Conservation of two lineages of peroxisomal (Type I) 3-ketoacyl-CoA thiolases in land plants, specialization of the genes in Brassicaceae, and characterization of their expression in Arabidopsis thaliana

Andrew A. G. Wiszniewski, Steven M. Smith and John D. Bussell*

ARC Centre of Excellence in Plant Energy Biology, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia

* To whom correspondence should be addressed: E-mail: John.Bussell@uwa.edu.au

Received 3 August 2012; Revised 3 August 2012; Accepted 24 August 2012

Abstract

Arabidopsis thaliana has three genes encoding type I 3-ketoacyl-CoA thiolases (KAT1, KAT2, and KAT5), one of which (KAT5) is alternatively transcribed to produce both peroxisomal and cytosolic proteins. To evaluate the potential importance of these four gene products, their evolutionary history in plants and their expression patterns in Arabidopsis were investigated. Land plants as a whole have gene lineages corresponding to KAT2 and KAT5, implying conservation of distinct functions for these two genes. By contrast, analysis of synteny shows that KAT1 arose by duplication of the KAT2 locus. KAT1 is found in the Brassicaceae family, including in the genera Arabidopsis, Capsella, Thellungiella (=Eutrema) and Brassica, but not in the more distantly related Caricaceae (order Brassicales), or other plants. Gene expression analysis using qRT-PCR and β-glucuronidase reporter genes showed strong expression of KAT2 during germination and in many plant tissues throughout the life cycle, consistent with its observed dominant function in fatty acid β-oxidation. KAT1 was expressed very weakly while KAT5 was most strongly expressed during flower development and in seedlings after germination. Isoform-specific qRT-PCR analysis and promoter β-glucuronidase reporters revealed that the two splicing variants of KAT5 have similar expression profiles. Alternative splicing of KAT5 to produce cytosolic and peroxisomal proteins is specific to and ubiquitous in the Brassicaceae, and possibly had an earlier origin in the order Brassicales. This implies that an additional function for KAT5 arose between 43 and 115 mybp. We speculate that this KAT5 mutation was recruited for a cytosolic function in secondary metabolism.

Key words: 3-ketoacyl-CoA thiolase, β-oxidation, Brassicaceae, evolution, flowering, germination, peroxisome.

Introduction

Germination and seedling establishment in oilseed species such as Arabidopsis thaliana require peroxisomal β-oxidation to degrade the seed storage lipids that fuel this stage of development (Eastmond and Graham, 2001; Graham, 2008). As well as its role during seed germination, β-oxidation is a significant pathway for the synthesis of the hormones jasmonic acid (JA) and indole-3-acetic acid (IAA) in plants (Baker et al., 2006), and is required for the turnover of fatty acids in plant cells during development and senescence (Yang and Ohlrogge, 2009). Three core enzymes catalyse peroxisomal β-oxidation: acyl-CoA oxidase (ACX), multifunctional protein (MFP, which can exhibit hydratase, dehydrogenase, epimerase and isomerase activities), and L-3-ketoacyl-CoA thiolase (KAT).

The ACX and MFP gene families, represented by six and two genes, respectively, have been extensively characterized in Arabidopsis (Graham, 2008; Arent et al., 2010). By contrast, of
the three KAT genes, only the role of KAT2 has been investigated in detail (Hayashi et al., 1998; Germain et al., 2001; Footitt et al., 2007a). Thiolas of two types are distinguished. Type 1 enzymes (KAT: EC 2.3.1.16) are typically peroxisomal and catalyze the thiolsis of acetyl-CoA units from the thiol end of the fatty acyl-CoA during fatty acid catabolism. Type 2 enzymes (acytely-CoA acetyltransferase or ACAT; EC 2.3.1.9) are typically cytosolic and are involved in acetoacetyl CoA synthesis in the mevalonate biosynthesis pathway. The three KAT genes in Arabidopsis were designated as KAT1 (At1g04710), KAT2 (At2g33150), and KAT5 (At5g48880) based on the chromosome on which they are located (Germain et al., 2001). Subsequent analysis has determined that while both KAT1 and KAT2 encode single peroxisome targeted proteins, KAT5 encodes the cytosolic KAT5.1 and the peroxisomal KAT5.2 isoforms. The KAT5.2 transcript differs from KAT5.1cyt in the 5′ region, with an additional exon encoding a peroxisome targeting signal type 2 (PTS2), and alternate transcription and translation start sites (Carrie et al., 2007).

Genes involved in lipid mobilization, including β-oxidation, the glyoxylate cycle, and gluconeogenesis, are expressed coordinately during early seedling growth in Arabidopsis, with transcript levels and enzyme activities peaking at 48 h after the commencement of germination (Ryoltt et al., 2001). Of the three Arabidopsis KAT genes, expression of KAT2 is dominant during this stage of the life cycle, with KAT2 transcript levels far more abundant than KAT1 and KAT5 (Germain et al., 2001; Kamada et al., 2003). Genes of lipid mobilization decline in expression level in the late stages of seedling establishment, a process associated with peroxisome matrix remodelling as the function of the organelle changes from one primarily concerned with oil mobilisation to metabolism associated with photosynthesis (Ryoltt et al., 2001; Kamada et al., 2003; Pracharoenwattana and Smith, 2008; Lingard et al., 2009). There is growing evidence for the functional importance of β-oxidation in reproductive tissue and seed development (Richmond and Bleeker, 1999; Ryoltt et al., 2003, 2006; Chia et al., 2005; Footitt et al., 2007b; Schilmiller et al., 2007). Expression of the Arabidopsis KAT gene has been detected in reproductive tissue (Kamada et al., 2003), however, detailed analysis has not been reported, nor have KAT5.1cyt and KAT5.2px transcript expression patterns been distinguished. To characterize the function of members of the KAT gene family further, the phylogenetic relationships of the genes in sequenced plant genomes was investigated and a detailed analysis of their patterns of expression in Arabidopsis thaliana was conducted. Comparative genomics has highlighted the dynamic expansion, specialization, and contraction of plant genomes (Wang et al., 2011; Rutter et al., 2012) and evolution of the KAT gene family in the Brassicaceae provides an excellent and specific example of this.

Materials and methods

Bioinformatics

KAT protein sequences were retrieved from the collection of sequenced plant genomes at Phytozone v8.0 (http://www.phytozone.org) and as individually submitted sequences deposited at NCBI (http://www.ncbi.nlm.nih.gov). Thellungiella parvula (Eutrema parvulum) KAT sequences were obtained from the Compare Genomes site (http://genomevolution.org/CoGe/index.pl). Sequence visualization and manipulations were done using the Geneious (v5.4.6) package (Drummond et al., 2011). Multiple sequence alignment of full-length predicted protein sequences was made in Geneious using the MAFFT Alignment plug-in (Katoh et al., 2002) and phylogenetic analysis of this alignment was done using the PHYML plugin (Guindon and Gascuel, 2003) with the WAG substitution model and 1000 bootstrap replicates. Synteny analysis was done using the CoGe and SynMap (http://genomevolution.org/CoGe/index.pl) comparative genome analysis tools (Lyons et al., 2008).

GUS reporter analysis

KAT gene promoter sequences, including the 5′-UTR and intergenic DNA upstream of the start codons, were amplified by PCR from wild-type Col-0 genomic DNA for Gateway cloning. Primer sequences are given in Supplementary Table S1 at JXB online. Amplified promoter fragments extended approximately 2kb upstream from the start codons (2156 bp for the KAT2 promoter, 2151 bp for KAT5.1cyt and 2125 bp for KAT5.2px) or until the adjacent gene (KAT1, 1010bp) as detailed in Figure 3A. Promoters were cloned into the GUS/GFP reporter plasmid pHGWFS7 (Karimi et al., 2002). Agrobacterium strain CV3130-C58 harbouring the pHGWFS7-KAT promoter vectors was used to transform Arabidopsis Col-0 plants using the floral dip method of Clough and Bent (1998). Transformed plants were screened for homozygosity over three generations by selection on hygromycin. Plants were grown for GUS staining under continuous light conditions to control for the diurnal regulation of promoter activity. Seedlings were grown on half-strength MS media (without sucrose). Flowers and siliques were removed from 6-week-old soil-grown plants and stained for GUS expression (Weigel and Glazebrook, 2002).

Quantitative RT-PCR

Aerial tissue samples (about 100 mg) were taken from 5-week-old Arabidopsis Col-0 plants grown in soil under continuous light conditions, and root samples were taken from hydroponically grown plants. Germinating seed samples utilized approximately 20 mg of dry seed that had been spread on plates containing half-strength MS media then stratified for 48 h before being transferred to continuous light. Samples were ground into a fine powder with a mortar and pestle pre-cooled with liquid nitrogen. Germinating seed RNA was extracted using the RNAqueous Kit with Plant RNA Isolation Aid (Ambion) and included LiCl precipitation. RNA from other plant tissues was isolated using an RNAqueous Kit with Plant RNA Isolation Aid (Ambion) and therefore could not bound introns. Primer sequences are listed in Supplementary Table S1 at JXB online. qRT-PCR was performed on a Roche LC480. The reaction volume was 5 μl and included 1× LightCycler 480 SYBR Green 1 Master (Roche), 0.5 μl diluted cDNA, and 0.1 μl of 20 μM primers. Cycle conditions were: 95 °C for 10 min; 45 cycles of 95 °C for 20 s, 60 °C for 20 s, and 72 °C for 20 s. Melt curve analysis of real-time PCR products was performed to verify amplification of a single product. Crossing point values were calculated under high confidence. Four biological replicates per tissue sample were examined, with at least two technical replicates of each real-time PCR. The average crossing point value of two technical replicates was used to calculate expression relative to an internal reference gene adjusted by primer efficiencies. Two reference genes were tested: ACT2 (At1g18780) and the clathrin adaptor complex subunit (CACS; At5g46630), identified by Czechowski et al. (2005) as a more suitable reference gene for transcript normalization. For analysis, CACS was used for normalization; the average relative expression and standard error for the four biological replicates are shown.
Results

*KAT2 and KAT5 isoforms are conserved in higher plants*

KAT protein sequences were obtained from Phytozome v8.0 to investigate the evolutionary history of the gene family. The Phytozome database includes genome sequences of 31 plant species that encompass major clades of plant (Viridiplantae) evolution, including green algae, mosses, spike mosses (lycophytes), monocot grasses, eudicots, and dicots. Interrogation of this database using *Arabidopsis* KAT2 (At3g33150) in a BLAST query yielded 82 protein sequences. This number was reduced to 71 hits (see Supplementary Table S2 at JXB online) after exclusion of proteins with obviously poor matches and partial alignments. KAT sequences from *Brassica napus* and *Thellungiella parvula* (which are not represented in Phytozome v8.0) were obtained from NCBI and CoGe, respectively, to yield a total of 76 KAT protein sequences from 33 species.

There was at least one KAT isoform encoded in each of the 33 genomes, with the majority (29/33) encoding two or more (see Supplementary Table S2 at JXB online). Notably, both green alga species (*Volvox carteri* and *Chlamydomonas reinhardtii*) only possessed a single gene. The exon–intron structure of the genes is highly conserved in land plants, with almost all of the KAT coding sequences from higher plants being assembled from 14 exons that span similar regions of the protein in each species (see Supplementary Table S2 at JXB online). In land plants, \(\beta\)-oxidation is a peroxisomal process and KAT proteins would thus be expected to localize to that organelle. Accordingly, 74/76 proteins possess a predicted peroxisome targeting signal 2 (PTS2) close to the N-terminus, the only exceptions being one of the two *Selaginella moellendorffii* isozymes (see Supplementary Table S2 at JXB online). KAT proteins from the Poales isozyme they contained. The taxonomic orders of the species are indicated on the tree for clarity. KAT proteins from the Poales formed two discrete clusters, one embedded in the KAT5 clade, and the second resolved basal to KATs of all angiosperms. We nominally label this later as KAT2 representatives of the Poales. Thus, of the 29 species of higher plants, 27 were represented by two KAT isoforms corresponding to KAT2 and KAT5. The two exceptions (*Medicago truncatula* and *Aquilegia coerulea*; see Supplementary Table S2 at JXB online) have only one annotated KAT gene. Whether the absence of another KAT gene from these genomes is genuine or due to incomplete sequencing or annotation remains to be determined. Interestingly, both moss genomes also encode two KAT proteins, but unlike those of higher plants these were not phylogenetically separated on the tree into KAT2 or KAT5 isoforms.

**KAT1 is a duplication of KAT2 unique to the Brassicaceae lineage**

The Brassicales appear to have experienced further duplication of KAT genes. The KAT2 cluster (Fig. 1) also contains six species (*Arabidopsis thaliana*, *A. lyrata*, *Capsella rubella*, *Thellungiella halophila*, *T. parvula*, and *Brassica rapa*) with a third KAT gene (KAT1) in addition to the KAT2 and KAT5 paralogues (see Supplementary Table S2 at JXB online). AtKAT1 does not appear to have orthologues other than in the Brassicaceae, suggesting that it is a duplication of KAT2 peculiar to that lineage. Indeed, there is significant synteny between the regions of *A. thaliana* chromosomes 1 and 2 surrounding the KAT1 and KAT2 genes. A fragment of about 2.2 Mbp on chromosome 2 shows nearly complete correspondence of genes with span of 1.3 Mbp on chromosome 1 (see Supplementary Fig. S1A at JXB online). It is likely that the duplication between chromosomes 1 and 2 was followed by a localized segmental inversion on one of the chromosomes, because about 170 kb is inverted with respect to the rest of the duplicated region (see Supplementary Fig. S1B at JXB online). In contrast, there is no synteny between the AtKAT2 or AtKAT1 genomic regions and that of AtKAT5 (not shown).

These syntenic chromosome arrangements are also found in *A. lyrata*, *T. parvula*, and *C. rubella* around genes corresponding to KAT1 and KAT2 (not shown). BrKAT1 (*B. rapa*) is likely to have had the same origin at AtKAT1, but the situation is less clear because the Brassica lineage underwent whole genome triplication after the split from the *Arabidopsis* lineage (Mun et al., 2009). There are three clear orthologues of AtKAT2 in the *B. rapa* genome (Fig. 1; see Supplementary Table S2 at JXB online). Genomic regions around these *B. rapa* KAT2 genes display significant synteny with each other and with the AtKAT2 genomic region (see Supplementary Table S3 at JXB online). In addition, there are three other regions displaying synteny with these, only one of which has retained a KAT gene. This locus, accession Bra030586 (see Supplementary Table S2 and S3 at...
Fig. 1. Phylogenetic analysis of the plant 3-ketoacyl-CoA thiolase (KAT) family proteins. Multiple sequence alignment of full-length KAT protein sequences predicted from sequenced plant genomes was made using MAFFT (Katoh et al., 2002) and phylogeny estimated using PHYML (Guindon and Gascuel, 2003). Numbers at the nodes are per cent bootstrap agreement from 1000 replicates. Major
Evolution of plant thiolase genes

JXB online) groups with AtKAT1 in the phylogenetic analysis (Fig. 1). These latter three regions correspond to approximately the same genes duplicated between KAT2 and KAT1 in A. thaliana (see Supplementary Fig. S1B at JXB online). This evidence suggests that the duplication occurred before the Brassica–Arabidopsis split (and also before whole genome triplication in Brassica). It is also interesting to note that there are six regions in the B. rapa genome that display synteny with the AtKAT5 region (see Supplementary Table S3 at JXB online). All correspond to approximately the same span of the A. thaliana genome (~620 Mbp), but in B. rapa only one of the six regions has retained a KAT5 gene. There are no such regions of synteny between CpKAT2 and other genomic regions in Carica papaya, a more distantly related species in the order Brassicales. The separation of Brassicaceae, which have maintained a KAT1 orthologue, from Caricaceae, which have not, can be dated to between 43 and 115 mybp (see inset, Fig. 1); these ages thus provide the upper and lower bounds for the timing of the segmental duplication.

Dual targeting of KAT5 is conserved in Brassicales

AtKAT5 is alternately transcribed to produce two species of mRNA, KAT5.1_ort and KAT5.2_ort (Fig. 3A). These possess alternative start (Met) codons that direct the encoded proteins to the peroxisome and the cytosol, respectively (Carrie et al., 2007). Further investigation of the plant KAT gene family revealed that all KAT5 orthologues from the Brassicales (represented by two families in Phytozome; Brassicaceae and Caricaceae), but not other plant orders, have potential alternative start codons in the same relative positions (Fig. 2). Thus, A. thaliana, A. lyrata, Capsella rubella, T. halophila, T. parvula, B. rapa, and B. napus (Brassicaceae) and Carica papaya (Caricaceae) KAT5 orthologues putatively encode KAT5.1 and KAT5.2 isoforms. Examination of PASA assembled ESTs (via Phytozome v8.0) provided, where data were available, evidence for alternate transcription of KAT5 for each of these species (see Supplementary Table S2 at JXB online).

Form of dicotyledonous plant thiolase proteins. Consensus sequences were derived from four groups of sequences comprising dicotyledonous KAT2 or KAT5 and Brassicales KAT2 or KAT5 sequences. AtKAT2 and AtKAT5 are included as samples of specific sequences. Typical plant KAT proteins are about 460 amino acids; the first approximately 200 residues are shown. Highly conserved residues (present in >90% of sequences) in each of the four consensus sequences are shown in capital letters, while moderately conserved residues (present in 50–90% of sequences) in lower case. Non-conserved positions are depicted by a period (.) and gaps introduced into the alignment by a dash (-). Symbols used to indicate residues with strongly similar properties on the Gonnet PAM 250 matrix are: # (N, D, Q, or E); % (F or Y); ! (I or V). Residues conserved between the different consensus sequences are highlighted with black shading and grey shading indicates consensus sequence residues that are not identical but have similar properties. The red highlighting indicates residues that clearly distinguish KAT5. The blue box shows the PTS2 near the N-terminus of the proteins and the yellow highlighting the alternative initial Met that is conserved in all Brassicales KAT5 proteins.

Examination of alignments of plant thiolase proteins highlights differences between KAT2 and KAT5 and between KAT5s of Brassicales and other dicotyledonous species (Fig. 2). Groups of Brassicales KAT sequences that were identified as KAT2 or KAT5 in Fig. 1 were aligned separately using Multalin (http://multalin.toulouse.inra.fr/multalin/multalin.html). Similarly, other dicotyledonous plant KAT2 or KAT5 sequences were aligned (i.e. excluding those from Brassicales) and the four resulting consensus sequences compared (Fig. 2). After the conserved PTS2 (residues 7–15 in Fig. 2), all taxa have a linker that shows little conservation (residues 25–30). The sequence following this (residues 31–51) is highly conserved in all KAT2 proteins and less so in KAT5. Moreover, Brassicales KAT5s differ substantially from the KAT2s and the KAT5s of other dicots in
this region and it includes the alternate Met described above and a four amino acid deletion relative to all other KAT proteins. Interestingly, although C. papaya (a basal Brassicales taxon) possesses the alternate Met in its KAT5 protein, it does not have the four amino acid deletion and the sequence surrounding the second Met is more similar to the typical KAT5 (and KAT2) consensus (not shown). Unfortunately, evidence for or against alternative transcription of CpKAT5 is not available because the EST collection is small and does not include the 5’ ends of the transcripts. A second region that distinguishes between KAT2 and KAT5 isoforms is found in residues 173–192. Here, KAT2 sequences are highly conserved and there is little conservation in the KAT5 sequences (Fig. 2).

Based on protein sequence similarity and EST data, we suggest that alternative transcription and protein localization of KAT5 isoforms is unique to the order Brassicales and originated with a mutation to make a second Met in a lineage ancestral to C. papaya (i.e. at least 115 mybp; Fig. 1) followed by further mutation and deletions between 115 and 43 mybp to yield the typical Brassicales KAT5 sequence.

Patterns of KAT gene expression by promoter::GUS reporters

Eleven KAT1-GUS, nine KAT2-GUS, ten KAT5.1-GUS, and ten KAT5.2-GUS independent transformants were obtained and screened for GUS activity. Lines giving GUS staining patterns representative of the group were selected for more detailed analysis. The promoter-reporter construct KAT5.2-GUS includes the KAT5 promoter up to the first start codon joined to the GUS reporter gene; KAT5.1-GUS includes the 5’ UTR and first exon of KAT5.2 and thus includes the full putative promoter that might be expected to drive independent transcription of KAT5.1 (Fig. 3A).

GUS expression in 2–6-d-old seedlings indicated low levels of expression for KAT1-GUS and KAT5.2-GUS, with higher levels for KAT2-GUS and KAT5.1-GUS (Fig. 3B). In KAT1-GUS lines, activity was nearly absent but some GUS staining was observed in the root tip (indicated by the arrow in Fig. 3B). KAT2-GUS was expressed strongly in the cotyledons and hypocotyls, as well as in the root tip, with this activity diminishing significantly by day four. KAT5.1-GUS expression was observed in the cotyledons, but appeared to be stronger near the apical meristem and in the root at day two. Expression in apical parts and the root tip diminished significantly during the five days analysed, but was maintained in the rest of the root. KAT5.1-GUS activity was also observed in the new leaves forming at the meristem. KAT5.2-GUS had lower expression levels with some staining of the petioles and also just behind the root tip during the first three days of germination. Seedlings grown for 16 d on MS medium were also stained for GUS (see Supplementary Fig. S2 at JXB online). Evidence for promoter activity was not seen in KAT1-GUS lines. KAT2-GUS and KAT5.1-GUS were expressed in the roots, petioles, and new leaves of the rosette and KAT5.2-GUS promoter activity was mostly limited to petioles (see Supplementary Fig. S2 at JXB online). Collectively, these results suggest that there are genomic elements in the region between the two KAT5 start codons that contribute significantly to transcription of the gene and that potentially drive expression of the isoforms to different levels.

KAT2 and KAT5 show differential expression patterns in reproductive tissue

In flowers, as in seedlings, KAT1-GUS promoter expression was essentially absent at all stages (Fig. 4). KAT2-GUS expression was absent during early flower development, and was first observed at early stages in young anthers and petals, corresponding to stage 12 of flower development. As the flowers matured, strong staining was observed in the anther filament and petals as well as the tip of the gynoecium and developing ovules. KAT5.1-GUS reported activity from the earliest stages of flower development, primarily in anthers, but was not present beyond stage 12 of flower development. Staining in KAT5.2-GUS lines was seen during the middle stages (~9–12) of flower development and was largely absent at the later stages. Promoter analysis in siliqua tissue indicated activity in the developing seeds for KAT2-GUS, KAT5.1-GUS, and KAT5.2-GUS, but not for the KAT1-GUS lines (Fig. 4).

Publicly available Arabidopsis thioglase gene expression data retrieved from the BAR eFP browser (http://bar.utoronto.ca/) (Winter et al., 2007) broadly agrees with the experimental data from promoter analysis (see Supplementary Fig. S3 at JXB online). KAT1 transcript levels were uniformly low. KAT2 transcript abundance vastly exceeded KAT1 and KAT5, with expression highest in germinating seeds (and also in dry seed and senescent leaves; see Supplementary Fig. S3A, B at JXB online). In flowers and developing seed, KAT2 expression was initially relatively low, and increased in maternal tissues from about stage 12 of flower development. KAT2 expression increases appreciably in the style once developing seeds have passed the torpedo stage (see Supplementary Fig. S3C, D, E at JXB online). As seen in the GUS reporter lines, KAT5 transcript abundance follows a complementary distribution to KAT2, with expression highest during early stages of flower and seed development (see Supplementary Fig. S3C, D, F at JXB online). In terms of the total abundance of transcript, KAT5 is substantially lower in most tissues (see Supplementary Fig. S3A, B at JXB online). However, KAT5 expression level matches KAT2 in early flower and seed development, after which KAT2 transcript abundance increases several-fold while KAT5 decreases substantially (see Supplementary Fig. S3D at JXB online).

As data derived from microarray analysis does not allow differentiation between KAT5.1cyt and KAT5.2px transcripts, qRT-PCR was used to detail the microarray expression data further (Fig. 5). Firstly, in germinating seeds (Fig. 5A), the KAT2 expression level was high and spiked 12 h after stratification but decreased again by the time seeds began to germinate at 48 h. KAT1 expression was much lower, but followed a similar pattern to KAT2. KAT5 expression was examined by use of primers located towards the 3’ end of the transcript, which assessed total transcript level (KAT5total), and by two alternative primer pairs at the 5’ end that were able to discriminate KAT5.1cyt and KAT5.2px. KAT5total levels, initially very low, began to rise significantly by 48 h when seeds were
germinating. *KAT5.2px* followed a similar pattern to *KAT5_total* and *KAT5.1cyt* remained barely detectable during the first 48 h of germination (Fig. 5A). In tissues of mature plants, *KAT1* expression was low and relatively uniform (Fig. 5B). *KAT2* and *KAT5total* were expressed at comparable levels in most tissues, although *KAT5 total* expression exhibited significantly higher peaks in flowers and siliques. *KAT5.2px* transcript was expressed at low levels in most plant tissues, with substantially higher expression in flowers and siliques, and *KAT5.1cyt* was expressed exclusively in the flowers and siliques (Fig. 5B).

Together with the GUS reporter and microarray data, these observations suggest a role for both *KAT2* and *KAT5* in fertility, in both flowers and siliques, with a role for *KAT5* primarily during early flower and embryo development and *KAT2* during seed filling and maturation.

**Discussion**

**Evolution of KAT genes**

Ancient Viridiplantae taxa such as green algae have a single *KAT* gene, whereas almost all sequenced land plant genomes (mosses and higher plants) possess at least two *KAT* genes. Phylogenetic analysis of predicted KAT protein sequences suggests that duplication of *KAT* genes may have been associated with the adoption of a land habitat and that two *KAT* genes have been selectively maintained ever since. In angiosperms the encoded proteins cluster into two groups that may be defined by their similarity to *A. thaliana* KAT2 and KAT5. Where a genome encodes three or more KAT proteins, the extra isozymes cluster with the respective KAT2 or KAT5 from that species suggesting that they are
Fig. 4. KAT:GUS promoter reporter activity in flowers and siliques. Flowers and siliques were removed from soil-grown KAT:GUS reporter plants at various stages of flower and silique development, and stained for GUS activity. The length of the scale bar is 0.5 mm.

Fig. 5. Quantitative RT-PCR analysis of KAT transcript abundance in Arabidopsis thaliana. (A) KAT transcript abundance during germination. (B) Relative expression of KAT transcripts in different plant organs. The abundance of KAT1, KAT2, KAT5total, KAT5.1cyt, and KAT5.2px transcripts was measured and normalized to reference gene (CACS) transcript levels. Error bars represent standard error from the mean (n=4).
more recent lineage-specific duplications (Fig. 1). For example, in the KAT2 clade, *Populus trichocarpa*, *Linum usitatissimum*, and *Glycine max* show duplicates clustered together, and there are three closely related *Brassica rapa* KAT2 proteins. KAT1 in the Brassicaceae appears to be an older duplication in that it was found in six species from four Brassicaceae genera (*Arabidopsis*, *Capsella*, *Thellungiella*, and *Brassica*) and there is extensive synteny between genomic sequences surrounding KAT1 and KAT2 in these species (see Supplementary Fig. S1 and Table S3 at JXB online). Despite its relatively high rate of divergence (branch lengths between KAT1s in Fig. 1 are long compared with those between KAT2s in the Brassicales), and very low levels of expression (Figs 3–5; see Supplementary Figs S2 and S3 at JXB online), *KAT1* has persisted in the Brassicaceae for at least 43 million years suggesting a selective advantage in maintaining this gene.

**Roles for thiolases during the plant life cycle**

In *A. thaliana*, both KAT2 and KAT5 are expressed in distinct (although not mutually exclusive) patterns. KAT2 is dominant during seed germination and its expression is co-ordinated with that of other β-oxidation genes, including MFP2, PXA1, ACX1, LACS6, and LACS7 (Fulda et al., 2004), during the first 2–3 d of seedling establishment when the bulk of storage lipid breakdown occurs. Indeed, the kAT2-1 mutant can germinate, but does not establish unless supplied with sugar in the medium (Germain et al., 2001). KAT2 expression is high throughout the life cycle with other peaks of expression during later stages of flower and seed development, and in senescence. By comparison, KAT5 expression is relatively minor during germination and strongest in young flowers and early in seed development.

**Strong expression of KAT2 and KAT5 in inflorescences and siliques** (Figs 4, 45; see Supplementary Fig. S3 at JXB online) and compromised reproductive success in mutants of core β-oxidation genes including kAT2, cts, acx1 acx5, and lacs6 lacs7 (Footitt et al., 2007a,b; Schlimmer et al., 2007) suggests a role for β-oxidation in reproduction. β-oxidation potentially contributes to reproductive success via its roles in fatty-acid turnover and/or synthesis of the hormones JA and IAA. Auxin can be synthesized via tryptophan-dependent and -independent pathways, and can also be derived from stored forms; either IAA conjugates (amino acid, sugar or peptide), or from indole-3-butyric acid (IBA) via one cycle of β-oxidation (Baker et al., 2006). Similarly, JA biosynthesis involves the precursor, 3-oxo-2-(2’-[Z]-pentenyl)-cyclopentane-1-octanoic acid (OPC8), undergoing three cycles of β-oxidation to produce JA (Baker et al., 2006).

Many mutants disrupted in JA synthesis or signalling (including acx1 acx5) display male sterility, usually in the form of defective pollen development, that can be rescued by exogenously supplied JA (Schlimmer et al., 2007). kAT2 mutants, however, are not sterile, presumably due to compensation of KAT function by KAT5 (Afitlhile et al., 2005). KAT2 appears to be the dominant thiolase in IBA metabolism: kAT2 mutant seedlings have an IBA-resistant phenotype while kAT1 and kAT5 knockouts do not (Wiszniewski et al., 2009). It is likely that β-oxidation of IBA to IAA contributes significantly to levels of free IAA in seedlings (Strader et al., 2011). In flowers, IAA generated in the stamens controls flower development by promoting anther growth and suppressing petal development (Aloni et al., 2006). Anther elongation in cts and kAT2-1 mutants is inhibited, but it can be rescued by exogenous supply of 1-naphthaleneacetic acid (NAA; a synthetic auxin) or IAA (but not by JA) (Footitt et al., 2007b). High KAT2 expression in anther filaments and KAT5 expression in anthers (Fig. 4) is consistent with an essential role for β-oxidation-mediated hormone metabolism in flower development.

During fertilization, pollen tube growth is extremely rapid and lipids may provide some of the energy required to drive this growth. Lipid bodies and peroxisomes accumulate in developing and mature *Arabidopsis* (Kuang and Musgrave, 1996) and olive (Rodriguez-Garcia et al., 2003) pollen. They then diminish in number during pollen tube growth. *In vitro* germination tests of pollen from kAT2-1 and cts mutants revealed reduced germination frequency and shorter pollen tube length, but these could be rescued by exogenous sucrose supply, suggesting lipid degradation via β-oxidation may contribute to the growth of germinating pollen tubes *in vivo* (Footitt et al., 2007b).

β-oxidation is also implicated during seed maturation. While fatty acid accumulation in developing embryos originates materially as photosynthetically fixed carbon, β-oxidation of stored oil may provide respiratory substrates following the breaking of the trophic connection with maternal tissue (Baud et al., 2002). In fact, a 10% decrease in seed oil content has been observed to occur in *B. napus* embryos late in their development (Eastmond and Rawsthorne, 2000) and β-oxidation, the glyoxylate cycle, and gluconeogenesis are all active in developing embryos (Eastmond and Graham, 2001; Chia et al., 2005). A lower respiration rate in kAT2-1 mutant ovules compared with the wild type suggests that β-oxidation is important for carbon flow into sugars via gluconeogenesis and respiration (Footitt et al., 2007a). This is corroborated by strong expression of KAT2 and KAT5 in silique tissue (Figs 4, 45; see Supplementary Fig. S3 at JXB online) implying a role for both thiolases in reproductive success.

**A specific function for thiolases in Brassicales?**

The apparent coincident origin and maintenance in the Brassicales (but not other plant orders) of a third KAT gene (KAT7; Fig. 1) and a KAT5 orthologue that is alternatively transcribed to produce cystosol- and peroxisome-targeted proteins (Fig. 2) raises the question as to whether these events were related and facilitated a specific function for thiolases in Brassicales. The first product of the phenylpropanoid pathway, trans-cinnamic acid, represents a branch point to either flavonoids or benzenoid metabolism (Boatright et al., 2004). Intriguingly, KAT5 coexpresses with genes of flavonoid metabolism (Carrie et al., 2007). Benzoic acid (BA) synthesis from *trans*-cinnamic acid can occur cytosolically or in peroxisomes. In the first case, hydration is followed by retro-aldol cleavage (to release an acetic acid molecule) and dehydrogenation of benzaldehyde to yield BA. BA can then be activated to form benzoyl-CoA. In peroxisomes, CoA-activated *trans*-cinnamic acid undergoes one cycle of β-oxidation: hydration followed by dehydrogenation and KAT-mediated thiolysis to yield acetyl-CoA and benzoyl-CoA. Both of these pathways have been shown in petunia to contribute to the plant BA pool.
Nine 3-ketoacyl-CoA thiolases (KATs) and acetoacetyl-CoA thiolases (ACATs) encoded by five genes in Arabidopsis thaliana are targeted either to peroxisomes or cytosol but not to mitochondria. Plant Molecular Biology 63, 97–108.


Hayashi M, Toriyama K, Kondo M, Nishimura M. 1998. 2,4-Dichlorophenoxybutyric acid-resistant mutants of Arabidopsis have defects in glyoxysomal fatty acid β-oxidation. The Plant Cell 10, 183–195.


