JARQ 50 (4), 307 - 317 (2016) http://www.jircas.affrc.go.jp

Molecular Cloning, Expression and Characterization of a Squalene Synthase Gene from Grain Amaranth (*Amaranthus cruentus* L.)

Young-Jun PARK[4], Kazuhiro NEMOTO[1], Mineo MINAMI[1], Kenichi MATSUSHIMA[1], Tomohide NOMURA[1], Jun-ichi KINOSHITA[1] and Tomotaro NISHIKAWA[2, 3]*

[1] Graduate School of Agriculture, Shinshu University (Kami-ina, Nagano 399-4598, Japan)
[2] Genetic Resources Center, National Institute of Agrobiological Sciences (Tsukuba, Ibaraki 305-8602, Japan)
[4] College of Science & Technology, Korea University (Seongbuk-Gu, Seoul 136-705, Korea)

Abstract

A gene encoding squalene synthase from grain amaranth was cloned and characterized. The full-length cDNA was 1805-bp long and contained a 1248-bp open reading frame encoding a protein of 416 amino acids with a molecular mass of 47.6 kDa. Southern blot analysis revealed that the *A. cruentus* genome contained a single copy of the gene. Comparison of the cDNA and genomic sequences indicated that the amaranth *SQS* gene had 12 introns and 13 exons. All 13 exons contributed to the coding sequence. The predicted amino acid sequence of the SQS cDNA shared high homology with those of SQSs from several other plants. It contained five conserved domains that are believed to represent crucial regions of the active site. We conducted qRT-PCR analyses to examine the expression pattern of the *SQS* gene in seeds at different developmental stages and in several tissues. The amaranth *SQS* gene was expressed late in seed development and played a role in the perisperm, and mainly expressed in stem and root tissues.

Discipline: Genetic resources
Additional key words: Gene expression, molecular phylogeny

Introduction

Squalene synthase (SQS) (farnesyl-diphosphate: farnesyl-diphosphate farnesyltransferase, EC 2.5.1.21) is an endoplasmic reticulum membrane enzyme that catalyzes a two-step reaction. In the first step, two farnesyl diphosphate (FPP) molecules are condensed to form an intermediate, presqualene diphosphate (PSPP). In the second reaction step, PSPP is reductively rearranged into squalene in the presence of NADPH and Mg^{2+} (Nakashima et al. 1995).

Previous studies have described the isolation and characterization of SQS, which plays a role in controlling sterol biosynthesis, from rats and yeast, and highly purified preparations of SQS have been obtained from these organisms (Karst & Lacroute 1977, Shechter et al. 1992, Tozawa et al. 1999). Nakashima et al. (1995) were the first to report on an SQS from plants. They cloned an *Arabidopsis* SQS cDNA and characterized the recombinant enzyme via prokaryotic expression. Since then, cDNAs for SQSs from several other plants have been isolated and characterized (Hata et al. 1997, Devarenne et al. 1998, Hayashi et al. 1999, Uchida et al. 2009).

The genus *Amaranthus* includes more than 60-70 wild or weedy species and some cultivated species (Sauer 1967). This genus, which originated in the New World, is an ancient crop that was already under cultivation 5000-7000 years ago (Sauer 1967). Ancient amaranth grains still used to this day include the three species *Amaranthus caudatus, Amaranthus cruentus*, and *Amaranthus hypochondriacus*. The potential of both grain (seed) and vegetable (leaf) amaranths as food resources has been reviewed extensively (Shukla et al. 2006). Amaranth grains have agronomic po-

Present address:
[3] Research Policy Planning Division, Ministry of Agriculture, Forestry and Fisheries (Chiyodaku, Tokyo 100-8950, Japan)
*Corresponding author: e-mail tomotaro@affrc.go.jp
Received 11 August 2015; accepted 20 November 2015.

tential due to high levels of the essential amino acid lysine in the protein. Therefore, this crop is among the New World super grains and gaining favor among health-conscious consumers. There is also a great deal of interest in amaranth oils as a potentially rich source of squalene. Recent research showed that the squalene content of olive, wheat germ, and rice bran oils ranged from 0.1 to 0.7% (He et al. 2002), which is not sufficient for these oils to be considered viable resources for squalene. Amaranth oil contains 2.4 to 8.0% squalene, almost eight times the amount found in common vegetable oil. As such, it is a more viable resource (Sun et al. 1997, He et al. 2002). According to previous reports, the squalene concentration per total seed weight in the cultivated amaranths was 0.43% in *A. cruentus* and 0.73% in *A. hypochondriacus* (He et al. 2002).

To date, there have been no reports on the cloning and characterization of a gene encoding SQS from *Amaranthus*, although there have been some reports on methods of extracting and purifying squalene from *Amaranthus* seeds (He et al. 2002, 2003, Westerman et al. 2006). In this study, we first isolated and characterized a cDNA clone for a grain amaranth *SQS* gene. We then monitored the expression pattern of the amaranth *SQS* gene in seeds at different developmental stages and in different tissues.

Materials and methods

1. Plant materials

Grain amaranth (*A. cruentus*, accession no. PI 433228) was grown in a greenhouse for 90 days. We observed the development process of amaranth seeds from the pollination of female flowers to the mature seeds. Developing grains were collected at several time points (day after pollination [DAP]) during maturation and divided into six developmental stages: initial (1-3 DAP; ca. 0.03 mg), early (3-5 DAP; ca.0.12 mg), early late (4-8 DAP; ca. 0.26 mg), middle (8-12 DAP; ca. 0.42 mg), mid-late (10-15 DAP; ca. 0.54 mg), and late (20 DAP; ca. 0.65 mg), based on their external morphology, fresh weight, and size (Park & Nishikawa 2012). Fresh weight values are the average of 10 seeds. The leaves, petioles, stems, and roots were collected from seedlings at the four- and six-leaf stages. Samples were immediately frozen in liquid nitrogen and stored at −80°C until use.

2. Isolation of total RNA and genomic DNA

Total RNA was isolated from plant tissue samples using a RNeasy Mini kit (Qiagen) as per the manufacturer's instructions. Genomic DNA was extracted from young leaves using the cetyl trimethyl ammonium bromide (CTAB) method of Murray & Thompson (1980) with some modifications. Isolated RNA and DNA were quantified with a Nanodrop ND-1000 spectrophotometer (Nano-Drop Technologies, Wilmington, DE, USA) and/or by separation on 1.4% (w/v) agarose-formaldehyde gels, which were visualized under ultraviolet light after staining with ethidium bromide. Finally, RNA was stored at −80°C and DNA at 4°C until use.

3. Cloning of the amaranth *SQS* gene

Alignment of the predicted cDNA sequences of *SQS* genes from *Arabidopsis thaliana* (accession no. NM_119630), *Nicotiana tabacum* (accession no. NTU60057), *Capsicum annuum* (accession no. AF124842), and *Solanum tuberosum* (accession no. AB022599) identified several conserved domains. The inferred amino acid sequences were obtained from the NCBI GenBank database (http://www.ncbi.nlm.nih.gov/). Two conserved regions were used to design sense (SQ-DP-F1: 5′-CATGTGGTACRAARGAGTACAAGGTTCT-3′) and antisense (SQ-DP-R1: 5′- TCCTCAAGCTTGTYARCATATTT-3′) degenerate primers to amplify the amaranth *SQS* gene. The first-strand cDNA was constructed using the SMARTer RACE cDNA Amplification kit (Clontech) and then by using the degenerate primers, a 400-bp cDNA fragment (ArSQ) was amplified and cloned. Based on the results of polymerase chain reaction (PCR) analysis, the 400-bp cDNA fragment was used to design specific primers for *SQS* amplification. RT-PCR, and 3′- and 5′-rapid amplification of cDNA end (RACE) were performed using the specific primers (Table 1). PCR amplification was carried out using *ExTaq* polymerase (0.5 U/μl), 2.5 mM dNTPs, 10 × PCR buffer (Takara), and 0.5 μM of each primer. Some amplified fragments were extracted from agarose gels using the QIAquick Gel Extraction kit (QIAGEN) as per the manufacturer's instructions. All cDNA-amplified fragments were ligated into the pGEM-T-Easy vector using the TA cloning kit (Promega) and then transformed into *Escherichia coli* strain JM109 cells. Finally, based on analysis of the cDNA sequences, we designed specific primers (Table 2) and the *SQS* gene was amplified and cloned from genomic DNA.

4. Southern blot analysis

Amaranth genomic DNA (10 μg) was digested at 37°C with *Eco*RI and *Sal*I. The digested products were separated on a 1 % agarose gel, and blotted onto a Hybond-N+ nylon membrane (GE Healthcare). The probe (*ArSQSp*) was amplified by PCR from *SQS* cDNA using the following specific primer pair: sense primer 5′-GAGGATGATACAAGCATAGCTACAGATGT-3′ and antisense primer 5′-AGTCTGTCTTATCAATAACTCTTGCGGT-3′. The blot was probed with the 742-bp *ArSQp* PCR product (corresponding to nucleotide positions 238- to 979-bp; accession no. AB691229) labeled using the ECL Direct Nucleic Acid Labeling and Detection System (GE Healthcare) as per the manufacturer's instructions. The membrane was prehybrid-

Table 1. Primers used for amplifying and/or sequencing of cDNA fragments

Primers	Forward and reverse PCR primer sequences (5'→3')	Direction	Amplification
SQ-DP-F1	AAGCAGATCCCKCCKKMKCCACAYTGGGG	F	Homologous fragment
SQ-DP-F2	CTTGAYACTGTYGAGGATGATACYAGCAT	F	Homologous fragment
SQGSP-F1	TAATGCAGGTTTGGAGGATCTTGCATCG	F	5' RACE
SQGSP-F2	ATGTCGCATGTTTTGGCCTCGGGAGATA	F	5' RACE
SQGSP-F3	ATTACCAAAAGGATGGGTGCGGGAAT	F	5' RACE
SQGSP-F4	GCAGGACTCGTGGGTTTAGGGTTGTC	F	5' RACE
SQGSP-F5	TCTGTGCTATTCCACAAATCATGGCCATTG	F	5' RACE
SQGSP-F6	ACAGACTCGATGCCTGATGTTTATGGTG	F	5' RACE
SQGSP-F7	ACATGTCGGCTCTGCGGGATCATGCT	F	5' RACE
SQGSP-R1	CCGATGCAAGATCCTCCAAACCTGCATT	R	3' RACE
SQGSP-R2	CCATTCCCGCACCCATCCTTTTGGTAAT	R	3' RACE
SQGSP-R3	CATATCTCCCGAGGCCAAAACATGC	R	3' RACE
SQGSP-R4	TGCCATTCCCGCACCCATCCTTTTGG	R	3' RACE
SQGSP-R5	ACAGTGTCGAGGGCTCGAAGGACCAAGT	R	3' RACE
SQ-DP-R3	TGCAARRAYWATRGCCAGTAKRAT	R	Homologous fragment
SQ-DP-R4	TCAGCAURCARGAAAARTCAWARAAAGCACCATA	R	Homologous fragment
UPM	CTAATACGACTCACTATAGGGCAAGCAGTGGTATCAACGCAGAGT	F and/or R	5' RACE and 3' RACE in contech
NUP	CTAATACGACTCACTATAGGGC	F and/or R	5' RACE and 3' RACE in contech

Table 2. Primer sequences and annealing temperatures used for amplification of DNA fragments

Primer pairs	Forward and reverse PCR primer sequences (5'→3')	Amplified region	Annealing temperature
SQ-F1a/SQ-R1b	F: ATTGCATCCAACTTCGAATGAATCGAAT	5' UTR-Intron 1	60°C
	R: AGTTCCGTATCATTCCCATGGTAATGAA		
SQ-F2a/SQ-R1	F: ACGGTCAGTAATAAACGGTGGTAATG	Intron 1-Exon 2	60°C
	R: ACAGTGTCGAGGGCTCGAAGGACCAAGT		
SQ-F2b/SQ-R2a	F: AGTCTCAAAAGTCAAAATTGAATTCACT	Intron 1- Intron 3	60°C
	R: ATGACTGAAACCTAAGGGCCACTATA		
SQ-F3a/SQ-R3a	F: ATCGCGAATGGCACTTTTCTTGTGAGTA	Exon 3- Intron 4	60°C
	R: TGAGTAGTAACGATCAGTTTTCATCAGTGA		
SQ-F4/SQGSP-R3	F: TCTCATGGATGAGTTTCATCAAGTCTCA	Exon 4-Exon 7	60°C
	R: CATATCTCCCGAGGCCAAAACATGC		
SQGSP-F1/SQ-R5a	F: TAATGCAGGTTTGGAGGATCTTGCATCG	Exon 6-Intron 8	60°C
	R: ACAGTGGCTAACAATTGATAAGGGTGAT		
SQGSP-F7/SQ-R6a	F: ACATGTCGGCTCTGCGGGATCATGCT	Exon 8- Intron 9	60°C
	R: AGAGCTTGAATTAGTCACGTCTAGAGT		
SQ-F4b/SQ-R8	F: ATATAATCCGATTCGAGTCCGAGGTA	Intron 9- Intron 11	60°C
	R: TAGTGTTTTCATGGCGTTTGGATCAT		
SQGSP-F6/SQ-R10	F: ACAGACTCGATGCCTGATGTTTATGGTG	Exon 10- 3' UTR	60°C
	R: ATCTCTTAAGTGATACTGATATATATGCT		

ized at 42°C for 1 h in hybridization buffer supplemented with 0.5 M NaCl and 5% blocking agent. Hybridization was carried out overnight at 42°C in roller bottles. After hybridization, the membrane was washed twice with 0.4% sodium dodecyl sulfate (SDS)/0.5× saline-sodium citrate (SSC) buffer for 15 min at 42°C. Finally, the membrane was washed twice with 2× SSC for 5 min per wash at room temperature. After incubation with ECL detection reagents,

hybridization products were detected by exposure of the membrane to Hyperfilm ECL X-ray film (GE Healthcare).

5. Quantitative real-time PCR analysis

Grain amaranth cDNA was synthesized from 1 µl total RNA (50 ng) using the High Capacity RNA-to-cDNA kit (Applied Biosystems, USA). Quantitative real-time PCR (qRT-PCR) analysis was conducted using an ABI Prism 7900HT sequence detection system and software version 2.2 (Applied Biosystems). The gene-specific primers were designed based on the nucleotide sequence of *A. cruentus SQS* (accession no. AB691229). The *A. tricolor Actin* gene (accession no. EF452618) was used as an internal constitutively expressed control (reference gene). The PCR primers used for real-time quantification of *SQS* were as follows: sense primer [5′-GACTCGTGGGTTTAGGGTTGTC-3′] and antisense primer [5′- TGGAAAGATCATCCGATGCA -3′]. The primers used for the control reaction using *Actin* were as follows: sense primer [5′-GTATGCAAGTGGTC-GTACTACAGG-3′] and antisense primer [5′-ATCTTCG-TAGGGTAATCAGTCAGG-3′]. All primers were designed using Primer Express 1.5 software (Applied Biosystems) following the supplier's guidelines for primer design. Transcript abundance was determined using Power SYBR Green PCR Master Mix (Applied Biosystems) as recommended by the manufacturer. One microliter of cDNA was used as a template for PCR. The PCR cycling conditions consisted of an initial polymerase activation step at 95°C for 10 min, followed by 40 cycles at 95°C for 15 s and at 58°C for 1 min. For each sample, reactions were set up in triplicate to ensure reproducibility of the results. Each sample was normalized using values obtained for the level of *Actin* mRNA.

6. Semi-quantitative RT-PCR analysis

Total RNA isolated from the tissue samples was used for semi-quantitative RT-PCR analysis to measure RNA expression levels of the *SQS* gene. First-strand cDNA was prepared with the High Capacity RNA-to-cDNA kit (Applied Biosystems) and subjected to PCR with appropriate gene-specific primers. We used the following primers: (sense) [5′-ATTACCAAAAGGATGGGTGCGGGAAT-3′] and (antisense) [5′-CCGATGCAAGATCCTCCAAACCT-GCATT-3′]. Total RNA (2 µg) was reverse-transcribed and amplified in 25-µl reaction mixtures. The RT-PCR cycling conditions were as follows: reverse transcription at 50°C for 60 min, amplification for 30 cycles of denaturation at 98°C for 10 s, annealing at 62°C for 30 s, and extension at 72°C for 30 s. Each RT-PCR reaction was performed at least twice. Finally, all products were cloned and sequenced to confirm that they were fragments of the intended gene. The cDNA for *Actin gene* (accession no. EF452618) was used as a control for RNA integrity and loading accuracy.

7. Sequencing and phylogenetic analyses

Sequencing of cDNA and genomic DNA was performed using the BigDye Terminator Cycle Sequencing kit (version 3.1, Applied Biosystems) on a 3100 Genetic Analyzer (Applied Biosystems) as per the manufacturer's instructions. Sequences were assembled using the CodonCode Aligner software to produce a full-length gene (CodonCode, Dedham, MA, USA). Protein molecular weight calculations were performed with Geneious Pro 5.5.7 (Biomatters Ltd.). BLAST analysis searches were made online with default settings (www.ncbi.nlm.gov/BLAST/). Sequences of SQS proteins from diverse plant species were obtained from the NCBI database for phylogenetic analysis. A multiple sequence alignment was conducted using ClustalX (Thompson et al. 1997). The sequence of the glycogen synthase of *E. coli* was used as an outgroup. Phylogenetic analysis was conducted using the neighbor joining (NJ) method with PAUP* 4.0 (Swofford 1988). Bootstrap confidence values were obtained from 1,000 replicates.

Results

1. Isolation and characterization of SQS cDNA clones

The complete sequence of the full-length SQS cDNA was determined from five positive cDNA clones by 3′- and 5′-RACE. Fig. 1 shows the sequence of the cDNA clone. The amaranth SQS cDNA was 1,805-bp long, and included a 1248-bp open reading frame (ORF) from 295 bp to 1542 bp of the sequence, a 294-bp 5′ untranslated region (UTR), and a 263-bp 3′ UTR including a 30-bp poly (A) tail. A typical eukaryotic polyadenylation signal (AATATA) was found upstream from the poly (A) tail. The amaranth SQS cDNA was 75.9, 75.1, 75.1, 74.9, 74.7, 74.5, 74.0, 72.0, 67.5, and 67.2% identical to that of the SQS in *Glycyrrhiza glabra* (D86409), *Euphorbia tirucalli* (AB433916), *Lotus japonicas* (AB102688), *Panax ginseng* (AB010148), *N. tabacum* (accession no. NTU60057), *C. annuum* (accession no. AF124842), *S. tuberosum* (accession no. AB022599), *A. thaliana* (accession no. NM_119630), *Oryza sativa* (NM_001058160), and *Zea mays* (NM_001111369), respectively. The nucleotide sequence data reported in this study was deposited in the DNA Data Bank of Japan (DDBJ), European Molecular Biology Laboratory (EMBL), and GenBank nucleotide sequence databases under accession no. AB691229.

2. Isolation and characterization of SQS genomic clones

The SQS genomic DNA was 7047 nucleotides long, of which 5242 nucleotides were introns. Alignment of the genomic sequence with the cDNA sequence showed that the genomic DNA consisted of 13 exons (195, 43, 90, 76, 70, 144, 105, 147, 76, 89, 92, 46, and 75 bp) and 12 introns

```
1     AGGATATCCCACCATTCTTTTCTTAAATTGCATCCAACTTCGAATGAATCGAATCCCAGAAGCCAGAATTTGTTCATCTTCACTTACATA

91    ATTAAAATATAAAAAAAATCAGTTATTCTTATAAAAACCCCTTTTGTAATTTTCAATTTGGTCAATATTTCAACTGATTTCCTGAGATTC

181   TCCTTACACCACAAAAGCGTGTTACAAATCTTTTACAGAATCAGCATATTATTGTATAAATTTACGCCTTAAATTGAGAAATTGAACATT

271   ACGCATTGTTGAAAATTGATCAAAATGGGGAGTTTAGGAGCAATTTTGAAACACCCAGATGAATTTTACCCACTTTTGAAGTTGAAGATG
                              M  G  S  L  G  A  I  L  K  H  P  D  E  F  Y  P  L  L  K  L  K  M

361   GCTGTTAAAGAAGCTGAGAAACAGATCCCATCTGAATCTCATTGGGGGTTTTGTTACTCTATGCTTCATAAAGTTTCCAGAAGTTTTGCT
       A  V  K  E  A  E  K  Q  I  P  S  E  S  H  W  G  F  C  Y  S  M  L  H  K  V  S  R  S  F  A

451   CTTGTTATTCAGCAACTTGGCACTGAGCTTCGCAATGCTGTATGTGTCTTTTACTTGGTCCTTCGAGCCCTCGACACTGTAGAGGATGAT
       L  V  I  Q  Q  L  G  T  E  L  R  N  A  V  C  V  F  Y  L  V  L  R  A  L  D  T  V  E  D  D

541   ACAAGCATAGCTACAGATGTCAAATTGCCTATCCTGAAAGCTTTTTATCAGCACATATATGATCGCGAATGGCACTTTTCTTGTGGCACA
       T  S  I  A  T  D  V  K  L  P  I  L  K  A  F  Y  Q  H  I  Y  D  R  E  W  H  F  S  C  G  T

631   AAACACTACAAAGTTCTCATGGATGAGTTTCATCAAGTCTCAACTGCTTTTTTGGAGCTTGAAAGAGGTTACCAGCTCGCAATTGAGGAT
       K  H  Y  K  V  L  M  D  E  F  H  Q  V  S  T  A  F  L  E  L  E  R  G  Y  Q  L  A  I  E  D

721   ATTACCAAAAGGATGGGTGCGGGAATGGCAAAATTCATCTGTCAGGAGGTAGAAACAGTTAGCGACTATGACGAATACTGCCACTATGTA
       I  T  K  R  M  G  A  G  M  A  K  F  I  C  Q  E  V  E  T  V  S  D  Y  D  E  Y  C  H  Y  V

811   GCAGGACTCGTGGGTTTAGGGTTGTCTAAGCTTTTCCATAATGCAGGTTTGGAGGATCTTGCATCGGATGATCTTTCCAATTCGATGGGT
       A  G  L  V  G  L  G  L  S  K  L  F  H  N  A  G  L  E  D  L  A  S  D  D  L  S  N  S  M  G

901   TTATTTCTTCAGAAAACCAACATTATCCGTGATTACCTCGAAGATATAAATGAGATTCCAAAATGTCGCATGTTTTGGCCTCGGGAGATA
       L  F  L  Q  K  T  N  I  I  R  D  Y  L  E  D  I  N  E  I  P  K  C  R  M  F  W  P  R  E  I

991   TGGAGTAAATATGTCAACAAGCTTGAGGACCTGAAATATGAGGAGAACTCTGTGAAGGCAGTTCAATGTTTAAATGACATGGTAACAAAT
       W  S  K  Y  V  N  K  L  E  D  L  K  Y  E  E  N  S  V  K  A  V  Q  C  L  N  D  M  V  T  N

1081  GCTTTATTGCATGTGGAAGATTGCCTAAAGTACATGTCGGCTCTGCGGGATCATGCTATATTCCGGTTCTGTGCTATTCCACAAATCATG
       A  L  L  H  V  E  D  C  L  K  Y  M  S  A  L  R  D  H  A  I  F  R  F  C  A  I  P  Q  I  M

1171  GCCATTGGAACTTTAGCTTTATGCTACAACAACGTTGAAGTCTTCAGAGGTGTGGTTAAAATGAGACGTGGTCTCACCGCAAGAGTTATT
       A  I  G  T  L  A  L  C  Y  N  N  V  E  V  F  R  G  V  V  K  M  R  R  G  L  T  A  R  V  I

1261  GATAAGACAGACTCGATGCCTGATGTTTATGGTGCCTTCTACGATTTTGCCTGCATGATAAAACCAAAGGTCGACAAGAATGATCCAAAC
       D  K  T  D  S  M  P  D  V  Y  G  A  F  Y  D  F  A  C  M  I  K  P  K  V  D  K  N  D  P  N

1351  GCCATGAAAACACTAAGTAGAATCGATGCTATCGAGAAGATTTGTAGGGACTCCGGCACTCTGAACAAAAGGAAACTGCACATCACAAGC
       A  M  K  T  L  S  R  I  D  A  I  E  K  I  C  R  D  S  G  T  L  N  K  R  K  L  H  I  T  S

1441  ACCAAGTCGGCATACACTCCAATCATGGTGATGGTGTTATTCATTGTATTGGCAATCTTCTTCAATCGTTTATCAGAATCCAACAGAATG
       T  K  S  A  Y  T  P  I  M  V  M  V  L  F  I  V  L  A  I  F  F  N  R  L  S  E  S  N  R  M

1531  ATTAATAACTAAGCTCTTGGGTATGTTTAGCACGTCTTGATTTGGATATTACGTACGCAAAAGCTACTAGCTACCTGCTATCTGCCAATA
       I  N  N  *

1621  AATTTTAACATTTACCGCTACATTTTTTCAGGCATGACGTCTTATTTACCGCTACTTTATCTAGGTGTGAAACCATGGCTTGTTTGATTA

1711  ATTATTTTGTAATATACTCCTATTTGGTAGTATGTAGCATATATATCAGTATCACTTAAGAGATGAAAAAAAAAAAAAAAAAAAAAAAA

1801  AAAAA
```

Fig. 1. Nucleotide and deduced amino acid sequences of amaranth *SQS* gene. ATG start codon and TAA stop codon are enclosed in boxes. Potential polyadenylation sites are underlined.

(1212, 448, 602, 423, 202, 117, 417, 302, 1276, 82, 78, and 83 bp), and that the assembled exon regions were completely identical to the cDNA sequence. The assignment of introns for the amaranth *SQS* gene was reinforced by the fact that all introns followed the universal GT/AG rule. The exons and introns had a total guanine-cytosine (GC) content of 41.1 and 33.1%, respectively.

3. Amaranth SQS protein

Fig. 1 shows the deduced amino acid sequence of amaranth SQS. The 1248 nucleotide ORF of the amaranth SQS cDNA encoded a protein containing 416 amino acids with a theoretical molecular mass of 47.6 kDa. NCBI protein-protein BLAST analysis showed that the deduced amino acid sequence of amaranth SQS had 86% (349/406) identities and 94% (382/406) positives with the SQS of *P. tunicoides*. Fig. 2 shows the alignment of the amino acid sequence of the SQS protein from amaranth against those of other plant species. The amaranth SQS showed broad and high local identities to those of other plant species. The amaranth SQS showed 80.2, 79.4, 78.8, 78.5, 78.2 78.5, 77.2, 74.3, 70.2, and 70.1% amino acid sequence identities with the SQS in *G. glabra, E. tirucalli, L. japonicas, C. annuum, N. tabacum, S. tuberosum, P. ginseng, A. thaliana, O. sativa*, and *Z. mays*, respectively. The amaranth SQS protein contained five domains, which are conserved sequences in other SQS proteins. Three of the five domains (II, III and IV) were highly conserved, while domains I and V and were moderately conserved within the amaranth SQS. The conserved domains II, III, and IV are believed to represent crucial regions of the active site (Hata et al. 1997). These domains are essential for the two half-reactions catalyzed by the enzyme (Kim et al. 2011). Domain V shows low sequence identity among the known SQS proteins; however, the amino acid sequence of this domain is very hydrophobic in all SQS enzymes, including the amaranth SQS. The high level of sequence identity and the presence of conserved domains found in other SQSs as reported thus far suggest that the amaranth cDNA isolated in this study encodes a SQS.

4. Genomic DNA blot analysis

DNA isolated from grain amaranth was used for genomic DNA blot analysis. Genomic DNA was digested with *Eco*RI and *Sac*I, and hybridized with a 742-bp fragment of *ArSQp* (Fig. 3). *Eco*RI digestion generated two fragments of 6.5 and 3.8 kb. Digestion with *Sac*I yielded only one fragment (2.3 kb). A cut site for *Eco*RI restriction enzyme was detected in intron 10 of the *SQS* gene. These results showed that a single gene encoding *SQS* was present in the genome of *A. cruentus*.

5. Expression patterns of *SQS* during seed development

The transcript expression pattern of the *SQS* gene was investigated using total RNA isolated from seeds at different stages of development (Fig. 4a). There were low levels of *SQS* transcripts at the early stage of seed development, and the levels remained low until the middle developmental stage. The expression of *SQS* increased rapidly to reach a peak at the mid-late developmental stage, and then declined dramatically. This pattern of expression was consistent with the results of RT-PCR analyses. All RNA samples generated a fragment of the expected size (183-bp). The amaranth *SQS* was expressed at low levels during the initial to middle stages of seed development, and its expression level increased at the mid-late development stage.

6. Expression patterns of *SQS* in different tissues

The tissue-specific expression of amaranth *SQS* was determined by quantifying its mRNA in total RNA isolated from the leaves, petioles, stems, and roots of seedlings at the four- and six-leaf stages (Fig. 4b). Using qRT-PCR and RT-PCR analysis, we detected amaranth *SQS* transcripts in some of the tissues at the six-leaf stage, but in none of the tissues from plants at the four-leaf stage. *SQS* transcripts accumulated in almost equal amounts in stems and roots, while a lower level accumulated in leaves and petioles during seedling development at the four- to six-leaf stages.

Discussion

Amaranthus, a potentially rich source of squalene, is becoming a model plant for research in the oilseed crop area. Isolation and characterization of a *SQS* gene from grain amaranth is the first step towards understanding regulation of the squalene biosynthetic pathway. Genes encoding SQS have been isolated from many sources, such as fungi (Jennings et al. 1991), bacteria (Lee & Poulter 2008), animals (McKenzie et al. 1992), and plants (Nakashima et al. 1995, Hata et al. 1997, Devarenne et al. 1998, Hayashi et al. 1999, Akamine et al. 2003, Huang et al. 2007, Uchida et al. 2009, Kim et al. 2011). To our knowledge, this is the first report on the molecular cloning and expression analysis of an *SQS* gene from grain amaranth. The recent molecular cloning of *SQS* genes from other plant species facilitated our isolation of *SQS* cDNA from *A. cruentus*. The full-length amaranth cDNA of *SQS* gene included a 1,248-bp ORF that encoded a predicted polypeptide of 416 amino acids (Fig. 1) with a predicted mass of 47.6 kDa. The result is similar to that of SQSs in other plants, including *A. thaliana* (Nakashima et al. 1995), *Taxus cuspidata* (Huang et al. 2007), and *P. ginseng* (Kim et al. 2011).

According to molecular evidence, in some organisms there is only one copy of the *SQS* gene in the genome, while other organisms have multiple copies in their genome. Mammals and yeasts have a single copy in their genomes (Robinson et al. 1993, Hata et al. 1997). Among plant species, a single *SQS* gene was reported in *O. sativa* (Hata et al. 1997), *T. cuspidata* (Huang et al. 2007), and *E. tirucalli* (Uchida et al. 2009). Likewise, it appears that there is a single copy of the *SQS* gene in the grain amaranth genome, as suggested by Southern analysis (Fig. 3). On the other hand, two or three *SQS* genes were reported in *G. glabra*

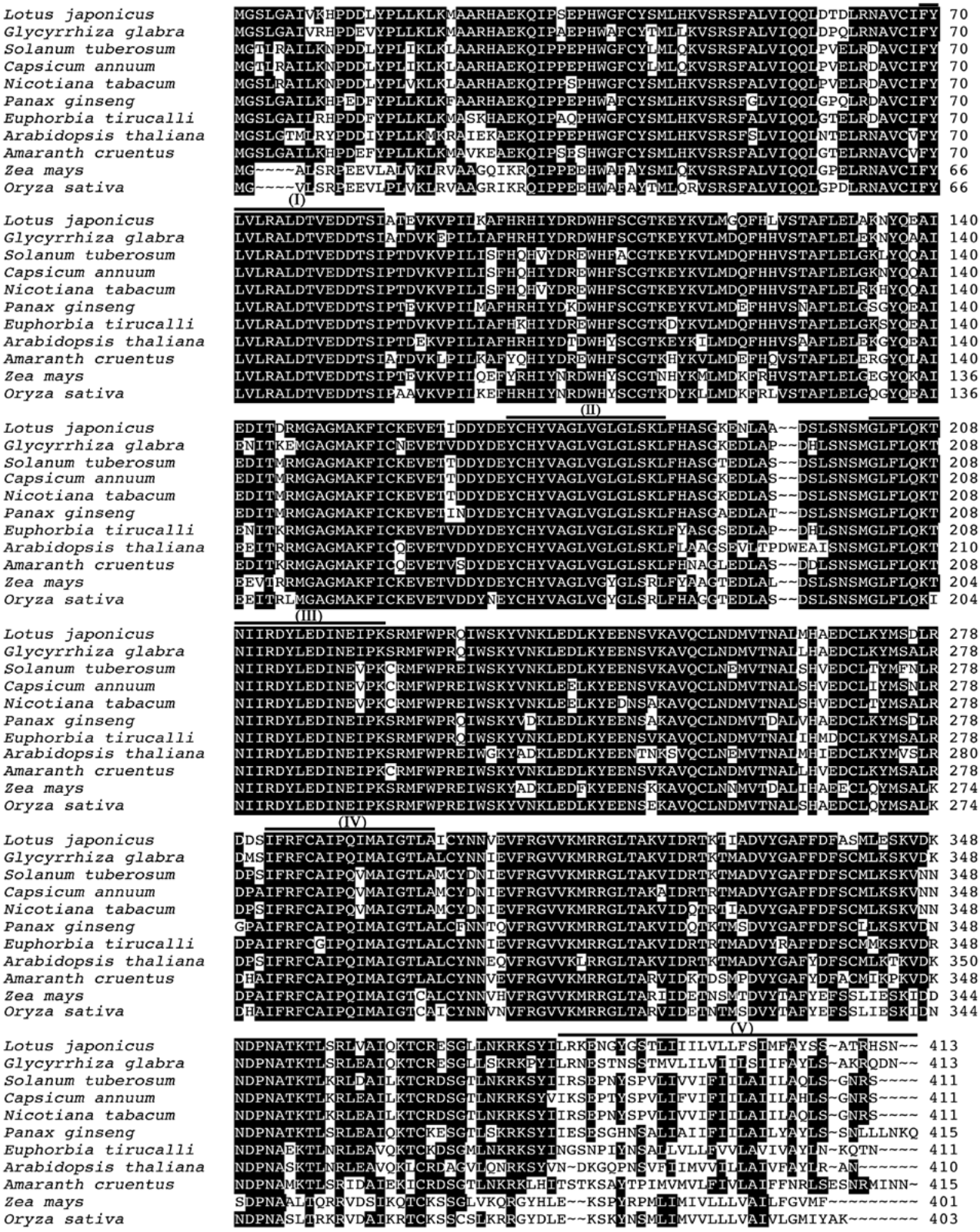

Fig. 2. Alignment of deduced amino acid sequences of SQS proteins. Deduced amino acid sequence of grain amaranth aligned with those of other species. Aligned sequences are from *A. cruentus* (AB691229), *A. thaliana* (NM_119630), *C. annuum* (AF124842), *E. tirucalli* (AB433916) *G. glabra* (D86409), *L. japonicas* (AB102688), *N. tabacum* (NTU60057) *O. sativa* (NM_001058160), *P. ginseng* (AB010148), and *Z. mays* (NM_001111369). Conserved domains are underlined. Identical residues are enclosed in black boxes.

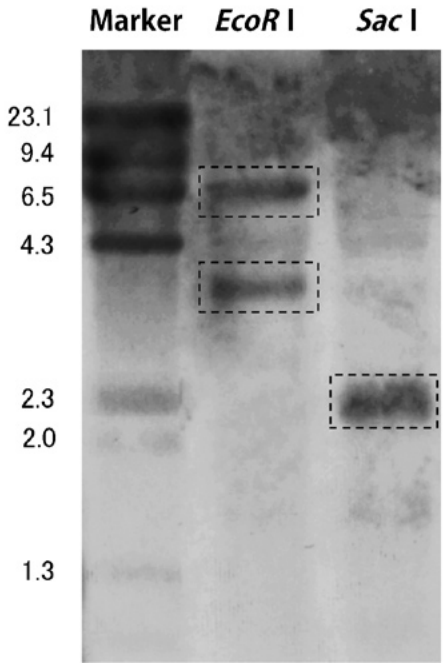

Fig. 3. Southern blot analysis of *A. cruentus* genomic DNA. Genomic DNA (15 µg) was digested with *Eco*RI and *Sac*I. Hybridization was carried out using ECL-labeled *ArSQSp* cDNA.

(Hayashi et al. 1999), *N. tabacum* (Devarenne et al. 2002), *A. thaliana* (Busquets et al. 2008), and *P. ginseng* (Kim et al. 2011). These observations were also reflected in the phylogenetic relationships inferred from the NJ tree constructed from the sequence alignment data (Fig. 5). We analyzed the relationships among the amino acid sequences of SQS proteins using PAUP*4.0 software. According to the phylogenetic tree, SQSs were clearly classified into two groups. The two mammalian SQSs formed one group, and were closely related to each other. The other group contained the plant SQSs, and was separated into two subgroups: monocots and dicots, suggesting that *SQS* genes evolved independently after the divergence of monocots and dicots (Hata et al. 1997). This result is also consistent with the overall genetic differences between dicots and monocots as reflected in the Angiosperm Phylogeny Group classification (APGIII 2009), which represents monocot/dicot divergence in the evolution of angiosperms. In these classes, amaranth was sister to the subgroup consisting of other dicots, and the amaranth SQS was most closely related to the *A. thaliana* SQS. This characteristic is very similar to that of the starch synthase family (*GBSSI, SSSI, SSSII*, and *SBE*) recently characterized in this species (Park & Nishikawa 2012). Differences in the sequences of genes between amaranth and other angiosperms are a general characteristic of this crop, reflecting differences in their genetic background.

Next, we investigated the tissue-specific expression of *SQS* in grain amaranth by qRT-PCR and RT-PCR. Previous studies showed that *SQS* transcripts and protein in higher plants were mainly localized in storage organs (Akamine et al. 2003, Busquets et al. 2008, Kim et al. 2011), but their localization in amaranth was unknown. We analyzed the expression of the *SQS* gene in amaranth seeds at various developmental stages. As shown in Fig. 4a, there were low levels of *SQS* transcripts at the initial stage of seed development, but the levels increased rapidly at the mid-late developmental stage before declining at the late developmental stage. These findings showed that the amaranth *SQS* is a late-expressed gene that is rapidly expressed at the mid-late stage of seed development.

We conducted a detailed analysis of the *SQS* expression pattern in different tissues of amaranth (Fig. 4b). *SQS* (or *SQS1*) activity and mRNA distribution in different tissues have been analyzed in several plants. According to previous observations, this gene is apparently widely expressed in many tissues, particularly in main roots. For example, in Korean ginseng (Kim et al. 2011) and *Arabidopsis* (Busquets et al. 2008), *SQS* mRNA is expressed preferentially in the leaves, petioles, and main roots. In the Fabaceae (Suzuki et al. 2002, Akamine et al. 2003) and rice (Hata 1997), a significant level of *SQS* mRNA was detected in the root tissues. However, in *E. tirucalli, SQS* transcripts accumulated in almost equal amounts in the stems and stalked leaves, while there were lower levels of transcripts in the roots (Uchida et al. 2009). In amaranth, there was a significant level of *SQS* transcripts in the roots, similar to that observed in other plants. The highest levels of *SQS* transcripts were in the stems, with lower levels in the leaves and petioles. Notably, amaranth *SQS* transcripts were not detected in leaf tissues during seedling development at the four- to six-leaf stages. This interesting pattern may indicate that amaranth *SQS* expression is not specific to leaf tissues. This hypothesis is supported by the data on squalene concentration in leaf lipid extracts. He & Corke (2003) reported that the squalene concentration in leaf lipid extracts averaged 0.26%, ranging from trace levels to 0.77%, which was much lower than the levels in seeds. We observed that the *SQS* mRNA levels in stems and roots increased rapidly during the four- to six-leaf stage of development (see Fig. 4b). Therefore, our results showed that the expression levels of *SQS* in stem and root tissues are significantly higher than those in leaf tissues.

In conclusion, this is the first report on the cloning and characterization of the *SQS* gene in grain amaranth. The nucleotide sequence information classified it as a *SQS*. Southern blot analysis revealed that there was a single copy of the *SQS* gene in the genome of *A. cruentus*. This gene showed late expression in the perisperm, and was expressed mainly in stem and root tissues. This study provides useful

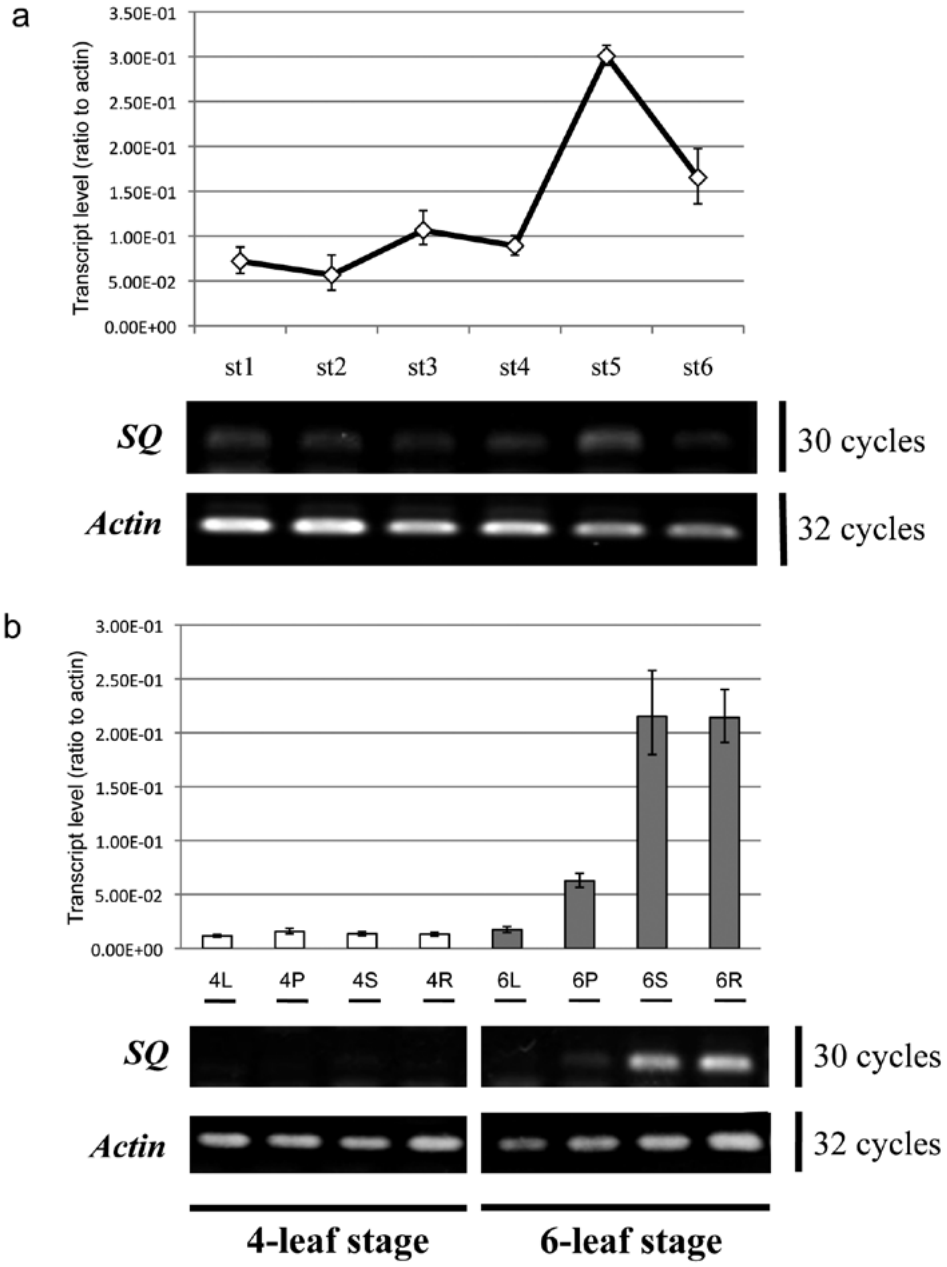

Fig. 4. Expression pattern of *SQS* gene in *A. cruentus*.

a) Expression pattern of *SQS* gene during seed development. Total RNAs from amaranth seeds were extracted at different developmental stages and expression pattern of *SQS* gene was analyzed by (a) qRT-PCR and (b) RT-PCR.

b) Expression pattern of *SQS* gene in various tissues of *A. cruentus*. Total RNAs from various tissues (L, leaf; P, petiole; S, stem; R, root) were extracted from plants at the four- and six-leaf stages. Expression pattern of *SQS* gene was analyzed by (a) qRT-PCR and (b) RT-PCR. *Actin* mRNA was used as a reference to normalize expressions in all samples.

information about the molecular characterization of the *SQS* clone isolated from grain amaranth. A basic understanding of these characteristics will contribute to further studies on the amaranth *SQS*.

References

[APG III] The Angiosperm Phylogeny Group (2009) An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Bot J Linn*

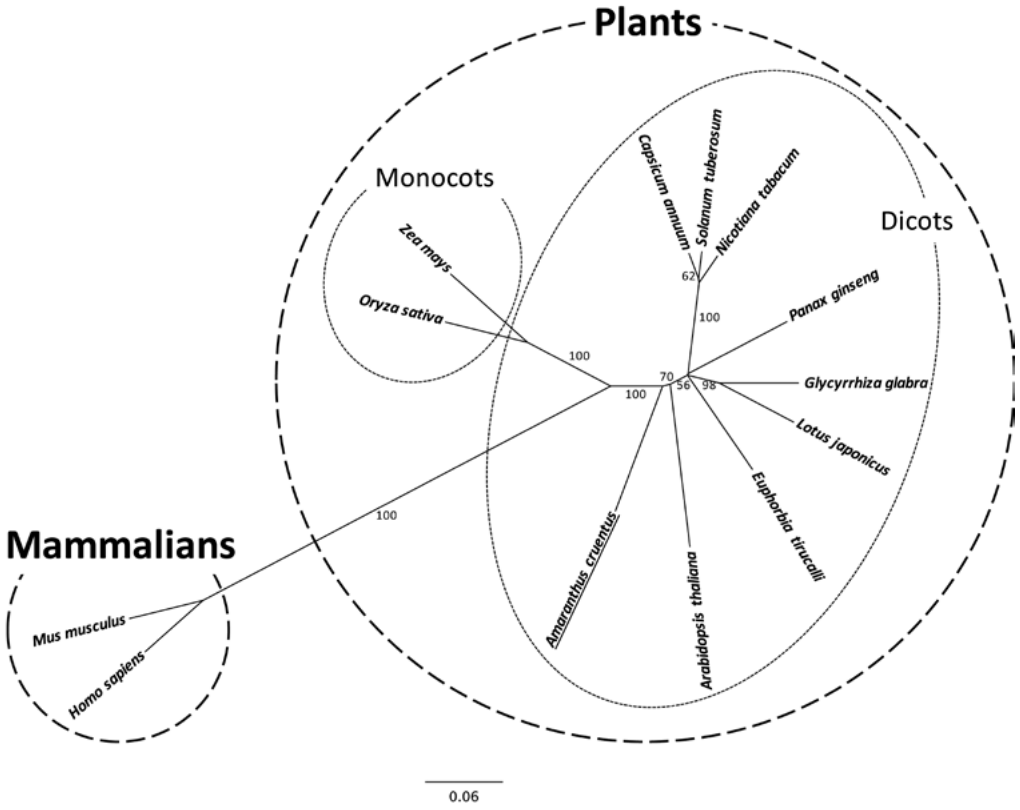

Fig. 5. Phylogenetic relationships of squalene synthases from various species. Tree was constructed using the neighbor-joining method. Amaranth squalene synthase is underlined. Scale bar represents evolutionary distance. Bootstrap support values from 1,000 replicates are shown at each node. GenBank IDs for squalene synthase genes are as follows: *A. cruentus*, AB691229; *A. thaliana*, NM_119630; *C. annuum*, AF124842; *E. tirucalli*, AB433916; *G. glabra*, D86409; *H. sapiens*. L06070; *L. japonicas*, AB102688; *M. musculus*, D29016; *N. tabacum*, NTU60057; *O. sativa*, NM_001058160; *P. ginseng*, AB010148; *S. tuberosum*, AB022599; *Z. mays*, NM_001111369.

Soc, **161**:105-121.

Akamine, S. et al. (2003) cDNA cloning, mRNA expression, and mutational analysis of the squalene synthase gene of Lotus japonicus. *Biochim Biophys Acta*, **1626**, 97-101.

Busquets, A. et al. (2008) Arabidopsis thaliana contains a single gene encoding squalene synthase. *Plant Mol Biol*, **67**, 25-36.

Devarenne, T. P. et al. (1998) Molecular characterization of tobacco squalene synthase and regulation in response to fungal elicitor. *Arch Biochem Biophys*, **349**, 205-215.

Hata, S. et al. (1997) cDNA cloning of squalene synthase genes from mono- and dicotyledonous plants, and expression of the gene in rice. *Plant Cell Physiol*, **38**, 1409-1413.

Hayashi, H. et al. (1999) Molecular cloning and characterization of two cDNAs for *Glycyrrhiza glabra* squalene synthase. *Biol Pharm Bull*, **22**, 947-950.

He, H. P. et al. (2002) Extraction and purification of squalene from *Amarnathus* grain. *J Agric Food Chem*, **50**, 368-372.

He, H. P. et al. (2003) Supercritical carbon dioxide extraction of oil and squalene from *Amaranthus* Grain. *J Agric Food Chem*, **51**, 7921-7925.

He, H. P. & Corke, H. (2003) Oil and squalene in amaranthus grain and leaf. *J Agric Food Chem*, **51**, 7913-7920.

Huang, Z. et al. (2007) Molecular cloning and characterization of the yew gene encoding squalene synthase from *Taxus cuspidate*. *J Biochem Mol Biol*, **40**, 625-635.

Jennings, S. M. et al. (1991) Molecular cloning and characterization of the yeast gene for squalene synthetase. P*roc Natl Acad Sci USA*, **88**, 6038-6042.

Karst, F. & Lacroute, F. (1977) Ergosterol biosynthesis in Saccharomyces cerevisiae: mutants deficient in the early steps of the pathway. *Mol Gene Genet*, **154**, 269-277.

Kim, T. D. et al. (2011) Expression and functional characterization of three squalene synthase genes associated with saponin biosynthesis in *Panax ginseng*. *Plant Cell Physiol*, **52**, 125-137.

Lee, S. & Poulter, C. D. (2008) Cloning, solubilization, and characterization of squalene synthase from *Thermosynechococcus elongatus* BP-1. *J Bacteriol*, **190**, 3808-3816.

McKenzie, T. L. et al. (1992) Molecular cloning, expression, and characterization of the cDNA for the rat hepatic squalene synthase. *J Biol Chem*, **267**, 21368-21374.

Murray, M. & Thompson, W. (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res*, **8**, 4321.

Nakashima, T. et al. (1995) Cloning, expression and characterization of cDNAs encoding *Arabidopsis thaliana* squalene synthase. *Proc Natl Acad Sci USA*, **92**, 2328-2332.

Park, Y. J. & Nishikawa, T. (2012) Characterization and expression analysis of the starch synthase gene family in grain amaranth (*Amaranthus cruentus* L.). *Genes Genet Syst*, **87**, 281-289.

Robinson, G. W. et al. (1993) Conservation between human and fungal squalene synthetases: Similarities in structure, function, and regulation. *Mol Cell Biol*, **113**, 2706-2717.

Sauer, J. D. (1967) The grain amaranths and their relatives: a revised taxonomic and geographic survey. *Ann Missouri Bot Gard*, **54**, 103-137.

Shechter, I. et al. (1992) Solubilization, purification, and characterization of a truncated form of rat hepatic squalene synthetase. *J Biol Chem*, **267**, 8628-8635.

Shukla, S. et al. (2006) Mineral profile and variability in vegetable amaranth (*Amaranthus tricolor*). *Plant Food Hum Nutr*, **61**, 23-28.

Sun, H. et al. (1997) Fractionation of Squalene from Amaranth Seed Oil. *J Assoc Offic Anal Chem*, **74**, 413-418.

Suzuki, H. et al. (2002) A genomic approach to the early stages of triterpene saponin biosynthesis in *Medicago truncatula. Plant J*, **32**, 1033-1048.

Swofford, D. (1988) *PAUP*: phylogenetic analysis using parsimony and other methods, version 4.0 (test ver. 61-64)*. Sinauer Associates Publishers, Sunderland.

Thompson, J. et al. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res*, **24**, 4876-4882.

Tozawa, R. et al. (1999) Embryonic lethality and defective neural tube closure in mice lacking squalene synthase. *J Biol Chem*, **274**, 30843-30848.

Uchida, H. et al. (2009) Cloning and characterization of a squalene synthase gene from a petroleum plant, *Euphorbia tirucalli* L. *Planta*, **229**, 1243-1252.

Westerman, D. et al. (2006) Extraction of Amaranth seed oil by supercritical carbon dioxide. *J Supercrit Fluids*, **37**, 38-52.