Natural Resources, 2022, 13, 65-76
https://www.scirp.org/journal/nr
ISSN Online: 2158-7086
ISSN Print: 2158-706X

Tyrosine Aminotransferase Gene (*SmTAT*) Revealed Genetic Diversity and Phylogeny of Cultivated *Danshen* (*Salvia miltiorrhiza*) Populations

Guanrong Li[1*], Ruihua Ren[1], Deying Kong[2], Jie Feng[1], Yanyan Yin[1], Fang Liao[3*]

[1]College of Agronomy and Biotechnology, Southwest University, Chongqing, China
[2]Technology Center, Chongqing Customs, Chongqing, China
[3]Animal, Plant and Foodstuff Inspection Center, Tianjin Customs, Tianjin, China
Email: *grli1963@126.com, 1298622650@qq.com, 125353759@qq.com, fengjie192@163.com, yinyanyan954550975@163.com, *liaoxyfang@163.com

How to cite this paper: Li, G., Ren, R.H., Kong, D.Y., Feng, J., Yin, Y.Y. and Liao, F. (2022) Tyrosine Aminotransferase Gene (*SmTAT*) Revealed Genetic Diversity and Phylogeny of Cultivated *Danshen* (*Salvia miltiorrhiza*) Populations. *Natural Resources*, **13**, 65-76.
https://doi.org/10.4236/nr.2022.133005

Received: February 7, 2022
Accepted: March 20, 2022
Published: March 23, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

Chinese traditional medicine *Danshen* is the radix of the perennial herbs of *Salvia miltiorrhiza* Bunge, which has a variety of pharmacological effects and is traditionally and extensively applied clinically to treat cardiovascular disorders. In this research, the genomic genes for tyrosine aminotransferase (*TAT*) of 38 cultivated populations of *Danshen* in China were cloned and bioinformatic analyses were conducted to reveal its genetic diversity and phylogeny. The full-length *SmTAT* was 2296 - 2444 bp including 6 exons (encoding 411 amino acids) and 5 introns. Overall, the *SmTAT* genes in cultivated *Danshen* populations are highly conserved with a relative low level of genetic diversity. The spliced exons (1236 bp) had 23 SNP variations with a rate of 1.86%, of which 22 occurred in the white flower *S. miltiorrhiza* Bge.f.alba population (W-SCHY-W-1) and led to 5 amino acid variations. The entire 290 SNP variations with a rate of 24% in the 5 introns occurred exclusively in W-SCHY-W-1. Phylogenetic trees based on the full-length, combined introns, the spliced exons, and the deduced amino acid sequences of *SmTAT* all showed a two-clade basic structure with W-SCHY-W-1 uniquely standing alone. The *SmTAT* gene of the white flower population (W-SCHY-W-1) is unique and especially rich in variations. The first time clarified genomic *SmTAT* gene structure and genetic diversity in cultivated *Danshen* populations laid an excellent foundation for further studies on the biosynthesis of bioactives and the molecular breeding of *Danshen* as well as in plant tyrosine metabolism.

Keywords

Danshen (*Salvia miltiorrhiza* Bge), Cultivated Population, Tyrosine Aminotransferase Gene (*TAT*), Genetic Diversity, Phylogeny

1. Introduction

Chinese traditional herbal medicine, *Danshen*, is the radix of perennial herbs of *Salvia miltiorrhiza* Bunge of the family Labiatae. It has been traditionally and extensively used in clinical practice to treat various ailments such as cardiovascular, cerebrovascular, hyperlipidemia, and acute ischemic stroke diseases [1] [2] [3] [4] [5].

Danshen owns abundant germplasm resources and many cultivated populations in China. In recent years, with the increase in market demand, the often-chaotic introduction of varieties in field cultivation and the nonstandard field management cause some confusion and result in poor quality of the herbal medicine. Extensive researches have been conducted on its cultivation, germplasm resources protection, and molecular identification.

However, research progress in the functional genes for pharmacologically active constituents has been slow. There are two major classes of pharmacologically effective components in *Danshen*, the hydrophilic salvianolic acids, and the lipophilic tanshinones [6]. So far, the key genes for the biosynthesis of effective components in *S. miltiorrhiza*, of 4-hydroxycinnamate coenzyme A ligase [6], 4-hydroxyphenylpyruvate reductase [7], cinnamic acid 4-hydroxylase [8], tyrosine aminotransferase [9], 3-hydroxy-3-methylglutaryl coenzyme A reductase [10], and phenylalanine ammonia-lyase genes [11], have been characterized and studied respectively.

The biosynthetic pathway of rosmarinic acid consists of two parallel phenylalanine and tyrosine branches [12]. Tyrosine aminotransferase (*TAT*) (EC 2.6.1.5) is the rate-limiting step in the tyrosine branch, which catalyzes the formation of 4-hydroxyphenylpyruvate from tyrosine and ultimately the biosynthesis of salvianolic acids. Various other structurally diverse natural compounds are also derived from the tyrosine metabolic pathway, among which tocopherols, plastoquinone, and ubiquinone are essential to plant survival [13].

So far, our knowledge of the plant tyrosine metabolism pathway remains rudimentary, and genes encoding the pathway enzymes have not been fully defined, despite that the tyrosine aminotransferase genes have been cloned in a few other plants such as *Coleus blumei* (AJ458993), *Arabidopsis thaliana* [14], *Glycine max* (AAY21813), *Medicago truncatula* (DQ006809) as well as *S. miltiorrhiza*, and functionally studied by overexpression of single gene and coexpression of several in *S. miltiorrhiza* hairy root cultures [15]. The structure and the genetic diversity of tyrosine aminotransferase genes among the various cultivated populations of *S. miltiorrhiza* are still unknown.

In this research, the genomic tyrosine aminotransferase genes of the 38 cultivated populations of *S. miltiorrhiza* from the major cultivation regions of China, were for the first time cloned by walking technology, and its sequences were analyzed bioinformatically to understand the genomic structure, genetic diversity, and phylogeny of the cultivated *S. miltiorrhiza* populations.

2. Materials and Methods

Plant materials

Seeds of 38 cultivated *Danshen* populations were collected from three major seed industries representing more than 30 regions of China (Table 1); uniform seeds preliminarily selected according to size, color and shape were used for sowing in Southwest University Agricultural Station, Chongqing; and morphologically representative single plants of each population were used for extraction of genomic DNAs.

Primer design

Two pairs of primers for cloning of the genomic *SmTAT* genes of cultivated

Table 1. The cultivated *S. miltiorrhiza* populations used in this study.

No.	Production Region	Code	Source	No.	Production region	Code	Source
1	Changchun, Jilin	V-JLCC-V-1	HDSI	20	Guangdong	W-GD-V-2	TDSI
2	Zunyi, Guizhou	V-GZZY-V-1	HDSI	21	Jiangsu	W-JS-V-2	TDSI
3	Shuyang Jiangsu	V-JSSY-V-1	HDSI	22	Yantai, Shandong	S-SDYT-V-1	HDSI
4	Lijiang, Yunnan	V-YNLJ-V-1	HDSI	23	Fangcheng, Henan	S-HNFC-V-1	HDSI
5	Guangdong	V-GD-V-1	HDSI	24	Changsha, Hunan	S-HNCS-V-1	HDSI
6	Chongqing	V-CQ-V-1	HDSI	25	Juxian, Shandong	S-SDJX-V-1	HDSI
7	Shandong	V-SD-V-2	TDSI	26	Jiuquan, Gansu	S-GSJQ-V-1	HDSI
8	Guizhou	V-GZ-V-2	TDSI	27	Nemeng	S-NM-V-1	HDSI
9	Jiangsu	V-JS-V-2	TDSI	28	Guangxi	S-GX-V-1	HDSI
10	Beijing	V-BJ-V-V-1	FHSI	29	Anguo, Hebei	S-HBAG-V-1	FHSI
11	Anguo, Hebei	V-HBAG-V-1	FHSI	30	Quanjiao, Anhui	B-AHQJ-V-1	HDSI
12	Longxi, Gansu	V-GSLX-V-1	FHSI	31	Zhongjiang, Sichuan	B-SCZJ-V-1	HDSI
13	Jingmen Hubei	W-HBJM-V-1	HDSI	32	Shandong	B-SD-V-2	TDSI
14	Xi'an, Shaanxi	W-SXXA-V-1	HDSI	33	Sichuan	B-SC-V-2	TDSI
15	Shenyang, Liaoning	W-LNSY-V-1	HDSI	34	Jiangsu	B-JS-V-2	TDSI
16	Luoyuan, Fujian	W-FJLY-V-1	HDSI	35	Guangdong	B-GD-V-2	TDSI
17	Shandong	W-SD-V-2	TDSI	36	Mediteranean	CYSWC-DZH-V-4	FHSI
18	Sichuan	W-SC-V-2	TDSI	37	Mediteranean	Q-DZH-V-4	FHSI
19	Guizhou	W-GZ-V-2	TDSI	38	Hongyuan, Sichuan	W-SCHY-W-1	SC

Notes: HDSI—Hengda Seed Industry, China; TDSI—Tongda Seed Industry, China; FHSI—Fenghong Seed Industry, China; SC—Self-Collected.

Table 2. Primers designed for the cloning of the genomic *SmTAT* genes of the cultivated populations.

Primer code	Sequence (5′→3′)	Length (nt)	Annealing Temp (Ta)	Reference Accession	Position in Reference
TAT-FP1	TTCCGTGTGAATGCTCTATG	21	55	EF192320.1	926
TAT-RP1	AGGAAACGAACTTAGCCAGA	21			1670
TAT-FP2	GAAGGAGAGCGGGAAGAGAGT	20	60		1364
TAT-RP2	GAGTGCCGTTCACAGAAAGAC	21			3630

S. miltiorrhiza populations were designed based on the reference accession (EF192320.1) with Primer 5 (Table 2).

Genomic DNA extraction

Total leaf genomic DNAs were extracted with Qiagen DNeasy Plant Mini Kit (Multi Sciences, Hangzhou, China) according to the manufacturer's instructions. The purity was assessed by agarose gel electrophoresis followed by Goldview staining, and the quantity was determined spectrophotometrically by Shimadzu UV mini-1240. Purified genomic DNAs were dissolved in 10 mmol/L Tris-HCl buffer and stored at −70˚C.

PCR amplification

About 1.0μg genomic DNA templates were amplified with primer pairs TAT-FP1/RP1 and TAT-FP2/RP2 respectively in a reaction mixture of 50 μL: 1.1 × T3 Super PCR Mix 36.0 - 44 μL, 10 μmol/L primers each 2.0 μL (final concentration 0.4 μM) in Biometra TGRADIENT thermocycler (Biometra GmbH, Germany) with the programme: initial-denaturation at 98˚C for 3 min followed by 35 cycles of denaturation at 98˚C for 10 s, annealing at 55˚C/60˚C for 10 s and elongation at 72˚C for 5 - 15 s, and a final extension at 72˚C for 2 min.

Amplified products were electrophoresed in 1% agarose gel and visualized with Goldview stain. And after recovery and purification, they were bidirectionally sequenced by dideoxy chain termination with ABI Prism 310 Genetic Analyzer (Applied Biosystems, Foster City, USA) and manually spliced and checked.

Sequence data processing

The BLAST confirmed two segments of the *SmTAT* gene sequences of the 38 cultivated populations of *S. miltiorrhiza* were spliced with Vector NTI Advance11. The spliced whole sequences were manually checked to ensure the quality of sequences and BLAST analyzed to confirm the gene of interest. The spliced whole *SmTAT* gene sequences were deposited in GenBank. Sequences were aligned with Vector NTI Advance11 to identify the nucleotide variation sites. Phylogenetic trees were constructed with MEGA X based on Neighbor-Joining (NJ) with a bootstrap value of 1000.

3. Results and Analyses

Structural features of the SmTAT genes of the cultivated S. miltiorrhiza populations

The *SmTAT* genes of the 38 cultivated populations were successfully ampli-

fied by the walking primers designed (TAT-FP1/RP1 and TAT-FP2/RP2). Agarose gel electrophoresis showed distinct single bands of about 700 bp and 2000 bp respectively with the two primer pairs.

All the obtained *SmTAT* gene sequences of the 38 cultivated populations of *S. miltiorrhiza* were BLAST confirmed and submitted to GenBank (Accession numbers shown in Table A1).

The sequence of accession EF192320.1 was used as the reference to demarcate the *SmTAT* gene sequences. Results showed that the full-length of genomic *SmTAT* gene was 2296 - 2444 bp, consisting of 6 exons with a total length of 1236 bp among all the tested populations, and 5 introns with a total length of 1060 bp for 37 populations except for W-SCHY-W-1, whose total intron length increased to 1208 bp, most of the increased 148bp distributed in introns 1 - 4; the corresponding exons or introns are equal in size for the majority (37 populations) except for W-SCHY-W-1. The total length of the *SmTAT* gene for the majority (37 populations) was 2296 bp, while for W-SCHY-W-1, 2444 bp (Table 3).

Nucleotide variations in the introns of the SmTAT genes

The 5 intron sequences of *SmTAT* of the 38 cultivated populations were aligned and compared. Results showed that there were 290 nucleotide variations with a rate of 24%, and all occurred in the population W-SCHY-W-1(Table 4). Statistics showed that there were 110 variations (11conversions, 21 transversions, 4 deletions and 74 insertions) in intron 1, 13 variations (4 transversions and 9 insertions) in intron 2, 120 variations (19 conversions, 21 transversions, 18 deletions and 62 insertions) in intron 3, 34 variations (2 conversions, 2 transversions, 30 insertions) in intron 4, and 13 variations (2 transversions, 2 transversions, 7 deletions and 2 insertions) in intron 5, accounting for variation rates of 26.3%, 15.7%, 30.9%, 45.3% and 13.5% respectively (Table 4).

Nucleotide variation in exons of the SmTAT genes

Alignment of the spliced 6 exon sequences of all the 38 cultivated *S. miltiorrhiza* populations showed that there were 23 nucleotide variation sites, of a variation rate of 1.86%, among which 12 conversions, 11 transversions. Most variations (22) occurred in population W-SCHY-W-1 and were distributed mainly in exons 2-4 (Table 5).

Amino acid variations in the deduced amino acid sequences of SmTAT

The spliced exon sequences of the *SmTAT* gene were 1236 bp in length with a complete reading frame of 1233 bp encoding 411 amino acid residues. The deduced amino acid sequences are highly conserved. All showed the aminotransferases family-I pyridoxal-phosphate attachment site (SLSKRWLVPGWRLG)

Table 3. Structure of the genomic *SmTAT* genes of the cultivated *S. miltiorrhiza* populations.

Population	E1	E2	E3	E4	E5	E6	E_{total}	I1	I2	I3	I4	I5	I_{total}	$SmTAT_{total}$
W-SCHY-W-1	254	340	219	220	90	113	1236	488	92	432	105	91	1208	2444
Remaining 37	254	340	219	220	90	113	1236	418	83	388	75	96	1060	2296

Note: E and I for exons and introns respectively.

Table 4. Nucleotide variation sites in introns of *SmTAT* genes of *S. miltiorrhiza* populations.

Site in intron	I1	22	25	26	27	33	41	51	58	68	92	101	108	130	131	132	133	134	135	136	137	138	145	147	161	180	197	215	218	239	248
Population																															
W-SCHY-W-1		**C**	**C**	**T**	**T**	**A**	**A**	**C**	**G**	**T**	**G**	**C**	**A**	**G**	**T**	**A**	**G**	**A**	**G**	**G**	**A**	**T**	**G**	**A**	**A**	**C**	**T**	**A**	**C**	-	-
Remaining 37		G	-	-	-	T	T	A	T	G	C	T	G	-	-	-	-	-	-	-	-	-	A	G	G	G	G	G	T	T	T
		249	258	268	280	281	282	283	284	285	286	287	288	289	290	292	292	293	294	302	303	310	317	332	337	339	355	356	357	358	359
W-SCHY-W-1		-	-	**G**	**G**	**G**	**A**	**G**	**C**	**C**	**T**	**C**	**G**	**C**	**G**	**A**	**T**	**A**	**G**	**A**	**T**	**T**	**G**	**A**	**T**	**C**	**G**	**T**	**A**	**G**	**A**
Remaining 37		T	T	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	A	A	T	G	C	G	-	-	-	-	-
		360	361	362	363	372	381	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	424	426	434	437	442
W-SCHY-W-1		**T**	**G**	**A**	**C**	**C**	**A**	**G**	**G**	**G**	**A**	**T**	**T**	**G**	**C**	**G**	**C**	**T**	**C**	**T**	**A**	**C**	**G**	**T**	**G**	**T**	**T**	**C**	**T**	**C**	**T**
Remaining 37		-	-	-	-	T	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	G	A	G	-
		443	444	445	446	447	448	449	455	456	470	471	472	473	474	475	476	477	478	482	485	I2	13	15	16	17	18	19	20	21	22
W-SCHY-W-1		**A**	**A**	**A**	**C**	**A**	**T**	**C**	**T**	**C**	**T**	**G**	**A**	**T**	**G**	**C**	**A**	**T**	**C**	**T**	**C**		**A**	**G**	**A**	**T**	**T**	**T**	**G**	**C**	**T**
Remaining 37		-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	G		C	-	-	-	-	-	-	-	-
		23	51	80	8 1		14	15	16	17	18	19	20	21	25	26	27	28	29	30	31	32	33	34	35	36	37	38	41	66	75
W-SCHY-W-1		**C**	**T**	**G**	**T**	I3	**T**	**C**	**A**	**T**	**A**	**T**	**C**	**T**	**A**	**T**	**G**	**T**	**A**	**T**	**T**	**A**	**A**	**A**	**T**	**T**	**T**	**C**	**A**	**A**	**C**
Remaining 37		-	G	C	A		-	-	-	-	-	-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	T
		76	77	78	84	95	99	100	101	102	103	104	105	106	107	108	109	110	111	112	117	150	170	184	186	188	191	196	205	206	207
W-SCHY-W-1		**T**	**G**	**G**	**G**	**T**	**T**	**T**	**A**	**T**	**C**	**A**	**C**	**A**	**G**	**A**	**G**	**T**	**C**	**C**	**A**	**G**	**G**	-	**A**	**G**	**G**	**T**	**A**	**A**	**T**
Remaining 37		C	C	C	C	C	-	-	-	-	-	-	-	-	-	-	-	-	-	-	C	C	T	T	C	C	A	C	-	-	-
		209	227	228	229	230	231	232	233	241	250	260	264	266	268	284	285	289	293	305	311	322	323	324	325	326	327	328	329	330	331
W-SCHY-W-1		**G**	**G**	**G**	**T**	**G**	**A**	**T**	**G**	-	**G**	**C**	**A**	**A**	**G**	**A**	**G**	**G**	**A**	**C**	**T**	**T**	**A**	**T**	**T**	**G**	**A**	**T**	**A**	**T**	**C**
Remaining 37		C	C	-	-	-	-	-	-	T	A	T	G	G	A	-	-	G	G	T	C	-	-	-	-	-	-	-	-	-	-
		332	333	334	339	349	356	357	358	364	374	376	377	379	380	381	382	384	386	388	389	392	393	398	401	402	403	404	405	406	407
W-SCHY-W-1		**C**	**G**	**G**	**G**	**G**	**T**	**C**	**A**	**A**	**C**	-	-	-	-	-	-	-	**T**	**A**	**T**	**G**	**A**	**T**	-	-	-	-	-	-	-
Remaining 37		-	-	-	A	C	-	-	-	T	G	C	T	A	G	T	A	A	A	T	G	T	G	A	A	T	G	A	A	A	T
		425	438	439	440	441		33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	53	65	66	67	68
W-SCHY-W-1		**C**	**A**	**T**	-	-	I4	**T**	**G**	**C**	**A**	**C**	**A**	**C**	**A**	**C**	**G**	**C**	**G**	**C**	**T**	**G**	**C**	**C**	**T**	**G**	**T**	**T**	**A**	**C**	**G**	**G**
Remaining 37		G	T	C	G	A		G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-
		69	70	71	72	73	74	75	76	80	86		16	20	30	31	45	46	47	48	49	50	62	75	83						
W-SCHY-W-1		**T**	**C**	**A**	**T**	**T**	**T**	**T**	**T**	**G**	**A**	I5	**G**	**C**	**A**	**A**	-	-	-	-	-	-	**A**	-	**G**						
Remaining 37		-	-	-	-	-	-	-	-	A	G		A	A	-	-	A	T	A	T	T	C	G	A	T						

Notes: nucleotide variation sites are shown in bold; "-" represent deletions.

Table 5. Nucleotide variation sites in exons of *SmTAT* sequences of *S. miltiorrhiza* populations.

	E1	E2									E3						E4				E5	E6	
Population	249	273	288	306	365	423	484	526	592	593	607	610	646	673	712	778	820	844	847	897	1090	1129	1177
W-SCHY-W-1	**T**	**G**	**C**	**C**	**C**	**C**	**G**	**A**	**G**	**G**	**A**	**G**	**C**	**T**	**G**	**C**	**G**	**C**	**C**	**G**	**A**	**T**	A
13 populations*	C	A	G	G	G	T	A	C	C	A	G	C	G	C	C	T	T	T	A	A	G	C	**G**
Remaining 24	C	A	G	G	G	T	A	C	C	A	G	C	G	C	C	T	T	T	A	A	G	C	A

Notes: Nucleotide variation sites are shown in bold; *: V-JS-V-2, V-HBAG-V-1, B-SC-V-2, S-HNCS-V-1, W-SXXA-V-1, Q-DZH-V-4, V-GZZY-V-1, CYSWC-DZH-V-4, V-JSSY-V-1, B-GD-V-2, S-HBAG-V-1, V-GZ-V-2 and W-GZ-V-2.

and an Arg385 that fixes the α-carboxylate of the incoming amino acid or α-ketoacid [9]. Comparison of the deduced amino acid sequences of the *SmTAT* of the 38 cultivated *S. miltiorrhiza* populations revealed that 37 populations were in consensus and 5 variations (E102D, G122A, N197K, K198E, and E299G) all occurred in population W-SCHY-W-1. Among them, N197K, K198E, and E299G were probably the most significant variations in TAT of population W-SCHY-W-1 from the remaining 37 populations.

Phylogenetic analyses of SmTAT gene and its deduced amino acid sequences

Phylogenetic trees based on the full-length *SmTAT* genomic sequences and the spliced 6 exons were nearly identical. Population W-SCHY-W-1 stands alone, while the other 37 populations cluster in another clade, which is branched into 2 subclades, one consisting of 13 populations (V-JS-V-2, V-HBAG-V-1, B-SC-V-2, S-HNCS-V-1, W-SXXA-V-1, Q-DZH-V-4, V-GZZY-V-1, CYSWC-DZH-V-4, V-JSSY-V-1, B-GD-V-2, S-HBAG-V-1, V-GZ-V-2, and W-GZ-V-2), and the other, of 24 populations (**Figure 1(a)**, **Figure 1(b)**).

Phylogeny analyses on the combined introns sequences (from intron 1 to 5) and the deduced amino acid sequences of *SmTAT* showed that population W-SCHY-W-1 stands alone and all the 37 populations cluster in one clade (**Figure 1(c)**, **Figure 1(d)**).

Basically, a two-clade topological structure was demonstrated in the phylogenetic trees based on the full-length *SmTAT*, combined introns, the spliced exons, and the deduced amino acid sequences. Population W-SCHY-W-1 is uniquely standing alone in all the phylogenetic trees.

4. Discussion

The spliced exon sequences of the *SmTAT* gene were 1236 bp in length with an ORF of 1233 bp encoding 411 amino acid residues, consistent with the similar report [9]. The deduced amino acid sequences of SmTAT are highly conserved. All showed the aminotransferases family-I pyridoxal-phosphate attachment site and an Arg385 that fixes the α-carboxylate of the incoming amino acid or α-ketoacid. The high conservation and overall low level of diversity of *SmTAT* genes demonstrated in this research suggests the necessity to further conserve its wild resources and to identify novel genetic resources materials as well as to

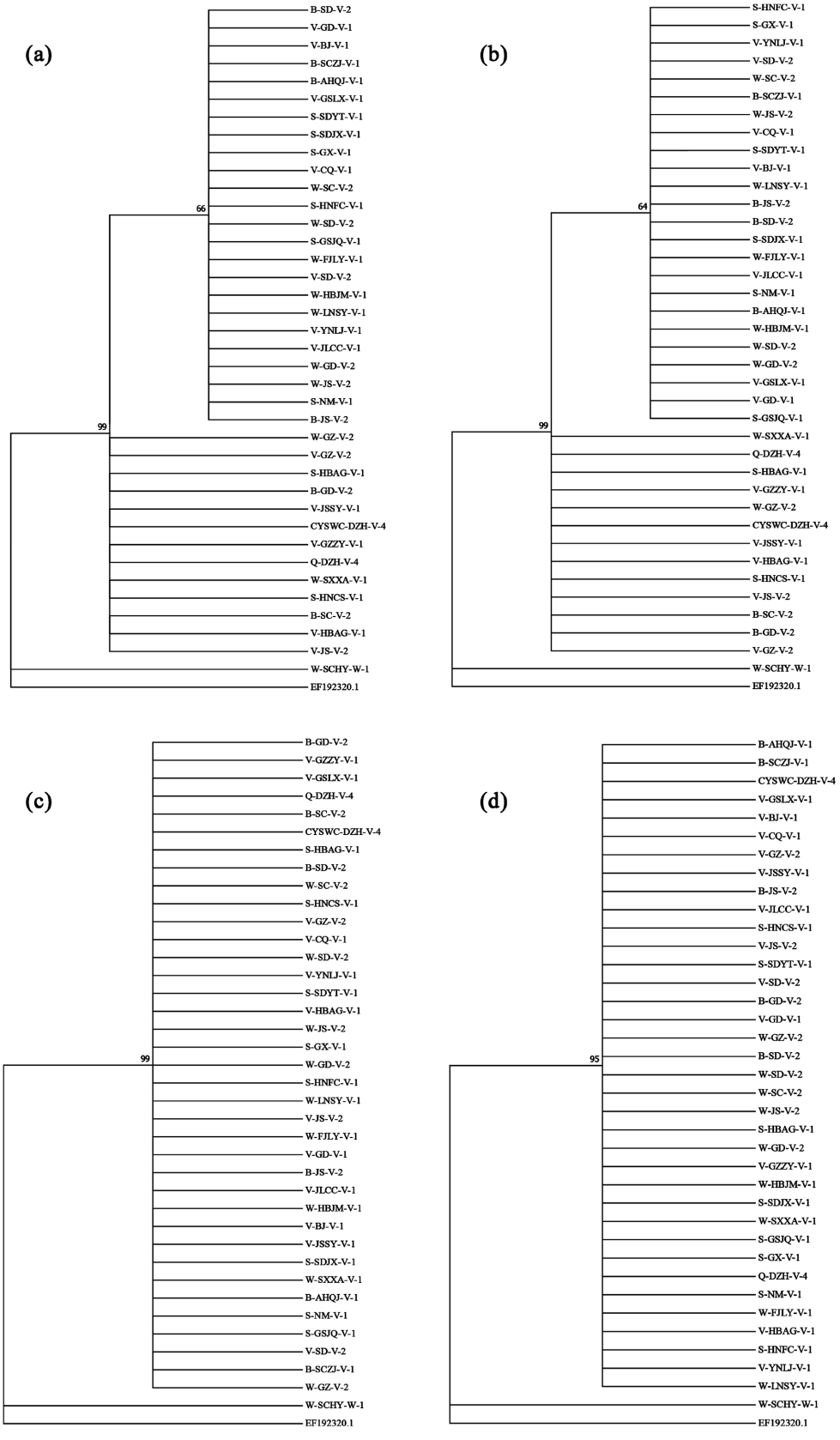

Figure 1. Phylogenetic trees based on the full-length (a), the spliced exons (b), combined introns (c), and the deduced amino acid sequences (d) of *SmTAT*.

accelerate the breeding of this important Chinese medicinal plants.

Comparison of the deduced amino acid sequences of the *SmTAT* gene of the 38 the cultivated *S. miltiorrhiza* populations revealed that 37 populations are identical and all the 5 variations occurred in population W-SCHY-W-1, of which N197K, K198E, and E299G are probably the most significant variations in TAT of population W-SCHY-W-1.

The white flower *S. miltiorrhiza* Bge.f.alba is a varietae or forma of *S. miltiorrhiza* Bge. Usually, the flower color of *S. miltiorrhiza* Bge is purple, while that of *S. miltiorrhiza* Bge.f.alba is white. There have been comparative reports of the medicinal value of the purple and white flower *Danshen*. One report showed two more bioactive ingredients in *S. miltiorrhiza* Bge.f.alba [16]. Another found that the contents of some trace element in white flower *Danshen* were higher than those in purple flower *Danshen* [17]. Still another showed that the phenolic acids contents in white flower *Danshen* were about two times higher than those in purple flower *Danshen* [18]. Also found is that most parts of *S. miltiorrhiza* Bge. f. alba plant had higher contents of bioactives than *S. miltiorrhiza* Bge [19]. The white flower *S. miltiorrhiza* Bge. f. alba had special pharmacological effect for treatment of thromboangiitisobiterans [19]. The crude drug of *S. miltiorrhiza* Bge.f. alba was found to increase cerebral blood flow significantly, reduce neuronal apoptosis, and promote neuronal regeneration in rats with cerebral ischemia/reperfusion impairment [20].

The rare white flower *S. Miltiorrhiza* Bge.f.alba is generally valued more over the purple flower majority *S. Miltiorrhiza* Bunge. The extensive variations in the genomic *SmTAT* gene of the white flower *Danshen* population W-SCHY-W-1 found in this research probably is an important genetic marker for white flower *Danshen*. Further elucidation of the structures and functions of *SmTAT* and other functional genes involved in tyrosine metabolic pathway in *S. miltiorrhiza* Bge.f.alba would be very beneficial in understanding its special pharmaceutical effects and for the acceleration of its breeding.

5. Conclusions

The successfully walking technologically cloned full-length genomic *SmTAT* was about 2296 bp and consisted of 6 exons and 5 introns. The spliced exon sequence of SmTAT gene was 1236 bp in length, encoding a complete reading frame of 411 amino acids. All the 23 SNP variation sites (1.86%) occurred in the white flower W-SCHY-W-1 population. The only 5 amino acid variations were located in population W-SCHY-W-1.

The 5 introns of *SmTAT* had 290 SNP variation sites, which were located in W-SCHY-W-1only, with a variation rate of 24% far greater than that in the spliced exons, indicating the faster evolution of the introns. Phylogenetic trees based on the full-length genomic *SmTAT*, the spliced exons, combined introns, and the deduced amino acid sequences all showed a two-clade structure with population W-SCHY-W-1 standing alone, which represented a special popula-

tion as regard to the TAT gene. The uniquely extensive variation in the genomic *SmTAT* gene of W-SCHY-W-1 is probably an important genetic marker for the white flower *Dansen* and a valuable molecular breeding target.

Further comparative and functional studies on *SmTAT* in relation to both the elucidation of plant tyrosine metabolism and the biosynthesis of the pharmacological ingredients in the currently cultivated *S. miltiorrhiza* Bge populations especially *S. miltiorrhiza* Bge.f.alba would be very revealing and valuable in *SmTAT* based molecular breeding.

Data Availability Statement

All the *SmTAT* sequence data of the 38 cultivated populations of *Danshen* has been submitted in GenBank and accession numbers has been provided in **Table A1**.

Declaration of Funding

This work was funded by National Key R & D Program of China (Project/Subject SN 2017YFF0205305 and 2017YFF0210303) of the Ministry of Science and Technology of the People's Republic of China.

Ethical Statement

This research did not involve any animal or human participants.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Zhou, L.M., Zuo, Z. and Chow, M.S.S. (2005) *Danshen*: An Overview of Its Chemistry, Pharmacology, Pharmacokinetics and Clinic Use. *The Journal of Clinic Pharmacology*, **45**, 1345-1359. https://doi.org/10.1177/0091270005282630

[2] Jiao, P., Chang, Q., Yang, M.F., Zhang, F., Zhao, X.M. and Xia, Z.L. (2007) Effects of *Salvia miltiorrhiza* Bge.f.alba Root Preparation on the Proliferation and Apoptosis of Gastric Cancer Cells. *World Journal of Gastroenterology*, **15**, 820-823.

[3] Chen, F.Y., Guo, R. and Zhang, B.K. (2015) [Advances in Cardiovascular Effects of Tanshinone II (A)]. *China Journal of Chinese Materia Medica*, **40**, 1649-1653.

[4] Maione, F. and Mascolo, N. (2016) *Danshen* and the Cardiovascular System: New Advances for an Old Remedy. *Seminars in Thrombosis and Hemostasis*, **42**, 321-322. https://doi.org/10.1055/s-0036-1580086

[5] Cao, H.M., Wu, J., Jia, L.Q. and Yang, G.L. (2017) Progress in Study of Tanshinone ⅡA's Pharmacological Effects on Cardiovascular System. *World Chinese Medicine*, **12**, 1718-1722.

[6] Zhao, S.J., Hu, Z.B., Liu, D. and Leung, F.C.C. (2006) Two Divergent Members of 4-Coumarate: Coenzyme A Ligase from *S. miltiorrhiza* Bunge: cDNA Cloning and Functional Study. *Journal of Integrative Plant Biology*, **48**, 1355-1364. https://doi.org/10.1111/j.1744-7909.2006.00302.x

[7] Xiao, Y., Di, P., Chen, J.F., Liu, Y., Chen, W.S. and Zhang, L. (2009) Characteriza-

tion and Expression Profiling of 4-Hydroxyphenylpyruvate Dioxygenase Gene (Smhppd) from *Salvia miltiorrhiza* Hairy Root Cultures. *Molecular Biology Reports*, **36**, 2019-2029. https://doi.org/10.1007/s11033-008-9413-2

[8] Huang, B., Duan, Y., Yi, B., Sun, L., Lu, B., Yu, X., Sun, H., Zhang, H. and Chen, W. (2008) Characterization and Expression Profiling of Cinnamate 4-Hydroxylase Gene from *S. miltiorrhiza* in Rosmarinic Acid Biosynthesis Pathway. *Russian Journal of Plant Physiology*, **55**, Article No. 390. https://doi.org/10.1134/S1021443708030163

[9] Huang, B., Yi, B., Duan, Y., Sun, L., Yu, X., Guo, J. and Chen, W. (2008) Characterization and Expression Profiling of Tyrosine Aminotransferase Gene from *S. miltiorrhiza* (Dan-shen) in Rosmarinic Acid Biosynthesis Pathway. *Molecular Biology Reports*, **35**, 601-612. https://doi.org/10.1007/s11033-007-9130-2

[10] Liao, P., Zhou, W., Zhang, L., Wang, J., Yan, X.M., Zhang, Y., Zhang, R., Li, L., Zhou, G.Y. and Kai, G.Y. (2009) Molecular Cloning, Characterization and Expression Analysis of a New Gene Encoding 3-Hydroxy-3-Methylglutaryl Coenzyme A Reductase from *S. miltiorrhiza*. *Acta Physiologiae Plantarum*, **31**, 565-572. https://doi.org/10.1007/s11738-008-0266-z

[11] Song, J. and Wang, Z. (2009) Molecular Cloning, Expression and Characterization of a Phenylalanine Ammonia-Lyase Gene (*SmPAL*1) from *S. miltiorrhiza*. *Molecular Biology Reports*, **36**, Article No. 939. https://doi.org/10.1007/s11033-008-9266-8

[12] Trócsányi, E., György, Z. and Zámboriné-Németh, É. (2020) New Insights into Rosmarinic Acid Biosynthesis Based on Molecular Studies. *Current Plant Biology*, **23**, Article ID: 100162. https://doi.org/10.1016/j.cpb.2020.100162

[13] Xu, J.J., Fang, X., Li, C.Y., Yang, L. and Chen, X.Y. (2019) General and Specialized Tyrosine Metabolism Pathways in Plants. *aBIOTECH*, 1, 97-105. https://doi.org/10.1007/s42994-019-00006-w

[14] Nakamura, Y., Sato, S., Kaneko, T., Kotani, H., Asamizu, E., Miyajima, N. and Tabata, S. (1997) Structural Analysis of *Arabidopsis thaliana* Chromosome 5. III. Sequence Features of the Regions of 1,191,918 bp Covered by Seventeen Physically Assigned P1 Clones. *DNA Research*, **4**, 401-414. https://doi.org/10.1093/dnares/4.6.401

[15] Xiao, Y., Zhang, L., Gao, S.H., Saechao, S.K., Di, P., Chen, J.F. and Chen, W.S. (2011) The *c4h*, *tat*, *hppr* and *hppd* Genes Prompted Engineering of Rosmarinic Acid Biosynthetic Pathway in *S. miltiorrhiza* Hairy Root Cultures. *PLoS ONE*, **6**, e29713. https://doi.org/10.1371/journal.pone.0029713

[16] Li, Z.T., Yang, B.J. and Ma, G.E. (1991) Chemical Studies of *Salvia miltiorrhiza* f. alba. *Acta Pharmaceutica Sinica*, **26**, 209-213.

[17] Qi, Y.X., Yang, Z.X., Li, K., Cao, M.L. and Gao, Y.S. (2004) Determination and Comparison of Trace Elements in *Salvia miltiorrhiza* Bge.f.alba and *Salvia miltiorrhiza* Bge. *Journal Taishan Medical College*, **25**, 589-590.

[18] Hao, G.P., Bian, G.P. and Zhang, Y.Y. (2006) Isolation of High-Quality Genomic DNA from Dried Leaves of Taishan *S. miltiorrhiza* Bge.f.alba. *Chinese Traditional Herbal Drugs*, **30**, 855-857.

[19] Hang, L., Wang, J.R., Yang, D.F., Shu, Z.M. and Liang, Z.S. (2008) Distribution Traits of Bioactives in different Parts of *Salvia miltiorrhiza* Bunge. and *Salvia miltiorrhiza* Bunge.f.alba. *Journal of Northwest A & F University*, **36**, 217-222.

[20] Zhang, Q., Sun, Y., Bo, B. and Hao, H. (2010) Effects of *Salvia miltiorrhiza* Bge.f. alba on Neuronal Regeneration following Cerebral Ischemia/Reperfusion. *Neural Regeneration Research*, **5**, 1066-1077.

Appendix

Table A1. GenBank accession numbers for the genomic *SmTAT* genes of the 38 cultivated populations of *S. miltiorrhiza.*

No.	Population code	GenBank Accession	No.	Population code	GenBank Accession
1	V-JLCC-V-1	MW674906	20	W-GD-V-2	MW674885
2	V-GZZY-V-1	MW674905	21	W-JS-V-2	MW674886
3	V-JSSY-V-1	MW674902	22	S-SDYT-V-1	MW674890
4	V-YNLJ-V-1	MW674904	23	S-HNFC-V-1	MW674891
5	V-GD-V-1	MW674909	24	S-HNCS-V-1	MW674892
6	V-CQ-V-1	MW674910	25	S-SDJX-V-1	MW674916
7	V-SD-V-2	MW674908	26	S-GSJQ-V-1	MW674912
8	V-GZ-V-2	MW674901	27	S-NM-V-1	MW674895
9	V-JS-V-2	MW674900	28	S-GX-V-1	MW674894
10	V-BJ-V-1	MW674903	29	S-HBAG-V-1	MW674893
11	V-HBAG-V-1	MW674907	30	B-AHQJ-V-1	MW674911
12	V-GSLX-V-1	MW674914	31	B-SCZJ-V-1	MW674915
13	W-HBJM-V-1	MW674881	32	B-SD-V-2	MW674898
14	W-SXXA-V-1	MW674882	33	B-SC-V-2	MW674896
15	W-LNSY-V-1	MW674884	34	B-JS-V-2	MW674897
16	W-FJLY-V-1	MW674883	35	B-GD-V-2	MW674899
17	W-SD-V-2	MW674887	36	CYSWC-DZH-V-4	MW674913
18	W-SC-V-2	MW674888	37	Q-DZH-V-4	MW674917
19	W-GZ-V-2	MW674889	38	W-SCHY-W-1	MW773118