Mol Genet Genomics (2016) 291:739–752
DOI 10.1007/s00438-015-1142-3

ORIGINAL ARTICLE

Evolution of the KCS gene family in plants: the history of gene duplication, sub/neofunctionalization and redundancy

Hai-Song Guo[1] · Yan-Mei Zhang[1] · Xiao-Qin Sun[1] · Mi-Mi Li[1] · Yue-Yu Hang[1] · Jia-Yu Xue[1]

Received: 30 July 2015 / Accepted: 28 October 2015 / Published online: 12 November 2015

Abstract Very long-chain fatty acids (VLCFAs) play an important role in the survival and development of plants, and VLCFA synthesis is regulated by β-ketoacyl-CoA synthases (KCSs), which catalyze the condensation of an acyl-CoA with malonyl-CoA. Here, we present a genome-wide survey of the genes encoding these enzymes, KCS genes, in 28 species (26 genomes and two transcriptomes), which represents a large phylogenetic scale, and also reconstruct the evolutionary history of this gene family. KCS genes were initially single-copy genes in the green plant lineage; duplication resulted in five ancestral copies in land plants, forming five fundamental monophyletic groups in the phylogenetic tree. Subsequently, KCS genes duplicated to generate 11 genes of angiosperm origin, expanding up to 20–30 members in further-diverged angiosperm species. During this process, tandem duplications had only a small contribution, whereas polyploidy events and large-scale segmental duplications appear to be the main driving force. Accompanying this expansion were variations that led to the sub- and neofunctionalization of different members, resulting in specificity that is likely determined by the 3-D protein structure. Novel functions involved in other physiological processes emerged as well, though redundancy is also observed, largely among recent duplications. Conserved sites and variable sites of KCS proteins are also identified by statistical analysis. The variable sites are likely to be involved in the emergence of product specificity and catalytic power, and conserved sites are possibly responsible for the preservation of fundamental function.

Keywords KCS genes · Evolution · Expansion · Functional divergence · Redundancy

Communicated by S. Hohmann.

H.-S. Guo and Y.-M. Zhang contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-015-1142-3) contains supplementary material, which is available to authorized users.

✉ Yue-Yu Hang
hangyueyu@cnbg.net

✉ Jia-Yu Xue
xuejy@cnbg.net

[1] Institute of Botany, Jiangsu Province and Chinese Academy of Sciences, 1 Qianhuhoucun Road, Nanjing 210014, Jiangsu, China

Introduction

A diversity of very long-chain fatty acids (VLCFAs), fatty acids that are longer than 18 carbons (C18), and their derivatives are produced by plant cells, and these molecules play multiple roles in survival and development. For example, VLCFAs are required in all plant cells for the production of sphingolipids, and in specific cells for the synthesis of other VLCFA derivatives such as cuticular waxes and suberin, which serve as the protective barrier of roots and leaves. In addition, VLCFAs serve as energy storage in seeds of some plant lineages like Brassicaceae (Post-Beittenmiller 1996; Haslam and Kunst 2013). Due to the involvement of VLCFAs in diverse and important metabolic pathways, as well as extensive industrial application in the manufacture of plastics, medicine and clothing, the synthesis of VLCFAs has been thoroughly investigated. The VLCFA synthesis pathway consists of a sequence of four core reactions catalyzed by a multi-enzyme acyl-CoA elongase system (Fehling and Mukherjee 1991). The cycle

of reactions involving multiple enzymes begins with the condensation of an acyl-CoA precursor with malonyl-CoA to form β-ketoacyl-CoA, followed by reduction, dehydration, and a second reduction, resulting in the production of an acyl-CoA that is extended by two carbons (Fehling and Mukherjee 1991; Haslam and Kunst 2013). Among the enzyme complexes, β-ketoacyl-CoA synthase (KCS), which catalyzes the condensation reaction, is the rate-limiting enzyme of the entire complex process. KCS determines whether FAs will be elongated and the amount and type of VLCFAs that are produced (Lechelt-Kunze et al. 2003).

The first gene encoding a KCS enzyme was isolated from a screen of mutant *Arabidopsis thaliana* plants deficient in VLCFAs and then genetically characterized (James and Dooner 1990; Lemieux et al. 1990; Kunst et al. 1992). The mutant line showed significantly decreased production of C20:0, C20:1 and C22:1 in seeds, and the gene was, therefore, named fatty acid elongation 1 (FAE1) (Kunst et al. 1992; James et al. 1995). Subsequently, other paralogous genes in *A. thaliana* were also identified and functionally characterized through mutant analysis and heterologous expression in yeast. KCS1 catalyzes the elongation of five VLCFAs, including C20:0, C20:1, C22:0, C24:0 and C26:0 (Trenkamp et al. 2004; Blacklock and Jaworski 2006; Tresch et al. 2012), KCS2 catalyzes the production of C20:0, C24:0 and C26:0 (Trenkamp et al. 2004; Paul et al. 2006; Tresch et al. 2012), and CUT1/CER6 and CER60 both generate products longer than C24 (Trenkamp et al. 2004; Tresch et al. 2012).

Homologous genes from other plant species have also been identified. For example, a jojoba (*Simmondsia chinensis*) gene was able to successfully complement canola (*Brassica napus*) and *A. thaliana* FAE1 mutants (Lassner et al. 1996), and a gene isolated from nasturtium (*Tropaeolum majus*) resulted in an eightfold increase in C22:1 production in transgenic *A. thaliana* seed oil (Mietkiewska et al. 2004). In addition, three genes identified from the early land plant *Marchantia polymorpha* were found to catalyze the elongation of saturated C18:0, C20:0 and C22:0 (Kajikawa et al. 2003a, b). With so many genes isolated and their functional diversity characterized, it is not difficult to infer that the genes encoding β-ketoacyl-CoA synthases should belong to a gene family with multiple members. Naturally, these genes are named as KCS genes, in line with the name of the condensing enzymes.

The functional characterization of KCS genes indicates that the members of this gene family have diverse catalytic activities. Some are related to very specific VLCFAs; for example, AtKCS18/AtFAE1 catalyzes the synthesis of erucic acid (C22:1) (Kunst et al. 1992; James et al. 1995). In contrast, others are related to general products; for example, AtKCS1, −2, −11, −18 and −20 all regulate the synthesis of eicosanoic acid (C22:0) (Trenkamp et al. 2004; Blacklock and Jaworski 2006; Paul et al. 2006; Tresch et al. 2012). However, the mechanism by which these genes evolved different functions remains poorly understood. Evolutionarily, the entire gene family could be traced back to a single gene in the common ancestor of the plant lineage because *Chlamydomonas reinhardtii* and *Volvox carteri*, two algae with sequenced genomes (Merchant et al. 2007; Prochnik et al. 2010), each encode only one KCS gene. However, after plants colonized the terrestrial environment, this single gene expanded into three genes, as isolated from the liverwort *M. polymorpha* (Kajikawa et al. 2003a, b), which represents the first diverging lineage of land plants (Qiu et al. 1998). In *A. thaliana* (an angiosperm), this family expanded to 21 genes (Costaglioli et al. 2005). Thus, KCS genes must have undergone multiple rounds of gene duplication, accompanied by functional divergence and renovation, whereby mutated duplicates were selected by changing environments, resulting in the diversity and specificity of products we observe today.

It is of interest to determine the history of how KCS genes duplicated and diverged from one gene to a gene family during the course of plant evolution. Previous evolutionary studies have attempted to address this question using only 21 *A. thaliana* KCS genes; however, two phylogenetic studies reported conflicting evolutionary relationships among the 21 genes (Costaglioli et al. 2005; Joubes et al. 2008). Clearly, analysis in one species alone is not sufficient to trace the evolutionary history of KCS genes, as the process of duplication and divergence at different key nodes could not be detected without the inclusion of other species. In the present study, we collected KCS genes from 26 sequenced nuclear genomes and two transcriptomes of plants occupying key systematic positions, as well as other functionally characterized genes, and reconstructed the evolutionary history of the KCS gene family in plant lineages to reveal the molecular basis of the functional divergence of different KCS genes. Additionally, the current nomenclature of functionally characterized KCS genes does not follow the evolutionary history, which has resulted in misunderstandings in the relationships of KCS genes among species.

Results

Genome-wide identification and distribution of KCS genes

Through BLAST and HMM (Hidden Markov Model) searches of 26 genomes and two transcriptomes, covering all major clades of land plants, a total of 475 KCS genes were identified (Fig. 1) in the green plant lineage. All genes identified contain two domains, a

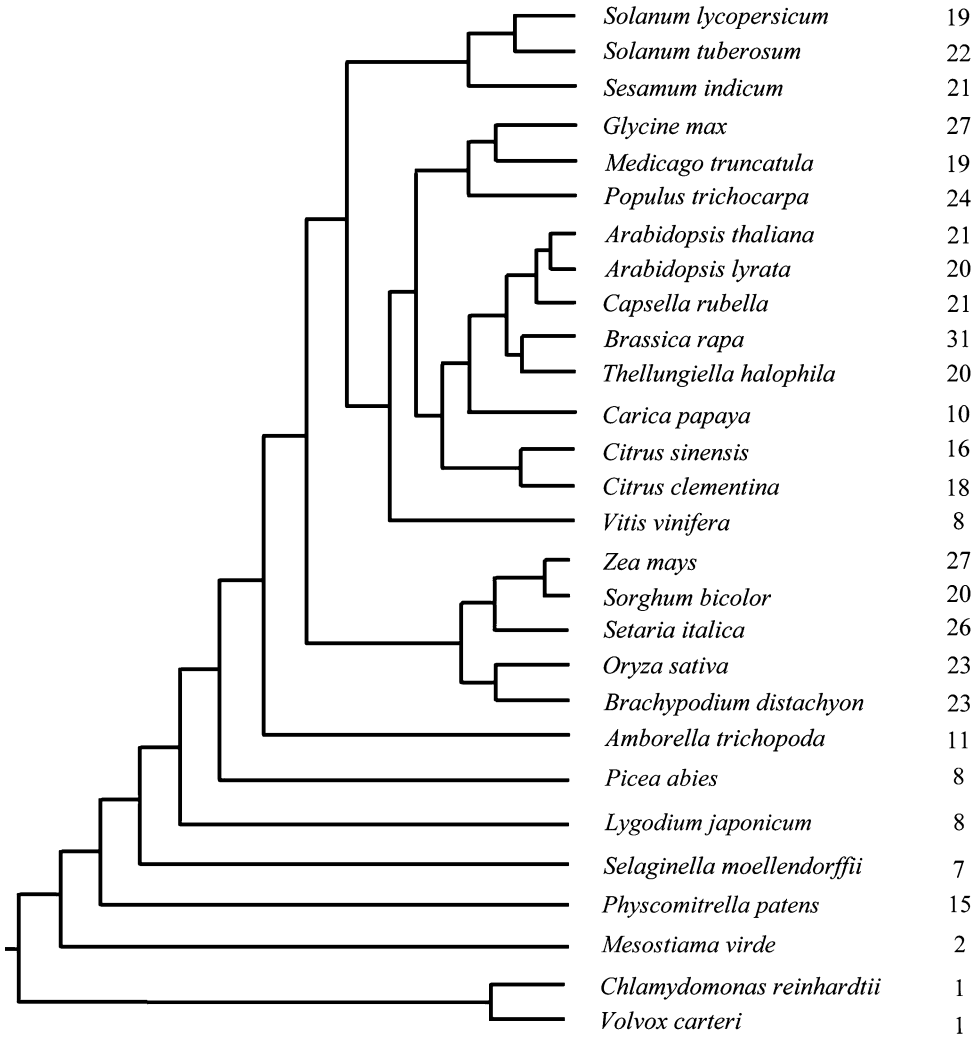

Fig. 1 The phylogeny of the 28 plants used in the study (Angiosperm Phylogeny Website) (Stevens 2015) and the number of KCS genes identified from the genomes or transcriptomes of these plants

3-Oxoacyl-[acyl-carrier protein (ACP)] synthase III C domain and a Type III polyketide synthase-like protein domain. Among these species, two green algae, *C. reinhardtii* and *V. carteri*, possess only one gene each, and two less advanced early land plants—the moss *Physcomitrella patens* and the lycophyte *Selaginella moellendorffii*—possess 15 and 7 genes, respectively, significantly fewer than angiosperms. The gymnosperm *Picea abies* has 8 KCS genes; thus, no obvious expansion is observed in this species. In contrast, angiosperms possess on average 20 genes, with *Brassica rapa* having the most (31 genes), likely due to a recent genome triplication. Nonetheless, exceptions were also observed. For example, three angiosperms, the basal *Amborella trichopoda*, basal rosid grapevine (*Vitis vinifera*) and papaya (*Carica papaya*), all contain only approximately 10 genes. In general, a history of KCS gene expansion during plant evolution was observed, as the gene number continued to increase from the ancestral plant, that is, from green algae to derived angiosperms.

KCS genes are widely distributed throughout genomes but are uneven among chromosomes. In *A. thaliana*, all five chromosomes harbor KCS genes, but seven of the 21 KCS genes are located on Chr (chromosome) 1, the maximum number (33.3 %); Chr 3, with only one gene detected (4.8 %), has the fewest. In other species, KCS genes are not found all chromosomes. For instance, Chr 7, 8 and 12 of rice, Chr 8 of *M. truncatula*, and Chr 1 and 7 of tomato do not carry KCS genes.

Phylogenetic analysis reveals the evolutionary history of plant KCS genes

All KCS genes described above and six functionally characterized genes from the liverwort *M. polymorpha* (Kajikawa et al. 2003a, b), jojoba (Lassner et al. 1996), nasturtium (Mietkiewska et al. 2004) and meadowfoam (*Limnanthes alba*) (Cahoon et al. 2000) were used for our comprehensive phylogenetic analysis; the tree was rooted by two Chlorophyte KCS genes (Fig. 2 and Figure S1). The

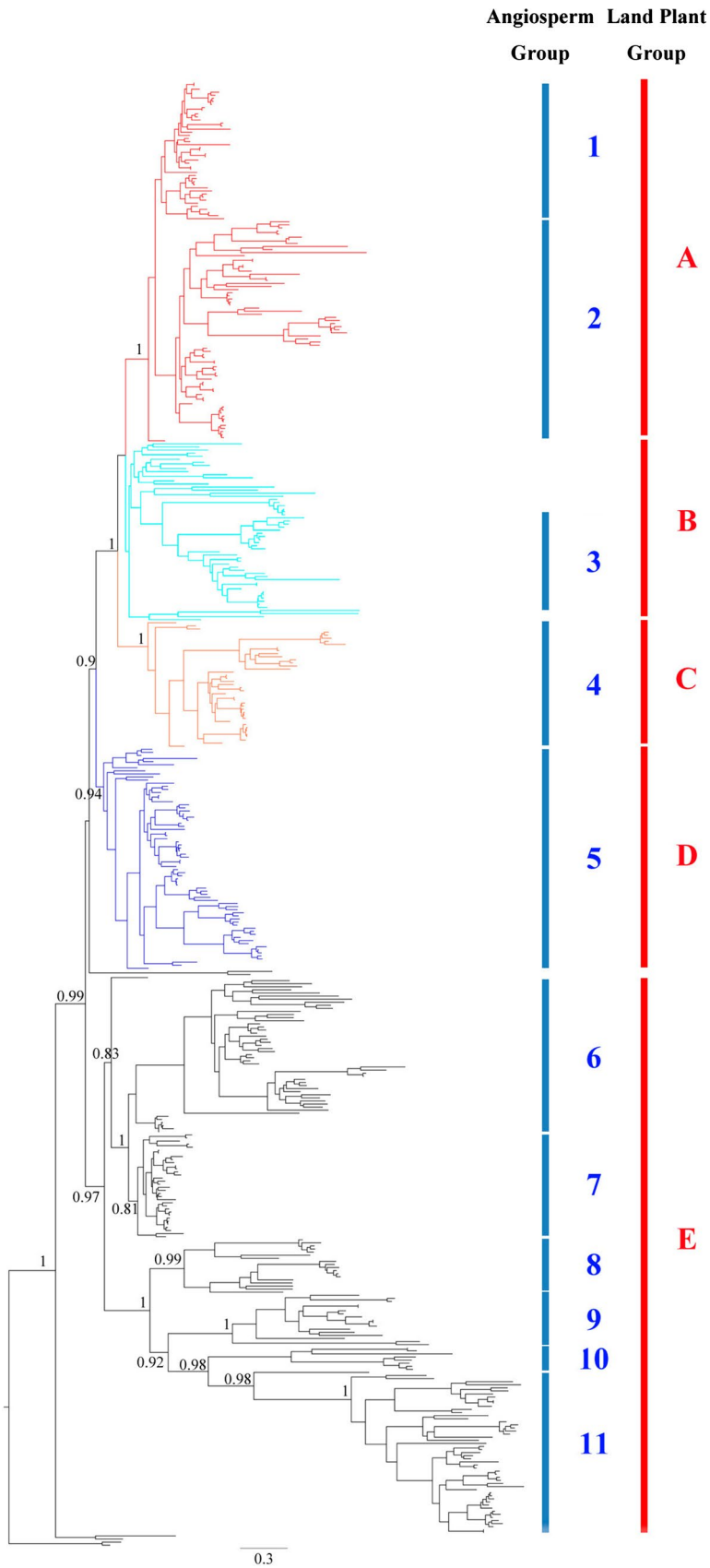

Fig. 2 The phylogenetic tree of KCS gene family in the plant lineage. Five monophyletic land plant groups *A* (*red*), *B* (*light blue*), *C* (*orange*), *D* (*dark blue*) and *E* (*black*) and 11 angiosperm groups (1–11) are indicated. The *scale bars* represent numbers of amino acid substitutions per site. The support values (SH-aLRT value) for deep nodes are indicated

phylogeny demonstrated a very clear process of how KCS gene expansion occurred from a single gene to a family of approximately 20 members in angiosperms.

In the phylogenetic tree, two Charophyta genes, likely recent duplicates in *Mesostigma viride*, are placed near the root and are sister to all other KCS genes. All land plant KCS genes show five basic monophyletic groups (Land Plant Group A, B, C, D and E), either including genes from the liverwort *M. polymorpha* or showing a sister relationship to groups including them, which indicates that at least 5 ancestral KCS genes had already arisen at the time of the origin of land plants. Thus, the shift from water to terrestrial environment resulted in the first dramatic duplication event for KCS genes, a fivefold increase in gene number in the common ancestor of all land plants. However, no other obvious expansion was observed before the origin of seed plants. Utilizing the genes from the basal angiosperm *A. trichopoda*, 11 smaller monophyletic groups could be further identified (Angiosperm Group 1–11), identifying 11 ancestral KCS genes in the common ancestor of all extant flowering plants. Therefore, after the transition from an aquatic to a terrestrial environment, the KCS gene number doubled again as angiosperms emerged. However, the 11 ancestral gene clusters only reached half of the size of the average KCS family in angiosperms. Therefore, further duplication events must have occurred to complement the gene number, and they were found to have occurred at different systematic nodes in different angiosperm lineages. For Brassicaceae, almost all extant KCS genes had emerged early, with the gene number reaching approximately 20 as this family split from other Brassicale families. The exception was *B. rapa*, which represents plants with recent polyploidy events, such that the *B. rapa* genome contains more genes than unduplicated relatives in this family. In another rosid lineage, KCS genes had expanded to nearly 20 at the early stage of Fabaceae evolution. In monocots, the Poaceae had also expanded to over 20 KCS genes upon the split of this grass family. Lineage-specific expansions were common. For example, Brassicaceae species expanded Angiosperm Group 5–7 genes, whereas other species show only one to two genes in this group. Poplar (*Populus trichocarpa*), potato (*Solanum tuberosum*) and tomato (*S. lycopersicum*) underwent more extensive duplications within Angiosperm Group 2, and in Angiosperm Group 6, an ancestral gene was inherited only in Poaceae or probably monocots and expanded to 11 genes in *Brachypodium distachyon*.

Of the five ancient land plant groups, Group E contributed the most to the increase in gene number, developing into 6 monophyletic angiosperm groups. Comparatively, Group A duplicated only once, resulting in two sister groups (Angiosperm Group 1 and 2). The other land plant groups (Group B, C and D) did not duplicate at all within

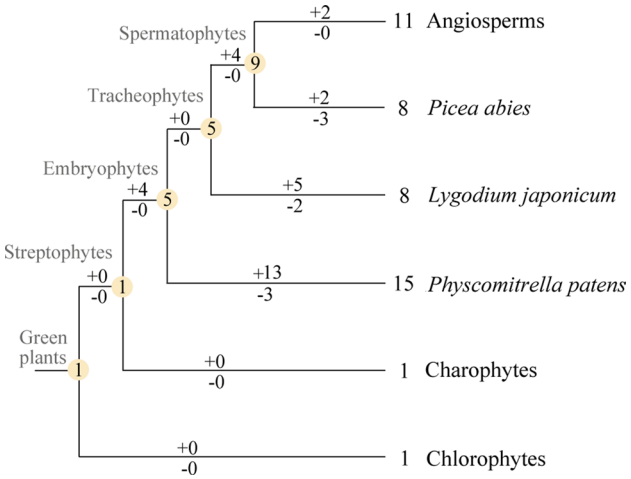

Fig. 3 KCS gene number variation at deep evolutionary nodes

approximately 300 million years from the origin of embryophytes to the split of angiosperms (Qiu et al. 1998, 1999; Lewis and McCourt 2004; Jiao et al. 2011; Wodniok et al. 2011; Zeng et al. 2014). Likely due to a lack of ancient duplications, Group B, C and D contain the fewest number of genes in angiosperms, even at present (53, 47 and 71 genes, respectively), while Group E, which experienced the most rampant ancient duplications, contains the most genes (181 genes), accounting for 38.1 % of this family. More careful observation of the phylogenetic tree revealed that the six angiosperm groups within Land Plant Group E exhibit longer branch lengths, indicating that functional divergence might have occurred among the six groups.

In general, the overall tendency of KCS genes during plant evolution is expansion, but gene loss events are also observed in the phylogenetic tree. These gene loss events occurred at a number of evolutionary nodes, and together with gene duplication, contribute to the variation in gene number among lineages and species. Notable examples of ancient gene loss events could be observed in Angiosperm Group 6 and 8; one was likely inherited only in monocots and lost in the other lineages, and the other was reciprocally lost in monocots but inherited by other lineages. Recent loss events were also typical, as observed in the group sister to Brassicaceae KCS18/FAE1 genes: compared with Brassicaceae, the sister group lost one gene each in *A. thaliana* and *Thellungiella salsuginea*.

Inferring ancient duplication and loss events of KCS genes at deep evolutionary nodes

Our primary survey and phylogenetic analysis presented above suggested an expanding pattern of KCS genes in the green plant lineage, with a relatively smaller number of gene loss events. To elucidate the ancestral variation

Table 1 Conservative sites of KCS genes among five land plant groups and the entire gene family

Frequency	Group A	Group B	Group C	Group D	Group E	All genes
>90 %	197	150	143	205	70	81
>70 %	330	276	277	311	150	198
>50 %	416	354	415	425	246	311

in gene number, gene losses and gains at deep divergence nodes (Fig. 3) were reconciled using Notung software (Stolzer et al. 2012). It is clear that before the emergence of terrestrial plants, the KCS gene remained as a single gene; however, the ancestral single gene was duplicated at the origin of embryophytes and expanded into a family of five genes in the common ancestor of all land plants, though only two genes were inherited by bryophytes. This family appears to have remained silent for a long period, even during the arrival of vascular plants. Nonetheless, the emergence of seed plants brought another sharp increase in KCS gene numbers, with nine ancestral genes present in the genome of the common ancestor of spermatophytes. Of the nine ancestral genes, the gymnosperm *P. abies* inherited only six, whereas angiosperms inherited them all, and quickly expand to 11 genes as this lineage evolved.

Amino acid polymorphism sites and conserved sites among KCS proteins

Using WebLogo (Crooks et al. 2004), we obtained a graphical representation of the pattern of KCS proteins, which provided a richer and more precise description of polymorphism sites and conserved sites (Figure S2). The relative frequency of corresponding amino acids is reflected by the height of the symbol at that position. In general, all KCS proteins appear to be conserved, and they were likely under strong purifying selection over the course of their evolution. However, the N-terminal region appears to be more diverse than the C-terminal region. Detailed statistics collectively revealed 81 sites, showing a frequency >90 % with one type of amino acid at each position; only eight sites are located within the first 300 positions of the alignment, whereas 73 sites are located among the 422 positions behind the alignment. Therefore, these 81 sites represent the most conserved sites of the KCS proteins. Our statistical analysis also found 198 sites with a frequency of >70 % and 311 sites with >50 %. These sites are also clustered at the terminal regions of the protein (Table 1, Table S1). Accordingly, these sites are likely important sites of for catalysis.

All KCS proteins consist of two domains, a 3-Oxoacyl-[acyl-carrier-protein (ACP)] synthase III C domain near the N-terminus and a Type III polyketide synthase-like protein domain at the C-terminus. The former is composed of approximately 300 amino acids, and within it, we found 44 sites showing a frequency >90 %, 125 sites >70 % and 214 sites >50 %. The latter contains only 85 amino acids, and within this domain, 37 sites show a frequency >90 %, 60 sites >70 % and 77 sites >50 %. It is easy to conclude that most conserved sites are distributed within the two domains, which cover all sites (81/81) with a frequency >90 % and the majority of sites >70 % (185/198) and >50 % (291/311). The transmembrane region of the KCS protein is located at the beginning of the protein, consisting of the first 100 amino acids, before the N-terminal domain. This region is the most variable: only 15 sites with a frequency >50 % were observed, but sites of >70 % and >90 % were not found.

With regard to different groups, Group A and D have more conserved sites at all three levels of frequency, and these groups were probably under strong purifying selection. Conversely, Group E has the least conserved sites (Table 1), even less than the average number of all genes from this family, suggesting a more diverse evolutionary pattern in this group.

Intron–exon structure of the KCS gene family

Introns, similar to coding regions, also contain valuable information. Indeed, the absence/presence, position and phase of introns are often regarded as important parameters in phylogenetics and gene family evolution (Qiu et al. 1998; Wang et al. 2009; Xue et al. 2010, 2012). The term 'intron phase' refers to the position of an intron within a codon; phase 0, 1 or 2 corresponds to an intron located before the first base, after the first base and after the second base, respectively, of a codon. In our previous studies of large gene families, intron positions and phases were found to be indicative of and to assist in the identification of subfamilies (Xue et al. 2012).

An overwhelming number of KCS genes are intronless, including genes from green algae and bryophytes, suggesting that an intronless structure is the most ancestral status in the plant lineage. A small number of KCS genes were found to contain introns, which should be the consequence of intron insertions during later evolution (Fig. 4). A phase 0 intron is shared by all genes from Angiosperm Group 2, indicating an insertion event no later than the emergence of angiosperms. In contrast, genes from Angiosperm Group 4 showed a more complicated process of intron gain. One intron (phase 0) is shared by all members, and it was thus

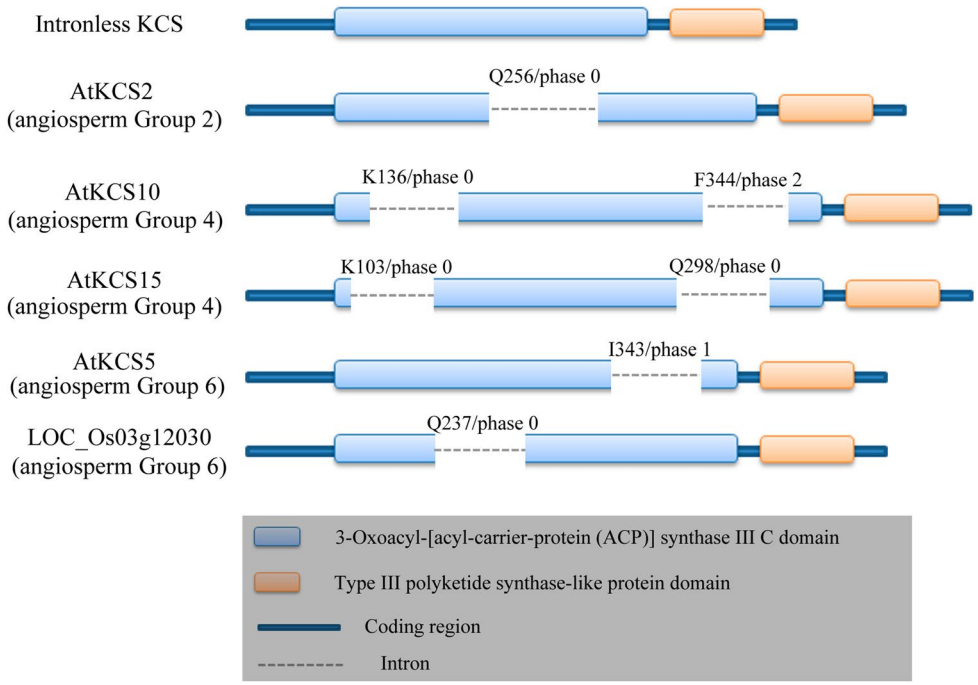

Fig. 4 The positions and phases of introns in KCS genes

obtained before the origin of angiosperms, similar to the intron of Group 2. However, Group 4 genes underwent an extra duplication in dicots after the split of monocots, which resulted in two monophyletic branches with only dicot genes, as represented by AtKCS10 (phase 2) and AtKCS15 (phase 0). The two sister groups each show another specific intron, which was very likely gained independently and accompanied by a duplication. Other new introns, the result of recent insertions, were also detected. Genes of Brassicaceae and Poaceae from Angiosperm Group 7 each have one characteristic intron (phase 1 and 0, respectively) that was likely gained as the two subfamilies diverged. Therefore, these introns are informative and should be helpful in determining the evolutionary history of KCS genes.

Discussion

The evolutionary patterns and gene gain and loss events of the KCS gene family

The elongation of fatty acids shares a basic mechanism across different eukaryotic phyla. KCS genes encode enzymes that catalyze the rate-limiting condensation reaction and accordingly regulate the synthesis of VLCFAs. Because they can be found in green algae and some other protists such as the Mycetozoa species *Dictyostelium discoideum* and the Heterokontophyta species *Phaeodactylum tricornutum*, these genes originated extremely early, suggesting that these genes could date back to an early stage in eukaryotic evolution. Notably, KCS genes remain as a single gene without expansion in all eukaryotic clades except plants.

The evolution of plant KCS genes is characterized by a history of gene duplications, expanding from a single ancestral gene into a gene family containing approximately 20–30 members in angiosperms. This is a lengthy process involving multiple rounds of duplications at different evolutionary stages. Duplications began as plants became terrestrial, prior to which all plants only carried only a single gene in the genome. The duplications at this stage involved a first duplication of the single ancestral gene, a subsequent duplication of each duplicated gene, and the last duplication of one of the four genes. Thus, duplication at this node created 5 ancestral genes, shaping the basic phylogenetic topology of KCS genes in land plants. At another key node—the emergence of angiosperms—the KCS genes had doubled, forming 11 angiosperm groups. However, the total 11 angiosperm genes were not evenly distributed with regard to the 5 ancestral genes; in fact, the process of generating 11 angiosperm genes was biased. Three of the 5 ancestral genes (Land Plant Group B, C and D) did not duplicate, one (Land Plant Group A) duplicated once, and the last (Land Plant Group E) duplicated vigorously, creating six ancestral angiosperm genes. Thereafter, with the radiation and divergence of angiosperms, independent duplications occurred in different lineages, generating specific duplication events. For example, Brassicaceae species specifically expanded to seven genes in Angiosperm Group 5, including KCS18/FAE1, suggesting that KCS18/FAE1 originated during the split

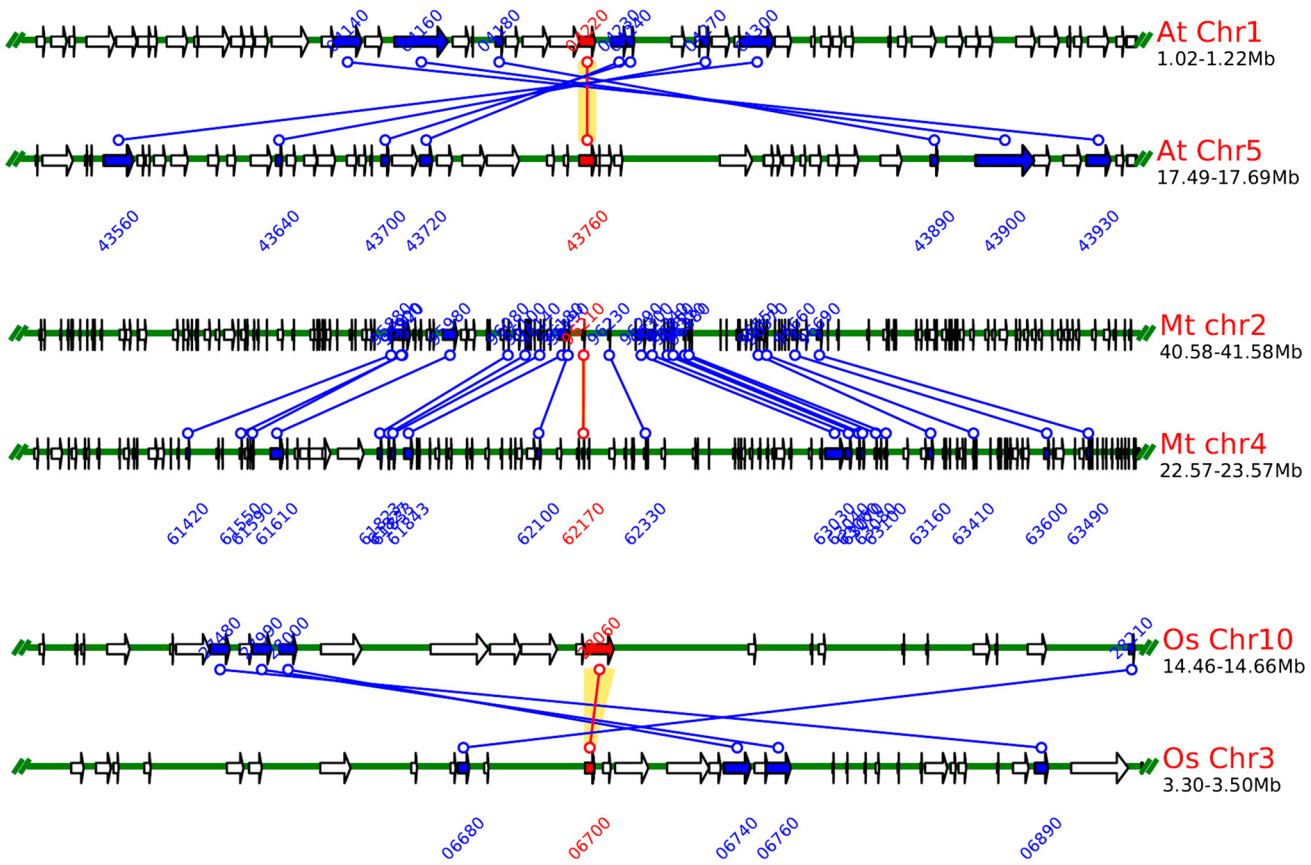

Fig. 5 Examples of the detailed locations of representative pairs of genes duplicated in recent polyploidy events in the syntenic regions (AT1G04220 and AT5G43760; Medtr2g096210 and Medtr4g062170; LOC_Os10g28060 and LOC_Os3g06700). At, *Arabidopsis thaliana*; Mt, *Medicago truncatula*; Os, *Oryza sativa*; chr, chromosome. *Arrows* illustrate the presence and orientation of syntenic paralogous genes, which are connected by *lines*

of Brassicaceae. Therefore, as KCS genes identified from non-Brassicaceae species, such as FAE1 from jojoba and nasturtium, are actually not orthologous to the *A. thaliana* or Brassicaceae FAE1 gene, their nomenclature is not evolutionarily accurate. The naming of 21 *A. thaliana* KCS genes by algebraic order is widely accepted (Joubes et al. 2008; Haslam and Kunst 2013) and thus sets a good example for the nomenclature of other KCS genes. To accurately name a new KCS gene, the nomenclature of the evolutionarily closest *A. thaliana* gene should be followed by establishing a phylogeny with all *A. thaliana* KCS genes. For example, as the jojoba and nasturtium genes are closest to *A. thaliana* KCS2 and KCS11, respectively, they should be named *S. chinensis* KCS2 and *T. majus* KCS11, with the species name incorporated.

Interestingly, all duplications appear to have stopped at the family level. Recent duplication within a species is difficult to observe in the phylogenetic tree, except, of course, for species with recent polyploidy events. Our further genomic analysis of paralogous KCS genes revealed that these genes are usually located within syntenic genomic blocks (for examples, see Fig. 5), which are the result of large-scale duplication events such as whole-genome duplications (WGDs) or segmental chromosomal duplications. Although some ancient WGDs or segmental chromosomal duplications are no longer identifiable due to long-term genomic variation, the duplication mode could be investigated in the opposite way; thus, tandem duplicates could be identified. Therefore, a survey of tandem-arrayed KCS genes was performed for all 26 genomes, with the results showing that only a small proportion of KCS genes exhibit tandem arrays (Table S2), different from plant resistance genes, which primarily expanded via recent tandem duplications (Xue et al. 2012; Shao et al. 2014; Wu et al. 2014). For example, only five of 21 KCS genes in *A. thaliana* and six of 23 KCS genes in rice show tandem arrays; six genomes do not contain any tandem KCS genes, including soybean with 26 KCS genes. The highest proportion of tandem genes is found in *B. distachyon* (eight out of 23). Therefore, the dominating duplication mode of KCS genes appears to be WGDs or segmental duplications, as opposed to tandem duplications at neighboring loci.

Table 2 Functionally characterized KCS genes and their VLCFA products

Subfamily		Gene name	Species	Product
Land plant group	Angiosperm group			
A	1	AtKCS11	*Arabidopsis thaliana*	20:0,20:1,22:0
	1	TmFAE	*Tropaeolum majus*	22:0,22:1,24:0
	1	LimFAE	*Limnanthes douglasii*	20:1D5
	2	AtKCS2	*Arabidopsis thaliana*	20:0,**22:0,24:0**,26:0
	2	AtKCS20	*Arabidopsis thaliana*	20:0,**22:0,24:0**,26:0
	2	ScFAE1	*Simmondsia chinensis*	22:1
B		MpFAE2	*Marchantia polymorpha*	22:0
		MpFAE3	*Marchantia polymorpha*	20:0,22:0
	3	AtKCS1	*Arabidopsis thaliana*	**20:0,20:1,22:0**,24:0,26:0
D	5	AtKCS9	*Arabidopsis thaliana*	24:0
	5	AtKCS17	*Arabidopsis thaliana*	**20:0**,20:1,22:0,**24:0**,26:0
	5	AtKCS18	*Arabidopsis thaliana*	**20:0,20:1,22:0,22:1,24:0**,24:1,26:0
E	6	AtKCS5	*Arabidopsis thaliana*	20:1,24:0,**26:0,28:0**,30:0
	6	AtKCS6	*Arabidopsis thaliana*	24:0,26:0,28:0

Products in bold were detected in independent studies. Functionally characterized KCS genes were collected from several previous studies (Kunst et al. 1992; James et al. 1995; Lassner et al. 1996; Cahoon et al. 2000; Kajikawa et al. 2003a, b; Mietkiewska et al. 2004; Trenkamp et al. 2004; Blacklock and Jaworski 2006; Paul et al. 2006; Tresch et al. 2012; Haslam and Kunst 2013)

Functional divergence, renovation and redundancy of KCS genes

KCS genes have grown into such a large gene family through duplications since plants invaded land. After a long evolutionary history, different family members have accumulated many mutations, leading to functional divergence and renovation and resulting in the specificity we observe today. Unfortunately, the functional characterization of KCS genes to date has been insufficient, with most efforts performed in *A. thaliana* and few studies in other plants. Nonetheless, some clues can be observed from the gene functions summarized via this limited sampling (Table 2, Haslam and Kunst 2013). Genes from angiosperms and the early-diverged liverwort *M. polymorpha* all show catalytic activities toward the synthesis of VLCFAs, suggesting that catalyzing the synthesis of VLCFAs should be the ancestral function of this gene family. The diversity of VLCFA products, encompassing almost all saturated and monounsaturated VLCFAs naturally produced by plants, also indicates that the ancestral gene resulting in this gene family should be able to, or at least had the potential to, evolve the capacity to catalyze all VLCFAs. As duplications created an increasing number of KCS gene copies, the important task of regulating diverse VLCFAs could be assigned to a number of duplicates, with each duplicate focusing on fewer products. This would increase the efficiency of VLCFA biosynthesis, similar to the strategy of division and cooperation of labor in modern industry. Evolutionarily, subfunctionalization refers to the process by which duplicated copies only inherit partial functions of an ancestral gene. Our analysis found that on occasion, genes from different groups may have a similar function; for example, a jojoba gene of Land Plant Group A and *A. thaliana* AtKCS18/AtFAE1 of Land Plant Group D both show C22:1 production activity, whereas other genes from the same groups may not. For example, none of the three *A. thaliana* genes (AtKCS2, −11 and −20 from Land Plant Group A) exhibit C22:1 activity. However, it is rather difficult to explain the phenomenon merely by a limited number of sequence mutations. Indeed, many mutation sites arose during such a long period of time, making it difficult to determine precisely which nucleotide/amino acid mutation(s) should be responsible for a specific functional change. Nevertheless, a study of protein structure provided another angle for deciphering this problem: 3-D modeling simulations revealed a discrepancy of binding pockets between KCS proteins, suggesting a hypothesis that the shape and size of binding pockets are decisive for substrates and, therefore, are likely to play a key role in product specificity (Joubes et al. 2008). As the binding pocket consists of a number of amino acids, these sites should all contribute to the generation of product specificity, which is not determined by any single site.

Neofunctionalization has also occurred in certain KCS genes, as they were found to lack VLCFA-catalyzing activities, yet they are related to other physiological processes or organ development in plants, such as a negative regulator of stomatal development that responds to CO_2 concentrations (AtKCS13/AtHIC) (Gray et al. 2000). Theoretically, genes with completely new functions should accumulate

more mutations, especially nonsynonymous mutations. As we observed in the phylogenetic tree, KCS13 genes should have originated as the family Brassicaceae diverged; thus, the gene is specific to Brassicaceae species. The branch covering KCS13 genes is unusually long in length, suggesting a large number of amino acid changes caused by nonsynonymous mutations, and these mutations accumulated very soon, even before the early speciation of this family. Therefore, other genes with longer branch lengths, such as KCS7, −14, −15, −19 and −21, may also have undergone neofunctionalization. Coincidently, these genes have been expressed in yeast but none showed activity in the elongation of VLCFAs, similar to AtKCS13. Among all groups, Land Plant Group E shows longer branch lengths, suggesting the possibility of functional renovation in this group.

Sub- and neofunctionalization represent the positive side of duplicated genes toward adaption, though the negative side is always present. Functional redundancy was discovered as well. AtKCS5 and 6 other genes duplicated upon the divergence of Brassicaceae, showing very similar expression patterns and catalytic activities (Joubes et al. 2008), except for a tenfold disparity in expression among them (Fiebig et al. 2000). Another example includes AtKCS2 and 20 other genes, which are paralogs that duplicated at the early evolution of Brassicaceae; these genes display redundancy in suberin and cuticular wax biosynthesis (Lee et al. 2009). A deeper look at the two pairs of redundant genes revealed that each pair maintains very high identity (amino acid identity >85 %) and a close relationship (evolutionary distance <20), comparable to orthologous KCS gene pairs (Sun et al. 2013). Therefore, amino acid identity and distance were calculated between all paralogous gene pairs, and accordingly, other potentially redundant gene pairs are suggested (Table S3). Similarly, AtKCS13 and 14, which are the products of one tandem duplication event alone in *A. thaliana*, show 98.5 % identity and a distance of only 0.9. Of the 26 species, *B. rapa*, *P. trichocarpa* and *G. max* exhibit the highest percentage of potentially redundant genes, 10, 13 and 20 pairs, respectively, likely because of recent polyploidy events.

Thus, redundancy exists together with sub- and neofunctionalization in the evolution of gene families, and there appears to be a subtle balance between them: sometimes redundant genes are needed to maintain genetic robustness (Dufton 1986; Wagner 1996; Nowak et al. 1997; Vavouri et al. 2008), whereas fitness cost will be evaluated from time to time. When the maintenance of redundant genes becomes a too heavy a burden for plants as the habitat changes, the balance could be easily broken, leading to the loss of redundant genes to ease this burden, a process that could occur very quickly. We investigated the KCS9 gene, which arose with the origin of the Brassicaceae and participates in the co-regulation of suberin and cuticular wax biosynthesis (Kim et al. 2013). Due to a recent whole-genome triplication in the genus *Brassica*, all *Brassica* species have extra copies of KCS9. For example, *B. rapa* has two copies of KCS9 (Bra037304 and Bra001974) with very high sequence similarity; thus, the two genes are likely functionally redundant. Interestingly, we found that in some *Brassica* species (several variants of *B. oleracea*), one copy of the KCS9 gene had already become pseudogenized, with frameshift mutations that resulted in a shortened ORF while the other copy remain intact (unpublished).

The conservation of sites and gene activity

Although the presence of a certain KCS gene may result in the specificity of its VLCFA products, the output relies on the activity level of the gene. Gene activity is interrelated to variation in coding sequences, including nonsynonymous mutations and indels, because they both induce amino acid changes. Indels often lead to frameshift mutations and a premature stop of translation and thus are likely to cause a complete loss of activity; whereas amino acid changes alter protein activity as well. Through only a few sequence changes at important sites, the production of a certain VLCFA could increase, decrease or even cease. Efforts have also been made to identify such sites. Sun et al. surveyed the erucic acid content of seed oil from 94 Brassicaceae species (Sun et al. 2011) and obtained FAE1 fragments from these species by PCR. By performing phylogenetic and biostatistical analyses, the authors finally detected 16 fixed mutation sites in low and high erucic acid species (Sun et al. 2013). A similar approach was coincidently adopted by another study that surveyed over 500 *A. thaliana* ecotypes. This study performed further sequence structure–function relationship analysis and proposed a 25 % reduction in the size of the catalytic binding pocket caused by a valine to leucine change (Jasinski et al. 2012).

In this study, we performed a thorough statistical analysis of conserved sites and detected 81 amino acid sites with a frequency of over 90 %, 198 sites with a frequency of over 70 %, and 311 sites with a frequency of over 50 % in this gene family (Table 1). The sites detected by the studies above mostly fall into the below 50 % frequency group, which possibly indicates that the variable sites, below 50 % frequency, are likely to be involved in the emergence of product specificity and catalytic power; while the extremely conserved sites, over 90 % frequency, are possibly essential functional sites for KCS proteins. The genes could entirely lose function similarly to frameshift mutations when these sites change. Overall, further experimental tests focusing on these conserved and variable positions should verify our hypothesis.

The evolutionary significance of three classes of enzymes related to VLCFAs elongation in plants: KCS, ELO and CER2-like proteins

Condensing enzymes catalyze the first reaction in fatty acid elongation and determine the final fatty acid product. In addition to KCS proteins, two other proteins, ELO and CER2-like protein, are also present in plants (Haslam and Kunst 2013). There remains the question of whether it is necessary for plant genomes to encode three different classes of condensing enzymes, and the answer is: probably yes. KCSs appear to mainly produce saturated and monounsaturated VLCFAs of 16–28 carbons but rarely polyunsaturated products or longer chain lengths. ELO proteins were first discovered and characterized in the yeast *Saccharomyces cerevisiae* (Toke and Martin 1996; Oh et al. 1997) and later in animals (Jakobsson et al. 2006). Although homologs of ELO proteins do exist in plants, their functions remain poorly understood. Only one study has demonstrated that two *Physcomitrella patens* ELO-like proteins are indispensable for the synthesis of polyunsaturated VLCFAs in this moss (Eiamsa-Ard et al. 2013). Thus, ELOs might have a role in the elongation of polyunsaturated VLCFAs in plants. CER2-like proteins were first characterized in *A. thaliana*, and they have all been found involved in the elongation of fatty acids longer than C28 (Haslam et al. 2012, 2015). Therefore, it appears that these three classes of condensing enzymes produce different VLCFA products, with no functional redundancy among them.

The evolution of the KCS gene family is essentially a history of gene duplications, developing from a single gene in the common ancestor of plants to a gene family of 20–30 members in angiosperms. The duplication mode appears to largely be WGDs and large-scale segmental duplications, whereas tandem duplications are rare. Accompanied by the expansion of this gene family, variations led to sub- and neofunctionalization of different members, resulting in the generation of product specificity and the emergence of completely new functions involved in other physiological processes. However, redundancy is also observed, mostly among recent duplicates. Statistical analysis indicates that the variable sites of KCS proteins are likely to be involved in the emergence of product specificity and catalytic power, and the conserved sites are possible responsible for the preservation of fundamental function.

Materials and methods

Database

The genomic sequences, annotations and gene models for 26 genomes were used in this study. File of all species were downloaded from the Phytozome database (http://phytozome.jgi.doe.gov/pz/portal.html) (Goodstein et al. 2012); but *S. indicum*, *A. trichopoda* and *P. abies*. Files of *S. indicum*, *A. trichopoda* and *P. abies* were downloaded from Sinbase (http://ocri-genomics.org/Sinbase/manual.html), (Zhang et al. 2013), *Amborella* genome database (http://www.amborella.org) (Amborella-Genome-Project 2013) and Spruce Genome Project (http://congenie.org) (Nystedt et al. 2013), respectively. For two transcriptomes data used in this study, *Mesostigma viride* was downloaded from NCBI (National Center for Biotechnology Information) (Ju et al. 2015) and *Lygodium japonicum* from Ljtrans DB (http://bioinf.mind.meiji.ac.jp/kanikusa) (Aya et al. 2015).

Identification of KCS genes

The KCS genes in the 26 genomes were identified as described in our previous study (Shao et al. 2014). Briefly, HMM searcher were first performed for protein sequences of each genome using the hmmer3.0 (http://hmmer.org) in default parameter settings, with the amino acid sequences of the 3-Oxoacyl-[acyl-carrier protein (ACP)] synthase III C domain (Pfam: PF08541) and the Type III polyketide synthase-like protein domain (Pfam: PF08392) as queries, respectively. The amino acid sequences were then used to run a BLASTp search against all protein sequences in each genome, with the threshold expectation value set to 0.0001. All hits obtained using HMM and BLAST searches were then merged together, and the redundant hits were removed. The remaining sequences were further subjected to the online Pfam analysis (http://pfam.sanger.ac.uk/) to further exclude sequences that do not contain a 3-Oxoacyl-[acyl-carrier protein (ACP)] synthase III C domain and a Type III polyketide synthase-like protein domain at the *E* value of 10^{-20}. Finally, the identified genes were further examined manually to eliminate spurious hits.

Sequence alignment, phylogenetic analysis, synteny analysis, and detection of amino acid polymorphism sites and conserved sites

Sequence alignment and phylogenetic analysis were performed as previously described (Shao et al. 2014). Amino acid sequences of KCS genes were extracted and used for alignment by ClustalW (Edgar 2004) with default settings (Thompson et al. 1994), and then manual adjusted in MEGA5.0 (Tamura et al. 2011). The derived alignment was then subjected to visional inspections and manual adjustments to improve quality. Phylogenetic analysis was conducted by a maximum likelihood method using the Phyml program (Guindon and Gascuel 2003) integrated in the seaview package (Gouy et al. 2010). The

reliability of internal nodes was evaluated by calculating SH-aLRT (Shimodaira–Hasegawa approximate likelihood ratio test) values (Guindon et al. 2010). This new evaluating method has been shown to provide a compelling alternative to those standard bootstrap methods, offering not only speed advantage, but also excellent levels of accuracy and power, especially for large datasets (Anisimova et al. 2011). The gene loss and gain at each divergence node were reconciled using Notung software (Stolzer et al. 2012).

MCScanX, a package developed by the Plant Genome Duplication Database (PGDD, http://chibba.agtec.uga.edu/duplication/), was adopted to perform synteny examination of paralogous genes among genomes through BLASTp searches (Wang et al. 2012; Lee et al. 2013). The amino acid polymorphism sites and conserved sites were analyzed by WebLogo (http://weblogo.berkeley.edu/logo.cgi) (Crooks et al. 2004), through which sequence logos were generated according to alignment.

Identification of tandem-arrayed KCS genes

In each genome, for all identified KCS genes, their genomic locations were determined by retrieving relevant information from the downloaded GFF annotation files. Thereafter, we examined the occurrence numbers of KCS genes on different chromosomes/scaffolds. The criterion used previously for *M. truncatula* NBS gene family (Ameline-Torregrosa et al. 2008) was adopted for KCS gene cluster assignment in this study. If two neighbored KCS genes were separated by no more than 250 kb on chromosome/scaffold, then these two genes are regarded as gene members of the cluster. Based on this criterion, tandem-arrayed KCS genes in each genome were identified.

Acknowledgments We gratefully acknowledge the editor-in-chief Dr. Hohmann and two anonymous reviewers for their constructive suggestions to improve the manuscript, Dr. Zhu-Qing Shao at Nanjing University for his assistance in the bioinformatics analysis, and Dr. Yang Liu at the University of Connecticut for insightful discussions. This work was supported by grants from the National Natural Science Founding of China to JYX (31300190), XQS (31200177 and 31470448), the Natural Science Founding of Jiangsu Province (BK20130565) to JYX, and Funds of Jiangsu Province Key Laboratory for Plant Ex Situ Conservation to XQS (QIAN201202) (http://www.nsfc.gov.cn, http://www.jskjjh.gov.cn, http://www.cnbg.net/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Compliance with ethical standards

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

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Amborella-Genome-Project (2013) The *Amborella* genome and the evolution of flowering plants. Science 342:1241089

Ameline-Torregrosa C, Wang BB, O'Bleness MS, Deshpande S, Zhu H, Roe B, Young ND, Cannon SB (2008) Identification and characterization of nucleotide-binding site-leucine-rich repeat genes in the model plant *Medicago truncatula*. Plant Physiol 146:5–21

Anisimova M, Gil M, Dufayard JF, Dessimoz C, Gascuel O (2011) Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. Syst Biol 60:685–699

Aya K, Kobayashi M, Tanaka J, Ohyanagi H, Suzuki T, Yano K, Takano T, Matsuoka M (2015) De novo transcriptome assembly of a fern, *Lygodium japonicum*, and a web resource database, Ljtrans DB. Plant Cell Physiol 56:e5

Blacklock BJ, Jaworski JG (2006) Substrate specificity of *Arabidopsis* 3-ketoacyl-CoA synthases. Biochem Biophys Res Commun 346:583–590

Cahoon EB, Marillia EF, Stecca KL, Hall SE, Taylor DC, Kinney AJ (2000) Production of fatty acid components of meadowfoam oil in somatic soybean embryos. Plant Physiol 124:243–251

Costaglioli P, Joubes J, Garcia C, Stef M, Arveiler B, Lessire R, Garbay B (2005) Profiling candidate genes involved in wax biosynthesis in *Arabidopsis thaliana* by microarray analysis. Biochim Biophys Acta 1734:247–258

Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. Genome Res 14:1188–1190

Dufton MJ (1986) Genetic code redundancy and its differential influence on the evolution of protein interiors versus exteriors. J Theor Biol 122:231–236

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Eiamsa-Ard P, Kanjana-Opas A, Cahoon EB, Chodok P, Kaewsuwan S (2013) Two novel *Physcomitrella patens* fatty acid elongases (ELOs): identification and functional characterization. Appl Microbiol Biotechnol 97:3485–3497

Fehling E, Mukherjee KD (1991) Acyl-CoA elongase from a higher plant (*Lunaria annua*): metabolic intermediates of very-long-chain acyl-CoA products and substrate specificity. Biochim Biophys Acta 1082:239–246

Fiebig A, Mayfield JA, Miley NL, Chau S, Fischer RL, Preuss D (2000) *Alterations* in CER6, a gene identical to CUT1, differentially affect long-chain lipid content on the surface of pollen and stems. Plant Cell 12:2001–2008

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, Rokhsar DS (2012) Phytozome: a comparative platform for green plant genomics. Nucleic Acids Res 40:D1178–D1186

Gouy M, Guindon S, Gascuel O (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221–224

Gray JE, Holroyd GH, van der Lee FM, Bahrami AR, Sijmons PC, Woodward FI, Schuch W, Hetherington AM (2000) The HIC signalling pathway links CO2 perception to stomatal development. Nature 408:713–716

Guindon S, Gascuel O (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst Biol 52:696–704

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol 59:307–321

Haslam TM, Kunst L (2013) Extending the story of very-long-chain fatty acid elongation. Plant Sci 210:93–107

Haslam TM, Manas-Fernandez A, Zhao L, Kunst L (2012) *Arabidopsis* ECERIFERUM2 is a component of the fatty acid elongation machinery required for fatty acid extension to exceptional lengths. Plant Physiol 160:1164–1174

Haslam TM, Haslam R, Thoraval D, Pascal S, Delude C, Domergue F, Fernandez AM, Beaudoin F, Napier JA, Kunst L, Joubes J (2015) ECERIFERUM2-LIKE proteins have unique biochemical and physiological functions in very-long-chain fatty acid elongation. Plant Physiol 167:682–692

Jakobsson A, Westerberg R, Jacobsson A (2006) Fatty acid elongases in mammals: their regulation and roles in metabolism. Prog Lipid Res 45:237–249

James DW Jr, Dooner HK (1990) Isolation of EMS-induced mutants in *Arabidopsis* altered in seed fatty acid composition. Theor Appl Genet 80:241–245

James DW, Lim E, Keller J, Plooy I, Ralston E, Dooner HK (1995) Direct tagging of the *Arabidopsis* FATTY ACID ELONGATION1 (FAE1) gene with the maize transposon Activator. Plant Cell 7:309–319

Jasinski S, Lecureuil A, Miquel M, Loudet O, Raffaele S, Froissard M, Guerche P (2012) Natural variation in seed very long chain fatty acid content is controlled by a new isoform of KCS18 in *Arabidopsis thaliana*. PLoS One 7:e49261

Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW (2011) Ancestral polyploidy in seed plants and angiosperms. Nature 473:97–100

Joubes J, Raffaele S, Bourdenx B, Garcia C, Laroche-Traineau J, Moreau P, Domergue F, Lessire R (2008) The VLCFA elongase gene family in *Arabidopsis thaliana*: phylogenetic analysis, 3D modelling and expression profiling. Plant Mol Biol 67:547–566

Ju CL, Van de Poel B, Cooper ED, Thierer JH, Gibbons TR, Delwiche CF, CC R (2015) Conservation of ethylene as a plant hormone over 450 million years of evolution. Nat Plants 14004:1–7

Kajikawa M, Yamaoka S, Yamato KT, Kanamaru H, Sakuradani E, Shimizu S, Fukuzawa H, Ohyama K (2003a) Functional analysis of a beta-ketoacyl-CoA synthase gene, MpFAE2, by gene silencing in the liverwort *Marchantia polymorpha* L. Biosci Biotechnol Biochem 67:605–612

Kajikawa M, Yamato KT, Kanamaru H, Sakuradani E, Shimizu S, Fukuzawa H, Sakai Y, Ohyama K (2003b) MpFAE3, a beta-ketoacyl-CoA synthase gene in the liverwort *Marchantia polymorpha* L., is preferentially involved in elongation of palmitic acid to stearic acid. Biosci Biotechnol Biochem 67:1667–1674

Kim J, Jung JH, Lee SB, Go YS, Kim HJ, Cahoon R, Markham JE, Cahoon EB, Suh MC (2013) *Arabidopsis* 3-ketoacyl-coenzyme a synthase9 is involved in the synthesis of tetracosanoic acids as precursors of cuticular waxes, suberins, sphingolipids, and phospholipids. Plant Physiol 162:567–580

Kunst L, Taylor DC, UE W (1992) Fatty acid elongation in developing seeds of *Arabidopsis thaliana*. Plant Physiol Biochem 30:425–434

Lassner MW, Lardizabal K, Metz JG (1996) A jojoba beta-Ketoacyl-CoA synthase cDNA complements the canola fatty acid elongation mutation in transgenic plants. Plant Cell 8:281–292

Lechelt-Kunze C, Meissner RC, Drewes M, Tietjen K (2003) Flufenacet herbicide treatment phenocopies the fiddlehead mutant in *Arabidopsis thaliana*. Pest Manag Sci 59:847–856

Lee SB, Jung SJ, Go YS, Kim HU, Kim JK, Cho HJ, Park OK, Suh MC (2009) Two *Arabidopsis* 3-ketoacyl CoA synthase genes, KCS20 and KCS2/DAISY, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. Plant J 60:462–475

Lee TH, Tang H, Wang X, Paterson AH (2013) PGDD: a database of gene and genome duplication in plants. Nucleic Acids Res 41:D1152–D1158

Lemieux B, Miquel M, Somerville C, Browse J (1990) Mutants of *Arabidopsis* with alterations in seed lipid fatty acid composition. Theor Appl Genet 80:234–240

Lewis LA, McCourt RM (2004) Green algae and the origin of land plants. Am J Bot 91:1535–1556

Merchant SS, Prochnik SE, Vallon O, Harris EH, Karpowicz SJ, Witman GB, Terry A, Salamov A, Fritz-Laylin LK, Marechal-Drouard L, Marshall WF, Qu LH, Nelson DR, Sanderfoot AA, Spalding MH, Kapitonov VV, Ren Q, Ferris P, Lindquist E, Shapiro H, Lucas SM, Grimwood J, Schmutz J, Cardol P, Cerutti H, Chanfreau G, Chen CL, Cognat V, Croft MT, Dent R, Dutcher S, Fernandez E, Fukuzawa H, Gonzalez-Ballester D, Gonzalez-Halphen D, Hallmann A, Hanikenne M, Hippler M, Inwood W, Jabbari K, Kalanon M, Kuras R, Lefebvre PA, Lemaire SD, Lobanov AV, Lohr M, Manuell A, Meier I, Mets L, Mittag M, Mittelmeier T, Moroney JV, Moseley J, Napoli C, Nedelcu AM, Niyogi K, Novoselov SV, Paulsen IT, Pazour G, Purton S, Ral JP, Riano-Pachon DM, Riekhof W, Rymarquis L, Schroda M, Stern D, Umen J, Willows R, Wilson N, Zimmer SL, Allmer J, Balk J, Bisova K, Chen CJ, Elias M, Gendler K, Hauser C, Lamb MR, Ledford H, Long JC, Minagawa J, Page MD, Pan J, Pootakham W, Roje S, Rose A, Stahlberg E, Terauchi AM, Yang P, Ball S, Bowler C, Dieckmann CL, Gladyshev VN, Green P, Jorgensen R, Mayfield S, Mueller-Roeber B, Rajamani S, Sayre RT, Brokstein P, Dubchak I, Goodstein D, Hornick L, Huang YW, Jhaveri J, Luo Y, Martinez D, Ngau WC, Otillar B, Poliakov A, Porter A, Szajkowski L, Werner G, Zhou K, Grigoriev IV, Rokhsar DS, Grossman AR (2007) The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. Science 318:245–250

Mietkiewska E, Giblin EM, Wang S, Barton DL, Dirpaul J, Brost JM, Katavic V, Taylor DC (2004) Seed-specific heterologous expression of a nasturtium FAE gene in *Arabidopsis* results in a dramatic increase in the proportion of erucic acid. Plant Physiol 136:2665–2675

Nowak MA, Boerlijst MC, Cooke J, Smith JM (1997) Evolution of genetic redundancy. Nature 388:167–171

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hallman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Kaller M, Luthman J, Lysholm F, Niittyla T, Olson A, Rilakovic N, Ritland C, Rossello JA, Sena J, Svensson T, Talavera-Lopez C, Theissen G, Tuominen H, Vanneste K, Wu ZQ, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Thompson SL, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584

Oh CS, Toke DA, Mandala S, Martin CE (1997) ELO2 and ELO3, homologues of the *Saccharomyces cerevisiae* ELO1 gene, function in fatty acid elongation and are required for sphingolipid formation. J Biol Chem 272:17376–17384

Paul S, Gable K, Beaudoin F, Cahoon E, Jaworski J, Napier JA, Dunn TM (2006) Members of the *Arabidopsis* FAE1-like 3-ketoacyl-CoA synthase gene family substitute for the Elop proteins of *Saccharomyces cerevisiae*. J Biol Chem 281:9018–9029

Post-Beittenmiller D (1996) Biochemistry and molecular biology of wax production in plants. Annu Rev Plant Physiol Plant Mol Biol 47:405–430

Prochnik SE, Umen J, Nedelcu AM, Hallmann A, Miller SM, Nishii I, Ferris P, Kuo A, Mitros T, Fritz-Laylin LK, Hellsten U, Chapman J, Simakov O, Rensing SA, Terry A, Pangilinan J, Kapitonov V, Jurka J, Salamov A, Shapiro H, Schmutz J, Grimwood J, Lindquist E, Lucas S, Grigoriev IV, Schmitt R, Kirk D, Rokhsar

DS (2010) Genomic analysis of organismal complexity in the multicellular green alga *Volvox carteri*. Science 329:223–226

Qiu YL, Cho Y, Cox JC, Palmer JD (1998) The gain of three mitochondrial introns identifies liverworts as the earliest land plants. Nature 394:671–674

Qiu YL, Lee J, Bernasconi-Quadroni F, Soltis DE, Soltis PS, Zanis M, Zimmer EA, Chen Z, Savolainen V, Chase MW (1999) The earliest angiosperms: evidence from mitochondrial, plastid and nuclear genomes. Nature 402:404–407

Shao ZQ, Zhang YM, Hang YY, Xue JY, Zhou GC, Wu P, Wu XY, Wu XZ, Wang Q, Wang B, Chen JQ (2014) Long-term evolution of nucleotide-binding site-leucine-rich repeat genes: understanding gained from and beyond the legume family. Plant Physiol 166:217–234

Stevens P (2015) Angiosperm Phylogeny Website. In: Version 14 edn, St Louis. http://www.mobot.org/MOBOT/research/APweb/

Stolzer M, Lai H, Xu M, Sathaye D, Vernot B, Durand D (2012) Inferring duplications, losses, transfers and incomplete lineage sorting with nonbinary species trees. Bioinformatics 28:i409–i415

Sun XQ, Pang H, Guo JL, Peng B, Bai MM, Hang YY (2011) Fatty acid analysis of the seed oil in a germplasm collection of 94 species in 58 genera of Brassicaceae. Chem Ind For Prod 31:46

Sun X, Pang H, Li M, Peng B, Guo H, Yan Q, Hang Y (2013) Evolutionary pattern of the FAE1 gene in brassicaceae and its correlation with the erucic acid trait. PLoS One 8:e83535

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Toke DA, Martin CE (1996) Isolation and characterization of a gene affecting fatty acid elongation in *Saccharomyces cerevisiae*. J Biol Chem 271:18413–18422

Trenkamp S, Martin W, Tietjen K (2004) Specific and differential inhibition of very-long-chain fatty acid elongases from *Arabidopsis thaliana* by different herbicides. Proc Natl Acad Sci USA 101:11903–11908

Tresch S, Heilmann M, Christiansen N, Looser R, Grossmann K (2012) Inhibition of saturated very-long-chain fatty acid biosynthesis by mefluidide and perfluidone, selective inhibitors of 3-ketoacyl-CoA synthases. Phytochemistry 76:162–171

Vavouri T, Semple JI, Lehner B (2008) Widespread conservation of genetic redundancy during a billion years of eukaryotic evolution. Trends Genet 24:485–488

Wagner A (1996) Genetic redundancy caused by gene duplications and its evolution in networks of transcriptional regulators. Biol Cybern 74:557–567

Wang B, Xue J, Li L, Liu Y, Qiu YL (2009) The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts. Curr Genet 55:601–609

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH (2012) MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res 40:e49

Wodniok S, Brinkmann H, Glockner G, Heidel AJ, Philippe H, Melkonian M, Becker B (2011) Origin of land plants: do conjugating green algae hold the key? BMC Evol Biol 11:104

Wu P, Shao ZQ, Wu XZ, Wang Q, Wang B, Chen JQ, Hang YY, Xue JY (2014) Loss/retention and evolution of NBS-encoding genes upon whole genome triplication of *Brassica rapa*. Gene 540:54–61

Xue JY, Liu Y, Li L, Wang B, Qiu YL (2010) The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. Curr Genet 56:53–61

Xue JY, Wang Y, Wu P, Wang Q, Yang LT, Pan XH, Wang B, Chen JQ (2012) A primary survey on bryophyte species reveals two novel classes of nucleotide-binding site (NBS) genes. PLoS One 7:e36700

Zeng L, Zhang Q, Sun R, Kong H, Zhang N, Ma H (2014) Resolution of deep angiosperm phylogeny using conserved nuclear genes and estimates of early divergence times. Nat Commun 5:4956

Zhang H, Miao H, Wang L, Qu L, Liu H, Wang Q, Yue M (2013) Genome sequencing of the important oilseed crop *Sesamum indicum* L. Genome Biol 14:401