the evolutionary perspective.

Our phylogenetic analysis and available biochemical evidences indicate that, during divergent evolution of the superfamily from reversible, non-decarboxylative condensation to irreversible, decarboxylative condensations, three distinct developments had occurred. Firstly, the chemistry of more efficient decarboxylative condensation evolved from thermodynamically less favourable non-decarboxylative condensation. This was achieved in part by coupling decarboxylation of malonyl unit to the formation of a new carbon–carbon bond. Decarboxylation of malonyl-CoA is energy-releasing, as the synthesis of malonyl-CoA from acetyl-CoA requires one molecule of ATP. Secondly, the Asn/His residue of thiolases (e.g., Asn316 of *Z. ramigera* thiolase II or His343 of *E. coli* thiolase I) was fixed to His in non-thiolase enzymes of the superfamily (e.g., His298 of *E. coli* KAS I (C*H*H) and His244 of *E. coli* KAS III (C*H*N)) (Figs. 2 and 3). Lastly, the His of the thiolase C(N/H)*H* diad (His348 of *Z. ramigera* thiolase II) is replaced with Asn in the CH*N* triad-containing initiation enzymes, while it remained the same in the CH*H* triad-containing elongation enzymes. Although most of the Bayesian-inferred ancestral sequences at internodes have lower probabilities due to sequence divergence, ancestral sequences at the key positions agree with the above scenario (Fig. S5, Supplementary data). The ancestral amino acids at the first and second position of the triad (*C*(*N*/*H*)(N/H)) at all nodes (nodes 1–5 in Figs. 4 and S5) are Cys and His, respectively, while the ancestral amino acid at the third position of the triad (C(N/H)(*N*/*H*)) is His at nodes 4 and 5 or Asn at nodes 1 and 2. At node 3, where the Asn was introduced at the third position, His is inferred at lower probability (<80%).

It should be noted that the role of individual catalytic residue in the CHH and CHN triads has changed in different enzymes, reflecting the fact that different enzymes use slightly different mechanisms (Fig. 5). Briefly, the second His in the triads was proposed

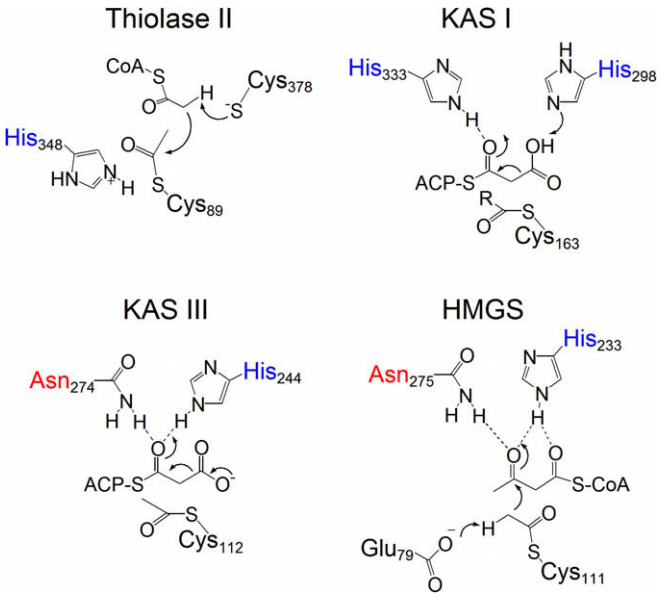

Fig. 5. Proposed roles for the catalytic residues of the enzymes in the thiolase superfamily. Each individual enzyme may use slightly different mechanism with catalytic residues playing different roles. In *Z. ramigera* thiolase II, His348 acts as a base to activate the nucleophile Cys89, and the base Cys378 deprotonates acetyl-CoA molecule for non-decarboxylative condensation (Modis and Wierenga, 2000). A similar mechanism was also proposed for yeast thiolase I (Mathieu et al., 1997). In *E. coli* KAS I, His298 plays a role in decarboxylation of the malonyl residue and extracts a proton from the carboxylic acid leaving group (Olsen et al., 2001). In KAS III-catalyzed reaction, both His244 and His274 function as an oxyanion hole for transition state stabilization during condensation reaction (Davies et al., 2000). His233 and His275 of *S. aureus* HMGS function similarly as oxyanion hole (Theisen et al., 2004), while His233 was also proposed to activate the nucleophilic Cys111 (Campobasso et al., 2004). Although not depicted correctly in this figure for simplicity, the thioester binding pockets of the enzymes are spatially similar relative to the active site residues. (Haapalainen et al., 2006).

to function as a base to activate the nucleophile Cys for HMGS (Campobasso et al., 2004) and CHS (Jez and Noel, 2000; Suh et al., 2000b). Hence, the His in the C*H*N triads of HMGS and CHS is structurally analogous to the Asn/His but functionally analogous to the His in thiolases of the C(N/H)H diad. On the contrary, the same His residues in the C*H*H triads of KAS I and KAS II were postulated to play a role in decarboxylation of the malonyl residue rather than in activating the active site nucleophile Cys (Olsen et al., 2001; McGuire et al., 2001; Zhang et al., 2006). Recent models suggested that His298 in *E. coli* KAS I act as a general base to extract a proton from the carboxylic acid leaving group (Fig. 5), while His303 in *Streptococcus pneumoniae* KAS II deprotonate a structured water molecule for nucleophilic attack on the C3 atom of malonate. Meanwhile, the His and Asn in the CH*H* and CH*N* triads seem to function similarly. His333 in *E. coli* KAS I (Olsen et al., 2001), His337 in *S. pneumoniae* KAS II (Zhang et al., 2006), Asn274 in *E. coli* KAS III (Davies et al., 2000), Asn275 in *Staphylococcus aureus* HMGS (Theisen et al., 2004), and Asn336 in *Medicago sativa* CHS (Jez et al., 2000) were all proposed to function as an oxyanion hole for stabilization of the charge on the oxo group of malonyl unit during condensation reaction. In KAS III-, HMGS-, and CHS-catalyzed condensation reactions, this stabilizing effect on the oxyanion during transition state was suggested to be also aided by the His of the triad. During evolution, individual catalytic machinery has been refined and optimized to extents that different catalytic residues are not readily replaced with their counterparts. For examples, the H333N mutant of *E. coli* KAS I had only 1.6% of the condensation activity of the wild-type enzyme (Price et al., 2001), and the N424H mutant of KCS and the N336H mutant of CHS were completely inactive (Ghanevati and Jaworski, 2002; Jez et al., 2000).

How the functional and structural divergence in the thiolase superfamily came to effect is an interesting subject in enzyme evolution. Some aspects of the divergence can reasonably be explained. During the evolution of decarboxylative condensing enzymes, the His residue of the triads (C*H*H and C*H*N) must have been recruited to catalyze the decarboxylation of malonyl units (at least in KAS I and KAS II). Non-decarboxylative condensation reaction requires a base to abstract a proton from the attacking acetyl unit. The base is Cys in thiolase and Glu in HMGS, and this difference may have been accidental. The fact that these bases are located at different positions in the respective active sites (Fig. 2) is readily understood in terms of different chemistry involved. The acetyl unit from which a proton is abstracted is the second substrate acetyl-CoA molecule in thiolase, while in HMGS the acetyl unit is already bound on the active site Cys (Fig. 5). During the evolution of HMGS from thiolase-like ancestor, it would have been easier to have a new base on the opposite side of the active site rather than to keep the same base and rebuild other parts of the active site to accommodate this different chemistry.

From the phylogenetic relationship revealed in this study and results obtained from previous structural and functional studies, a plausible evolutionary scenario can be put forward (Fig. 6). An ancestral enzyme that likely catalyzed non-decarboxylative condensation or a similar reaction would first have emerged in the ancestor cells. This enzyme had a catalytic CH diad and a His that played mainly a structural role at the active site (the C(H)H architecture). The chemistry of reversible non-decarboxylative condensation evolved to include a base (Cys) in thiolases, while the structural His was replaced with an Asn in some thiolases (the C(N/H)H*C* architecture). Then the first decarboxylative condensing enzyme evolved, while (i) malonyl-thioesters were recruited as a second substrate to couple the decarboxylation of the malonyl unit to the condensation reaction in order to drive the reaction in the direction of condensation, (ii) the structural His was assigned a catalytic role, and (iii) the base Cys was discarded (the CHH and CHN

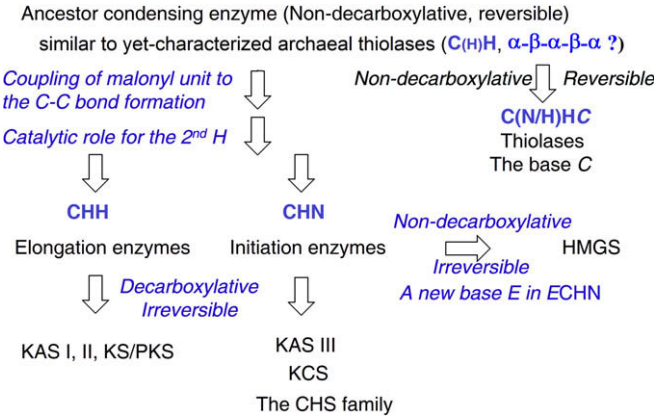

Fig. 6. The proposed evolutionary scenario of the thiolase superfamily.

triads). For some reason(s) yet to be identified, the CHH triad was adopted for chain elongation in KAS I and KAS II, while the CHN triad was instead adopted for KAS III- and HMGS-catalyzed initiation reactions. The thermodynamically favorable formation of HMG-CoA (condensation followed by hydrolysis) may have made decarboxylative condensation energetically unnecessary, and thus during the evolution of HMGS, non-decarboxylative condensation may have been retained with a different base (Glu). The lineage that led to KAS III may have been further diversified to utilize long-chain acyl-CoAs as starter substrates in plant KCS and various acyl-CoA esters including phenylpropanoid-CoAs as starter substrates in the CHS family in bacteria, fungi and plants. The CHS family has then undergone an extreme functional diversification.

3.4. Divergent evolution of the plant CHS family

Flavonoids may have played a significant role as UV protector during the early evolution of land plants. They are found in ~40% of liverworts and ~50% of mosses and ubiquitous in the posteriities of bryophytes including the ferns and higher plants (Markham, 1990). The development of CHS that catalyzes the first reaction of flavonoid biosynthesis should have paralleled the evolution of flavonoids. In most plant species, CHS and related non-CHS enzymes are encoded by multigene families that are products of gene duplication. Multiple copies of *chs* genes have been found in many species including *Petunia, Gerbera*, and *Ipomoea* species (Koes et al., 1989; Helariutta et al., 1996; Durbin et al., 2000). In addition, a growing number of diverse non-CHS enzymes are being found in different, unrelated plants. For example, STS has been found in whisk fern (*Psilotum nudum*), *Pinus* species, peanut (*Arachis hypogaea*), *Vitis* species, and rhubarb (*Rheum tataricum*), and VPS has been characterized from whisk fern and hop (*Humulus lupus*).

Both Bayesian and NJ phylogenetic trees constructed with over 100 sequences of the plant CHS family showed separate clustering of non-CHSs and CHSs and many poorly supported nodes. The NJ tree (Fig. 7) built using ClustalX was considered superior to other trees as the sequences from primitive plants were generally located basal to the sequences from higher plants. In the NJ tree, most angiosperm CHSs form one group and non-CHSs and enzymes of unknown functions form another group with many sub-branches, while bryophyte and gymnosperm enzymes form their own clusters. The observed separate clustering of CHSs and non-CHSs of angiosperms suggests that massive gene duplication/diversification of *chs* genes leading to various non-CHSs may have occurred before the divergence of the ancestor of angiosperms from gymnosperms. During subsequent evolution of

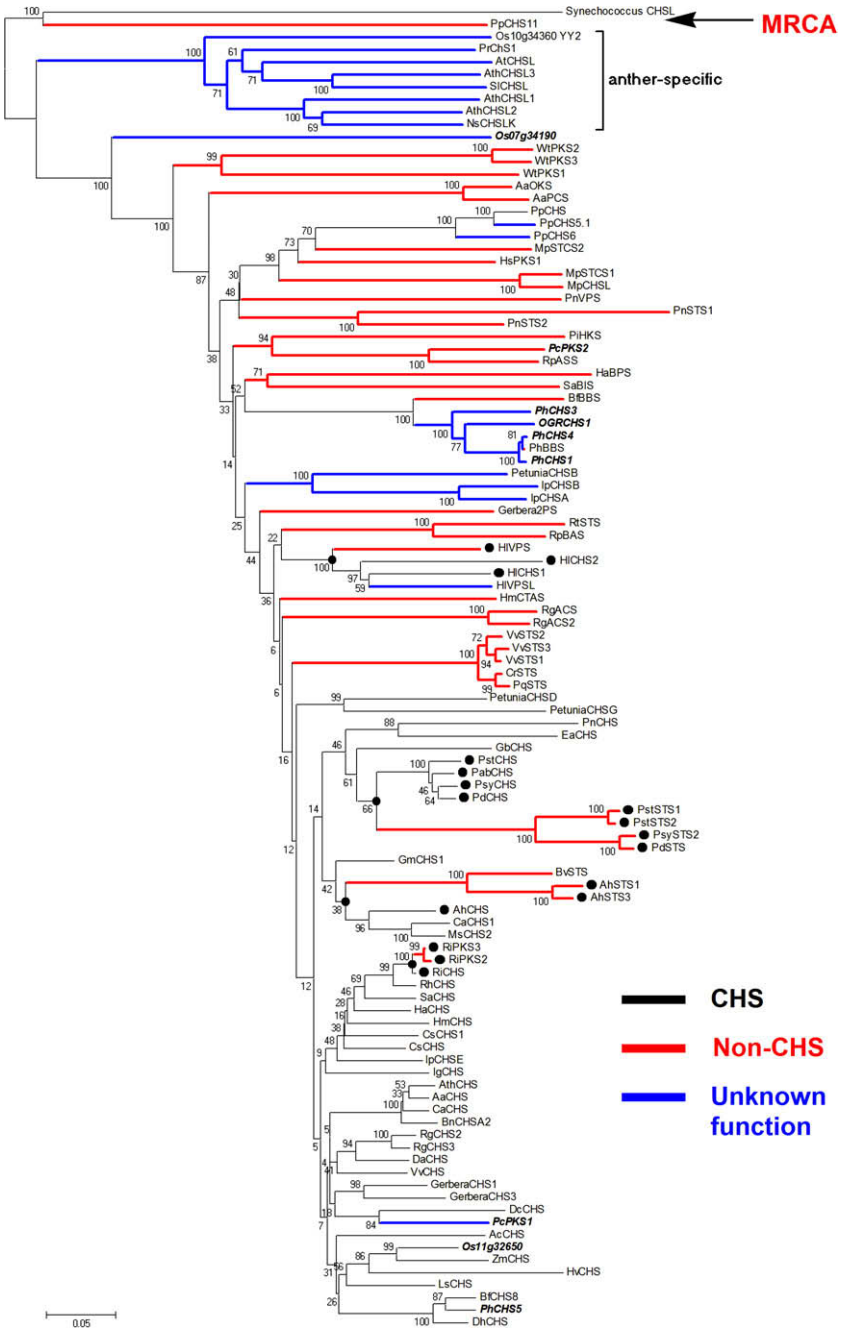

Fig. 7. Neighbor-joining phylogeny of selected enzymes of the plant CHS family. The NJ tree was built using ClustalX (1.83) that implements Kimura's method to correct for multiple substitutions. Numbers above branches indicate the bootstrap value of each clade in the tree. A cyanobacterial CHS-like enzyme from *Synechococcus* sp. WH8102 was also included in the analysis. Those non-CHSs that cluster with CHSs from the same or closely related plants are indicated with dots. PpCHS11 that appears as the most recent common ancestor (MRCA) to the plant CHS family is also indicated by an arrow, and those enzymes whose functions are predicted in the text are shown in bold italic. Accession Nos. of the enzymes included in the analysis are available as Supplementary data (Table S1).

angiosperms, deletion of non-CHS genes may have occurred, as dictated by each lineage's strategy in adaptation to environment. Non-CHS functions could have been abandoned in many species in favor of more advanced secondary metabolism that produce more potent metabolites. Exceptions to this general grouping of CHSs and non-CHSs of angiosperms are found in the enzymes of *Humulus lupulus* (Rosales), *A. hypogaea* (Fabales), and *Rubus idaeus* (Rosales) (marked with dots in Fig. 7). Thus, *H. lupulus* VPS, *A. hypogaea* STS, and *R. ideaeus* PKSs are found in same clusters with CHSs from the same plants, and these enzymes are likely the products of recent, independent gene duplication events. Similarly, phylogenetic evidence suggests that STSs of the order Fabales (*A. hypogaea* STS and *Bauhinia variegata* STS) have evolved from Fabales CHSs by gene duplication and diversification, independently from other STSs such as *Pinus* STSs or *Vitis* STSs (Tropf et al., 1994).

Based on a phylogenetic analysis of a smaller number of angiosperm enzymes, Huang et al. (2004) invoked repetitive gene duplication/loss of angiosperm *chs* genes and independent evolution of certain non-CHS enzymes. Thus, this study expands Huang's conclusions to the entire plant kingdom. It should be also noted that many branches in the non-CHS group have low bootstrap values, and some lineages in the tree may not be a true phylogeny. This

makes it difficult to study the evolutionary relationship among non-CHS enzymes and could be attributed in part to the fact that these non-CHS enzymes are from distantly related species. Currently, non-CHS enzymes are found dispersed throughout the plant kingdom and a few families such as *Asparagales* are overrepresented. Certainly more non-CHS sequences are needed to obtain more accurate phylogenetic tree in this regard.

A CHS-like enzyme (PpCHS11) of the moss, *P. patens*, (Jiang et al., 2006) is consistently positioned next to a cyanobacterial CHS-like enzyme in phylogenetic trees, separated from the cluster of other members of the *P. patens* CHS family of enzymes (Fig. 7). Thus, PpCHS11 appears to be the most recent common ancestor (MRCA) to the plant CHS family so far identified. PpCHS11 shows 25–40% amino acid sequence identity to other plant CHSs. Preliminary studies in our laboratory showed that PpCHS11 is similar to the cyanobacterial CHS-like enzyme in enzyme activity in spite of their low amino acid sequence identity (27%). When assayed *in vitro*, both recombinant PpCHS11 and the *Synechococcus* CHS-like enzyme prefer long-chain acyl-CoA esters such as palmitoyl-CoA as the starter unit. While PpCHS11 condenses long-chain acyl-CoA ester with malonyl-CoA molecules to produce alkylpyrones and possibly pentaketide alkylresorcinols depending on the starter unit, the *Synechococcus* enzyme produces alkylpyrones and tetraketide alkylresorcinols (unpublished data). Therefore, PpCHS11 seems more closely related to the recently characterized fungal alkylresorcylic acid synthase from *Neurospora crassa* (Funa et al., 2007), whereas the *Synechococcus* enzyme seems more similar to the bacterial alkylresorcinol synthase (ArsB) from *Azotobacter vinelandii* (Funa et al., 2006).

Alkylresorcinols and related compounds are found not only in bacteria and fungi but also in alga and plants (Kozubek and Tyman, 1999). However, their biosynthesis had been scarcely studied until recently, and algae remain as a ''missing link' in the CHS evolution because no algal CHS-like enzyme has been isolated. Therefore, research on the biosynthesis of alkylresorcinols in algae may yield important information on the early evolution of the CHS family. It is interesting to note that the apparent MRCA of the plant CHS family is an alkylresorcinol/pyrone synthase that uses long-chain acyl-CoA ester as substrate and closely related to cyanobacterial and fungal alkylresorcinol/pyrone synthases. It is expected that fatty acyl-CoA esters produced by the KAS enzymes were available prior to phenylpropanoid-CoA esters as substrates for ancestral enzymes of the plant CHS family. Notably, alkylresorcinol/pyrone synthase is mechanistically more related to STS than to CHS, since it catalyzes decarboxylative aldol cyclization as STS does (Funa et al., 2006). It has been generally thought that CHS is the prototype enzyme of the family and other non-CHSs including STS have evolved from CHS (Tropf et al., 1994). If the plant CHS family is monophyletic, it seems possible that the true MRCA may not be a CHS.

Located next to PpCHS11 is the cluster made of anther-specific CHS-like enzymes that share <40% amino acid sequence identity with other plant CHSs and are essential for male fertility in plants. This cluster contains both gymnosperm and angiosperm enzymes, suggesting an early specialization of these enzymes before divergence of gymnosperms and angiosperms and a strong evolutionary pressure against mutation (Ageez et al., 2005). Little is known about their functions but a role in the biosynthesis of exine (the outer layer of the wall of a spore or pollen grain) was proposed based on the observation that both *p*-coumaric acid and ferulic acid moieties are found in exine (Atanassov et al., 1998). Austin and Noel (2003) also noticed a similarity in active site residues of these anther-specific enzymes with those of a CHS from *Hordeum vulgare* that utilizes feruloyl-CoA as substrate. Once enzymatic function of an anther-specific CHS-like enzyme is characterized, more can be learned about its role in exine biosynthesis and its evolutionary history. Next to the anther-specific enzymes, *Aloe arborescens* enzymes are found to be basal to the rest of the family including enzymes from bryophytes, suggesting their appearance during earlier evolution of the family (Fig. 7). *A. arborescens* octaketide synthase (AaOKS) catalyzes consecutive condensations of malonyl-CoA with seven other molecules of malonyl-CoA to give an octaketide, while the pentaketide chromone synthase from the same plant (AaPCS) produces a pentaketide from five molecules of malonyl-CoA (Abe et al., 2005a,b). As malonyl-CoA is a common starter substrate of bacterial CHS-like enzymes, it is speculated that these enzymes may represent rare ''living fossils" of earlier enzymes of the CHS family.

Separate clustering of CHSs and non-CHSs in phylogenetic trees raises an interesting question of whether phylogenetic analysis can be used to infer enzymatic function of an enzyme of the CHS family. For bryophyte and gymnosperm enzymes, it will not be possible to deduce enzyme function from phylogenetic analysis since the CHS and non-CHS enzymes from related species form a single cluster, as seen in the moss and pine enzymes. On the contrary, for angiosperm enzymes, an emerging rule is that if an enzyme is found in the non-CHS cluster, it is most likely a non-CHS, and if found in the CHS cluster, it is not necessarily a CHS and may be a non-CHS as in the cases of *A. hypogaea* STS and *H. lupulus* VPS. Based on these arguments, a few predictions can be made. Firstly, a protein from *Polygonum cuspidatum* (PcPKS1), which is annotated as a putative STS in database, is likely a CHS, as it is found in the CHS cluster made of CHSs from related plants (Fig. 7). Meanwhile, PcPKS2, a protein found in the same plant is most likely a non-CHS enzyme, as it is found in the non-CHS cluster next to *Rheum palmatum* aloesone synthase (RpASS). Secondly, among *Phalaenopsis hybrida chs* genes reported by Han et al. (2006), PhCHS5 is likely a CHS, while others (PhCHS1, PhCHS2, PhCHS3, and PhCHS4) are likely non-CHSs. The authors predicted PhCHS3 to be an anther-specific CHS based on its expression in reproductive organs. However, our phylogenetic analysis predicts these four *Phalaenopsis* enzymes, along with another enzyme from an orchid Oncidium Gower Ramsey (OGRCHS1), to be bibenzyl synthases, as they form a single cluster with a known orchid bibenzyl synthase. Thirdly, several predictions can be made for *chs* genes of rice, *Oryza sativa*. The *O. sativa* genome contains as many as 12 expressed *chs* genes (Ouyang et al., 2007). Two of these *chs* gene products (Os11g32650 and Os07g11440) may be CHS as they are found clustered with CHSs from related plants (in Fig. 7, only Os11g32650 is shown). One protein (Os07g22850) is closely related to an anther-specific rice enzyme (Os10g34360 YY2, Hihara et al., 1996) and may have a similar function. The remaining eight rice *chs* genes belong to the non-CHS group, and one of them (Os07g34190) is shown to be basal to all the non-CHS cluster in Fig. 7. Thus, rice may be a rich source for a variety of non-CHS enzymes with interesting evolutionary implications.

A plausible evolutionary scenario for the CHS family that agrees with the observed data is as follows (Fig. 8). The ancestor bacterial CHS-like enzymes containing the CHN triad would have evolved from the KAS III lineage by acquiring abilities to catalyze multiple condensations and to cyclize polyketide intermediates. This would have been achieved in part by modifying the substrate binding pocket and by adopting the GFGPG loop that provides stereo-control during the cyclization (Suh et al., 2000a). These ancestral CHS-like enzymes in bacteria would still have utilized simple acyl-CoA esters such as malonyl-CoA as starter substrates. Ancient CHS-like enzyme(s) in cyanobacteria ancestral to the modern day cyanobacterial alkylresorcinol synthase may have been passed down to the plant lineage about 450 million years ago (Kenrick and Crane, 1997). Gene duplication/diversification events and the evolution of the phenylpropanoid pathway may have yielded diverse enzymes of the CHS family including the first CHSs in liverworts

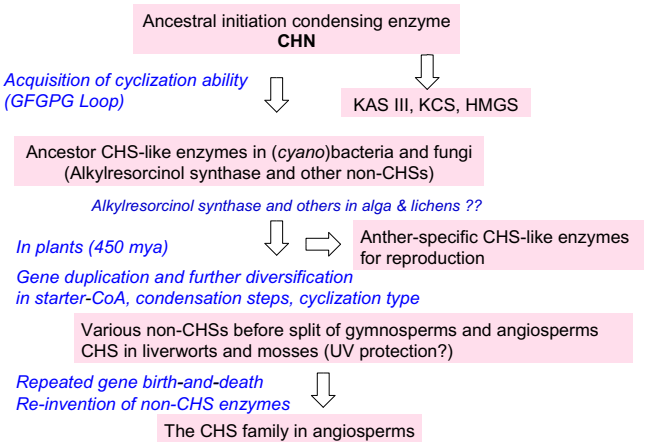

Fig. 8. The proposed evolutionary scenario of the CHS family.

and mosses. The plant CHS family has subsequently evolved through gene duplication-death and re-invention of non-CHS functions that may have occurred at different times throughout evolution. Functional diversification of the family was achieved by starter substrate specialization, by changing the number of condensation reactions, and by utilizing different cyclization mechanisms to provide plants with chemical arsenals to survive in a more complicated environment. Modern plants are thought to have evolved along the evolutionary line of cyanobacteria, unicellular algae, primitive multicellular algae, charophycean algae, lichens, bryophytes, and higher plants. As mentioned above, algae and lichens represent "missing links" in the evolution of the CHS family. It is tempting to suggest that alkylresorcinol/pyrone synthases that link the lineage of cyanobacterial alkylresorcinol synthase and functionally related alkylresorcinol/pyrone synthase (PpCHS11) in moss may still be present in these organisms. Alkylresorcinols are found in algae (Kozubek and Tyman, 1999; Zarnowski et al., 2000) and related alkylresorcinolic acids are well known components of lichen depsides (Elix, 1996), although lichen alkylresorcinolic moieties can conceptually be synthesized by type I PKS systems (Miao et al., 2001).

Acknowledgments

The authors thank W. Chapco, L. Lintott, and an anonymous reviewer for helpful comments on the manuscript. This research was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and the University of Regina.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2008.09.002.

References

Abe, I., Sano, Y., Takahashi, Y., Noguchi, H., 2003. Site-directed mutagenesis of benzalacetone synthase. J. Biol. Chem. 278, 25218–25226.
Abe, I., Oguro, S., Utsumi, Y., Sano, Y., Noguchi, H., 2005a. Engineered biosynthesis of plant polyketides: chain length control in an octaketide-producing plant type III polyketide synthase. J. Am. Chem. Soc. 127, 12709–12716.
Abe, I., Utsumi, Y., Oguro, S., Morita, H., Sano, Y., Noguchi, H., 2005b. A plant type III polyketide synthase that produces pentaketide chromone. J. Am. Chem. Soc. 127, 1362–1363.
Ageez, A., Kazama, Y., Sugiyama, R., Kawano, S., 2005. Male-fertility genes expressed in male flower buds of *Silene latifolia* include homologs of anther-specific genes. Genes Genet. Syst. 80, 403–413.
Atanassov, I., Russinova, E., Antonov, L., Atanassov, A., 1998. Expression of an anther-specific chalcone synthase-like gene is correlated with uninucleate microspore development in *Nicotiana sylvestris*. Plant Mol. Biol. 38, 1169–1178.
Austin, M.B., Noel, J.P., 2003. The chalcone synthase superfamily of type III polyketide synthases. Nat. Prod. Rep. 20, 79–110.
Austin, M.B., Bowman, M.E., Ferrer, J.L., Schroder, J., Noel, J.P., 2004. An aldol switch discovered in stilbene synthases mediates cyclization specificity of type III polyketide synthases. Chem. Biol. 11, 1179–1194.
Blacklock, B.J., Jaworski, J.G., 2006. Substrate specificity of *Arabidopsis* 3-ketoacyl-CoA synthases. Biochem. Biophys. Res. Commun. 346, 583–590.
Brand, S., Holscher, D., Schierhorn, A., Svatos, A., Schroder, J., Schneider, B., 2006. A type III polyketide synthase from *Wachendorfia thyrsiflora* and its role in diarylheptanoid and phenylphenalenone biosynthesis. Planta 224, 413–428.
Campobasso, N., Patel, M., Wilding, I.E., Kallender, H., Rosenberg, M., Gwynn, M.N., 2004. *Staphylococcus aureus* 3-hydroxy-3-methylglutaryl-CoA synthase: crystal structure and mechanism. J. Biol. Chem. 279, 44883–44888.
Davies, C., Heath, R.J., White, S.W., Rock, C.O., 2000. The 1.8 Å crystal structure and active-site architecture of beta-ketoacyl-acyl carrier protein synthase III (FabH) from *Escherichia coli*. Structure 8, 185–195.
Dawe, J.H., Porter, C.T., Thornton, J.M., Tabor, A.B., 2003. A template search reveals mechanistic similarities and differences in β-ketoacyl synthases (KAS) and related enzymes. Proteins 52, 427–435.
Durbin, M.L., McCaig, B., Clegg, M.T., 2000. Molecular evolution of the chalcone synthase multigene family in the morning glory genome. Plant Mol. Biol. 42, 79–92.
Elix, J.A., 1996. Biochemistry and secondary metabolites. In: Nash, T.H., III (Ed.), Lichen Biology. Cambridge University Press, Cambridge, pp. 154–180.
Ferrer, J.L., Jez, J.M., Bowman, M.E., Dixon, R.A., Noel, J.P., 1999. Structure of chalcone synthase and the molecular basis of plant polyketide biosynthesis. Nat. Struct. Biol. 6, 775–784.
Firn, R.D., Jones, C.G., 2000. The evolution of secondary metabolism—a unifying model. Mol. Microbiol. 37, 989–994.
Funa, N., Ozawa, H., Hirata, A., Horinouchi, S., 2006. Phenolic lipid synthesis by type III polyketide synthases is essential for cyst formation in *Azotobacter vinelandii*. Proc. Natl. Acad. Sci. USA 103, 6356–6361.
Funa, N., Awakawa, T., Horinouchi, S., 2007. Pentaketide resorcylic acid synthesis by type III polyketide synthase from *Neurospora crassa*. J. Biol. Chem. 282, 14476–14481.
Ghanevati, M., Jaworski, J.G., 2002. Engineering and mechanistic studies of the *Arabidopsis* FAE1 β-ketoacyl-CoA synthase, FAE1 KCS. Eur. J. Biochem. 269, 3531–3539.
Gross, F., Luniak, N., Perlova, O., Gaitatzis, N., Jenke-Kodama, H., Gerth, K., Gottschalk, D., Dittmann, E., Muller, R., 2006. Bacterial type III polyketide synthases: phylogenetic analysis and potential for the production of novel secondary metabolites by heterologous expression in pseudomonads. Arch. Microbiol. 185, 28–38.
Haapalainen, A.M., Merilainen, G., Wierenga, R.K., 2006. The thiolase superfamily: condensing enzymes with diverse reaction specificities. Trends Biochem. Sci. 31, 64–71.
Hall, B., 2006. Simple and accurate estimation of ancestral protein sequences. Proc. Natl. Acad. Sci. USA 103, 5431–5436.
Han, Y.-Y., Ming, F., Wang, W., Wang, J.-W., Ye, M.-M., Shen, D.-L., 2006. Molecular evolution and functional specialization of chalcone synthase superfamily from *Phalaenopsis* Orchid. Genetica 128, 429–438.
Heath, R.J., Rock, C.O., 2002. The Claisen condensation in biology. Nat. Prod. Rep. 19, 581–596.
Henikoff, S., Henikoff, J.G., 1992. Amino acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. USA 89, 10915–10919.
Helariutta, Y., Kotilainen, M., Elomaa, P., Kalkkinen, N., Bremer, K., Teeri, T.H., Albert, V.A., 1996. Duplication and functional divergence in the chalcone synthase gene family of Asteraceae: eolution with substrate change and catalytic simplification. Proc. Natl. Acad. Sci. USA 93, 9033–9038.
Hihara, Y., Hara, C., Uchimiya, H., 1996. Isolation and characterization of two cDNA clones for mRNAs that are abundantly expressed in immature anthers of rice (*Oryza sativa* L.). Plant Mol. Biol. 30, 1181–1193.
Hopwood, D.A., Sherman, D.H., 1990. Molecular genetics of polyketides and its comparison to fatty acid biosynthesis. Annu. Rev. Genet. 24, 37–66.
Huang, W., Jia, J., Edwards, P., Dehesh, K., Schneider, G., Lindqvist, Y., 1998. Crystal structure of beta-ketoacyl-acyl carrier protein synthase II from *E. coli* reveals the molecular architecture of condensing enzymes. EMBO J. 17, 1183–1191.
Huang, J.-X., Qu, L.-J., Yang, J., Yin, H., Gu, H.-Y., 2004. A preliminary study on the origin and evolution of chalcone synthase (CHS) gene in angiosperms. Acta Bot. Sin. 46, 10–19.
Jenke-Kodama, H., Sandmann, A., Muller, R., Dittmann, E., 2005. Evolutionary implications of bacterial polyketide synthases. Mol. Biol. Evol. 22, 2027–2039.
Jez, J.M., Ferrer, J.L., Bowman, M.E., Dixon, R.A., Noel, J.P., 2000. Dissection of malonyl-coenzyme A decarboxylation from polyketide formation in the reaction mechanism of a plant polyketide synthase. Biochemistry 39, 890–902.
Jez, J.M., Noel, J.P., 2000. Mechanism of chalcone synthase. pK_a of the catalytic cysteine and the role of the conserved histidine in a plant polyketide synthase. J. Biol. Chem. 275, 39640–39646.
Jiang, C., Schommer, C.K., Kim, S.Y., Suh, D.-Y., 2006. Cloning and characterization of chalcone synthase from the moss, *Physcomitrella patens*. Phytochemistry 67, 2531–2540.

Kenrick, P., Crane, P.R., 1997. The origin and early evolution of plants on land. Nature 389, 33–39.
Koes, R.E., Spelt, C.E., Mol, J.N.M., 1989. The chalcone synthase multigene family of *Petunia hybrida* (V30): differential, light-regulated expression during flower development and UV light induction. Plant Mol. Biol. 12, 213–225.
Kozubek, A., Tyman, J.H.P., 1999. Resorcinolic lipids, the natural non-isoprenoid phenolic amphiphiles and their biological activity. Chem. Rev. 99, 1–25.
Kroken, S., Glass, N.L., Taylor, J.W., Yoder, O.C., Turgeon, B.G., 2003. Phylogenomic analysis of type I polyketide synthase genes in pathogenic and saprobic ascomycetes. Proc. Natl. Acad. Sci. USA 100, 15670–15675.
Kubitzki, K., 1987. Phenylpropanoid metabolism in relation to land plant origin and diversification. J. Plant Physiol. 131, 17–24.
Kumar, S., Tamura, K., Nei, M., 2004. *MEGA3*: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief. Bioinform. 5, 150–163.
Kursula, P., Sikkila, H., Fukao, T., Kondo, N., Wierenga, R.K., 2005. High resolution crystal structures of human cytosolic thiolase (CT): a comparison of the active sites of human CT, bacterial thiolase, and bacterial KAS I. J. Mol. Biol. 347, 189–201.
Liu, T., Liu, S.J., Xue, Y., Ma, Y., Zhou, P., 2002. Purification and characterization of an extremely halophilic acetoacetyl-CoA thiolase from a newly isolated *Halobacterium* strain ZP-6. Extremophiles 6, 97–102.
Liu, B., Falkenstein-Paul, H., Schmidt, W., Beerhues, L., 2003. Benzophenone synthase and chalcone synthase from *Hypericum androsaemum* cell cultures: cDNA cloning, functional expression, and site-directed mutagenesis of two polyketide synthases. Plant J. 34, 847–855.
Liu, B., Raeth, T., Beuerle, T., Beerhues, L., 2007. Biphenyl synthase, a novel type III polyketide synthase. Planta 225, 1495–1503.
Lukacin, R., Schreiner, S., Matern, U., 2001. Transformation of acridone synthase to chalcone synthase. FEBS Lett. 508, 413–417.
Markham, K.R., 1990. Bryophyte flavonoids, their structures, distribution, and evolutionary significance. Proc. Phytochem. Soc. 29, 143–159.
McGuire, K.A., Siggaard-Andersen, M., Bangera, M.G., Olsen, J.G., von Wettstein-Knowles, P., 2001. β-Ketoacyl-[acyl carrier protein] synthase I of *Escherichia coli*: aspects of the condensation mechanism revealed by analyses of mutations in the active site pocket. Biochemistry 40, 9836–9845.
Mathieu, M., Modis, Y., Zeelen, J.P., Engel, C.K., Abagyan, R.A., Ahlberg, A., Rasmussen, B., Lamzin, V.S., Kuman, W.H., Wierenga, R.K., 1997. The 1.8 Å crystal structure of the dimeric peroxisomal 3-ketoacyl-CoA thiolase of *Saccharomyces cerevisiae*: implications for substrate binding and reaction mechanism. J. Mol. Biol. 273, 714–728.
Miao, V., Coeffet-LeGal, M.F., Brown, D., Sinnemann, S., Donaldson, G., Davies, J., 2001. Genetic approaches to harvesting lichen products. Trends Biotechnol. 19, 349–355.
Modis, Y., Wierenga, R.K., 1999. A biosynthetic thiolase in complex with a reaction intermediate: the crystal structure provides new insights into the catalytic mechanism. Structure 7, 1279–1290.
Modis, Y., Wierenga, R.K., 2000. Crystallographic analysis of the reaction pathway of *Zoogloea ramigera* biosynthetic thiolase. J. Mol. Biol. 297, 1171–1182.
Moore, B.S., Hertweck, C., Hopke, J.N., Izumikawa, M., Kalaitzis, J.A., Nilsen, G., O'Hare, T., Piel, J., Shipley, P.R., Xiang, L., Austin, M.B., Noel, J.P., 2002. Plant-like biosynthetic pathways in bacteria: from benzoic acid to chalcone. J. Nat. Prod. 65, 1956–1962.
Olsen, J.G., Kadziola, A., von Wettstein-Knowles, P., Siggaard-Andersen, M., Larsen, S., 2001. Structures of β-ketoacyl-acyl carrier protein synthase I complexed with fatty acids elucidate its catalytic machinery. Structure 9, 233–243.
Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., Thibaud-Nissen, F., Malek, R.L., Lee, Y., Zheng, L., Orvis, J., Haas, B., Wortman, J., Buell, C.R., 2007. The TIGR Rice Genome Annotation Resource: improvements and new features. Nucleic Acids Res. 35, D846–D851.
Paniego, N.B., Zuurbier, K.W., Fung, S.Y., van der Heijden, R., Scheffer, J.J., Verpoorte, R., 1999. Phlorisovalerophenone synthase, a novel polyketide synthase from hop (*Humulus lupulus* L.) cones. Eur. J. Biochem. 262, 612–616.
Peretó, J., Lopez-Garcia, P., Moreira, D., 2005. Phylogenetic analysis of eukaryotic thiolases suggests multiple proteobacterial origins. J. Mol. Evol. 61, 65–74.
Price, A.C., Choi, K.H., Heath, R.J., Li, Z., White, S.W., Rock, C.O., 2001. Inhibition of beta-ketoacyl-acyl carrier protein synthases by thiolactomycin and cerulenin. Structure and mechanism. J. Biol. Chem. 276, 6551–6559.
Ronquist, F., Huelsenbeck, J.P., 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Rost, B., Yachdav, G., Liu, J., 2004. The PredictProtein server. Nucleic Acids Res. 32, W321–W326.
Schmitt, I., Martin, M.P., Kautz, S., Lumbsch, H.T., 2005. Diversity of non-reducing polyketide synthase genes in the Pertusariales (lichenized Ascomycota): a phylogenetic perspective. Phytochemistry 66, 1241–1253.
Suh, D.-Y., Fukuma, K., Kagami, J., Yamazaki, Y., Shibuya, M., Ebizuka, Y., Sankawa, U., 2000a. Identification of amino acid residues important in the cyclization reactions of chalcone and stilbene synthases. Biochem. J. 350, 229–235.
Suh, D.-Y., Kagami, J., Fukuma, K., Sankawa, U., 2000b. Evidence for catalytic cysteine–histidine dyad in chalcone synthase. Biochem. Biophys. Res. Commun. 275, 725–730.
Theisen, M.J., Misra, I., Saadat, D., Campobasso, N., Miziorko, H.M., Harrison, D.H., 2004. 3-Hydroxy-3-methylglutaryl-CoA synthase intermediate complex observed in "real-time". Proc. Natl. Acad. Sci. USA 101, 16442–16447.
Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.
Todd, J., Post-Beittenmiller, D., Jaworski, J.G., 1999. *KCS1* encodes a fatty acid elongase 3-ketoacyl-CoA synthase affecting was biosynthesis in *Arabidopsis thaliana*. Plant J. 17, 119–130.
Tropf, S., Lanz, T., Rensing, S.A., Schroder, J., Schroder, G., 1994. Evidence that stilbene synthases have developed from chalcone synthases several times in the course of evolution. J. Mol. Evol. 38, 610–618.
Tsay, J.T., Oh, W., Larson, T.J., Jackowski, S., Rock, C.O., 1992. Isolation and characterization of the β-ketoacyl-acyl carrier protein synthase III gene (fabH) from *Escherichia coli* K-12. J. Biol. Chem. 267, 6807–6814.
von Wettstein-Knowles, P., Olsen, J.G., McGuire, K.A., Larsen, S., 2000. Fatty acid biosynthesis: fatty acid synthase. Biochem. Soc. Trans. 28, 601–607.
Walsh, C., 1979. Enzyme-catalyzed aldol and Claisen condensation. In: Freeman, W.H. (Ed.), Enzymatic Reaction Mechanisms, San Francisco, pp. 741–776.
Wanibuchi, K., Zhang, P., Abe, T., Morita, H., Kohno, T., Chen, G., Noguchi, H., Abe, I., 2007. An acridone-producing novel multifunctional type III polyketide synthase from *Huperzia serrata*. FEBS J. 274, 1073–1082.
Zarnowski, R., Suzuki, Y., Esumi, Y., Pietr, S.J., 2000. 5-*n*-Alkylresorcinols from the green microalga *Apatococcus constipatus*. Phytochemistry 55, 975–977.
Zhang, Y.M., Hurlbert, J., White, S.W., Rock, C.O., 2006. Roles of the active site water, histidine 303, and phenylalanine 396 in the catalytic mechanism of the elongation condensing enzyme of *Streptococcus pneumoniae*. J. Biol. Chem. 281, 17390–17399.