

TCS1a	MELATA	AGKVNEVLFMNRGE	ESS	YAQNS	SFTQQ	VASMA	QPALEN	AVETLFS	RDFHL	QAL	59					
HYC		MGKVNEVLFMNRGE	EITS	YAQNSA	FTQKV	ASMA	PALEN	AVETLFS	KDFHL	LQA	55					
CCT		MGKVNEVLFMNRGE	EITS	YAQNSA	FTQKV	ASMA	PALEN	AVETLFS	KDFHL	LQA	55					
TCS1		MGKVNEVLFMNRGE	EITS	YAQNSA	FTQKV	ASMA	PALEN	AVETLFS	KDFHL	LQA	54					
Consensus		gkvnevlfmnrgege		yaqns	ftq	vasma	palenavetlfs		dfhl							
TCS1a	N	AADLGCAAGPNTFAVISTIKRMEKKCRELN		CQ	LELQ	VYLN	DLFGN	DFNTL	FKGL	SS	118					
HYC	LN	AADLGCAAGPNTFAVISTIKRMEKKCRELY		CQ	LELQ	VYLN	DLFGN	DFNTL	FKGL	SS	115					
CCT	LN	AADLGCAAGPNTFAVISTIKRMEKKCRELY		CQ	LELQ	VYLN	DLFGN	DFNTL	FKGL	SS	115					
SDT	LT	AADLGCAAGPNTFAVISTIKRMEKKCRELY		CQ	LELQ	VYLN	DLFGN	DFNTL	FKGL	SS	114					
Consensus		aadlgcaagpntfavistikrmekkreln		cqtlelq	vylndlfgndfntlfkglss											
TCS1a	EV	IGNKCEEV	PCYVMG	VP	GSF	HGR	LFP	RNS	LHLV	HSSYSV	HWLTQAPKGLTSREGLALNK	178				
HYC	QV	IGNKCEEV	SCYVMG	VP	GSF	HGR	LFP	RNS	LHLV	HSSYSV	HWLTQAPKGLTSREGLALNK	175				
CCT	QV	IGNKCEEV	SCYVMG	VP	GSF	HGR	LFP	RNS	LHLV	HSSYSV	HWLTQAPKGLTSREGLALNK	175				
SDT	EV	IGNKCEEV	SCYVMG	VP	GSF	HGR	LFP	RNS	LHLV	HSSYSV	HWLTQAPKGLTSREGLALNK	174				
Consensus	sv	gnkceev	cyvmgvp	gsf	hgrlfp	rns	lhlvhssysvhwltqapkg	lts	reglaln	k						
TCS1a	GKIYISK	TSPPVV	REAYLSQ	FHEDFT	MFLNARSQ	EVVP	NGCMVLIL	GRQ	CS	DP	SDM	QSC	238			
HYC	GKIYISK	TSPPVV	REAYLSQ	FHEDFT	MFLNARSQ	EVVP	NGCMVLIL	GRQ	SS	DP	SEM	ESC	235			
CCT	GKIYISK	TSPPVV	REAYLSQ	FHEDFT	MFLNARSQ	EVVP	NGCMVLIL	GRQ	SS	DP	SEM	ESC	235			
SDT	GKIYISK	TSPPVV	REAYLSQ	FHEDFT	MFLNARSQ	EVVP	NGCMVLIL	GRQ	SS	DP	SEM	ESC	234			
Consensus	gkiyisk	tsppvv	reaylsq	fhedftmflnarsq	evvpngcmvlil	grq	sdps	m	sc							
TCS1a	FTWELL	AM	IAELVSQ	GLIDED	KLDTFN	PSY	FAS	LE	EVK	DIVERD	GSFTID	HIEGF	DLD	298		
HYC	FTWELL	AM	IAELVSQ	GLIDED	KLDTFN	PSY	WPS	LE	EVK	DIVERD	GSFTID	RLEGF	ELED	295		
CCT	FTWELL	AM	IAELVSQ	GLIDED	KLDTFN	PSY	WPS	LE	EVK	DIVERD	GSFTID	HLEGF	ELED	295		
SDT	STWELL	AM	IAELVSQ	GLIDED	KLDTFN	PSY	WPS	LE	EVK	DIVERD	GSFTID	HLEGF	ELED	294		
Consensus	stwell	am	aiaelvsqglidedkldtfn	psy	sl	evkdiver	dgsftid	egf	ld							
TCS1a	SV	EMQENDK	WVRG	E	KFTK	V	VRAFTE	PIIS	NQFG	PE	IMDKLYDKFTHI	V	VS	DL	EAKLPKTT	358
HYC	S	LEMQENDK	WVRG	D	KFAK	MVRAFTE	PIIS	NQFG	HE	IMDKLYDKFTHI	V	LS	DL	EAE	LPKTT	355
CCT	S	LEMQENDK	WVRG	D	KFAK	MVRAFTE	PIIS	NQFG	HE	IMDKLYDKFTHI	V	LS	DL	EAE	LPKTT	355
SDT	S	LEMQENDK	WVRG	D	KFAK	MVRAFTE	PIIS	NQFG	HE	IMDKLYDKFTHI	V	LS	DL	EAE	LPKTT	354
Consensus		emqendkwvrg	kf	k	vraftepiisnqfg	eimdklydkfthi	sdl	a	lpktt							
TCS1a	SIILV	LSKID														368
HYC	SIILV	LSKIV														365
CCT	SIILV	LSKIV														365
SDT	SIILV	LSKIV														364
Consensus		siilvlski														

Figure S5. Amino acid sequence analysis of caffeine synthase gene. The red box represents conserved motif. Green circle has a critical role in substrate recognition. Red circle represents different amino acid sequences.