

## Supplementary data

Characterization of terpene synthase gene family in *Artemisia annua* L

Weizhang Jia, Ying Wang\*, Xuesong Yu, Ling Zhang, Shunhui Liu, Zebo Huang\*

**Table S1.** GenBank accessions for screening, multiple sequence alignment and assembly of AaTPSs transcripts among nucleotid collection (nr/nt), expressed sequence tags (EST) and transcriptome Shotgun Assembly (TSA) databases.

Gene name	GenBank accession
AaSTS1	EZ313203, EZ229184, EZ381466, EZ341212, EZ258363
AaSTS2	EZ388063, EZ357842, EZ384994, EZ292080, EZ358562, EZ376417, EZ362869, EZ278160, EZ298064, EZ266508, EZ339040
Aa16267	EZ350327
Aa14765	GW334158
Aa02039	EZ212454, EZ153545, EZ299087
AaMTS1	EZ346397
AaMTS2	EZ388701, EZ144596, EZ374272, EZ328402
AaMTS3	EZ242654, EZ277194, EZ215976, EZ402158, EZ393089, EZ164598, EZ251571, EZ160235, EZ164598

**Table S2.** TPS proteins represented in phylogenetic analysis.

Species	Protein ID/Accession nos	Note
<i>Artemisia annua</i>	Q94G53	AaBPS
	Q9SPN0	AaLS (QH1)
	Q9SPN1	Aa LS (QH5)
	CAB94691	AaADS
	Q8SA63	AaCPS
	Q9LLR9	AaECS
	AAX39387	AaBFS
	ABE03980	AaGAS
	New	AaMTS1
	New	AaMTS2
	New	AaMTS3
	New	AaSTS1
	New	AaSTS2
	AAG24640	
	New*	Aa16267
	New*	Aa14765
New*	Aa02039	

---

<i>Arabidopsis thaliana</i>	At1g31950	
	At1g33750	
	At1g48800	
	At1g48820	
	At1g61120	AtTPS04
	At1g61680	AtTPS14
	At1g66020	
	At1g70080	
	At1g79460	
	At2g23230	
	At2g24210	AtTPS10
	At3g14490	
	At3g14520	
	At3g14540	
	At3g25810	AtTPS24
	At3g25820/At3g25830	AtTPS23/AtTPS27
	At3g29110	
	At3g29190	
	At3g29410	
	At3g32030	
	At4g02780	
	At4g13280	AtTPS12
	At4g13300	AtTPS13
	At4g15870	
	At4g16730	
	At4g16740	AtTPS03
	At4g20200	
	At4g20210	
	At4g20230	
	At5g23960	AtTPS21
	At5g44630	AtTPS11
	At5g48110	
	<i>Vitis vinifera</i>	GSVIVT01036308001
GSVIVT01036312001		VvTPS02
GSVIVT01036313001		VvTPS03
GSVIVT01036315001		VvTPS04
GSVIVT01036317001		VvTPS05
GSVIVT01036320001		VvTPS06
GSVIVT01036322001		VvTPS07
GSVIVT01036330001		VvTPS08
GSVIVT01036343001		VvTPS09
GSVIVT01036344001		VvTPS10
GSVIVT01036348001		VvTPS11
GSVIVT01036351001		VvTPS12
GSVIVT01036360001		VvTPS13
GSVIVT01036361001		VvTPS14
GSVIVT01036366001		VvTPS15
GSVIVT01036370001		VvTPS16

---

---

GSVIVT01036372001	VvTPS17
GSVIVT01036374001	VvTPS18
GSVIVT01036376001	VvTPS19
GSVIVT01014174001	VvTPS20
GSVIVT01014175001	VvTPS21
GSVIVT01014323001	VvTPS22
GSVIVT01014324001	VvTPS23
GSVIVT01014325001	VvTPS24
GSVIVT01014557001	VvTPS25
GSVIVT01014558001	VvTPS26
GSVIVT01014566001	VvTPS27
GSVIVT01014569001	VvTPS28
New	VvTPS29
GSVIVT01036367001	VvTPS30
GSVIVT01033458001	VvTPS31
GSVIVT01030646001	VvTPS32
GSVIVT01030647001	VvTPS33
GSVIVT01000401001	VvTPS34
GSVIVT01000402001	VvTPS35
New	VvTPS36
New	VvTPS37
New	VvTPS38
New	VvTPS39
New	VvTPS40
GSVIVT01013515001	VvTPS41
NEW	VvTPS42
GSVIVT01013517001	VvTPS43
GSVIVT01013518001	VvTPS44
New32	VvTPS45
GSVIVT01006399001	VvTPS46
GSVIVT01006642001	VvTPS47
New	VvTPS48
New	VvTPS49
New	VvTPS50
GSVIVT01000410001	VvTPS51
GSVIVT01000414001	VvTPS52
GSVIVT01005217001	VvTPS53
GSVIVT01005221001	VvTPS54
GSVIVT01005268001	VvTPS55
GSVIVT01005272001	VvTPS56
GSVIVT01006466001	VvTPS57
GSVIVT01006467001	VvTPS58
GSVIVT01006589001	VvTPS59
GSVIVT01006591001	VvTPS60
New	VvTPS61
GSVIVT01007468001	VvTPS62
GSVIVT01002718001	VvTPS63
GSVIVT01005218001	VvTPS64

---

---

	GSVIVT01005271001	VvTPS65
	GSVIVT01006465001	VvTPS66
	GSVIVT01001155001	VvTPS67
	GSVIVT01001153001	VvTPS68
	GSVIVT01036724001	VvTPS69
<i>Oryza sativa</i>	Os01g23530	
	Os01g42610	
	Os02g02930	
	Os02g17780	OsCPS1
	Os02g36020	
	Os02g36140	OsKSL7/OsDTC1
	Os02g36210	OsCPS2
	Os02g36220	OsKSL5
	Os02g36264	OsKSL6
	Os03g22634	
	Os03g24640	
	Os03g24680	
	Os03g24690	
	Os03g31430	
	Os04g09900	OsCPS4
	Os04g10060	OsKSL4
	Os04g26960	
	Os04g27070	
	Os04g27190	
	Os04g27340	
	Os04g27400	
	Os04g27540	
	Os04g27720	
	Os04g27790	
	Os04g52210	
	Os04g52230	OsKS1
	Os04g52240	
	Os07g11790	
	Os08g04500	
	Os08g07080	
	Os08g07100	
	Os11g28530	OsKSL8/OsDTC2
	Os12g30824	OsKSL10
<i>Phycomitrella patens</i>	Phypa_234290	PpCPS/KS
<i>Populus trichocarpa</i>	POPTR_0001s31550	
	POPTR_0001s31570	
	POPTR_0001s31580	
	POPTR_0001s44080	
	POPTR_0002s05300	
	POPTR_0004s02970	
	POPTR_0004s02990	
	POPTR_0004s03810	
	POPTR_0005s09830	

---

---

	POPTR_0005s23190
	POPTR_0007s02810
	POPTR_0007s02920
	POPTR_0007s07360
	POPTR_0007s07410
	POPTR_0008s08190
	POPTR_0008s08220
	POPTR_0011s03440
	POPTR_0011s14600
	POPTR_0015s05270
	POPTR_0015s09710
	POPTR_0017s06920
	POPTR_0019s01270
	POPTR_0019s01290
	POPTR_0019s01320
	POPTR_0019s01340
	POPTR_0019s03350
	POPTR_0019s03980
	POPTR_0019s03990
	POPTR_0019s06190
	POPTR_0019s06220
	POPTR_0092s00200
	POPTR_0121s00250
	POPTR_0408s00200
	XP_002305455
<i>Sorghum bicolor</i>	Sb01g015070
	Sb01g021990
	Sb01g032610
	Sb01g034700
	Sb01g035460
	Sb01g039090
	Sb04g001780
	Sb04g001800
	Sb04g001810
	Sb05g006470
	Sb05g019210
	Sb05g022320
	Sb06g002820
	Sb06g028210
	Sb06g028220
	Sb06g031270
	Sb07g003080
	Sb07g004470
	Sb07g004480
	Sb07g004485
	Sb07g005130
	Sb07g020980
	Sb07g025700

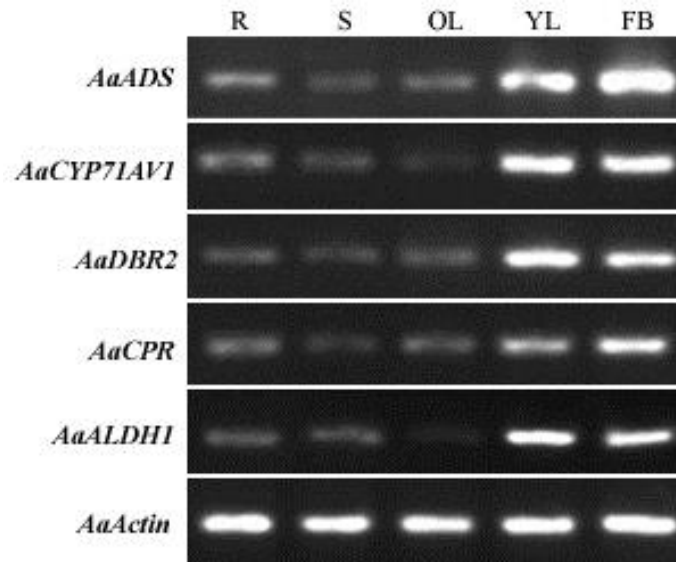
---

	Sb09g000990	
	XP_002466259	
<i>Selaginella moellendorffii</i>	Selmol_76432	
	Selmol_86551	
	Selmol_112927	
	Selmol_124329	
	Selmol_163980	
	Selmol_402349	
	Selmol_402351	
	Selmol_403761	
	Selmol_403764	
	Selmol_406214	
	Selmol_407280	
	Selmol_412139	
	Selmol_418910	
	Selmol_439448	
	XP_002971951	
<i>Abies grandis</i> (grand fir)	AAF61453	$\beta$ -phellandrene synthase
	AAB71084	myrcene synthase
	AAF61455	$\alpha$ -pinene/limonene synthase
	AAF61454	terpinolene synthase
	AAB70907	(-)-limonene synthase
	AAB71085	(-)-pinene synthase
	AAB70707	(-)-camphene synthase
	AAC05727	$\delta$ -selinene synthase
	AAC05728	$\gamma$ -humulene synthase
	AAC24192	<i>E</i> - $\alpha$ -bisabolene synthase
	AAB05407	abietadiene synthase
<i>Picea abies</i> (Norway spruce)	AF461459	(+)-3-carene synthase
	AY473624	(-)-limonene synthase
	AY473623	(-)-linalool synthase
	AY473622	(-)- $\alpha$ / $\beta$ -pinene synthase
	AY473626	myrcene synthase
	AY473627	<i>E,E</i> - $\alpha$ -farnesene synthase
	AY473625	longifolene synthase
	AY473619	<i>E</i> - $\alpha$ -bisabolene synthase
	AY473621	levopimaradiene/abietadiene synthase
	AY473620	isopimara-7,15-diene synthase
<i>Picea sitchensis</i> (Sitka spruce)	AAP72020	(-)-pinene synthase
<i>Pinus taeda</i> (loblolly pine)	AAO61228	(+)- $\alpha$ -pinene synthase
	AAO61225	(-)- $\alpha$ -pinene synthase
	AAO61227	$\alpha$ -terpineol synthase
	AAO61226	$\alpha$ -farnesene synthase
<i>Taxus brevifolia</i>	AAC49310	taxadiene synthase
<i>Ginkgo biloba</i>	AAL09965	levopimaradiene synthase

\*: the asterisk represents that the sequence was identified by Wang et al. (2009), there is no corresponding protein ID in GenBank.

**Table S3.** Nucleotide sequences of primers used in clone, RT-PCR and qRT-PCR.

Gene Name	Forward Primer (5' to 3')	Reverse Primer (5' to 3')	Fragment size (bp)	Application
$\beta$ -Actin	CCTGCTATGTATGTTGCCATCCA	CTCGGTAAGGATCTTCATCAATGCA	198	RT-PCR and qRT-PCR
AaADS	CTGTCTGAATGGGCTGTCTCTG	GTAATCAAGGTTTGGGCATACTCC	179	RT-PCR and qRT-PCR
AaCPS	CCTATGTACGAACAACAAGCATCTC	GCATAAGGTAGATTGTTTGGGACA	152	RT-PCR and qRT-PCR
AaECS	GGGTGATGTAATTGCAGATGACTC	ATTCAATAGTGGAGGCAACATGCT	149	RT-PCR and qRT-PCR
AaGAS	GATGCTTTGGCTTGGTATGAAAGTC	GCGTCAATAGCTTCCTTTTCTGAG	176	RT-PCR and qRT-PCR
AaBFS	TTTAAATGGGTGTCCTCGTATCCT	CTTTCCTTAGAATAGCATTTCGATGC	140	RT-PCR and qRT-PCR
AaAAG24640	GAAGCTGTTTCAGAGATGGTCCTT	CCACTAAGAGGTTTCGACAATTTTC	184	RT-PCR and qRT-PCR
Aa02039	GGGAAGAATCTGTTTTAGATGAAGC	CAAAAATAGCCTTGCCTCTACCAT	161	RT-PCR and qRT-PCR
Aa14765	GGCTTCCAAGACTAAATTCATTGC	GGTGCATTCTTTGGAGCATCA	178	RT-PCR and qRT-PCR
Aa16267	GCAAGGGTTTTCTTAACGAAGGT	TGTTTCATGTAATCTGGAAGCTCGT	155	RT-PCR and qRT-PCR
AaSTS1	GTTGGCATGGGTGATATAGCATC	TTCTTTCTCCTCCTTGACCAAC	129	Clone, RT-PCR and qRT-PCR
AaSTS2	GGGTTCTCACAAATCCACCACT	TGTAGCATTCAACAACAGACACGA	123	Clone, RT-PCR and qRT-PCR
AaBPS	TGGTACTTTCCTGAACCTCGAAC	AATCCTGCCACGCCTTTTTAAGG	194	RT-PCR and qRT-PCR
AaLS	AACAGATGGGATATAAAGGCGAT	CCTTACATAGATCTGCCCATGCT	163	RT-PCR and qRT-PCR
AaMTS1	CAGGCAGTTCACCAAGAAGATCT	CTCAGAAGAGTAAATTGAGGTCGGT	155	Clone, RT-PCR and qRT-PCR
AaMTS2	CCAATAGATGGGATATCAATGCGGT	GTTACATAAATCCACCCACCCTTTC	164	Clone, RT-PCR and qRT-PCR
AaMTS3	TGATGCTACCAATAGATGGGACATT	CCAGTTTGCTTCTACTACAAATGAC	196	Clone, RT-PCR and qRT-PCR
AaSQS	GCTTAAGTCCAAGGTTGACATGC	AGAGCAAGCAAAGCTGCACTGT	159	RT-PCR and qRT-PCR
AaCYP71AV1	GCAAGAGAGCTCGGTTAACTAGC	GTGCCTGCTCCAACATATCCAA	202	RT-PCR and qRT-PCR
AaDBR2	TGGTCGCTTTTGGTCGTTATTTTG	AAACTAGAGGAGTGACCCTTTGTC	164	RT-PCR and qRT-PCR
AaCPR	GCAGGAATCGCAAAGTGGATTTT	ATATCCGAAGCCTTCTGAGTCATC	151	RT-PCR and qRT-PCR
AaALDH1	TTTTGATCTCGCCACTCGTCATG	GCTATGGTCATATCATCCGTAACG	186	RT-PCR and qRT-PCR



**Fig S1.** Semi-quantitative RT-PCR analysis of artemisinin biosynthesis genes from *A. annua*. Expression levels were measured using the primers listed in Table S3. *AaADS* amorpho-4,11-diene synthase (accession no. CAB94691); *AaCYP71AV1* cytochrome P450 monooxygenase (DQ315671); *AaDBR2* double bond reductase (EU704257); *AaCPR* cytochrome P450 reductase (EF197890); *AaALDH1* aldehyde dehydrogenase (FJ809784). R Root; S Stem; OL Old leaf; YL Young leaf; FB Flower bud.