J Ind Microbiol Biotechnol (2010) 37:1171–1182
DOI 10.1007/s10295-010-0764-z

ORIGINAL PAPER

Cloning and characterization of squalene synthase gene from *Fusarium fujikuroi* (Saw.) Wr.

Rui-Yu Zhao · Wei Xiao · Hai-Li Cheng · Ping Zhu · Ke-Di Cheng

Received: 3 March 2010 / Accepted: 7 June 2010 / Published online: 29 June 2010

Abstract The gene encoding squalene synthase (GfSQS) was cloned from *Fusarium fujikuroi* (*Gibberella fujikuroi* MP-C) and characterized. The cloned genomic DNA is 3,267 bp in length, including the 5′-untranslated region (UTR), 3′-UTR, four exons, and three introns. A noncanonical splice-site (CA-GG, or GC-AG) was found at the first intron. The open reading frame of the gene is 1,389 bp in length, corresponding to a predicted polypeptide of 462 amino acid residues with a MW 53.4 kDa. The predicted GfSQS shares at least four conserved regions involved in the enzymatic activity with the SQSs of varied species. The recombinant protein was expressed in *E. coli* and detected by SDS–PAGE and western blot. GC–MS analysis showed that the wild-type GfSQS could catalyze the reaction from farnesyl diphosphate (FPP) to squalene, while the mutant mGfSQS (D82G) lost total activity, supporting the prediction that the aspartate-rich motif (DTXED) in the region I of SQS is essential for binding of the diphosphate substrate.

Keywords *Fusarium fujikuroi* · Squalene synthase · Recombinant protein · Conserved regions · GC–MS analysis

R.-Y. Zhao and W. Xiao contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s10295-010-0764-z) contains supplementary material, which is available to authorized users.

R.-Y. Zhao · W. Xiao · H.-L. Cheng · P. Zhu (✉) · K.-D. Cheng
Department of Biosynthesis of Natural Products, Institute of Materia Medica, Chinese Academy of Medical Sciences and Peking Union Medical College, 1 Xian Nong Tan Street, Beijing 100050, China
e-mail: zhuping@imm.ac.cn

R.-Y. Zhao · W. Xiao · H.-L. Cheng · P. Zhu · K.-D. Cheng
Key Laboratory of Biosynthesis of Natural Products, Ministry of Health of PRC, 1 Xian Nong Tan Street, Beijing 100050, China

R.-Y. Zhao · W. Xiao · H.-L. Cheng · P. Zhu · K.-D. Cheng
Key Laboratory of Bioactive Substances and Resources Utilization, Ministry of Education of PRC, 1 Xian Nong Tan Street, Beijing 100050, China

Introduction

Fusarium fujikuroi (*Gibberella fujikuroi* mating population C) is well known as a producer of carotenoids and large amounts of gibberellins, the industrial source for the various applications of gibberellins in agriculture and brewing. In this fungus, sterols and gibberellins (GAs) as well as carotenoids share all the early intermediates, up to farnesyl diphosphate (FPP) through the common mevalonate (MVA) pathway, which is catalyzed by acetoacetyl-CoA thiolase, 3-hydroxy-3-methylglutaryl-CoA (HMG-CoA) synthase, HMG-CoA reductase, MVA kinase, phosphomevalonate kinase, MVA-5-diphosphate decarboxylase, isopentenyl diphosphate (IPP) isomerase, geranyl diphosphate (GPP) synthase, and FPP synthase [17, 44] (Fig. 1), although their biosynthesis is physically separated from the beginning in different subcellular compartments [5]. From this point, the specific enzyme squalene synthase is used to produce squalene, the precursor of sterols and triterpenes, while the geranylgeranyl diphosphate (GGPP) synthases [encoded by gene *ggs1* and *gibA* (originally *ggs2*), respectively, of which *ggs1* encodes the enzyme responsible for GGPP synthesis for primary metabolism (biosynthesis of carotenoids, ubiquinone, non-cyclic diterpenes), and *gibA* encodes the GA pathway-specific GGPP synthase] are used to produce

GGPP, the common precursor of *ent*-kaurene (precursor of GAs) and phytoene (precursor of carotenoids) [2, 32]. Sterols are essential components of cell membranes and are synthesized from squalene. While GAs are dispensable secondary metabolites of the fungus and are derived from GGPP via the tetracyclic hydrocarbon *ent*-kaurene, a key intermediate that is formed by the two-step cyclization of GGPP via copalyl diphosphate (CPP). Gene disruption experiments demonstrated that the GA gene cluster of *F. fujikuroi* consists of at least seven genes, such as *orf3*, *P450-4*, *P450-1*, *P450-2*, *gibA* (originally *ggs2*), *gibB* (originally *cps/ks*), and *P450-3*, involved in GA biosynthesis [43]. The two branch pathways interfere with each other. Many genes, such as HMG-CoA synthase and HMG-CoA reductase, in the MVA pathway contain sterol-regulated element sequences mediating sterol-regulated transcription [8], which may influence the biosynthesis of the fungal gibberellins or carotenoids. On the other hand, downregulation of the squalene synthase in the sterol biosynthetic pathway will lead to the accumulation of FPP, which is redirected away from this pathway and toward the synthesis of other commercially important isoprenoids, as has been done in the engineered yeasts [27, 35]. Studies on squalene (precursor of sterols) and *ent*-kaurene (precursor of gibberellins) synthesis in *Fusarium fujikuroi* cell-free extracts showed that squalene appears to be synthesized 'by default' in various mycelial extracts, unless the original mycelia were engaged in gibberellin production, and in the latter case squalene was displaced by *ent*-kaurene as the main in vitro product of mevalonate [6]. Carotenoid biosynthesis branches out from GGPP through five different enzymatic reactions, in which the enzymes encoded by *carRA*, *carB*, *carX*, and *carT* are involved in the carotenoid biosynthesis [32].

Squalene synthase (SQS, EC 2.5.1.21) is a membrane-bound enzyme that catalyzes the first committed step for sterol and other triterpenoid biosynthesis, and is thought to play an important role in the regulation of isoprenoid biosynthesis in eukaryotes [11, 25]. It is a bifunctional enzyme that catalyzes the condensation of two molecules of FPP to form the presqualene diphosphate (PSPP) and then converts the PSPP to squalene in the presence of NADPH and Mg^{2+} (Fig. 1). In bacteria, these reactions constitute the first pathway-specific steps in the biosynthesis of hopanoids, which are pentacyclic triterpene lipids localized in bacterial membranes and exert many of the same stabilizing effects as membrane sterols in eukaryotes [19]. As a key enzyme in the regulation of isoprenoid biosynthesis, SQS encoding genes have been cloned from bacteria [19], yeasts [14, 22, 45], *Ganoderma lucidum* [46], protozoa and animals [13, 24, 33], human beings [31, 38], and plants [11, 12, 16, 25, 42]. But very little is known about the SQS gene in *F. fujikuroi*. In this report we describe the cloning, gene organization, heterologous expression, and functional analysis of the fungal SQS from *F. fujikuroi*.

Materials and methods

Isolation of the fungal DNA and RNA

The strain *Fusarium fujikuroi* N920 was cultured in MYG liquid medium (0.5% yeast extract, 0.5% maltose, 1% glucose, pH6.5) for 2 days at 26°C on a rotary shaker. The mycelia were filtered and ground into fine powder in liquid nitrogen. The genomic DNA was extracted by the genomic DNA isolation mini-prep method [28], and the total RNA was isolated by using Concert™ Plant RNA reagent (Invitrogen).

Genomic DNA cloning of *F. fujikuroi SQS* gene

To obtain the *SQS* gene from *F. fujikuroi*, the mRNA sequence of *Fusarium graminearum* (*Gibberella zeae* PH-1) hypothetical protein was retrieved from GenBank (accession number XM_389557), which shared 45–60% identities with those of the SQSs of *Saccharomyces cerevisae*, *Ganoderma lucidum*, etc. The following primers were designed and synthesized based on this mRNA sequence: forward primers: gf: 5′-ATGGGTTACCTTTACTACCTTCTACACC-3′, gf1: 5′-TCTGTCTAATGACGCCTCAGG-3′, gf2: 5′-GCTGGAGCACTTTGATGTTGTTA-3′; reverse primers: gr1: 5′-GCTCGGTGATAACAACATCAAAGTG-3′, gr2: 5′-ACTCGATAGCCTTTGTTTGTTGG-3′, gr3: 5′-AAGCTGGATGAGTGTTAGTTGAGTATG-3′, gr: 5′-CTACAATTCTTCGTGACCCGTAATCATG-3′. PCR amplification was performed in a 25-μl reaction volume containing 2.5 μl of 10× *Ex Taq*™ buffer (Mg^{2+} Plus), 2.5 mM dNTPs, 10 μM of each primer, 2.5 units of the enzyme (TaKaRa *Ex Taq*™), and 20 ng of the template genomic DNA of *F. fujikuroi* as described above. PCR conditions were as follows: initial denaturation at 95°C for 4 min, 30 cycles of 94°C for 30 s, 60°C for 30 s, and 72°C for 1 min, and a final extension at 72°C for 10 min. The PCR products were purified and cloned into pMD18-T vector (Takara, Japan) for sequencing.

To obtain the 5′ and 3′ flanking regions of the putative *SQS* gene, a genome walking approach was used with the Genome Walking Kit (Takara, Japan). The 5′-end region of the putative *SQS* gene was amplified by three rounds of thermal asymmetric interlaced PCR (TAIL PCR) with the template genomic DNA [21]. The specific primers were RSP1: 5′-GGCAGAGGTAGAAGAGAGTGATAGG-3′,

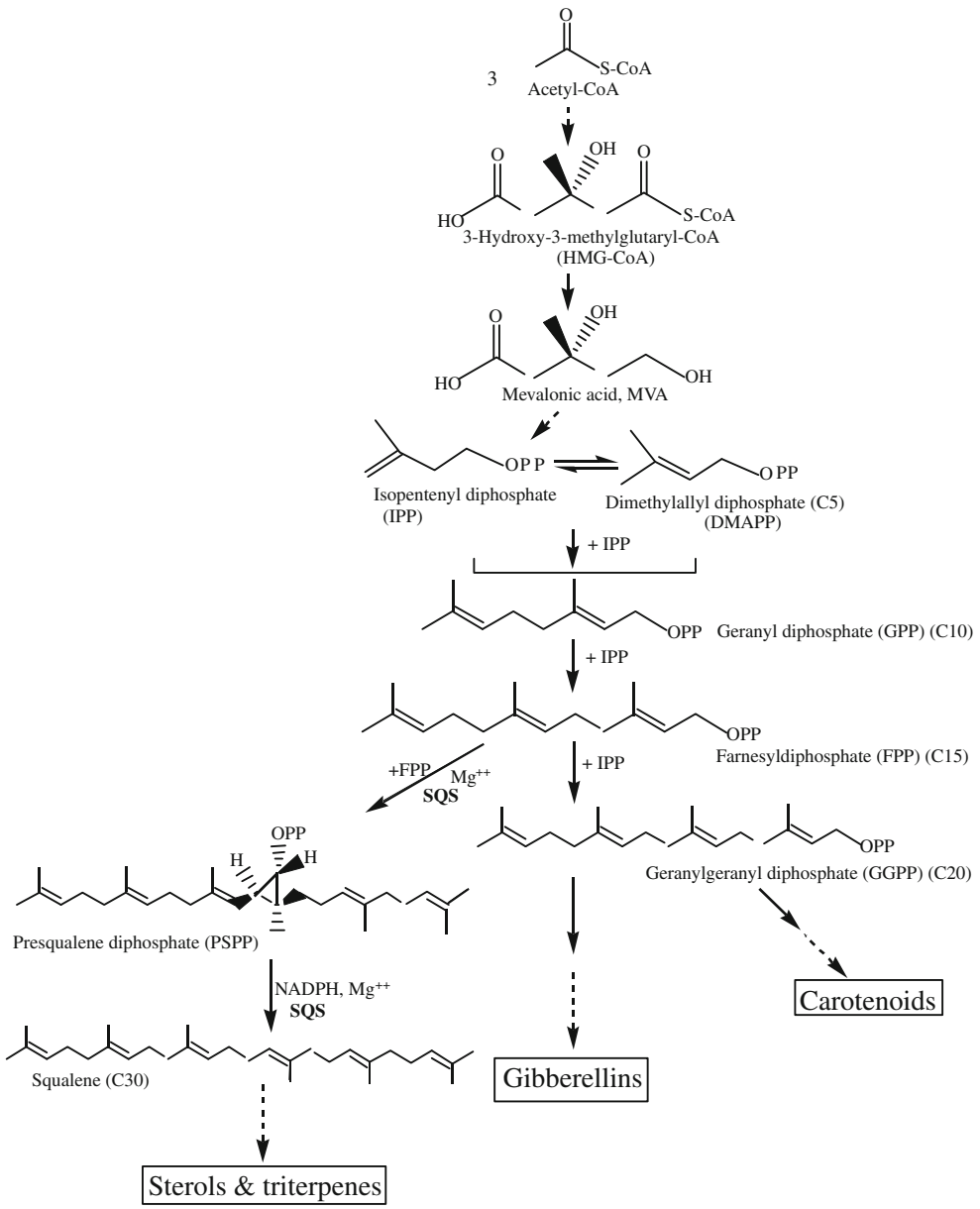

Fig. 1 Main steps in the biosynthesis of terpenoids in *Fusarium fujikuroi*. Squalene synthase (*SQS*) is a bifunctional enzyme that catalyzes the condensation of two molecules of farnesyl diphosphate (*FPP*) to form presqualene diphosphate (*PSPP*), and the subsequent reduction of PSPP to form squalene in the presence of NADPH [6, 25]

RSP2: 5′-CTCTGGTGACTCTGTGCTAGGATC-3′, and RSP3: 5′-CTTGCCAGTGAATGATCGATCGAAGC-3′. The amplification reactions were performed in a 50-μl reaction volume, with the specific primers mentioned above and the adaptor primers provided in the kit, and carried out according to the protocol. The same method was employed to amplify the 3′-end sequence of the gene, except that the following specific primers were used, FSP1: 5′-CAGAAGCTCTCTATGACTCAGGCC-3′, FSP2: 5′-GCATTTCTGGTCTCATGGTACGCC-3′, and FSP3: 5′-GAGCCAGATTCGACACCATCTTTAGG-3′. The PCR products were cloned into pMD18-T vector (Takara, Japan) and subjected to automated DNA sequencing.

Cloning of the cDNA of *F. fujikuroi SQS* gene

Single-stranded cDNA was synthesized from the total RNA by reverse transcription with ThermoScript™ RT-PCR system (Invitrogen). The gf and gr primers were used to amplify the cDNA, except that the *EcoR* I and *Not* I restriction sites were added to the 5′ end of each primer, respectively: gf-E: 5′-*GAATTC***ATG**GGTTACCTTT

ACTA CCTT CTACACC-3′ (*EcoR* I restriction site underlined and the translation start codon in bold); gr-N: 5′-*GCGGCCGC***CTA**CAATTCTTCGTGACCCGTAATCA TG-3′ (*Not* I restriction site underlined and the stop codon in bold). PCR conditions were the same as those described above. PCR products were purified and cloned into pMD18-T vector (Takara, Japan), and sequenced.

Bioinformatic analysis

The cloned sequences were first analyzed via BLASTn at the National Center of Biotechnology Information (NCBI) to aid selection of the most closely related reference sequences. Bioinformatic analysis of the genomic DNA sequence was performed by using Recognition of Regulatory Motifs with Statistics in the Softberry software (http://www.softberry.ru/berry.phtml) [34]. The promoter prediction was carried out using Neural Network Promoter Prediction (version 2.2) in the Berkeley Drosophila Genome Project (BDGP; http://www.fruitfly.org/seq_tools/promoter.html) [29], which has been applied to predict other promoters of both eukaryotes and prokaryotes [20, 23, 36].

Comparison of SQS protein sequences and phylogenetic analysis

SQS protein sequences of the representative species, such as *G. lucidum*, *S. cerevisiae*, *Arabiodopsis halianas*, and *Artemisia annua*, were obtained from published reports and aligned with MEGA 4.1.222.0 [40]. The aligned sequences were imported into PAUP* 4.0b10 [39], and were subjected to phylogenetic analysis. Maximum parsimony analysis (MP) was carried out with a heuristic search [39] under constraint of simple addition of sequences and tree bisection reconnection (TBR) branch swapping, with MaxTrees set to 100. All characters were unordered and equally weighted, with gaps treated as missing data. The confidence of the branches was measured by bootstrap analysis with 1,000 bootstrap replicates using heuristic search. Neighbor-joining analysis (NJ) was conducted with the uncorrected distance (p) model. Branch-swapping algorithm was TBR, steepest descent option not in effect, multrees option in effect. Support for internal nodes was estimated by 1,000 bootstrap replicates under the same model settings [39]. Only bootstrap values above 50% are shown, and only those above 70% were considered significant. Both the phylogenetic trees were rooted by using a bacterial SQS (*Staphylococcus aureus*, GenBank accession number YP_044561) as an outgroup.

Expression and preparation of the fungal recombinant SQS

A putative open reading frame (ORF) of the fungal *SQS* (*GfSQS*) cDNA was cut from the pMD18-T vector with *EcoR* I and *Not* I and subjected to the gel purification. The purified DNA fragment was subcloned into the *EcoR* I/*Not* I polycloning site of the pET-32a vector (Novagen, USA), and the recombinant plasmid pET-GfSQS was then transferred into *E. coli* BL21 (DE3).

E. coli BL21 (DE3) cells harboring the plasmid pET-GfSQS or pET-32a (mock transformant) were grown overnight at 37°C in Luria-Bertain (LB) medium containing ampicillin (100 μg ml^{-1}). A 500-μl aliquot of the overnight culture was added to 50 ml of fresh LB medium supplemented with 100 μg ml^{-1} ampicillin. When OD_{600} reached 0.8, cultures were induced by addition of isopropylthio-β-galactoside (IPTG) to a final concentration of 0.8 mM and grown for 17 h at 16°C. The bacterial cultures were collected with centrifugation at 10,000*g* for 5 min and washed with 50 mM phosphate buffer, pH 7.2. The pellet was resuspended in the phosphate buffer and sonicated for 5 min (10-s pulse each time, with 10-s intervals on ice). After centrifugation at 10,000*g* for 15 min, the supernatant was subjected to the SDS–PAGE, western blot, and the recombinant protein purification using HiTrap ChelatingHP (Amersham Biosciences, Sweden) as described by the manufacturer. To observe the expression of the recombinant protein by the SDS–PAGE, aliquots (1 ml) of the cultures were collected periodically (2-h intervals, from 1 to 17 h) during the cultivation process, and the supernatants of the sonicated cells were prepared as above. The amount of the soluble recombinant protein at different IPTG induction time was evaluated by BandScan software and BCA protein assay [37].

SDS–PAGE and western blot

Proteins were differentiated on 12% SDS–PAGE gel and transferred onto PVDF (polyvinylidene difluoride) for 2 h at 250 mA in 20 mM Tris, 150 mM glycine, and 20% (v/v) methanol. The membrane was incubated in blocking solution containing 5% (w/v) non-fat milk in TBST (Tris buffered saline with Tween 20) for 12 h at 4°C. The membrane was washed three times with TBST (5 min each time) and incubated with anti-His-Tag mouse monoclonal antibody in TBST for 1 h at room temperature. After washing with TBST, the membrane was incubated with horseradish peroxidase-conjugated goat anti-mouse IgG (H&L) secondary antibody and then detected with Amersham eECL kit and visualized with LAS-3000 (FuJiFilm, Japan).

Determination of GfSQS enzymatic activity

The SQS activity of the recombinant protein was measured on the basis of the conversion of farnesyl diphosphate (FPP) to squalene in the presence of NADPH and Mg^{2+}; 250 µl of the reaction mixture included 25 µM FPP (Sigma), 25 mM $MgCl_2$, 25 mM mercaptoethanol, 5 mM NADPH, and 125 µl of the enzyme solution. The reaction was carried out at 37°C for 5 h. After incubation, the mixture was extracted three times with 500 µl *n*-hexane. The *n*-hexane solution was concentrated overnight at room temperature until the total volume reached approx. 120 µl. The catalytic product squalene was detected by gas chromatography–mass spectrometry (GC–MS). Diluted authentic squalene (Sigma) was directly subjected to the GC–MS detection.

Squalene detection by GC–MS

Agilent GCMS-7890A/5975C (Agilent, USA) was fitted with an Agilent J&W HP-5 ms column (φ 0.25 mm × 30 m, 0.25-µm film thickness, P/N 19091S-433, USA) and a helium carrier (flow rate 1.2 ml min^{-1}), and operated at a scan range of *m*/*z* 20–550. The sample volume was 1 µl with a split ratio 10:1. The injector temperature was 250°C. The column temperature was maintained at 120°C for 3 min, elevated to 180°C at 15°C min^{-1}, and then to 260°C at 25°C min^{-1} for 25 min.

Results

Genomic DNA cloning of *F. fujikuroi SQS* gene

To obtain the fungal *SQS* gene, seven primers were designed based on the mRNA sequence of *F. graminearum* (*G. zeae* PH-1) hypothetical protein which shared certain similarities with the SQSs of other species. Twelve combinations of the primers were used to amplify the *SQS* gene with the genomic DNA of *F. fujikuroi* as the template. Only three combinations of primers (gf/gr, gf/gr1, and gf2/gr) yielded the specific DNA fragments of the anticipated sizes. The gf/gr and gf2/gr amplified DNA fragments were chosen for cloning and sequencing. Sequence analysis of the two DNA fragments revealed that the gf/gr amplified DNA fragment was 1,550 bp in length and harbored the whole length of the gf2/gr amplified DNA sequence. Based on the genome walking strategy, the putative fungal *SQS* (*GfSQS*) 5′-untranslated region (UTR) of 1,259 bp and 3′-UTR of 485 bp were determined (Fig. 2), respectively,

Fig. 2 The putative regulatory sequence of *SQS* gene in *F. fujikuroi*. 1,259 bp before initiation codon and 485 bp after termination codon (*underlined*) are presented. The transcription initiation site and termination site determined by RACE are shown in *larger font* and *boxed*. The putative transcriptional regulatory elements of SRE-1, Inv-SRE-1, Inv-Y-box, GC-box, Hap2/3/5, C/ERB, and GATA-1 are *labeled* and *underlined*. Five consensus elements in human *SQS* promoter homologous to these are presented

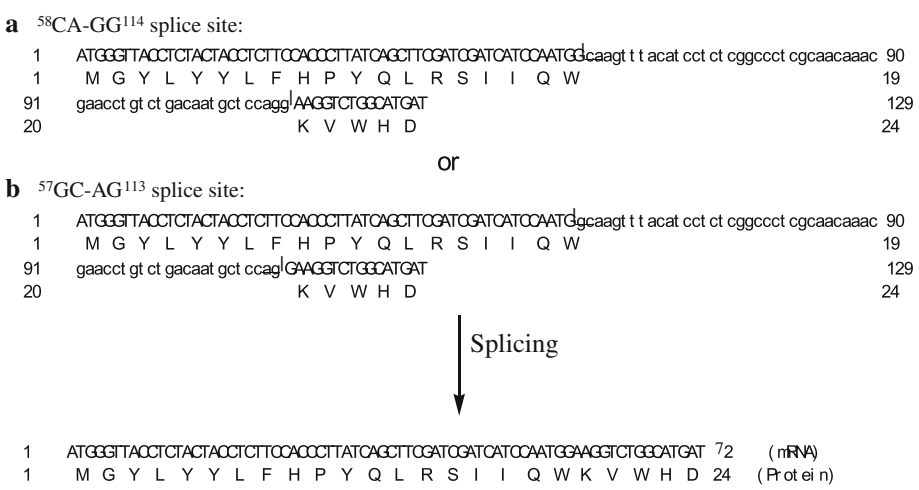

Fig. 3 Two putative splice sites of intron 1 in *GfSQS* gene. The exon and intron boundaries are shown by the *small vertical lines*; the first and last two nucleotides of the intron are *underlined* and the consensus sequence is *boxed*. **a** Showing the splice site at CA-GG (58–114); **b** showing the splice site at GC-AG (57–113). Both splicings result in the same mRNA sequence

and a total of 3,267 bp sequence (GenBank accession number EU275246) involving the putative open reading frame (ORF) of the GfSQS (462 amino acid residues) was obtained.

Bioinformatic analysis using Recognition of Regulatory Motifs with Statistics in the Softberry software revealed that both of the classic TATA box and the polyA site were not found, but a GC box (at position −933 to −928) and several other cis-elements were located (Fig. 2). Several types of consensus elements presumed to be related with the transcriptional regulation sites were documented, five of which were homologous to the consensus elements (HSS-SRE-1, Inv-SRE-3, Inv-Y-box, and SRE-1) of the human *SQS* promoter. These elements were reported to be responsible for sterol regulation [10]. Two putative CCAAT/enhancer binding protein (C/EBP) binding sites could be found in the 5′-UTR region. One is located at position −1,039 to −1,025 and the other at position −757 to −749. There was also an inverted Hap2/3/5 binding site with a sequence of CCAGT located at position −1,122 to −1,118, corresponding to the Hap2/3/4 binding site of the yeast *SQS* (*ERG9*) promoter, as both of the sites had a conserved CCAC(G)T motif [15]. A GATA-1 binding site was located at position −909 to −901. Only one sequence, which was located at position −954 to −905 with a score of 0.99, was found by using the NNPP v.2.2 for promoter prediction.

RT-PCR strategies were applied to isolate the putative ORF of *GfSQS* cDNA. The ORF is 1,389 bp in length and corresponded to a predicted polypeptide of 462 amino acid residues, with a predicted molecular weight of 53.4 kDa and p*I* of 5.17.

Comparison of the full *GfSQS* genomic DNA with the putative coding region of the *GfSQS* cDNA indicated a complex organization of four exons and three introns within the gene. It is noteworthy that the *GfSQS* genomic DNA contains a noncanonical intron (Fig. 3), with the splice site CA-GG (58–114 bp), or GC-AG (57–113 bp), instead of the canonical splice site (GT-AG).

Comparison of the putative GfSQS with other SQSs

The NCBI protein–protein BLAST showed that the deduced GfSQS amino acid sequence shared 92% identities with the hypothetical protein of *F. graminearum*. It also shared very broad and high local alignments and positives with other fungal species, such as *S. cerevisiae* (ACD03847.1), *Schizosacharomyces pombe* (NP_595363.1), *G. lucidum* (ABF57213.1), *Ustilago maydis* (CAA68054.1), and *Candida glabrata* (BAB12207.1), whereas GfSQS shared a relatively less identity to the SQSs of animals and plants.

Alignment of GfSQS with other five fungal SQSs displays at least four conserved regions (Fig. 4). All of these four regions are highly conserved even in plants and animals (data not shown). Among the different fungal species, the corresponding regions II, III, and IV are nearly identical. The region I is relatively less conservative, but it harbors a highly conserved aspartate-rich motif, whose relevance is discussed below. The site-directed mutagenesis of rat hepatic squalene synthase (RSS, EC 2.5.1.21) showed that the first Tyr residue (Tyr^{171}, corresponding to Tyr^{175} of GfSQS) in the region II is essential for the activity of RSS and is likely involved in the first reaction, and mutation of Tyr^{171} to the mutants Y171F, Y171S, and Y171 W completely abolished formation of PSPP or squalene from FPP [9]. This result also indicated that the phenyl ring is specifically required for activity and cannot be substituted with just either an aromatic or hydroxyl group [9]. The region III also contains an aspartate-rich motif (^{223}DIHED^{227} in GfSQS, which overlaps with the ^{219}DYLED^{223} sequence in human SQS, data not shown). In the RSS, the two Asp residues Asp^{219} and Asp^{223}

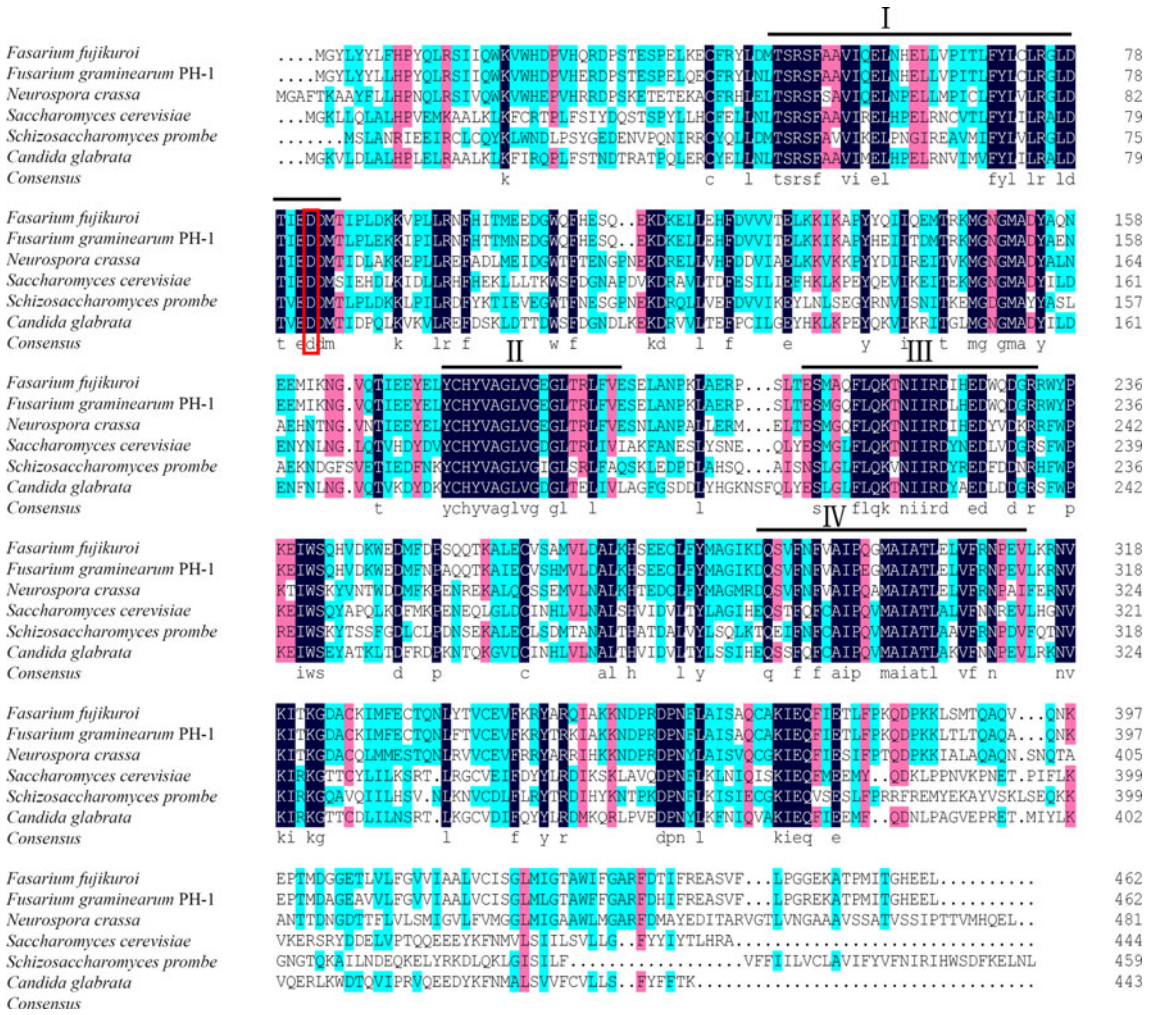

Fig. 4 Alignment of SQSs from *F. fujikuroi* and other five fungi. Sequences of other species are obtained from the GenBank (accession numbers: *F. graminearum* PH-1, XP_389557.1; *Neurospora crassa*, XP_959817.1; *Saccharomyces cerevisiae*, ACD03847.1; *Schizosaccharomyces pombe*, NP_595363.1; *Candida glabrata*, BAB12207.1). At least four highly similar regions are defined

(equivalent to Asp^{223} and Asp^{227} of GfSQS, respectively) are essential; not only does neutralization or reversal of charge (D219 N, D223 N, and D223 K) cause inactivation, so does the subtle modification of adding a single carbon atom to the side chain (D219E and D223E) [9]. The results are consistent with the participation of Asp^{219} and Asp^{223} in the binding of the diphosphates of two substrate FPP molecules via bridging Mg^{2+} ions [26]. The region IV is believed to be responsible for the catalysis of the second reaction and the likely NADPH binding motifs are the FC/VAIPQXMAIA/GTL (X = V/G/A) sequence found in this region and the VKIR/TKG sequence located downstream from the region [9, 19, 26]. The site-directed mutagenesis showed that the RSS mutant F288L (corresponding to Phe^{292} in the motif FVAIP of GfSQS) caused almost a complete loss of the second activity, but led to accumulation of significant amounts of PSPP even in the presence of NADPH [9]. The aspartate-rich motif (^{78}DTIED^{82} in GfSQS, relevant to ^{80}DTLED^{84} in human SQS, data not shown) in the region I was also predicted to participate in the binding of the substrates [26].

Amino acid residues in the C-terminal region exhibited a low level of sequence identity among all SQS proteins. This region is very hydrophobic in all SQS enzymes and may function as an anchor in the endoplasmic reticulum membrane. In accordance with this suggestion, Lee et al. [18] reported that the carboxyl-terminal deletion of residues 389–411 in the hot pepper SQS resulted in the accumulation of a functionally soluble SQS protein.

Phylogenetic analysis

A phylogenetic tree was constructed by using known SQS amino acid sequences from a wide range of different organisms, including plants, animals, protozoa, fungi, and bacteria, suggesting that the different forms of SQS were

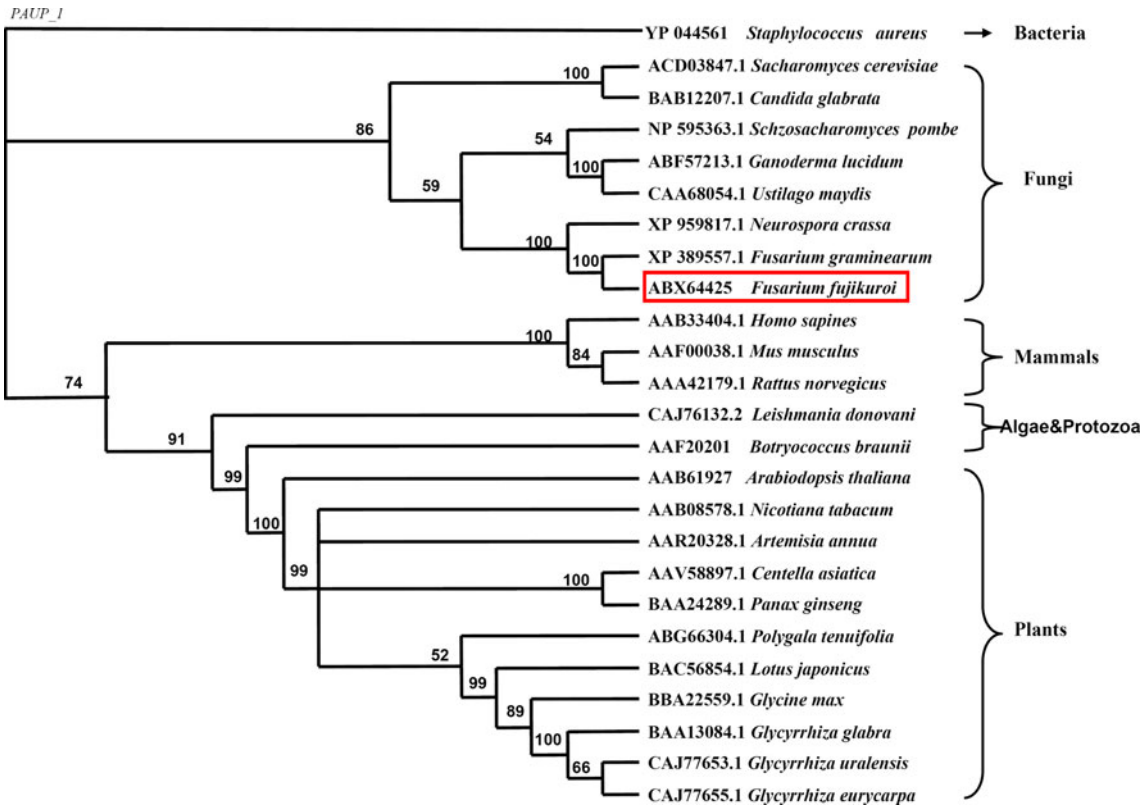

Fig. 5 Phylogenetic analysis of SQS amino acid sequences using the neighbor-joining (*NJ*) method. Twenty-four sequences from different species were retrieved from GenBank. The accession numbers are indicated in the front of each species. The numbers on each branch are the bootstrap values (shown >50%) obtained by NJ analysis of 1,000 resampled data sets. The SQS from a bacterial strain (*Staphylococcus aureus*, GenBank accession number YP_044561) was defined as an outgroup

evolved from a single ancestral gene. The fungal enzymes were clustered into one group, in which GfSQS was more closely related to the hypothetical protein of *F. graminearum* and that of *Neurospora crassa*, forming a distinct subgroup (Fig. 5).

Expression of *GfSQS* in *E. coli*

E. coli BL21 (DE3) cells harboring the plasmid pET-GfSQS or pET-32a (control) were cultured as described in the "Materials and methods". In the construction of the recombinant plasmid pET-GfSQS, a random clone of pET-GfSQS-31 was found to express the mutant GfSQS (D82G) with a spontaneous mutation (A → G) in coding 82 at position +302 in the GfSQS coding sequence, resulting in a substitution of aspartic acid (D) for glycine (G). Both the recombinant wild-type GfSQS protein from the transformant BL21(DE3)/pET-GfSQS-15 and the recombinant mutant GfSQS (mGfSQS) protein from the transformant BL21(DE3)/pET-GfSQS-31 were subjected to SDS–PAGE and western blot analysis. SDS–PAGE analysis showed that the same protein size (including the signal peptide encoded by the vector) was determined from the two transformants. This distinct band was not observed in the control at the corresponding position (Fig. 6a). Western blot analysis showed that both the GfSQS and the mGfSQS could specifically bind to anti-His-tag antibody (Fig. 6b). Each recombinant protein was purified to over 90% purity by nickel affinity chromatography (calculated by BandScan software) for the GC–MS analysis. The amount of the soluble recombinant wild-type GfSQS at different IPTG induction time is 70.2 mg l^{-1} (9 h), 90 mg l^{-1} (11 h), 118 mg l^{-1} (13 h), 92.7 mg l^{-1} (15 h), and 93.1 mg l^{-1} (17 h), respectively, calculated by BandScan software and BCA protein assay.

Determination of squalene by GC–MS

To confirm the *GfSQS* as a functional gene encoding squalene synthase, both the recombinant wild-type GfSQS and the mGfSQS were chosen for the GC–MS analysis, in which the mGfSQS harbored a D82G mutation. Figure 7 shows the GC–MS results. A comparison of the retention time and the corresponding full-scan mass spectra of the samples with those of authentic squalene (Sigma) confirms that the recombinant wild-type enzyme from BL21(DE3)/

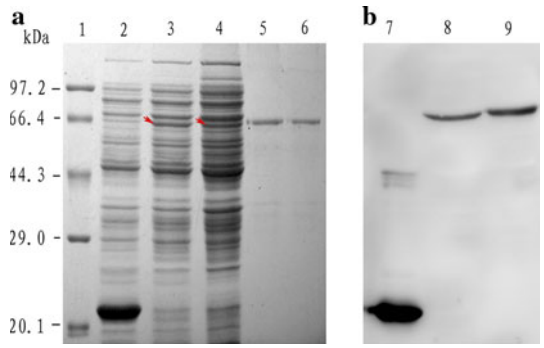

Fig. 6 Characterization of the recombinant wild-type GfSQS and the mGfSQS. **a** Silver-stained SDS–PAGE. **b** Western blot. *Lane 1* protein maker, *Lane 2* mock transformant BL21(DE3)/pET-32(a) (control), *Lane 3* transformant BL21(DE3)/pET-GfSQS-15, *Lane 4* transformant BL21(DE3)/pET-GfSQS-31, *Lane 7* mock transformant BL21(DE3)/pET-32(a) (exhibiting approx. 20 kDa signal peptide with 6× His-tag), *Lane 8* transformant BL21(DE3)/pET-GfSQS-15, *Lane 9* transformant BL21(DE3)/pET-GfSQS-31, *Lane 5* the purified recombinant wild-type GfSQS from BL21(DE3)/pET-GfSQS-15, *Lane 6* the purified recombinant mGfSQS from BL21(DE3)/pET-GfSQS-31. The *arrows* indicate the distinct protein band

pET-GfSQS-15 catalyzed a two-step cyclization reaction of FPP to squalene via PSPP. By contrast, no such product peak or the corresponding fragments were detected in the mock transformant. These results indicate that the wild-type *GfSQS* codes for squalene synthase of *F. fujikuroi*. In addition, the mGfSQS from BL21(DE3)/pET-GfSQS-31 did not show any SQS activity, demonstrating that the Asp^{82} in the aspartate-rich motif (^{78}DTIED^{82}) of the GfSQS region I is indispensable for the enzymatic activity.

Discussion

The plant pathogen *F. fujikuroi* is well known as a producer of gibberellin plant hormones. These gibberellins are supposed to be synthesized by the MVA pathway and share all the early intermediates with sterols, up to FPP. Many genes in the MVA pathway are regulated by sterols. Squalene is a key intermediate of sterols and is synthesized from FPP catalyzed by squalene synthase (SQS). SQS is commonly considered to be an incipient and crucial branch point enzyme and a potential regulatory point that controls carbon flux into the sterol biosynthesis. Downregulation of this enzyme will lead to the accumulation of FPP, which is redirected away from the sterol biosynthetic pathway, toward the synthesis of other commercially important isoprenoids.

Based on the mRNA sequence of a hypothetical protein of *F. graminearum* deposited in GenBank (accession number XM_389557), a gene encoding squalene synthase (GfSQS) was cloned from *F. fujikuroi* and characterized by different methods. Phylogenetic analysis showed that the GfSQS is clustered into the fungal SQS group and is more closely related to the hypothetical protein of *F. graminearum* and the SQS of *N. crassa*, which is in accordance with the classical taxonomy. GC–MS analysis confirmed that the wild-type protein (GfSQS) produced by the transformant BL21(DE3)/pET-GfSQS-15 could catalyze the formation of squalene from FPP.

Alignment of GfSQS with other fungal SQSs and comparison with different sources of SQSs showed that at least four highly conserved regions were present in the GfSQS. These consensus regions are predicted or even have been proven to be important for the SQS activity based on the kinetic studies with site-directed mutagenesis or the analysis of a crystal structure of human SQS [9, 26]. In human SQS, the crystal structure reveals five helices surrounding a central active cavity, of which one end is predominantly hydrophobic, and the other end is more hydrophilic and contains two conserved aspartate-rich motifs (^{80}DTLED^{84} and ^{219}DYLED^{223}, corresponding to ^{78}DTIED^{82} and ^{223}DIHED^{227}of GfSQS and present in the region I and the region III, respectively), which are predicted to bind the diphosphates of two substrate FPP molecules via bridging Mg^{2+} ions [1, 26]. One of these motifs, ^{219}DYLED^{223}, has been demonstrated to be essential for binding of the diphosphate units in FPP [9]. Our results again supported the prediction that the other aspartate-rich motif in the region I of SQS is also indispensable for the similar function, as the D82G mutation in this motif resulted in total loss of GfSQS activity.

Eukaryotic protein-coding genes are frequently interrupted by multiple introns which are removed at the donor and acceptor splice sites so that the adjacent exons are spliced. It has been reported that among the 22 most commonly found canonical (GT-AG) and noncanonical splice sites of human genes, the top four most represented donor–acceptor pairs (GT-AG, GC-AG, AT-AC, and GT-GG) accounted for 99.16% [4]. The vast majority of eukaryotic introns have the canonical splice site. Among the noncanonical introns over 90% of them have the GC-AG splice site [3]. Similar situations were found in *N. crassa* and *F. graminearum*. Apart from the majority of GT-AG introns, a frequency of 1.2 or 1.0% of GC-AG intron was identified in *N. crassa or F. graminearum*, respectively [30]. Thanaraj and Clark [41] reported that in humans, 5% of alternatively spliced introns are GC-AG introns and 60% of these GC-AG introns are alternatively spliced. In *Caenorhabditis elegans* the majority of GC-AG introns appear to be constitutively spliced and have no evolutionary constraints to prevent them from being GT-AG introns, while a subset of GC-AG introns is involved in alternative splicing and the C at the +2 position of these introns can have an important role in splicing regulation [7]. In this work we also discovered a noncanonical intron present in the *GfSQS* gene.

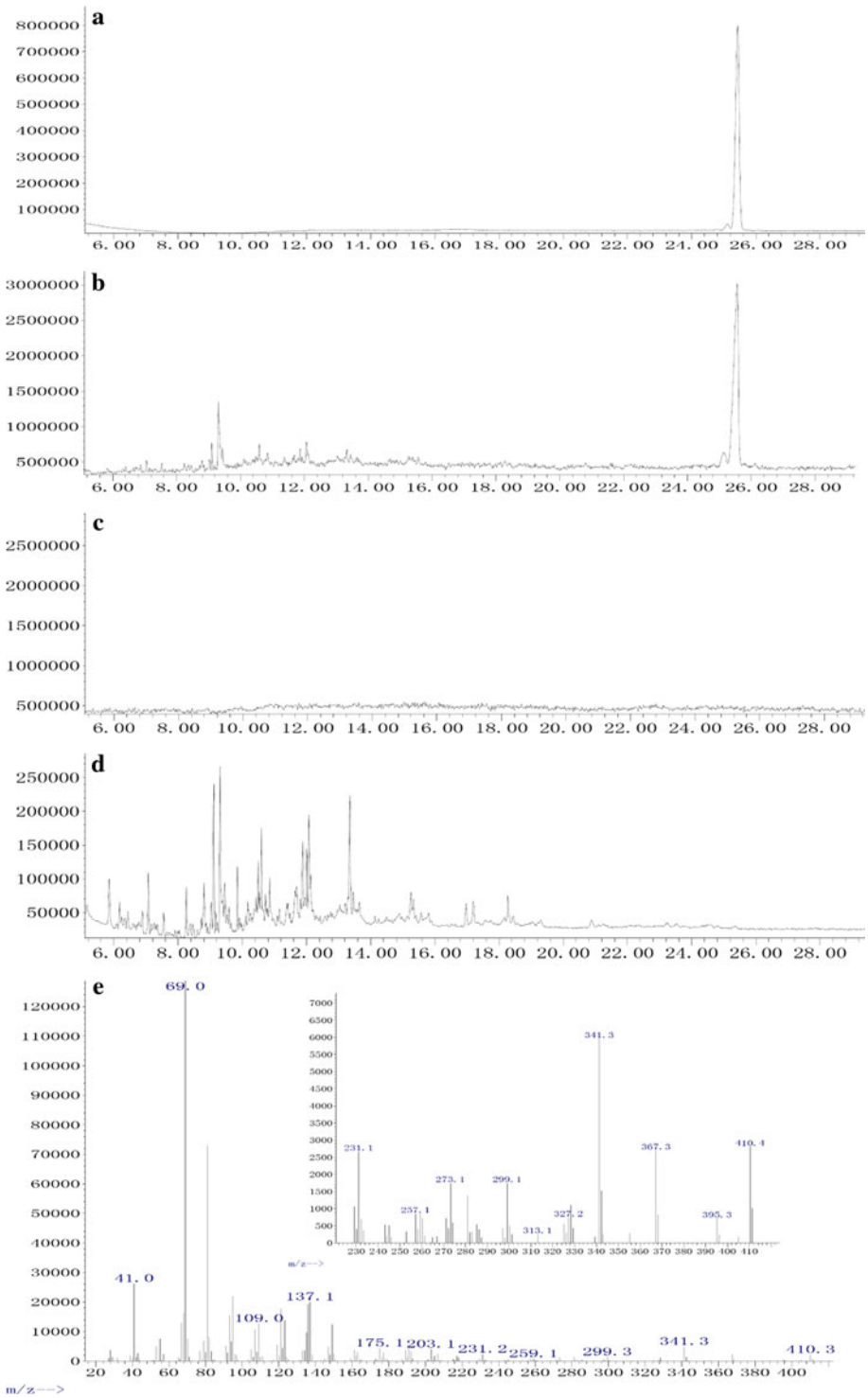

Fig. 7 GC-MS detection of the catalyzed product of the recombinant enzyme encoded by *GfSQS*. Total ion chromatograms (*TIC*) of the samples analyzed (**a** authentic squalene; **b** the sample extracted from an in vitro reaction mixture containing the purified recombinant wild-type GfSQS; **c** the sample extracted from an in vitro reaction mixture containing the purified recombinant mGfSQS; **d** the sample extracted from an in vitro reaction mixture containing the total proteins of the mock transformant). **e** Mass spectrum of peak at approx. 25.5 min in TIC of authentic squalene. The other MS data please see the "Electronic supplementary material"

Except the theoretical noncanonical splicing sites CA-AG (58–113) and GC-GG (57–114) present in the gene, which result in the truncated proteins, there are two possible noncanonical splicings to remove the intron from *GfSQS* gene: one is through the CA-GG splice site (58–114), and the other is through the GC-AG splice site (57–113), both of which produce the same mRNA sequence (Fig. 3). In view of the fact that the CA-GG splice site is rarely found in the eukaryotic genes [4], moreover, the G|GC AAGT consensus sequence (the exon and intron boundary is shown by the vertical line and the first two nucleotides of the intron are underlined) [30] is present at the donor splice signal, this intron is most likely to be removed from the pre-mRNA through the GC-AG splicing.

In conclusion, this is the first report on the gene cloning and characterization of squalene synthase of *F. fujikuroi*. Recombinant wild-type and mutant proteins were purified and subjected to the catalytic activity analysis by GC–MS. The wild-type GfSQS has normal activity of converting two molecules of FPP to squalene, while the mutant mGfSQS (D82G) loses total activity, supporting the prediction that the first aspartate-rich motif in the region I of SQS is essential for binding of the substrate FPP. In addition, a noncanonical intron was found in the *GfSQS* gene. Our results facilitate the investigation of the effects of regulation on isoprenoid metabolism in the fungus *F. fujikuroi* by up- or downregulation of *GfSQS* expression.

Acknowledgments We thank Professor Chen Ma and Mr. Tong Li at the National Research Center for Analysis of Drugs and Metabolites for the GC–MS analysis. This work was supported by National High Technology Research and Development Program ("863" Program) of China (No.2001AA234021) and Beijing Municipal Natural Science Foundation (No.5992012).

References

1. Ashby MN, Edwards PA (1990) Elucidation of the deficiency in two yeast coenzyme Q mutants. Characterization of the structural gene encoding hexaprenyl pyrophosphate synthetase. J Biol Chem 265:13157–13164
2. Bömke C, Rojas MC, Gong F, Hedden P, Tudzynski B (2008) Isolation and characterization of the gibberellin biosynthetic gene cluster in *Sphaceloma manihoticola*. Appl Environ Microbiol 74:5325–5339
3. Burset M, Seledtsov A, Solovyev VV (2000) Analysis of canonical and noncanonical splice sites in mammalian genomes. Nucleic Acids Res 28:4364–4375
4. Chong A, Zhang G, Bajic VB (2004) Information for the coordinates of exons (ICE): a human splice sites database. Genomics 84:762–766
5. Domenech CE, Giordano W, Avalos J, Cerdá -Olmedo E (1996) Separate compartments for the production of sterols, carotenoids and gibberellins in *Gibberella fujikuroi*. Eur J Biochem 239:720–725
6. Fernández-Martín R, Domenech C, Cerdá-Olmedo E, Avalos J (2000) Ent-kaurene and squalene synthesis in *Fusarium fujikuroi* cell-free extracts. Phytochemistry 54:723–728
7. Farrer T, Roller AB, Kent WJ, Zahler AM (2002) Analysis of the role of *Caenorhabditis elegans* GC-AG introns in regulated splicing. Nucleic Acids Res 30:3360–3367
8. Goldstein JL, Brown MS (1990) Regulation of the mevalonate pathway. Nature 343:425–430
9. Gu P, Ishii Y, Spencer TA, Shechter I (1998) Function-structure studies and identification of three enzyme domains involved in the catalytic activity in rat hepatic squalene synthase. J Biol Chem 273:12515–12525
10. Guan G, Dai PH, Osborne TF, Kim JB, Shechter I (1997) Multiple sequence elements are involved in the transcriptional regulation of the human squalene synthase gene. J Biol Chem 272:10295–10302
11. Hanley KM, Nicolas O, Donaldson TB, Smith-Monroy C, Robinson GW, Hellmann GM (1996) Molecular cloning, in vitro expression and characterization of a plant squalene synthetase cDNA. Plant Mol Biol 30:1139–1151
12. Huang Z, Jiang K, Pi Y, Hou R, Liao Z, Cao Y, Han X, Wang Q, Sun X, Tang K (2007) Molecular cloning and characterization of the yew gene encoding squalene synthase from *Taxus cuspidata*. J Biochem Mol Biol 40:625–635
13. Inoue T, Osumi T, Hata S (1995) Molecular cloning and functional expression of a cDNA for mouse squalene synthase. Biochim Biophys Acta 1260:49–54
14. Jennings SM, Tsay YH, Fisch TM, Robinson GW (1991) Molecular cloning and characterization of the yeast gene for squalene synthetase. Proc Natl Acad Sci U S A 88:6038–6042
15. Kennedy MA, Barbuch R, Bard M (1999) Transcriptional regulation of the squalene synthase gene (ERG9) in the yeast *Saccharomyces cerevisiae*. Biochim Biophys Acta 1445:110–122
16. Kim OT, Seong NS, Kim MY, Hwang B (2005) Isolation and characterization of squalene synthase cDNA from *Centella asiatica* (L.) urban. J Plant Biol 48:263–269
17. Lange BM, Rujan T, Martin W, Croteau R (2000) Isoprenoid biosynthesis: the evolution of two ancient and distinct pathways across genomes. Proc Natl Acad Sci U S A 97:13172–13177
18. Lee JH, Yoon YH, Kim HY, Shin DH, Kim DU, Lee IJ, Kim KU (2002) Cloning and expression of squalene synthase cDNA from hot pepper (*Capsicum annuum* L.). Mol Cells 13:436–443
19. Lee S, Poulter CD (2008) Cloning, solubilization, and characterization of squalene synthase from *Thermosynechococcus elongatus* BP-1. J Bacteriol 190:3808–3816
20. Lewin A, Mayer M, Chusainow J, Jacob D, Appel B (2005) Viral promoters can initiate expression of toxin genes introduced into *Escherichia coli*. BMC Biotechnol 5:19
21. Liu YG, Mitsukawa N, Oosumi T, Whittier RF (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J 8:457–463
22. LoGrasso PV, Soltis DA, Boettcher BR (1993) Overexpression, purification, and kinetic characterization of a carboxyl-terminal-truncated yeast squalene synthetase. Arch Biochem Biophys 307:193–199
23. Lu JP, Duan ZB, Liu TB, Lin FC (2007) Cloning, sequencing and expression analysis of the *NAR* promoter activated during hyphal stage of *Magnaporthe grisea*. J Zhejiang Univ Sci B 8:661–665
24. McKenzie TL, Jiang G, Straubhaar JR, Conrad DG, Shechter I (1992) Molecular cloning, expression and characterization of the cDNA for the rat hepatic squalene synthase. J Biol Chem 267:21368–21374
25. Nakashima T, Inoue T, Oka A, Nishino T, Osumi T, Hata S (1995) Cloning, expression, and characterization of cDNAs encoding *Arabidopsis thaliana* squalene synthase. Proc Natl Acad Sci U S A 92:2328–2332
26. Pandit J, Danley DE, Schulte GK, Mazzalupo S, Pauly TA, Hayward CM, Hamanaka ES, Thompson JF, Harwood HJ Jr (2000) Crystal structure of human squalene synthase. A key enzyme in cholesterol biosynthesis. J Biol Chem 275:30610–30617
27. Paradise EM, Kirby J, Chan R, Keasling JD (2008) Redirection of flux through the FPP branch-point in *Saccharomyces cerevisiae* by down-regulation squalene synthase. Biotechnol Bioeng 100:371–378
28. Raeder U, Broda P (1985) Rapid preparation of DNA from filamentous fungi. Lett Appl Microbiol 1:17–20
29. Reese MG, Eeckman FH (1995) Novel neural network algorithms for improved eukaryotic promoter site recognition. The Seventh International Genome Sequencing and Analysis Conference, Hilton Head Island
30. Rep M, Duyvesteijn RG, Gale L, Usgaard T, Cornelissen BJ, Ma LJ, Ward TJ (2006) The presence of GC-AG introns in

Neurospora crassa and other euascomycetes determined from analyses of complete genomes: implications for automated gene prediction. Genomics 87:338–347
31. Robinson GW, Tsay YH, Kienzle BK, Smith-Monroy CA, Bishop RW (1993) Conservation between human and fungal squalene synthetases: similarities in structure, function, and regulation. Mol Cell Biol 13:2706–2717
32. Rodríguez-Ortiz R, Limón MC, Avalos J (2009) Regulation of carotenogenesis and secondary metabolism by nitrogen in wild-type *Fusarium fujikuroi* and carotenoid-overproducing mutants. Appl Environ Microbiol 75:405–413
33. Sealey-Cardona M, Cammerer S, Jones S, Ruiz-Pérez LM, Brun R, Gilbert IH, Urbina JA, González-Pacanowska D (2007) Kinetic characterization of squalene synthase from *Trypanosoma cruzi*: selective inhibition by quinuclidine derivatives. Antimicrob Agents Chemother 51:2123–2129
34. Shahmuradov KA, Kolchanov NA, Solov'ev VV, Ratner VA (1986) Enhancer-like structures in middle repetitive sequences of the eukaryotic genomes. Genetika 22:357–367
35. Shimada H, Kondo K, Fraser PD, Miura Y, Saito T, Misawa N (1998) Increased carotenoid production by the food yeast *Candida utilis* through metabolic engineering of the isoprenoid pathway. Appl Environ Microbiol 64:2676–2680
36. Skoblov M, Shakhbazov K, Oshchepkov D, Ivanov D, Guskova A, Ivanov D, Rubtsov P, Prasolov V, Yankovsky N, Baranova A (2006) Human RFP2 gene promoter: unique structure and unusual strength. Biochem Biophys Res Commun 342:859–866
37. Smith PK, Krohn RI, Hermanson GT, Mallia AK, Gartner FH, Provenzano MD, Fujimoto EK, Goeke NM, Olson BJ, Klenk DC (1985) Measurement of protein using bicinchoninic acid. Anal Biochem 150:76–85
38. Summers C, Karst F, Charles AD (1993) Cloning, expression and characterisation of the cDNA encoding human hepatic squalene synthase, and its relationship to phytoene synthase. Gene 136:185–192
39. Swofford DL (2004) *PAUP**. Phylogenetic analysis using parsimony (and other methods) version 4.0b 10. Sinauer, Sunderland
40. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599
41. Thanaraj TA, Clark F (2001) Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. Nucleic Acids Res 29:2581–2593
42. Uchida H, Yamashita H, Kajikawa M, Ohyama K, Nakayachi O, Sugiyama R, Yamato KT, Muranaka T, Fukuzawa H, Takemura M, Ohyama K (2009) Cloning and characterization of a squalene synthase gene from a petroleum plant, *Euphorbia tirucalli* L. Planta 229:1243–1252
43. Voss T, Schulte J, Tudzynski B (2001) A new MFS-transporter gene next to the gibberellin biosynthesis gene cluster of *Gibberella fujikuroi* is not involved in gibberellin secretion. Curr Genet 39:377–383
44. Withers ST, Keasling JD (2007) Biosynthesis and engineering of isoprenoid small molecules. Appl Microbiol Biotechnol 73:980–990
45. Zhang D, Jennings SM, Robinson GW, Poulter CD (1993) Yeast squalene synthase: expression, purification, and characterization of soluble recombinant enzyme. Arch Biochem Biophys 304:133–143
46. Zhao MW, Liang WQ, Zhang DB, Wang N, Wang CG, Pan YJ (2007) Cloning and characterization of squalene synthase (SQS) gene from *Ganoderma lucidum*. J Microbiol Biotechnol 17:1106–1112