GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

Development and characterization of single nucleotide polymorphism markers in *Camellia sinensis* (Theaceae)

C.C. Zhang, L.Y. Wang, K. Wei and H. Cheng

Tea Research Institute of the Chinese Academy of Agricultural Sciences, National Center for Tea Improvement, Hangzhou, China

Corresponding author: H. Cheng
E-mail: chenghao@mail.tricaas.com

Genet. Mol. Res. 13 (3): 5822-5831 (2014)
Received February 21, 2013
Accepted September 2, 2013
Published April 14, 2014
DOI http://dx.doi.org/10.4238/2014.April.14.10

ABSTRACT. Tea is the second most popular non-alcoholic beverage in the world. In recent years, several molecular markers have been used in genetic studies of the tea plant. Yet, only a few single nucleotide polymorphisms (SNPs) have been reported. Here, we identified 818 putative SNPs from expressed sequence tag (EST) databases for the tea plant, which produced a frequency of 1 SNP/170 bp. A direct sequencing method was then used to verify 253 putative SNPs in genome DNA of 17 tea varieties. Fifty (20%) candidate and 299 new SNPs were identified. The haplotype polymorphism and nucleotide diversity of these markers ranged from 0 to 0.960 and 0 to 1.797%, respectively. Using these SNPs, the 17 varieties were classified into 2 groups by cluster analysis. The results indicate that *Camellia sinensis*-derived ESTs provide a valuable resource for SNP discovery. Furthermore, the abundance of SNPs in tea varieties is anticipated to generate the development of associated genetic studies, in addition to enhancing tea plant-breeding programs.

Key words: *Camellia sinensis*; Expressed sequence tag; Haplotype; Nucleotide diversity; Single nucleotide polymorphisms

INTRODUCTION

Tea is the second most popular beverage in the world, not only for its rich flavor and taste, but also for its benefits to human health (Wei et al., 2011). Therefore, the tea plant [*Camellia sinensis* (L.) O. Kuntze] is grown throughout Asia and Africa as one of the most important perennial cash crops. However, the tea plant is largely heterogeneous because of its high inbreeding depression and long periods of allogamy. In addition, an extended juvenile period and large genome size hinder in-depth research about its molecular genetics (Tanaka et al., 2006). Researchers have attempted to use various molecular markers in tea plant studies, including random amplified polymorphic DNA (RAPD), amplified fragment length polymorphism (AFLP), inter-simple sequence repeat (ISSR), and simple sequence repeat (SSR) (Paul et al., 1997; Liu et al., 2009; Ma et al., 2010). However, molecular genetic studies of the tea plant remain largely restricted by the limited number and availability of these markers. Furthermore, selectively neutral markers (e.g., RAPD and AFLP markers) are not always good candidates for mapping traits on chromosomes. Therefore, it is important to develop and characterize novel markers that are suitable for the construction of genetic maps and linkage analysis of functional traits in the tea plant.

Single nucleotide polymorphisms (SNPs) based on single-base pair mutations are the most abundant type of sequence variation in plants. The use of SNP markers has significantly facilitated relative genetic studies, because of their high abundance, stability, bi-allelic variation throughout the genome of diploid species, and automation for high-throughput analysis (Mochida et al., 2003). Moreover, because SNPs tend to occur in functional genomic regions, they are particularly valuable for characterizing genes associated with complex traits (Nelson et al., 2004; McCouch et al., 2010). However, to our knowledge, few tea plant SNPs have been reported to date.

Here, we validated the feasibility of developing SNP markers in the tea plant. In addition, these SNPs were used to study the genetic relationship of 17 tea varieties. The novel set of SNPs will enhance the number of molecular markers in the tea plant, and may promote the development of genetic studies and breeding programs in this plant.

MATERIAL AND METHODS

Mining candidate SNPs from expressed sequence tags (ESTs)

ESTs of the *C. sinensis* species complex were retrieved from the database of EST (dbEST) of the GenBank (http://www.ncbi.nlm.nih.gov/dbEST/index.html). All of the sequences were assembled using the DNAStar software (http://www.dnastar.com). A four-step scheme was followed to remove noise generated by sequencing errors and alignment procedures (Picoult-Newberg et al., 1999). First, only contigs with 4 or more reads were screened for subsequent analysis. Second, any mismatches that appeared in the low-quality region were removed. Third, indels were excluded from the analysis, because they were probably false positives. Fourth, to confirm that a nucleotide difference was an SNP, at least 2 reads aligning to the reference sequence were required to have a variant allele, while 2 other reads were required to have the reference allele.

Validation and characterization of SNPs

Primers were designed using the Primer Premier 5.0 software (http://www.premierbiosoft.com/). A collection of *C. sinensis* (Theaceae) species complex, including 7 individuals of *C. sinensis* var. *sinensis*, 6 of *C. sinensis* var. *assamica* (Masters) Kitamura, and 4 of *C. sinensis* var. *pubilimba* Chang (Table 1) were used for SNP genotyping. Genomic DNA was extracted from mature leaf of each individual by the cetyltrimethylammonium bromide (CTAB) method. The PCRs were performed in a 50-μL volume containing 80 ng template, 1 U Taq polymerase (TaKaRa, Dalian, Liaoning, China), 2 mM $MgCl_2$, 0.15 mM dNTPs, 5 μM of each primer, and 1X PCR buffer. The PCR program started with an initial denaturation step at 94°C for 4 min, followed by 35 cycles at 94°C for 30 s, appropriate annealing temperatures for 30 s, 72°C for 1 min, and a final extension step of 72°C for 10 min. Then, 2 μL of the PCR products were checked on 1% agarose gel stained with ethidium bromide, and photographed after UV transillumination. The amplicons with single bands were sequenced from both directions on an ABI3730XL DNA sequencer (Applied Biosystems, Foster City, CA, USA) in the Shenzhen Huda Genomics Institute (Shenzhen, China).

Table 1. Information of 17 tea accessions in this study.

Code	Accession name	Species	Origin
FD	Fuding Dabai	*Camellia sinensis* var. *sinensis*	Fuding, Fujian Province
BHZ	Baihaozao	*C. sinensis* var. *sinensis*	Anhua, Hunan Province
LJ	Longjing43	*C. sinensis* var. *sinensis*	Hangzhou, Zhejiang Province
TGY	Tieguanyin	*C. sinensis* var. *sinensis*	Anxi, Fujian Province
MSBH	Mingshan Baihao	*C. sinensis* var. *sinensis*	Mingshan, Sichuan Province
WNZ	Wuniuzao	*C. sinensis* var. *sinensis*	Yongjia, Zhejiang Province
HY	Huangyan	*C. sinensis* var. *sinensis*	Anxi, Fujian Province
CYBH	Changye Baihao	*C. sinensis* var. *assamica*	Menghai, Yunnan Province
JGHY	Jinggu Hongya Zhilicha	*C. sinensis* var. *assamica*	Jinggu, Yunnan Province
JGDB	Jinggu Dabaicha	*C. sinensis* var. *assamica*	Jinggu, Yunnan Province
LTXY	Lingtou Xiaoye	*C. sinensis* var. *assamica*	Qiongzhong, Hainan Province
YWDY	Taiwan Daye	*C. sinensis* var. *assamica*	Taiwan Province
XWQT	Xinwei Qunti1	*C. sinensis* var. *assamica*	Qiongzhong, Hainan Province
YJZJ	Yuanjiang Zhujiecha	*C. sinensis* var. *pubilimba*	Yuanjiang, Yunnan Province
YJN	Yuanjiang 5 Nuocha1	*C. sinensis* var. *pubilimba*	Yuanjiang, Yunnan Province
YJYJ	Yuanjiang Yangjiecha	*C. sinensis* var. *pubilimba*	Yuanjiang, Yunnan Province
BSC	Bansuicha	*C. sinensis* var. *pubilimba*	Ningming, Guangxi Province

Data analysis

Sequences obtained from each individual were aligned by the CodonCode Aligner software (http://www.Codoncod.com/aligner). SNPs were identified as overlapping nucleotide peaks. Several population genetic parameters were analyzed using the DNA sp5 software (http://www.ub.es/dnasp/), including the haplotype number (H_N), haplotype polymorphism (H_p), and nucleotide diversity (π) of the 17 tea plant varieties. A molecular phylogenetic dendrogram was constructed by the unweighted pair group method with arithmetic mean (UPGMA) method.

Determining putative function

The SNP-associated sequences were blasted onto the NCBI database (http://www.

ncbi.nlm.nih.gov) by the BLASTX program, to determine their putative functions. A threshold E-value of <1.00 E^{-10} was selected.

RESULTS AND DISCUSSION

A total of 12,757 ESTs, derived from the *C. sinensis* species complex, were retrieved from dbEST of the GenBank (http://www.ncbi.nlm.nih.gov/dbEST/index.html). The ESTs were aligned into 1513 tentative contigs and 2487 singletons, which represent an estimated total length of 140 kb. Of these contigs, 570 contained 4 or more sequences. In total, 818 putative SNPs were identified from these sequences (Figure 1A), with an average density of 1 SNP/170 bp. This density was close to that reported for barley (1 SNP/200 bp) (Rostoks et al., 2005) and citrus (1 SNP/164 bp) (Jiang et al., 2010).

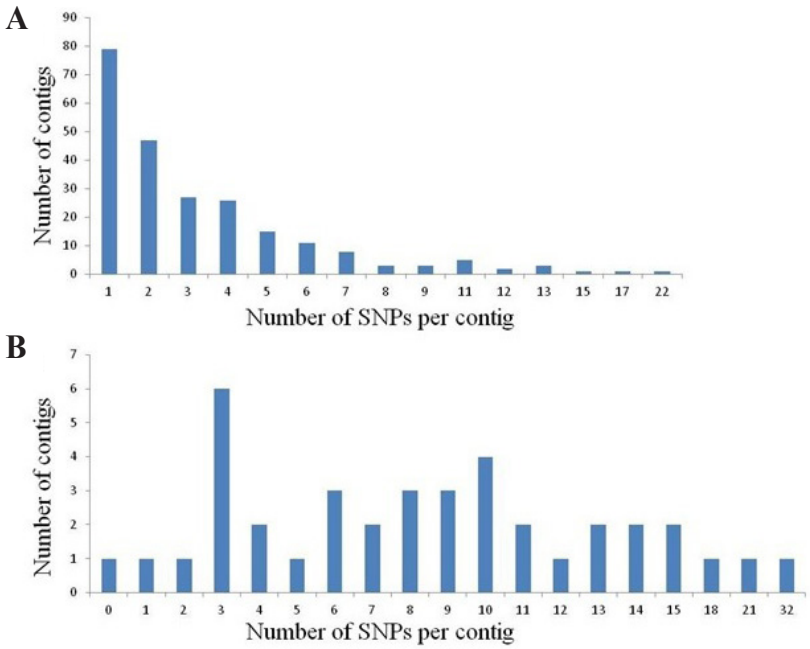

Figure 1. Distribution of putative and discovered SNPs in contigs. **A.** Distribution of 818 putative SNPs in contigs. **B.** Distribution of discovered SNPs per contigs.

Subsequently, 123 PCR primer pairs were designed to evaluate 253 randomly selected putative SNPs. Seventeen tea varieties were used for amplification (Table 1). Sixty-three primer pairs (51.2%) produced single bands, which were then sequenced in both directions. Ultimately, 39 primer pairs provided readable sequences. The length of the new sequences reached 18,557 bp. Thirteen sequences were found to be larger than the expected product sizes, indicating the presence of non-coding regions (Table 2). Furthermore, only 20% (50) SNPs predicted from ESTs were validated, which was slightly lower compared to that reported for tomato (27.2%) (Labate and Baldo, 2005) and the rubber tree (37%) (Pootakham et al., 2011).

Table 2. Characteristics of 39 new markers developed in *Camellia sinensis*.

Me	Primer sequences (5'-3')	Length (bp, anticipative/actual)	Number of SNPs		H_N	H_P	π (%)	Accession No.	Predict function (E ≤ 10)
			Total	Putative/validated/new					
Con60	F: CAGAACACTCCATGAAAAGTTTGAC R: AAGGCTGACAAGAAGGAATACGA	285/579	12	1/1/11	10	0.88	0.416	GE651638.1	Translation machinery-associated protein
Ccon142	F: AGTTATGGGTCGCACAGGA R: CTCCAGCTCAAGCATAAATGA	356/356	3	3/3/0	10	0.708	0.355	FE861601.1	40S ribosomal protein
Con182	F: AGGACGCAGTGAACTGTAATT R: GGTACGCCGTTAGAGTGAGT	339/339	0	1/0/0	1	0	0	GE651922.1	Mitochondrial protein
Con198	F: TTCAGTTGGTAGAAGTGATGAGACA R: GCCAAATTTTGATATTACGTTACAG	175/175	1	2/1/0	2	0.214	0.122	GH623805.1	Unknown
Con 216	F: AATCCCCTTCATGGCTATTATTATG R: CTGCATATTCTGATGATATGGGAAA	538/894	10	4/2/8	12	0.825	0.162	GE651034.1	Nonspecific lipid-transfer protein precursor
Con 229	F: CGAATGTCGCTATGCTGTC R: CAAGCCAAGTTCACCATCC	320/408	8	1/0/8	8	0.75	0.383	GE651354.1	Actin depolymerizing factor
con 245	F: GGTGAAGATCGCGTTAGGA R: CAATAGAGGATCGACCTGAATA	392/538	10	1/1/9	7	0.752	0.396	GE653036.1	Tonoplast intrinsic protein
Con259	F: CAAACCCTATAGTCTTCTTCGAGAT R: CAGATCAGATCAGATCACAGATTTG	551/551	3	2/2/1	5	0.412	0.119	GE650486.1	Rotamase cyclophilin
Con 263	F: AAGTTCAAGTTCCCTGGTC R: AAGCAGGAGTCCTTCAGTT	251/808	18	1/0/18	3	0.221	0.029	GE650902.1	Ribosomal protein
Con 279	F: TAGGGTTAGGGTTTGTGAAGAAAGA R: AAACTCCTGAGAAGCAGAACCAATA	459/607	15	1/1/14	2	0.258	0.044	GE652420.1	Histone 2
Con 335	F: GTTAGGCAAGATTTGGGATG R: CCAGTCATAAACAGTGGGAGA	307/307	8	1/0/8	2	0.138	0.046	GE652717.1	Auxin-repressed protein
Con339	F: GTTGAAGCCAAGATTAAGGAAG R: TATCCTCTAATCCTCACTCTTCTCA	360/360	3	4/3/0	4	0.535	0.254	FE861539.1	Unknown
Con343	F: AGTCCTTTCTTTGCTGAACTCTTTC R: CAGCTGCTATTATAGTTGGGTCAGA	418/418	7	3/1/6	9	0.783	0.307	GE652520.1	Chalcone synthase 2
Con 344	F: ACATTTGGAGCATTCATAGGG R: ATGTTGGCACCCACTATGAAA	296/798	32	3/0/32	11	0.869	0.311	GE650454.1	Tonoplast intrinsic protein
Con371	F: GCGGTGTAGGGTTGGTGAA R: AGTCTCGTGCGCTCCTCTT	259/259	13	2/2/11	16	0.916	1.179	FE861551.1	Histone H2A protein
Con 373	F: AAATTCGCAAGTACCAAAAGAGTAC R: TCCTCCTTGCCAACTGAATGT	240/354	5	2/0/5	4	0.57	0.47	GE650640.1	Histone H3 protein
Con 377	F: AGTCTTTGGGTTCAGTAAAGAGTGT R: CCCATACACTAAACAGCCAAAA	457/457	11	4/3/8	8	0.791	0.608	GE652758.1	Ferredoxin 3
Con380	F: TTCCATACCCAGTTCTGCTAA R: ATAAACCAACCGTCAAGGAAG	328/328	9	1/0/9	10	0.708	0.355	GE651109.1	Unknown
Con383	F: TACTCCGCAGCACAGGATAAA R: TAACAAAGCAGGAGCCCAACA	341/341	14	2/2/12	11	0.865	0.845	GE651159.1	Copper-binding protein 2
Con392	F: TATTCCACCAGACCAACAG R: TCAGCCAAAGTCCTTCC	525/525	21	1/1/20	8	0.819	0.527	GT969157.1	Polyubiquitin

Continued on next page

Table 2. Continued.

Me	Primer sequences (5'-3')	Length (bp, anticipative/actual)	Number of SNPs		H_N	H_P	π (%)	Accession No.	Predict function (E ≤ 10)
			Total	Putative/validated/new					
Con 394	F: AAATTCTCACAGATCGCAACAATCT R: AGAGAAGATGCACGGAAAAAATATG	308/308	4	1/1/3	5	0.575	0.235	FE861486.1	Unknown
Con463	F: AAATTACATCGCACTCCCAATC R: ATACTAGCCTTCACATGATCCATCT	442/442	7	4/3/4	15	0.888	0.560	GT968996.1	Heat shock protein
Con 507	F: CTCCATTGCCTTGATGCTGAT R: GCCTGCCAACAATCTCACTCA	275/275	3	1/1/2	2	0.059	0.022	GE650659.1	Glutaredoxin C4
Con 553	F: GAAAGGTCGCTGTTGTCTGTG R: TTGCCGGTACTCTTCTCCTTC	508/508	13	3/2/11	6	0.749	0.64	GE650655.1	S-adenosyl-l-homocysteine hydrolase 2
Con 610	F: ACTCGCAAACTGTTGTGGTGT R: ACTAAGCTCTCAAAATTCTTCCATT	172/172	6	6/5/1	2	0.515	1.797	FE861429.1	Ribosomal protein
Con 627	F: AAGGCAATAAATGACGAGGGA R: TCCAACCGCATCCTCAAAC	339/339	4	2/1/3	5	0.367	0.18	GE652354.1	Heat shock protein
Con 634	F: CGATCTGGCAAGGAGATTGAG R: CCACAAGAGTAACAAAACCAAAGTC	488/488	14	1/0/14	11	0.706	0.316	FE861350.1	GTP binding Elongation factor Tu family protein
Con 643	F: GTGGCACAAGATGATGTTTCC R: CAGTAAACGACCATATCCACGA	436/994	2	2/0/2	3	0.204	0.062	GE651380.1	Unknown
Con 649	F: AGAGATCTGAGAATGGATTCAACTG R: AAGAGGACAGGATTAATATTGGGTA	389/876	8	2/1/7	11	0.874	0.236	FE861635.1	Histone H2A protein
Con658	F: CCCACACATCTACAACTTGACTATC R: AGCAACAAACCCAAGACTCAA	596/596	9	2/1/8	14	0.854	0.429	GT969353.1	Glutathione S-transferase
Con676	F: GGATCAGAACGAATCTCGGTTAC R: CTAAGATGAAGCCGACGGAAC	395/395	6	2/0/6	6	0.672	0.273	GE651629.1	Prenylated rab acceptor
Con 694	F: CCTAAAATCTGTTTGTCTGTCTGTC R: CTTCTTTGGTAACAAGTGACTGTGA	406/590	10	2/1/9	6	0.784	0.486	FE861332.1	Histone H2A protein
Con 720	F: ACAAGGCGAGTACATCAAGACC R: AATCTGAATCAATCTCTCTCTAGCG	347/347	3	1/1/2	4	0.357	0.108	FE861394.1	Pleckstrin homology domain-containing protein
Con 936	F: GGATTCTGAGCCTTCTTGATT R: CTTGTATGATTCCTCACCCAAC	350/898	9	1/0/9	7	0.714	0.214	GE650773.1	14-3-3 protein
Con1033	F: CCAATGGTCCCTTTCCTG R: CCCGCAACGGTGTCAAG	363/363	11	5/4/7	15	0.96	1.046	GE650251.1	Chlorophyll a-b-binding protein
Con1839	F: GGCGGGTTTCCGTATCTC R: AAGAGGCAATAAACAGAGTACCATA	620/756	6	4/1/5	4	0.706	0.3	FJ648828.1	Aquaporin
Con2063	F: CCAAACCATATCACTACTCCCTC R: AGGTAGGCTGCCGTTAAGATA	281/281	15	4/2/13	15	0.907	0.875	GH738558.1	Chlorophyll a/b-binding protein
Con2190	F: ATTGCGGACTTTGTGAGGATG R: GTATTCACTCCACAACTGGTCAACT	234/234	3	2/2/1	2	0.515	0.222	GH738698.1	Plant basic secretory protein family protein
Con2496	F: AGCCATGAACTCCAATGTGCT R: TTCGAGGTATTCTTCGATGACG	292/292	10	2/1/9	7	0.558	0.335	CV014011.1	Heat shock protein

H_N = haplotype number; H_P = haplotype polymorphism; π = nucleotide diversity.

The lower verification rate in this study might be attributed to genetic differences among the germplasm resources used. Overall, 299 new SNPs were identified, including 215 in coding regions and 84 in non-coding regions (Table 2). Each contig contained 0 to 32 SNPs (Figure 1B). The average SNP frequency was 1 per 54 bp in coding regions and 1 per 49 bp in non-coding regions (Table 3), which was significantly higher compared to the SSR frequency in the tea plant (1 SSR/7.55 kb; Ma et al., 2012). The high frequency of SNPs presents high potential for the development of associated genetic mapping studies. For example, Taniguchi et al. (2012) reported 3 new reference genetic maps for tea using 1124 markers, including 441 SSRs. The average distances between markers was 1.93 cM [cultivar (cv) 'Sayamakaori'], 1.85 cM (cv 'Kana-Ck17'), and 4.35 cM (core map, constructed by merging the markers present in both parents), respectively. The exploration of tea SNP markers might contribute towards increasing the mapping density and facilitating the use of the tea plant in QTL location and map-based gene cloning.

Table 3. Summary of single nucleotide polymorphisms.

Parameter	Value	Comments
Number of EST loci sequenced	63	-
Number of good-quality sequences	39	-
Number of bases of sequence screened	18,557 bp	-
Number of bases of coding region	14,433 bp	-
Number of bases of non-coding region	4,124 bp	-
Number of SNPs	349	-
Mean SNP frequency	0.0188	1/53 bp
Frequency of SNP in coding region	0.0184	1/54 bp
Frequency of SNP in non-coding region	0.0204	1/49 bp

The 349 SNPs contained 114 (32.66%) transversions, 232 (66.48%) transitions and 3 (0.86%) tri-allelic SNPs (C/G/T, A/C/T, and A/C/G), resulting in an average transition and transversion ratio of 2.0 (Table 4). This bias might arise because transitions are more likely to preserve protein structures (Wakeley, 1996). In addition, 3 new tri-allelic SNPs were identified, which is of interest for identification purposes and for system evolution research (Westen et al., 2009).

Table 4. Numbers of transitions and transversions in different genomic regions of tea plant.

	Transitions		Transversions				Tri-allelic
	C/T	A/G	G/T	A/T	C/G	A/C	
In non-coding region	22	29	10	7	6	8	2
In coding region	110	71	23	22	22	16	1
In complete sequences	132	100	33	29	28	24	3

The population genetic parameters were analyzed using the DNAsp5 program, to examine the effectiveness of SNPs in cultivar identification. The π, H_N, and H_P ranged from 0 to 1.797%, 1 to 16, and 0 to 0.960, respectively. Con371, which contains 13 SNPs, has the largest number of haplotypes among the 39 amplicons. Fifteen of 17 varieties may be separated by this single marker (Figure 2), indicating that SNPs have high potential for use in the identification of tea plant cultivars.

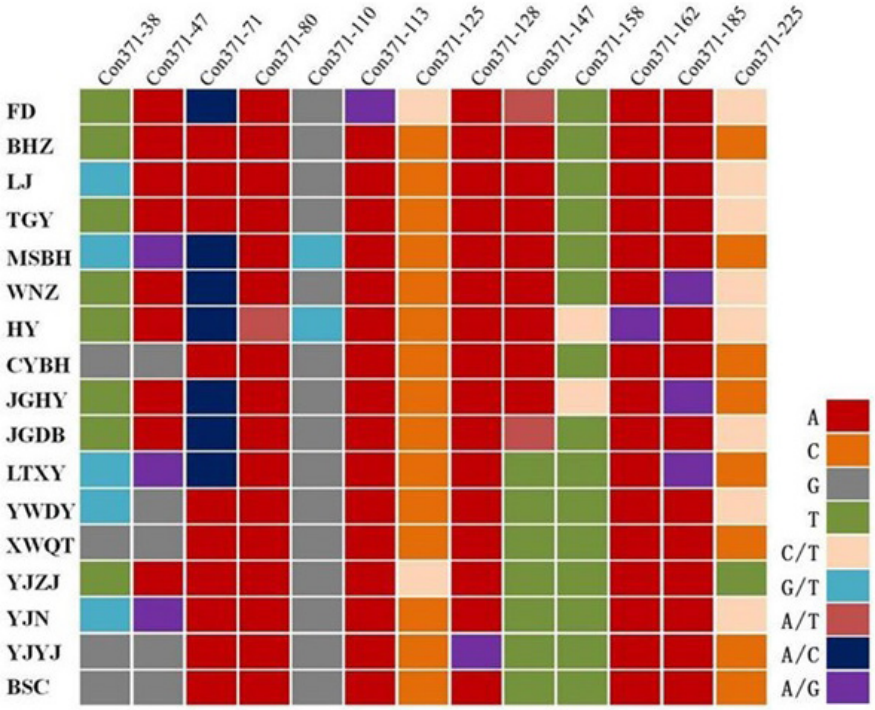

Figure 2. Fingerprint of 17 varieties constructed by marker con371.

To evaluate the genetic relationships between the 17 tea plant varieties, a dendrogram was constructed from all SNPs using the UPGMA method. The dendrogram was divided into 2 groups (Figure 3). All of the *sinensis* varieties were clustered into group A, which was consistent with previous findings by AFLP (Balasaravanan et al., 2003) and ISSR (Yao et al., 2008) markers. Thus, these SNPs are effective for genetic distance estimation. Group B contained 5 *assamica* varieties and 4 *pubilimba* varieties. The taxonomic position of *pubilimba* is still under debate; however, our results provide evidence that *pubilimba* had high similarity to other varieties in *C. sinensis* (Chen et al., 2000; Pandolfi et al., 2009). In contrast, 'BSC' was separated from the rest of the *pubilimba* varieties ('YJZJ', 'YJN', 'YJYJ'), which might be due to differences in geographic origin.

To determine the putative functions of SNP-associated genes, the sequences were blasted onto the NCBI database. Thirty-four sequences coding known proteins were obtained, including cyclophilin, chalcone synthase, ribosomal protein, and heat shock protein (Table 2). Some of these proteins are associated with important biological functions. For instance, con343 encodes a fragment of chalcone synthase. This enzyme plays an important role in catechin biosynthesis, which is closely correlated with tea quality (Kaundun and Matsumoto, 2003). In addition, con335 is in a gene encoding an auxin-repressed protein, which might be down-regulated by auxin, and responsible for the dormancy of the tea bud (Wang et al., 2011). Further studies of SNPs in these fields might provide novel insights about fundamental metabolism mechanisms, and ultimately promote the development of associated genetic studies, and enhance tea plant-breeding programs.

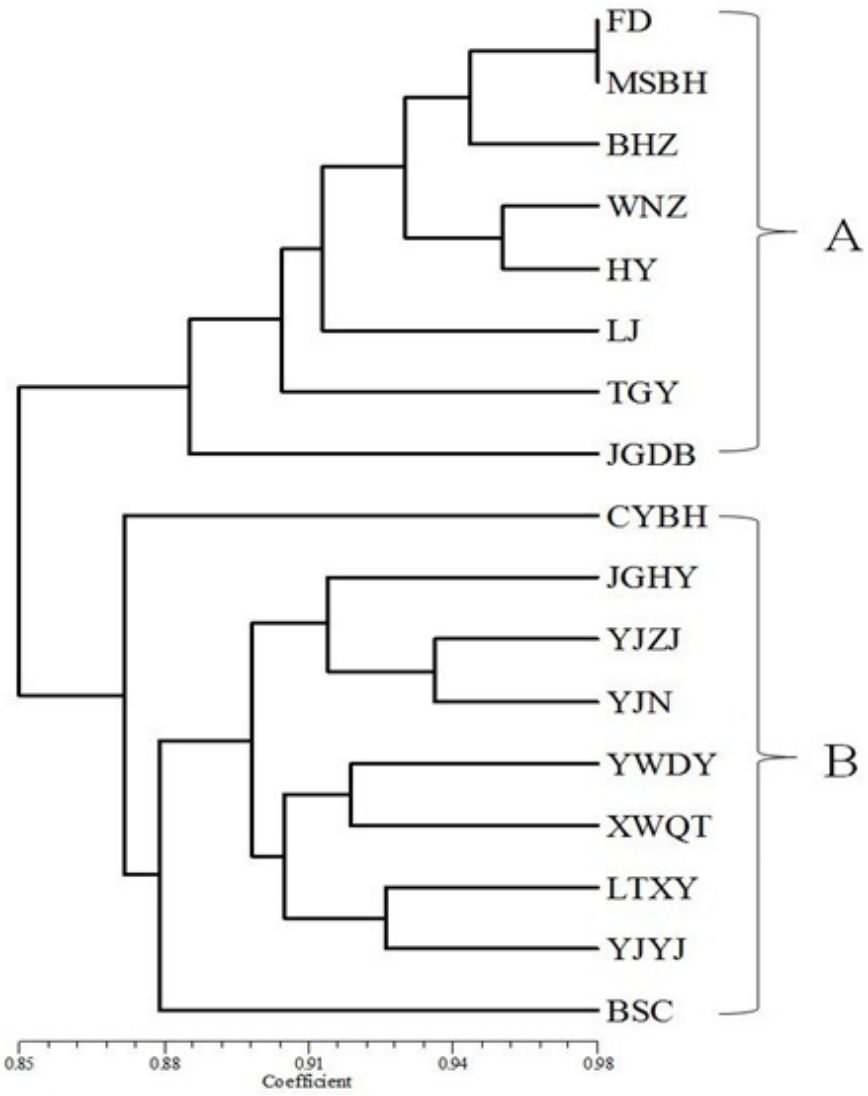

Figure 3. Genetic relationships of the 17 varieties constructed based on the data of all SNPs.

CONCLUSIONS

A total of 818 putative SNPs were identified from tea-derived dbEST. In total, 50 (20%) SNPs predicted from ESTs and 299 new SNPs were confirmed for the tea plant. The H_N, H_P, and π of 39 markers ranged from 1 to 16, 0 to 0.960, and 0 to 1.797%, respectively. These results indicate that it is feasible to identify SNPs from *C. sinensis*-derived ESTs. In conclusion, the new markers reported here could be used in cultivar identification, genetic map construction, and marker assisted-breeding of the tea plant.

ACKNOWLEDGMENTS

Research supported by the Modern Agro-Industry Technology Research System of China (#nycytx-23).

REFERENCES

Balasaravanan T, Pius PK, Kumar RR, Muraleedharan N, et al. (2003). Genetic diversity among south Indian tea germplasm (*Camellia sinensis*, *C. assamica* and *C. assamica* spp. *lasiocalyx)* using AFLP markers. *Plant Sci.* 165: 365-372.

Chen L, Yu FL and Tong QQ (2000). Discussions on phylogenetic classification and evolution of Sect. Thea. *J. Tea Sci.* 20: 89-94.

Jiang D, Ye QL, Wang FS and Cao L (2010). The mining of *Citrus* EST-SNP and its application in cultivar discrimination. *Agr. Sci. China* 9: 179-190.

Kaundun SS and Matsumoto S (2003). Development of CAPS markers based on three key genes of the phenylpropanoid pathway in tea, *Camellia sinensis* (L.) O. Kuntze, and differentiation between *assamica* and *sinensis* varieties. *Theor. Appl. Genet*. 106: 375-383.

Labate JA and Baldo AM (2005). Tomato SNP discovery by EST mining and resequencing. *Mol. Breed.* 16: 343-349.

Liu BY, Wang LY, Li YY, Hi W, et al. (2009). Genetic diversity in tea (*Camellia sinensis*) germplasms as revealed by ISSR markers. *Indian J. Agr. Sci.* 79: 715-721.

Ma JQ, Zhou YH, Ma CL, Yao MZ, et al. (2010). Identification and characterization of 74 novel polymorphic EST-SSR markers in the tea plant, *Camellia sinensis* (Theaceae). *Am. J. Bot.* 97: e153-e156.

Ma JQ, Ma CL, Yao MZ, Jin JQ, et al. (2012). Microsatellite markers from tea plant expressed sequence tags (ESTs) and their applicability for cross-species/genera amplification and genetic mapping. *Sci. Hortic.* 134: 167-175.

McCouch SR, Zhao K, Wright M, Tung CW, et al. (2010). Development of genome-wide SNP assays for rice. *Breed. Sci.* 60: 524-535.

Mochida K, Yamazaki Y and Ogihara Y (2003). Discrimination of homoeologous gene expression in hexaploid wheat by SNP analysis of contigs grouped from a large number of expressed sequence tags. *Mol. Genet. Genomics* 270: 371-377.

Nelson MR, Marnellos G, Kammerer S, Hoyal CR, et al. (2004). Large-scale validation of single nucleotide polymorphisms in gene regions. *Genome Res*. 14: 1664-1668.

Pandolfi C, Mugnai S, Azzarello E, Bergamasco S, et al. (2009). Artificial neural networks as a tool for plant identification: a case study on Vietnamese tea accessions. *Euphytica* 166: 411-421.

Paul S, Wachira FN, Powell W and Waugh R (1997). Diversity and genetic differentiation among populations of Indian and Kenyan tea (*Camellia sinensis* (L.) O. Kuntze) revealed by AFLP markers. *Theor. Appl. Genet*. 94: 255-263.

Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, et al. (1999). Mining SNPs from EST databases. *Genome Res.* 9: 167-174.

Pootakham W, Chanprasert J, Jomchai N, Sangsrakru D, et al. (2011). Single nucleotide polymorphism marker development in the rubber tree, *Hevea brasiliensis* (Euphorbiaceae). *Am. J. Bot.* 98: e337-e338.

Rostoks N, Mudie S, Cardle L, Russell J, et al. (2005). Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. *Mol. Genet. Genomics* 274: 515-527.

Tanaka J, Taniguchi F, Hirai N and Yamaguchi S (2006). Estimation of the genome size of tea (*Camellia sinensis*), *Camellia* (*C. japonica*), and their interspecific hybrids by flow cytometry. *Tea Res. J.* 101: 1-7.

Taniguchi F, Furukawa K, Ota-Metoku S, Yamaguchi N, et al. (2012). Construction of a high-density reference linkage map of tea (*Camellia sinensis*). *Breed. Sci.* 62: 263-273.

Wakeley J (1996). The excess of transitions among nucleotide substitutions: new methods of estimating transition bias underscore its significance. *Trends Ecol. Evol*. 11: 158-162.

Wang XC, Ma CL, Yang YJ, Yao MZ, et al. (2011). Cloning and expression analysis of auxin-repressed protein gene CsARP1 in tea plant (*Camellia sinensis*). *J. Nucl. Agr. Sci.* 25: 910-915.

Wei K, Wang LY, Zhou J, He W, et al. (2011). Catechin contents in tea (*Camellia sinensis*) as affected by cultivar and environment and their relation to chlorophyll contents. *Food Chem.* 125: 44-48.

Westen AA, Matai AS, Laros JF, Meiland HC, et al. (2009). Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples. *Forensic Sci. Int. Genet.* 3: 233-241.

Yao MZ, Chen L and Liang YR (2008). Genetic diversity among tea cultivars from China, Japan and Kenya revealed by ISSR markers and its implication for parental selection in tea breeding programmes. *Plant Breed.* 127: 166-172.

