

## Recombinant Squalene Synthase. A Mechanism for the Rearrangement of Presqualene Diphosphate to Squalene

Brian S. J. Blagg, Michael B. Jarsfer, Daniel H. Rogers, and C. Dale Poulter

Department of Chemistry  
University of Utah  
315 South 1400 East, RM 2020  
Salt Lake City, UT 84112

### Supplementary Information

#### Methods for Sequence Comparisons

The deduced amino acid sequences for squalene, phytoene and farnesyl diphosphate synthases were retrieved from the databases available through the National Center for Biochemical Information (NCBI). Table 5.1 lists the sources of the sequences used in this analysis.

Sequences were compared pairwise by the program Bestfit from GCG, and alignments were obtained on entire translated peptide sequences using PILEUP from GCG. The PHYLIP program (version 3.5) was used to infer phylogenetic relationships.<sup>1</sup> For determining the phylogenetic relationships of peptide sequences, sequence alignments were edited to remove gaps and to remove sections of sequence characterized by low homology. These manipulations increased the efficacy of analysis for the determination of phylogenetic relationships and decreased computation time. Phylogenetic relationships were obtained using the protein parsimony method using a bootstrapped data set containing 100 replicates. Consensus trees were assigned by majority rules consensus.

Secondary structure predictions were obtained on multiple sequence alignments of SQSase or PHYSase using the DSC<sup>2</sup> program from the Imperial Cancer Research Fund accessed at [http://bonsai.lif.icnet.uk/bmm/dsc/dsc\\_read\\_align.html](http://bonsai.lif.icnet.uk/bmm/dsc/dsc_read_align.html) and the PHD<sup>3-5</sup> program from EMBL accessed at <http://www.embl-heidelberg.de/predictprotein>. Secondary structure predictions were obtained on single amino acid sequences using the GOR<sup>6</sup> and CF<sup>7</sup> algorithms accessed through GCG and the NN-PREDICT<sup>8</sup> program accessed at <http://cmparm.ucsf.edu/~nomi/nnpredict.html>. Predictions were analyzed manually and were used to assign a consensus secondary structure.

1. Felsenstein, J. Phylogeny Inference Package (PHYLIP version 3.5).
2. King, R.D.; Sternberg, M.J.E. Identification and Application of the Concepts Important for Accurate and Reliable Protein Secondary Structure Prediction. *Protein Sci.* **1996**, *5*, 2298-2310.

3. Rost, B.; Sander, C. Prediction of Protein Structure at Better than 70% Accuracy. *J. Mol. Biol.* **1993**, *232*, 584-599.
4. Rost, B.; Sander, C.; Schneider, R. PHD – An Automatic Mail Server for Protein Secondary Structure Prediction. *CABIOS*, **1994**, *10*, 53-60.
5. Rost, B.; Sander, C. Combining Evolutionary Information and Neural Networks to Predict Protein Secondary Structure. *Prot. Sci.* **1994**, *19*, 55-72.
6. Garnier, J.; Osguthorpe, D.J.; Robson, B. Analysis of the Accuracy and Implications of Simple Methods for Predicting the Secondary Structure of Globular Proteins. *J. Mol. Biol.* **1978**, *120*, 97-120.
7. Chou, P.Y.; Fasman, G.D. Prediction of Protein Conformation. *Biochemistry* **1974**, *13*, 222-224.
8. Kneller, D.G.; Cohen, F.E.; Langridge, R. Improvements in Protein Secondary Structure Prediction by an Enhanced Neural Network. *Methods Enzymol.* **1990**, *214*, 171-182.

Supplementary Table1. A List of Proteins Used in Sequence Analysis

Protein <sup>†</sup>	Source	Description on source	Accession no.
SQS_HSAP <sup>a</sup>	Homo sapien	Human	X69141
SQS_RNOR <sup>a</sup>	Rattus norvegicus	Rat	Q02769
SQS_MMUS <sup>a</sup>	Mus musculus	Mouse	D29016
SQS_ATHA <sup>a</sup>	Arabidopsis thaliana	Mouse-ear cress	P53799
SQS_ATHB <sup>a</sup>	Arabidopsis thaliana		AF0043916
SQS_GGLA <sup>a</sup>	Glycyrrhiza glabra	Licorice source	G2144186
SQS_GGLB <sup>a</sup>	Glycyrrhiza glabra	Licorice source	G2144187
SQS_ZMAY <sup>a</sup>	Zea mays	Maize	G2463567
SQS_OSAT <sup>a</sup>	Oryza sativa	Rice	G2463565
SQS_PGIN <sup>a</sup>	Panax ginseng	Korean ginseng plant	G2780361
SQS_GMAX <sup>a</sup>	Glycine max	Soybean	AB007503
SQS_NTAB <sup>a</sup>	Nicotinium tabacum	Tobacco strain	U60057
SQS_NBEN <sup>a</sup>	Nicotinium benthamiana	Tobacco strain	P53800
SQS_SCER <sup>a</sup>	Saccharomyces cerevisiae	Baker's yeast	P29704
SQS_UMAY <sup>a</sup>	Ustilago maydis	Infectious yeast	Q92459
SQS_CALB <sup>a</sup>	Candida albicans	Infectious yeast	P78589
SQS_SPOM <sup>a</sup>	Schizosaccharomyces pombe	Fusion yeast	P36596
SQS_LMAJ <sup>c</sup>	Leishmania major	Pathogenic Trypanosome	U30455
SQS_SYNY <sup>c</sup>	Synechocystis sp.	Cyanobacteria Photosynthetic	D4006
PHY_CMEL <sup>a</sup>	Cucumis melo	Melon	P49293
PHY_CANN <sup>a</sup>	Capsicum annum	Green pepper	P37272
PHY_LESC <sup>a</sup>	Lycopersicum esculenta	Tomato	P08196
PHY_ATHA <sup>a</sup>	Arbidopsis thaliana	Watercress	P37271
PHY_DBAR <sup>a</sup>	Dunaliella bardawil	Flowering plant	U91900
PHY_NPSE <sup>a</sup>	Narcissus pseudonarcissus	Daffodil	P53797
PHY_ZMAY <sup>a</sup>	Zea mays	Maize	P49085
PHY_SGRI <sup>a</sup>	Streptomyces griseus	Gram-positive bacteria Non-photosynthestic	P54977
PHY_NCRA <sup>a</sup>	Neurospora Crassa	Non-photosynthetic fungus	L27652
PHY_SYNC <sup>a</sup>	Synechococcus sp.	Cyanobacteria Photosynthetic	P37269
PHY_SYNY <sup>a</sup>	Synechocystis sp.	Cyanobacteria Photosynthetic	P37294
PHY_TTHE <sup>a</sup>	Thermus thermophilus	Green non-sulfur bacteria	P37270
PHY_MXAN <sup>b</sup>	Myxococcus xanthus	Proteobacteria $\delta$ -division	S32170

Supplementary Table 1. continued

Protein	Source	Description of source	Accession no.
PHY_EHER <sup>a</sup>	<i>Erwinia herbicola</i>	Proteobacteria $\gamma$ -subdivision	P22872
PHY_EURE <sup>a</sup>	<i>Erwinia uredovora</i>	Proteobacteria $\gamma$ -subdivision	P21683
PHY_MTUB <sup>c</sup>	<i>Mycobacterium tuberculosis</i>	Gram-positive bacteria Non-photosynthetic	Z77165
PHY_MMAR <sup>b</sup>	<i>Mycobacterium marinum</i>	Gram-positive Non-photosynthetic	U92075
PHY_FLAV <sup>a</sup>	<i>Flavobacterium ATCC21588</i>	Green sulfur bacteria Non-photosynthetic	U62808
PHY_AAUR <sup>a</sup>	<i>Argobacterium aurantiarum</i>	Proteobacteria $\alpha$ -subdivision	P54975
PHY_RCAP <sup>a</sup>	<i>Rhodobacter capsulatus</i>	Proteobacteria $\alpha$ - subdivision, Facultative photosynthetic	P17056
PHY_RSPH <sup>a</sup>	<i>Rhodobacter sphaeroides</i>	Proteobacteria $\alpha$ - subdivision, Facultative Photosynthetic	P54905
PHY_MTHE <sup>c</sup>	<i>Methanobacterium thermoautotrophicum</i>	Archeabacteria	AE000934
DSQ_SAUR <sup>a</sup>	<i>Staphylococcus aureus</i>	Gram-positive Bacteria	A55548
PHY_RGEL <sup>b</sup>	<i>Rubrivivax gelatinus</i>	Proteobacteria $\beta$ -division	U87626
PHY_SPLA <sup>a</sup>	<i>Spirulina plantensis</i>	Cyanobacteria Photosynthetic	AB001284
PHY_RHIZ <sup>c</sup>	<i>Rhizobium sp NGR234</i>	Proteobacterium $\alpha$ -division	P55350
PHY_BJAP <sup>c</sup>	<i>Bradyrhizobium japonicum</i>	Proteobacterium $\alpha$ -division	X86552

<sup>1</sup>Proteins were identified in the databases by <sup>a</sup> biochemical identification, indicating that the product of the enzyme catalyzed reaction has been characterized; <sup>b</sup> genetic identification, product of reaction has not been characterized but gene is essential for the first step of carotenoid biosynthesis in the organism studied; <sup>c</sup> predicted, identity of protein is based solely on similarity to other proteins.



S6

Supplemental Table 2. continued

PROTEIN	SQS_ NBEN	SQS_ SCER	SQS_ UMAX	SQS_ CALB	SQS_ SPOM	SQS_ LMAJ	SQS_ SYN_ SQS
SQS_HSAP	45.6	42.8	42.5	43.4	42.8	39.7	31.6
SQS_RNOR	45.6	43.3	43.2	43.7	41.5	39.0	30.0
SQS_MMUS	45.6	44.1	42.8	43.9	42.3	41.3	29.9
SQS_ATHA	76.2	42.0	39.0	40.9	40.5	42.2	28.5
SQS_ATHB	70.2	38.5	37.6	39.0	37.9	39.7	27.7
SQS_GGLA	79.7	41.3	40.3	41.4	42.5	43.9	27.0
SQS_GGLB	81.0	42.2	40.2	44.1	40.8	42.5	30.2
SQS_ZMAY	67.5	40.0	37.6	43.7	42.5	44.7	29.0
SQS_OSAT	69.2	40.4	38.7	44.9	42.7	43.9	28.5
SQS_PGIN	81.9	40.0	40.9	43.6	41.4	42.6	31.4
SQS_GMAX	81.5	41.2	39.1	41.5	40.6	42.3	27.6
SQS_NTAB	98.5	41.2	38.9	44.4	42.0	45.1	29.3
SQS_NBEN	-	41.4	38.6	44.4	43.0	46.5	30.0
SQS_SCER	-	-	39.8	56.6	40.5	36.6	28.3
SQS_UMAY	-	-	-	42.0	47.1	36.3	30.0
SQS_CALB	-	-	-	-	44.8	38.2	27.8
SQS_SPOM	-	-	-	-	-	38.1	30.1
SQS_LMAJ	-	-	-	-	-	-	27.8
SQS_SYNY	-	-	-	-	-	-	-





S9

Supplemental Table 3. continued

PROTEIN	PHY_ EURE	PHY_ EHER	PHY_ AAUR	PHY_ FLAV	PHY_ NCAS
PHY_CMEL	29.5	27.8	33.7	29.0	25.0
PHY_CANN	30.4	27.8	29.9	30.5	23.0
PHY_LESC	29.5	26.7	29.2	28.4	21.5
PHY_ATHA	30.1	27.7	33.6	33.7	24.2
PHY_DBAR	28.8	32.4	31.3	29.1	23.9
PHY_NPSE	30.5	30.3	33.1	32.3	23.0
PHY_ZMAY	30.0	31.0	29.4	29.8	25.6
PHY_SYNC	32.8	30.4	31.3	31.9	28.3
PHY_SYNY	30.7	30.4	28.9	31.6	25.0
PHY_SPLA	32.1	31.4	32.7	30.9	28.7
PHY_MXAN	30.6	33.4	28.2	32.1	27.2
PHY_TTHE	32.5	34.3	31.5	31.3	26.7
PHY_MTUB	26.3	25.7	31.6	29.8	26.5
PHY_MTHE	23.4	26.2	24.6	24.8	25.1
DSQ_SAU	24.4	24.3	25.3	23.7	21.4
PHY_MMAR	29.8	29.1	26.8	30.6	26.3
PHY_SGRI	30.1	28.6	25.4	28.7	26.6
PHY_RHIZ	25.5	28.5	27.2	30.0	26.8
PHY_BJAP	24.7	29.7	30.4	29.5	28.3
PHY_RSPH	36.6	35.6	39.3	39.7	29.3
PHY_RCAP	36.2	32.6	38.0	39.2	30.1
PHY_RGEL	34.4	33.2	38.8	39.7	35.0
PHY_EURE	-	64.5	49.1	50.7	30.4
PHY_EHER	-	-	51.5	52.0	28.6
PHY_AAUR	-	-	-	80.0	35.6
PHY_FLAV	-	-	-	-	31.4
PHY_NCAS	-	-	-	-	-

SQS_SCER	.....MGKLLQ	LALHPVEMKA	ALKLKFQRT	L.....FSIY	DQSTSPYLLH	CFELNLNLSR	SFAAVIREL.	HPELRNCVTL	70
SQS_CALB	.....MGKFLQ	LLSHPTLKA	VIQFGFRQP	L.....HPGK	RDVNDKELGR	CYELNLNLSR	SFAAVIEEL.	HPELRDAVMI	70
SQS_UMAY	MGLLSYILLG	FTHP.SELRA	MIGYKWRDP	LNDIKANPQA	SGWDRQMRD	CWFLDLTSR	SFAAVIKEL.	KGDVSRVICL	78
SQS_SPOM	.....MS	LANRIEEIRC	LCQYKLV...	.NDLPSYGED	EN.VPQNIRR	CYQLLDMTSR	SFAAVIKEL.	PNGIREAVMI	66
SQS_RNOR	.....MEFVK	CLGHPEEFYN	LLFRM...	G	GRRNFIPKMD	RNSLSNSLKT	SFAAVIQAL.	DGDIRHAVCV	71
SQS_MMUS	.....MEFVK	CLGHPEEFYN	LLFRM...	G	GRRNFIPKMD	QDSLSSSLKT	SFAAVIQAL.	DGDIRHAICV	71
SQS_HSAP	.....MEFVK	CLGHPEEFYN	LVRFRI...	G	GKRKVPKMD	QDSLSSSLKT	SFAAVIQAL.	DGEMRNAVCI	71
SQS_OSAT	.....MG	VLSRPEEVL	P	LVKLRV...	A	AGRI...	KR	QIPPEHWAF	64
SQS_ZMAY	.....MG	ALSRPEEVL	A	LVKLRV...	A	AGQI...	KR	QIPPEHWAF	64
SQS_NTAB	.....MGSLRA	ILKNPDDL	Y	LVKLLK...	A	ARHA...	ARHA	QIPSPHWGF	68
SQS_NBEN	.....MGSLRA	ILKNPDDL	Y	LVKLLK...	A	ARHA...	ARHA	QIPSPHWGF	68
SQS_GGLA	.....MGSLGA	ILKHPDDL	Y	LVKLLK...	A	ARQA...	ARQA	QIPREPHWSF	68
SQS_GMAX	.....MGSLGA	ILKHPDDL	Y	LVKLLK...	A	ARNA...	ARNA	QIPPEPHWAF	68
SQS_GGLB	.....MGSLGA	IVRHPDEV	Y	LVKLLK...	A	ARHA...	ARHA	QIPAEPHWAF	68
SQS_PGIN	.....MGSLGA	ILKHPDEF	Y	LVKLLK...	A	ARHA...	ARHA	QIPPEPHWAF	68
SQS_ATHA	.....MGS LGT	MLRYPDDI	Y	LVKMKR...	A	IEKA...	IEKA	QIPPEPHWGF	68
SQS_ATHB	.....MGS LST	ILRHPDEL	Y	LVKLLK...	A	ITKA...	ITKA	QIPLEPHLAF	68
SQS_LMAJ	.....MGFFSDSVA	MMRVKWA	...	Q	MRSV...	KIQ	VPPEETDLRF	CYDIMNDVSR	63
SQS_SYNY	.....MS	GVD RMSLRN	ALKVLQETS	R	TFYIPISILP	D	QLLDAVAS	41	
CONSENSUS								SR xFxxVIxxL	

Supplemental Figure 1. Sequence alignment and consensus sequence for the predicted peptide sequences of 19 squalene synthases. The consensus sequence is based on amino acids that are present in > 90% of the sequences.

	REGION I I				REGION					
SQS_SCER	FYLILRALDT	IEDDMSIEHD	LKIDLLRHFH	EKLLLTKWSF	DGNA...	PDV	KDRAVLTDPE	SILIEFHKLK	PEYQEVIKE	146
SQS_CALB	FYLVLRALDT	IEDDMTIKSS	IKIPLLRREFD	TKLNTKNWTF	DYGG...	PNE	KDRTLVEFD	KILNVYHRLK	PQYQDIKS	146
SQS_UMAY	FYLVLRALDT	VEDDMTIPAQ	RKIPLLVNFY	KYLEQPGWNF	TESG...	PNE	KDRQLLVEFD	KVIAEYQLLD	VGYKTVISD	154
SQS_SPOM	FYLVLRLGLDT	VEDDMTLPLD	KKLPILRDFY	KTIEVEGWNF	NESG...	PNE	KDRQLLVEFD	WVIKEYLNL	EGYRNVISN	142
SQS_RNOR	FYLILRAMDT	VEDDMAISVE	KKIPLLRNPH	TFLYEPWRF	TE.S...	KEK	HRVLEDFP	TISLEFRNLA	EKYQTVIAD	145
SQS_MMUS	FYLVLRALDT	VEDDMSISVE	KKIPLLCNFH	TFLYDPEWRF	TE.S...	KEK	DRQVLEDFP	TISLEFRNLA	EKYQTVIDD	145
SQS_HSAP	FYLVLRALDT	LEDDMTISVE	KKVPLLNHFH	SFLYQPDWRF	ME.S...	KEK	DRQVLEDFP	TISLEFRNLA	EKYQTVIAD	145
SQS_OSAT	FYLVLRALDT	VEDDTSIPAA	VKVPILKEFH	RHIYNRDWHY	SC.G...	TKD	YKLLMDKFR	LVSTAFLELG	QGYQEAIEE	138
SQS_ZMAY	FYLVLRALDT	VEDDTSIPTE	VKVPILQEFY	RHIYNRDWHY	SC.G...	TNH	YKMLMDKFR	HVSTAFLELG	EGYQKAIEE	138
SQS_NTAB	FYLVLRALDT	VEDDTSIPTD	VKVPILISFH	QHVVYDREWHF	SC.G...	TKE	YKVLMDQFH	HVSTAFLELR	KHYQQAIED	142
SQS_NBEN	FYLVLRALDT	VEDDTSIPTD	VKVPILISFH	QHVVYDREWHF	SC.G...	TKE	YKVLMDQFH	HVSTAFLELR	KHYQQAIED	142
SQS_GGLA	FYLVLRALDT	VEDDTSIATD	VKVPILIAFH	RHIYDCDWHF	SC.G...	TKE	YKVLMDQFH	HVSTAFLELG	KNYQEAIED	142
SQS_GMAX	FYLVLRALDT	VEDDTSIETD	VKVPILIAFH	RHIYDRDWHF	SC.G...	TKE	YKVLMDQFH	HVSTAFLELG	KNYQEAIED	142
SQS_GGLB	FYLVLRALDT	VEDDTSIATD	VKEPILIAFH	RHIYDRDWHF	SC.G...	TKE	YKVLMDQFH	HVSTAFLELE	KNYQAAIEN	142
SQS_PGIN	FYLVLRALDT	VEDDTSIPTE	VKVPILMAFH	RHIYDKDWHF	SC.G...	TKE	YKVLMDQFH	HVSNAPLELG	SGYQEAIED	142
SQS_ATHA	FYLVLRALDT	VEDDTSIPTD	EKVPILIAFH	RHIYDGDWHF	SC.G...	TKE	YKVLMDQFH	HVSAAPLELE	SGYQEAIED	142
SQS_ATHB	FYLILRALDT	VEDDTSVPVE	IKVPILIAFH	RHIYDGDWHF	SC.G...	TKE	YKVLMDQFH	HVSAAPLELE	SGYQEAIED	142
SQS_LMAJ	FYLVLRALDT	LEDDMSVPVD	VKLKELPKFH	THTSMSWCM	SGVG...	EGR	ERELLAKYP	CVSREFKCLK	KEYQDVIAN	138
SQS_SYNY	AYLCMRAIDE	VEDHPDLLDAE	TKARILNQIS	LNLQSATENS	HSEDFIRGLD		PYRNLLPEVT	LRVGEWALLA	PESIAPRVWD	121
consensus	FYLxLRALDx	xEdDxxxxxxx	xKxxxxLxxxF						YxxxxIxx	

Supplemental Figure 1. continued

		REGION IV		REGION V					
SQS_SCER	ITEKMGNGMA	DYI.LDENYN	LNGLQTVHDY	DVYCHYVAGL	VGDGLTRLIV	IAKAFANESLY	SNE.QLYESM	GLFLQKTNII	224
SQS_CALB	ITFKMGNGMA	DYI.LDEEFN	VYGVATVEDY	NLYCHYVAGL	VGEGLTNLFV	LANFGDKTILT	ENNFAKADSM	GLFLQKTNII	224
SQS_UMAY	ITAKMGAGMA	SYIELSAK.G	PLKVAMWKHF	DLYCHFVAGL	VGEGLSRLFS	ESKLERPWL.	GHQLELSNHM	GLFLQKTNII	232
SQS_SPOM	ITKEMGDMA	YYASLAEKND	GFSVETIEDF	NKYCHYVAGL	VGIGLSRLFA	QSKLEDPDL.	AHSQAISNSL	GLFLQKVNII	221
SQS_RNOR	ICHRMGCGMA	EFLNKD....	...VTSKQDW	DKYCHYVAGL	VGIGLSRLFS	ASEFEDP.IV	GEDTECANSM	GLFLQKTNII	217
SQS_MMUS	ICHQMGCGMA	EFVDKD....	...VTSKQDW	DKYCHYVAGL	VGIGLSRLFS	ASEFEDP.IV	GEDIECANSM	GLFLQKTNII	217
SQS_HSAP	ICRRMGIGMA	EFLDKH....	...VTSEQEW	DKYCHYVAGL	VGIGLSRLFS	ASEFEDP.LV	GEDTERANSM	GLFLQKTNII	217
SQS_OSAT	ITRLMGAGMA	KFICKE....	...VETVDDY	NEYCHYVAGL	VGYGLSRLFH	AGGTED..LA	SD..SLSNSM	GLFLQKINII	207
SQS_ZMAY	VTRRMGAGMA	KFICKE....	...VETVDDY	DEYCHYVAGL	VGYGLSRLFY	AAGTED..LA	LD..SLSNSM	GLFLQKTNII	207
SQS_NTAB	ITMRMGAGMA	KFICKE....	...VETTDDY	DEYCHYVAGL	VGLGLSKLFH	ASGKED..LA	SD..SLSNSM	GLFLQKTNII	211
SQS_NBEN	ITMRMGAGMA	KFICKE....	...VETTDDY	DEYCHYVAGL	VGLGLSKLFH	ASEKED..LA	SD..SLSNSM	GLFLQKTNII	211
SQS_GGLA	ITKRMGAGMA	EFICKE....	...VETVDDY	DEYCHYVAGL	VGMGLSKLFH	ASGSED..LA	SD..YLSNSM	GLFLQKTNII	211
SQS_GMAX	ITKRMGAGMA	KFICKE....	...VETIDDY	DEYCHYVAGL	VGLGLSKLFH	ASGSED..LA	PD..DLSNSM	GLFLQKTNII	211
SQS_GGLB	ITMRMGAGMA	KFICNE....	...VETVDDY	DEYCHYVAGL	VGLGLSKLFH	ASGKED..LA	PD..HLSNSM	GLFLQKTNII	211
SQS_PGIN	ITRRMGAGMA	KFICKE....	...VETINDY	DEYCHYVAGL	VGLGLSKLFH	ASGAED..LA	TD..SLSNSM	GLFLQKTNII	211
SQS_ATHA	ITRRMGAGMA	KFICQE....	...VETVDDY	DEYCHYVAGL	VGLGLSKLFL	AAGSEV..LT	PDWEAISNSM	GLFLQKTNII	213
SQS_ATHB	ITKRMGAGMA	KFICKE....	...VETIDDY	DEYCHYAAGL	VGLGLSKIFI	ASELEI..LT	PDWKQISNST	GLFLQKTNII	213
SQS_LMAJ	ICERMANGMC	EFLKRP....	...VVTKDDY	NQYCHYVAGL	VGHGLTQLFA	RCGFEDP.SL	DDDLTSSNHM	GLFLQKTNII	210
SQS_SYNY	ATAAMSDRMA	FWALQN....	.WQIRTESEL	DQYTFVAGL	VGLLLSDLW.	..AWHDGTQT	NRSHAIGFGR	G...LQAVNIV	191
Consensus	xxxxmxxxGmA			YCHYVAGL	VGxxLxxL			GLFLQKTNII	

Supplemental Figure 1. continued





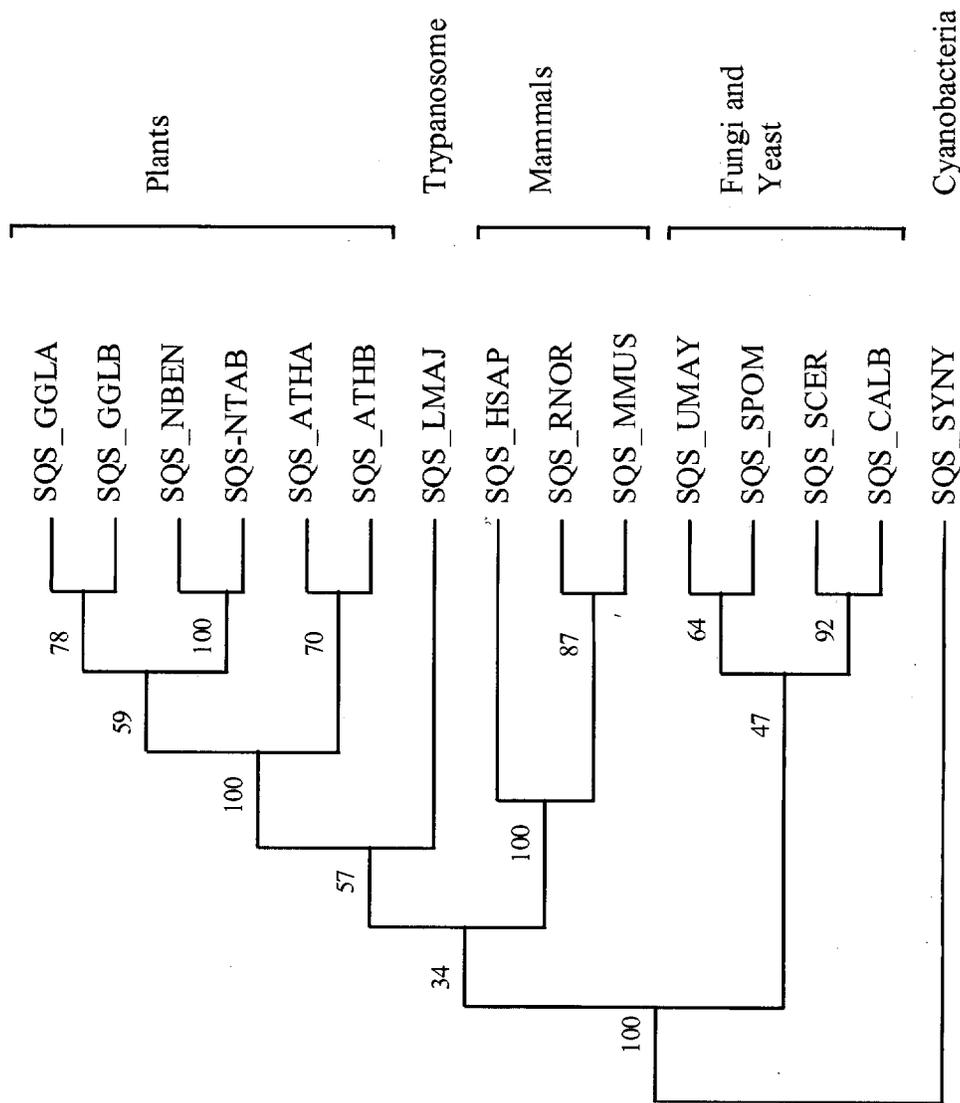
SQS_SCER	KLNIQISKIE	QFMEEMYQDK	.....LPP	NVKPNETPIF	LK.VKERSRY	406			
SQS_CALB	KIGIKGEIE	QYCEMIYPNK	QA.....LPP	SMKSLPENKF	TKIVASRESI	412			
SQS_UMAY	RWSVELARIE	QWCETYYPSPF	IAAAASEGKPT	DIRANALRTW	SESRRTQALI	LKQ..AKLING	SDASTLDAKS	VLEAAQNSAL	467
SQS_SPOM	KISIECGKIE	QVSESLFP..	.....RRFREMY	EKAYVSKLSE	QKKGNGTQKA	ILNDEQKELY	416		
SQS_RNOR	KAKQLISNIR	TQSLP.....	.....	NCQLISR..	.....SHY	382			
SQS_MMUS	KTKQVISKIR	TQNL.P.....	.....	NCQLISR..	.....SHY	382			
SQS_HSAP	KTRQIISTIR	TQNL.P.....	.....	NCQLISR..	.....SHY	382			
SQS_OSAT	LTRKRVDIAK	RTCKS.....	.....	SCSLKRRG	YDL..EKSKY	381			
SQS_ZMAY	LTQRRVDSIK	QTCKS.....	.....	SGLVKQRG	YHL..EKSPY	381			
SQS_NTAB	KTLKRLEAIL	KTCRD.....	.....	SGTLNKRK	SYIIRSEPNY	387			
SQS_NBEN	KTLKRLEVIL	KTCRD.....	.....	SGTLNKRK	SYIIRSEPNY	387			
SQS_GGLA	KTLRLEAIIQ	KTCRE.....	.....	SGLLNKRK	SLLRNGN.CY	386			
SQS_GMAX	KTLRLEAIIQ	KTCRE.....	.....	SGLLNKRK	SYIVNDESGY	387			
SQS_GGLB	KTLRLEAIIQ	KTCRE.....	.....	SGLLNKRK	PYLLRNESTN	387			
SQS_PGIN	KTLRLEAIIQ	KTCRE.....	.....	SGLLNKRK	SYIIESESGH	387			
SQS_ATHA	KTLNRLEAVQ	KLCRD.....	.....	AGVLQNRK	SYV.NDKGQP	388			
SQS_ATHB	KTLNRLETIK	KVCRE.....	.....	NGVLHKRK	SYV.NDETQS	388			
SQS_LMAJ	SIGNSLAAAI	ERIDL.....	.....	QLKCKQDV	SYTRSLLARY	385			
SQS_SYNY	.....	.....	.....	.....	.....	277			
CONSENSUS									

Supplemental Figure 1. continued

SQS_SCER	D.....	.DELVPTQOE	EYKFNWVLS	IILSVLLGFY	YIYTLHRA...	.....	444	
SQS_CANL	D.....	.LSVQRRIEP	GNFNCNVVLF	GIGALLISLI	YFVL.....	.....	446	
SQS_UMAY	DPRDLMTEDE	RAAQDKRDRD	QMWKFFLIIL	VGMVTFMGIV	ALITWEIVWM	WTMDTPDPLS	VYVKHAYYLV	537
SQS_SPOM	.....	.....	.....	.....	.....	.....	460	
SQS_RNOR	SPI.....	.....	.....	.....	.....	.....	416	
SQS_MMUS	SPI.....	.....	.....	.....	.....	.....	416	
SQS_HSAP	SPI.....	.....	.....	.....	.....	.....	417	
SQS_OSAT	NSM.....	.....	.....	.....	.....	.....	403	
SQS_ZMAY	RPM.....	.....	.....	.....	.....	.....	401	
SQS_NTAB	SPV.....	.....	.....	.....	.....	.....	411	
SQS_NBEN	SPV.....	.....	.....	.....	.....	.....	411	
SQS_GGLA	GST.....	.....	.....	.....	.....	.....	412	
SQS_GMAX	GST.....	.....	.....	.....	.....	.....	413	
SQS_GGLB	SST.....	.....	.....	.....	.....	.....	412	
SQS_PGIN	NSA.....	.....	.....	.....	.....	.....	415	
SQS_ATHA	NSV.....	.....	.....	.....	.....	.....	410	
SQS_ATHB	KAI.....	.....	.....	.....	.....	.....	413	
SQS_LMAJ	PGL.....	.....	.....	.....	.....	.....	414	
SQS_SYNY	.....	.....	.....	.....	.....	.....	277	

CONSENSUS

Supplemental Figure 1. continued



Supplemental Figure 2. Phylogenetic analysis based on squalene synthase amino acid sequences. Branches are not drawn to scale.

```

PHY_AAUR .....
PHY_FBAC .....
PHY_EURE .....
PHY_EHER .....
PHY_RCAP .....
PHY_RSPH .....
PHY_RGEL .....
PHY_BJAP .....
PHY_RHIZ .....
PHY_SGRI .....
PHY_MMAR .....
PHY_SYNY .....
PHY_SYNC .....
PHY_SPLA .....
PHY_LESC .....MSVALL WVVSP.CDVS .NGTSFMESV REGNRFDFS. ....SRHRN LVSNERINRG 48
PHY_CANN .....MSVALL WVVSP.CDVS .NGTGFLVSV REGNRFDFS. ....SRRN LACNERIKRG 48
PHY_NPSE .....MVVAIL RVVSA.IEIP .IRLGF....SEANWRFSS. ....PKYDN LGRKKSRLSV 44
PHY_CMEL .....MSLASS LVVSSNVELS PSSFGFLDSV RDGPQIPDSF RFSSRNRPVN LINKKQKWN 56
PHY_ATHA .....MSSVAVL WVATSSLNPD PMNCGLVRV LESSRLFSP. ....CQNQRNLKG 42
PHY_ZMAY .....MAII LVRAASPGLS AADSISHQGT LQCSTLLKTK RPAARRWMP. .... 43
PHY_DBAR .....MTLSMLDAR RMAQRSASSS SSFPIGSGTA PSRMSRICGI 39
PHY_MXAN .....
PHY_TTHE .....
PHY_MTUB .....
DSQ_SAU .....
PHY_MTHE .....
PHY_NCAS HALHLNNQON PPAMRVRVKV TGQVVLVALS VMGWNAAQVH QETSYLGLIL VWACPFLLAI WTLAGRFFILS LPWYATVILPM 180
CONSENSUS

```

Supplemental Figure 3. Sequence alignment of 26 phytoene synthases and a dehydroqualene synthase. The consensus sequence is based on amino acids that are present in > 90% of the sequences.



	REGION I		
PHY_AAUR	.....MS.....	DLVLTSTEA	TQGSQSFATA
PHY_FBAC	.....MT.....	DLTATSEAA	AQGSQSFQA
PHY_EURE	.....	.....M	AVGSKSFATA
PHY_EHER	.....MSQP.....	PLLDHATQTM	ANGSKSFATA
PHY_RCAP	.....MIAE.....	ADMEVCRELI	RTGYSFHA
PHY_RSPH	.....MIAS.....	ADLDACREMI	RTGYSFHA
PHY_RGEL	.....MP	SELQACRELM	RGSKSFHAA
PHY_BJAP	.....MTLE.....	ATSPGASYGS	TASGSSFYAA
PHY_RHIZ	.....MRAE.....	AAAH.ANHRS	TALGSSFYLG
PHY_SGRI	.....HDPALR.....	AAAYTRCRGLN	ARHGKTYFLA
PHY_MMAR	.....GDPRLR.....	EAYQOSRRIN	AASGRTFFLA
PHY_SYNY	.....E.....	EAYEICRQVT	AQYAKTFYLG
PHY_SYNC	.....S.....	EAYEECRQIT	ARYAKTFYLG
PHY_SPLA	.....S.....	DSYELCRQIT	AQYAKTFYLG
PHY_LESC	.....L.S.....	EAYDRCGEVC	AEYAKTFNIG
PHY_CANN	.....L.S.....	EAYDRCSEVC	AEYAKTFYLG
PHY_NPSE	.....L.K.....	DAYDRCGEVC	AEYAKTFYLG
PHY_CMEL	.....L.N.....	EAYDRCGEVC	AEYAKTFYLG
PHY_ATHA	.....LVG.....	EAYDRCGEVC	AEYAKTFYLG
PHY_ZMAY	.....K.....	EAYDRCGEVC	AEYAKTFYLG
PHY_DBAR	.....QGSLE.....	GAYERCGAVC	SEYAKTFYLG
PHY_MXAN	.....MSPPV	RGYVLAQVT	RHHAKSFFFA
PHY_TTHE	.....DKALVA.....	PDMKALLRVL	RAHSATFYLG
PHY_MTUB	.....MPASME.....	QAYRITESIT	RTAARNFYYG
DSQ_SAUR	.....MTEIE.....	MNFKYCHKIM	KKHKSFSYA
PHY_MTHE	.....MTMMD.....	LIDGKIYSIF	KRGSKTYFYS
PHY_NCAS	.....LMGDIT.....	THMLLQSRLL	NTSRDYDLERI
	.....EGLREAVERL	RLKRSRFFLA	NSLFSGRRLI
CONSENSUS			

YxxxxR xxDxxxxD

F

Supplemental Figure 3. continued

PHY_AAUR	NDPQARLDGL	RADTLAAL..	.....	QCGPVTTPF	AC..AVARRH	DFQAWPMDL	IEGFAMD...	113	
PHY_FBAC	GDPQARLGAL	RADTLAAL..	.....	HEDGPMSPPF	AALRQVARRH	DFPDLWPMDL	IEGFAMD...	115	
PHY_EURE	QTPEQRMLQL	EMKTRQAY..	.....	AGSQMHEPAF	AAFQEVAMAH	DIAPAYAFDH	IEGFAMD...	106	
PHY_EHER	EEATQRLARL	RTLTLAAF..	.....	EGAEMQDPAF	AAFQEVALTH	GITPRMALDH	LDGFAMD...	119	
PHY_RCAP	....A.VLKL	GDRLEDIY..	.....	AGRPRNAPSD	RAFAAVVEEF	EMPRALPEAL	IEGFAMD...	111	
PHY_RSPH	VKRRR.VLSL	RDRLDLVY..	.....	GGRPRNAPAD	RAFAAVVEEF	EMPRALPEAL	LEGLAWD...	118	
PHY_RGEL	....A.MAEL	RTRLDGVY..	.....	AGTPAPIAAD	RALACTVHRY	GVPRVLLDAL	LEGFLWD...	112	
PHY_BJAP	....AALQAW	RTDIDALY..	.....	..QGNPPRL	KDYLASVKTFF	GLKREDFLAI	VDGMEMD...	109	
PHY_RHIZ	....AALQQW	RDHIDALY..	.....	..QCVPPRL	KDYLASVTTFF	GLKREDFLAV	VDGMEMD...	108	
PHY_SGRI	....RALLGL	EAQLEAGL..	.....	RGARAAEPI	RALADTAARY	TIDHRHFTDF	LSSMRSD...	127	
PHY_MMAR	....QRLQQL	AQFFSG..	.....	.ADHPDNPVL	AAVSHTRARY	GIGADLFEDF	LESMRMD...	125	
PHY_SYNY	....ETLDHW	ERRLEGIF..	.....	.AGQPQDDAD.	VALVDTLETFF	PLDIQFFRDM	IAGQRM...	152	
PHY_SYNC	....ATLDW	ERRLERLF..	.....	.AGEPEDDCD.	VALVDTLARY	PLDIQFFRDM	IEQRM...	123	
PHY_SPLA	....ETLDHW	EEQLESIF..	.....	.AGHPIEPVD.	VALVDTLGRF	PLDIQFFRDM	IAGQRM...	123	
PHY_LESC	....AALDRW	ENRLEDVF..	.....	.NGRPFDMLD.	GALSDTVSNF	PVDIQFFRDM	IEGMRM...	224	
PHY_CANN	....AALDRW	EDRLEDVF..	.....	.SGRPFDMLD.	AALSDTVSKF	PVDIQFFRDM	IEGMRM...	236	
PHY_NPSE	....SALDRW	EARLEDLF..	.....	.AGRPFDMFD.	AALSDTVSRF	PVDIQFFRDM	VEGMRM...	232	
PHY_CMEL	....TALDRW	EARLEELF..	.....	.QGRPFDMLD.	AALADTVTKF	PVDIQFFKDM	IEGMRM...	240	
PHY_ATHA	....MALDRW	EARLEDLF..	.....	.RGRPFDMLD.	AALADTVARY	PVDIQFFRDM	IEGMRM...	236	
PHY_ZMAY	....TALDRW	EKRLEDLF..	.....	.TGRPYDMLD.	AALSDTISRFF	PIDIQFFRDM	IEGMRSD...	222	
PHY_DBAR	....QALDRW	EERLEAMF..	.....	.QGKPYDELD.	AALDTLTSKY	PLEIQFFRDM	IEGMRM...	233	
PHY_MXAN	....VRLARA	RQRVAELYLP	MPELASRELG	PPADR.VKGSE	AATPWDAREF	RIPEQFFQDL	ISGMEMD...	143	
PHY_TTHE	....EALEAW	WAGVERAY..	.....	.RGRPLAEWEK	G.LAWALERW	DIPFEAF.LHM	REGFLTD...	120	
PHY_MTUB	....DAIRKS	LDNIDDS..	.....	....SDPVL	VALADAARRF	PVPIAMFAEL	IDGARME...	109	
PHY_SOUR	KEDIQSIEKY	PYEHFFF..	.....	....QSDRRIM	MALQVAQHK	NIAFQSFYNL	IDTV.YK...	113	
PHY_MTHE	....DGFYEF	VERYRRAY..	.....	.SGETGDVVV	DSFVELSERK	NFEDEWTEAF	LRSMEMD...	119	
PHY_NCAS	....REVLSW	TAKLNHFLDL	HYKDATATED	PKKKAERIDA	YIKTAFPPCA	YQALHLLPTH	ILPPKPLYDL	IKGFEMDSQF	408

CONSENSUS

Supplemental Figure 3. continued

	REGION II		REGION III					
PHY_AAUR	V. EARDVRTLDD	VLEYSYHVAG	IVGVMMARVM	GV.....RDDPV	LDRACDLGLA	FQL.TNIARD	170	
PHY_FBAC	V. ADREYRSLDD	VLEYSYHVAG	VVGVMMARVM	GV.....QDDAV	LDRACDLGLA	FQL.TNIARD	172	
PHY_EURE	V. REAQYQLDD	TLRYCYHVAG	VVGLMMAQIM	GV.....RDNAT	LDRACDLGLA	FQL.TNIARD	163	
PHY_EHER	V. AQTRYVTFED	TLRYCYHVAG	VVGLMMARVM	GV.....RDERV	LDRACDLGLA	FQL.TNIARD	176	
PHY_RCAP	A. EGRWHTLSD	VQAYSARVAA	AVGAMMCVLM	RV.....RNPDA	LARACDLGLA	MQM.SNIARD	168	
PHY_RSPH	A. VGRSYDSFSG	VLDYSARVAA	AVGAMMCVLM	RV.....RDPDV	LARACDLGLA	MQM.TNIARD	175	
PHY_RGEL	A. DGRRYDTIAD	VEAYGARVAG	TVGAAMALIM	GV.....RSPQA	LARACELGVA	MQF.TNIARD	169	
PHY_BJAP	VP QDIRAPDMAT	LDLYCDRVAS	AVGRLSRVVF	GL.....PEEDG	IQLAYHLGRA	LQL.TNILRD	167	
PHY_RHIZ	VL QDIRAPKMAT	LDLYCDRVAS	AVGRMSRVVF	GL.....SEEDG	IALAHHLGRA	LQL.TNILRD	166	
PHY_SGRI	LT VG.GYASYEE	LGRYMHGSA	VIGLQMLPVL	GTV.....VAREEA	APHAAALGVA	FQL.TNFLRD	186	
PHY_MMAR	LT IT.DYPRDA	LNTYMRGSAE	AIGLQLLPIL	GTV.....VPAREA	MPYAAALGRA	FQL.TNFIRD	184	
PHY_SYNY	LY RS.RYQTFEE	LDLYCYRVAG	TVGLMSSAVL	GVDTGNGQAP	WQPD.AVIYP	QEEAIALGVA	NQL.TNILRD	221
PHY_SYNC	LL QN.RYSTFED	LNLYCYRVAG	TVGLMSQPVV	GIESTNSRAP	WDPT.TPPDP	TQEALALGIA	NQL.TNILRD	192
PHY_SPLA	LS RN.RYNTFDE	LNLYCYRVAG	TVGLMSLAVM	GTAEPDLSVP	WNRDQSIYYP	KEEAIALGIA	NQL.TNILRD	193
PHY_LESC	LR KS.RYKNFDE	LYLYCYVVAG	TVGLMSVPIM	GIAP.....ESKATTESV	YNAALALGIA	NQL.TNILRD	287	
PHY_CANN	LR KS.RYRNFDE	LYLYCYVVAG	TVGLMSVPIM	GIAP.....ESKATTESV	YNAALALGIA	NQL.TNILRD	299	
PHY_NPSE	LK KS.RYKNFDE	LYLYCYVVAG	TVGLMSVPVM	GIAP.....ESLAEAESV	YNAALALGIA	NQL.TNILRD	295	
PHY_CMEL	LR KS.RYKNFDE	LYLYCYVVAG	TVGLMSVPVM	GIAP.....ESQASTESV	YNAALALGIA	NQAPPNLRD	304	
PHY_ATHA	LK KS.RYQNFDD	LYLYCYVVAG	TVGLMSVPVM	GIDP.....KSKATTESV	YNAALALGIA	NQL.TNILRD	299	
PHY_ZMAY	LR KT.RYNNFDE	LYMYCYVVAG	TVGLMSVPVM	GIAT.....ESKATTESV	YSAALALGIA	NQL.TNILRD	285	
PHY_DBAR	LF KS.RYTFDE	LYEYCYRVAG	TVGLMTPVVM	GVDPSY....	YRAALALGTA	NQL.TNILRD	296	
PHY_MXAN	LT KH.RYATWEE	LDLYCYRVAG	VVGLMLTPVL	GC.....SDAAA	VEPAADLGRA	MQM.TNILRD	200	
PHY_TTHE	LG PV.RLGTEAE	LLRYCYQVAG	TVGRMMAPIA	G.....GGKEA	EARAVKLGQA	MQM.TNILRD	176	
PHY_MTUB	ID WT.GCRDFDE	LIVYCRRGAG	TIGKLCLSIF	G.....PVSTAT	SRVAEQLGIA	LQ.QTNILRD	166	
PHY_SAU	VN ILQCLKRTLE	LFGYCYGVAG	RRSSIDADF.	.....SDHETHQT	YDVARRLGES	LQL.INILRD	171	
PHY_MTHE	IT VS.SYSTMSA	LEEYLLGSSE	VVGLFMASIM	GL.....HTDS	YHHARYLGRA	MQY.VNFIRD	175	
PHY_NCAS	TFHGTSDSTD	LQYPIADDDK	LVGELCIALI	.....IYHCLP	DMSDTQKREL	LQY.VNIARD	483	
CONSENSUS		YxxxxxAX	xxxG		AxxxLGXA	xQxxxxNxxRD		

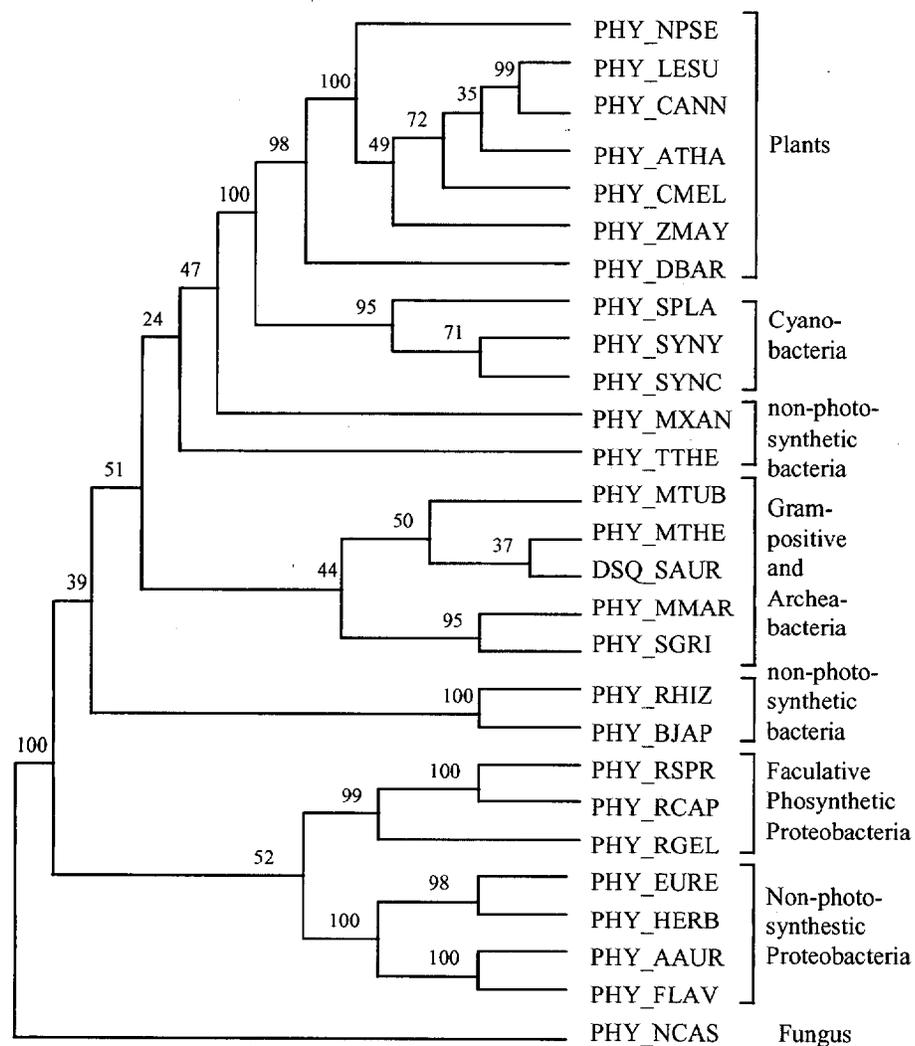
Supplemental Figure 3. continued



PHY_AAUR	LRI...YRA	IGLRIRKGGP	EAYRQRISTS	KAAKIGLIGI	GGWDVADHAC	RGRRVA.TGL	WTRPHHA	301
PHY_FBAC	LRI...YRA	IGTRIRQGGP	EAYRQRISTS	KAAKIGLLAR	GGLDAAAARL	RGGEISRDLG	WTRPRA	303
PHY_EURE	KQV...YRK	IGVKVEQAGQ	QAWDQRQSTT	TPEKLTLLLA	ASGQALTSRM	RAHPPRPAHL	WQRPL	296
PHY_EHER	RSV...YRE	IGIKVKAAGG	SAWDRRQHTS	KGEKIAMLMA	APGQVIRAKT	TRVTPRPAGL	WQRPV	309
PHY_RCAP	GKI...YAA	IGAEVAKAKY	DNITRRAHTT	KGRKLIWLVN	SAMSATATSM	LPLSPRVH.A	KPEPEVAHLV	309
PHY_RSPH	RHI...YAG	IGDEIARNGY	DSVTRRAFTT	RRQKLWLVGN	SSTRAALSPF	GPGCATLH.A	APEPEVAFLV	316
PHY_RGEL	RWS...TAE	IGKRLERDGL	DSVNRVVVP	ARRKAALMAR	AVRRL	QHARARLHLD	AALPAIQYLV	313
PHY_BJAP	PRIMSKYHS	ILDLLIARGF	NAPREPVRS	KVTRILILLR	YAFI	.....	.....	279
PHY_RHIZ	PTIMSKCYRA	ILDLLIRGF	AAPREPVRT	NLTKRAILFR	YALML	.....	.....	279
PHY_SGRI	AFVL...YSG	ILDIAEADGY	AVLHRRRAVVP	RR.RRAAVAL	DGLVRRAVAVR	TVCRRRADAV	DASRSAELSR	337
PHY_MMAR	ALT...YSE	ILDRIEGDDF	AVVGRRAATVG	TA.RRLRVAG	AGLIRA	.....	RSEYGAA	322
PHY_SYNY	ALML...YKG	ILDVIEANNY	NVFNRRAYVP	TP.KKLLYLP	VAWLRAQVL	.....	.....	337
PHY_SYNC	SIML...YRE	ILDVIEQNNY	DVFRKRAYVP	TW.RKLCSLP	VAMLRATVL	.....	.....	308
PHY_SPLA	ALML...YRQ	ILDEIERNEY	DVFNQRAYVP	TW.KKMMCLP	LAQLRARVL	.....	.....	309
PHY_LESC	SLVL...YRK	ILDEIEANDY	NNFTKRAYVS	KS.KKLIALP	IYAKSLVP	PTKTASLQR	.....	412
PHY_CANN	SLLL...YRR	ILDEIEANDY	NNFTKRAYVS	KP.KKLIALP	IYAKSLVP	.STRT	.....	419
PHY_NPSE	SLLL...YRQ	ILDEIEANDY	NNFTKRAYVS	KV.KRLAALP	LAYGKSLILP	LSLRPPSLSK	A	423
PHY_CMEL	SLLL...YRQ	ILDEIEANDY	DNFTKRAYVS	KA.KKILALP	MAYGRALLGP	S	.....	422
PHY_ATHA	SLLL...YRR	ILDEIEANDY	NNFTKRAYVG	KV.KKIAALP	LAYAKSVLKT	SSRSLI	.....	423
PHY_ZMAY	SLLL...YRQ	ILDEIEANDY	NNFTKRAYVG	KG.KKLLALP	VAYGKSLLLP	CSLRNGQT	.....	410
PHY_DBAR	ALIL...YRQ	ILDSIEKNDY	DNFSMRAYVP	KA.KKFTSLP	MALFRAMVPQ	NQNK	.....	425
PHY_MXAN	MGAI...YAD	ILRDIEARDY	DVFSARAHVT	TR.RKLLALAS	AAMVRAAVL	PAPQGEVRMP	LLPTGAGG	336
PHY_TTHE	AALQ...YRG	ILDKLRLSGY	DNLGRRHLK	AW.ERALLLP	KAFLAARFPP	RPEGSP	.....	298
PHY_MTUB	MSGI...YRR	QLALIRASPA	VVYDRRISLS	GL.KKAQVAA	AALASSVTCC	PAHGPLPADL	GSHPSH	303
DSQ_SAUR	ARI...YIE	ILGRS	.....	.....	.....	.....	.....	254
PHY_MTHE	ADM...YMW	TSRIIERDPM	IVYSRKVKPS	GG.RVVSGAV	FNMFRMLMGPL	RVSRGP	.....	299
PHY_NCAS	VENYMAIGRV	LREKKEGTVF	VRMEGRATVP	KRRRLSTLLR	ALYEQ	.....	.....	601

CONSENSUS

Supplemental Figure 3. continued



Supplemental Figure 4. Phylogenetic analysis of I phytoene synthases and a dehydrosqualene synthase..