

贵州省主要栽培茶树品种指纹图谱构建与遗传结构分析

乔大河, 郭 燕, 杨 春, 李 燕, 陈 娟, 陈正武

(贵州省农业科学院茶叶研究所, 贵阳 550006)

摘要: 贵州是中国目前茶树种植面积最大的省份, 同时也拥有丰富的茶树种质资源。对品种真实性以及遗传结构的准确鉴定是种质识别以及杂交亲本选配的关键。本研究以包含贵州省内主要栽培品种和新选育品系的 54 份茶树资源为材料, 利用 35 对 SSR 引物进行遗传鉴定, 共检测到 317 个等位基因位点, 多态信息含量介于 0.22~0.92 之间, 平均 0.73; 利用 *PI* 和 *PIsibs* 两个参数评估了每对引物的鉴别力, 表明任意两对引物组合都能够区分本研究中的茶树材料, 并从中筛选出 5 对核心引物构建了 54 份材料的指纹识别图谱, 这 5 对核心引物组合足够用于更大群体的个体鉴定。利用 GBS 方法对 54 份材料进行了全基因组 SNP 挖掘, 共鉴定到 698117 个高质量 SNP; 基于 SNP 进一步进行系统进化分析、主成分分析以及遗传结构分析, 54 份材料被划分为 4 个类群, 育成品种与新选育品系间遗传背景呈现高度重叠; 并据此推测出了前期贵州茶树育种的一般模式, 表明贵州前期茶树育种过多依赖福鼎大白茶以及云南地方种质资源, 预示着在后续育种实践中, 加大对省外优良品种以及贵州省内古茶树资源的杂交利用力度, 将有助于茶树种质创新以及新品种的选育。

关键词: 茶树; 指纹识别; SSR; SNP; 遗传结构

Fingerprinting Construction and Genetic Structure Analysis of the Main Cultivated Tea Varieties in Guizhou Province

QIAO Da-he, GUO Yan, YANG Chun, LI Yan, CHEN Juan, CHEN Zheng-wu

(Institute of Tea, Guizhou Academy of Agricultural Science, Guiyang 550006)

Abstract: Guizhou has the largest tea planting area in China, which also has the abundant tea plant germplasm resources. Accurate identification of the authenticity and genetic structure of the cultivars is the pivotal to germplasms identification and parents selection in hybridized breeding. In this study, 54 tea plant germplasms including the cultivars or improved cultivars in Guizhou province were used as materials, the genetic identification of them was conducted using 35 SSR markers, firstly. A total of 317 alleles were detected, yielding the polymorphic information content (PIC) ranged from 0.22 to 0.92, with a mean value of 0.73. The genetic differentiation of each SSR marker was evaluated by *PI* and *PIsibs*, which showed that any two pairs of primers could distinguish the tea plant germplasms in this study. Five core SSR markers were selected to generate the fingerprinting of these tea plant resources, which provided the resolution sufficient for identification of the individuals in larger groups. Additionally, by using the genotyping-by-sequencing (GBS) approach, the genome-wide SNP were identified with the 54 cultivars or improved cultivars, and a total of 698117 high quality SNP were identified. The phylogenetic analysis, principle component analysis (PCA) and genetic structure analysis of the 54 tea plant samples were performed. The 54 tea plant germplasms were divided into 4 sub-populations,

收稿日期: 2018-08-01 修回日期: 2018-08-27 网络出版日期: 2018-11-12

URL: <http://kns.cnki.net/kcms/detail/11.4996.S.20181109.1731.008.html>

第一作者研究方向为茶树资源与育种, E-mail: dahe10466@163.com

通信作者: 陈正武, 研究方向为茶树资源与育种, E-mail: zwchentea@163.com

基金项目: 贵州省 - 贵州农科院省院联合基金项目 (LH[2015]7075); 贵州省科技支撑计划项目 ([2017]2557); 贵阳市 - 贵州省农科院院地合作项目 ([2014]7)

Fundation project: The Guizhou Province-Guizhou Academy of Agricultural Sciences Cooperation Project (LH[2015]7075), The Science and Technology Support Project of Guizhou Province ([2017]2557), The Guiyang City-Guizhou Academy of Agricultural Sciences Cooperation Project ([2014]7)

and the genetic backgrounds of the cultivars and the improved cultivars were highly overlapped. According to the results, the general breeding model of tea plants in Guizhou in the previous was speculated, which revealed that the germplasms utilization of early tea plants breeding in Guizhou depended too much on Fudingdabaicha and the local tea plant germplasm resources of Yunnan province, and it was also shown that increasing the hybrid utilization of the excellent tea plant cultivars in other provinces and the ancient tea plant resources in Guizhou would contribute to the germplasm innovation and the breeding of new tea plants in the subsequent breeding practice.

Key words: tea plant (*Camellia sinensis* (L.) Kuntze); fingerprinting; SSR; SNP; genetic structure

茶是世界上三大非酒精类饮品之一,作为茶树的起源地和茶文化的发祥地,我们国家具有悠久的饮茶与茶树种植历史^[1]。茶树[*Camellia sinensis* (L.) Kuntze] 是隶属于山茶科山茶属的一种常绿多年生木本植物,目前已在世界上 50 余个国家广泛种植,而同时作为世界上最大的产茶国,中国的云南、贵州、四川、湖南、安徽、河南、浙江、福建等 20 余省份都有茶树种植^[2]。截至 2016 年底,贵州的茶园面积已达 $4.4 \times 10^5 \text{ hm}^2$ (http://zzys.agri.gov.cn/chaye_cx.aspx), 位居全国第一,但是超过 60% 的种植品种为福鼎大白茶,品种结构单一。随着近年来贵州省茶园面积的不断增加,科研机构和茶叶生产企业在从省外引进优良茶树品种的同时逐渐重视起对省内优良品种的开发,并加大了对省内茶树种质资源尤其是大量古茶树资源的育种利用力度。

不同于一般的作物,茶树生育期长、高度自交不亲和、杂交结实率低、基因组大,杂交育种以及生物技术育种在茶树中的应用要远远落后于一般作物^[3-4]。长期以来茶树育种多采用从自然授粉群体中单株选择或者从优良品种中采集种子进行后代选择的方式,这也造成了育成品种的遗传背景复杂;另外,茶树育种周期长,一个新品种的培育需要 20 年左右的时间,这也限制了茶树育种的进程^[5]。杂交育种(包括远缘杂交)是创造新变异、培育新品种的重要方法,采用人工控制的、有目的的杂交育种并进行分子标记辅助选择是茶树育种的发展趋势,在这个过程中对双亲的准确鉴定是育种成败的关键^[5]。而传统的表型鉴定已不能满足于种质资源评价以及育种利用的实际需要^[6]。

基于 DNA 序列多态性的分子标记得益于其稳定、便捷、高效等特点已广泛应用于植物种质资源鉴定、评价以及育种利用^[7-9]。简单序列重复(SSR, simple sequence repeat) 是基因组上广泛分布的一

种多态性序列形式,基于此的 SSR 标记多态性高、重复性好、易于检测且具有共显性,已成为品种指纹图谱构建、种质资源遗传多样性分析、分子标记辅助选择等方面的主流分子标记^[10]。除在玉米^[9]、小麦^[11]、水稻^[12]等一般禾本科植物的相关研究中广泛应用以外,目前在包括茶树在内的木本植物中都发挥着重要作用。比如在茶树中,张成才等^[13]利用 35 对 SSR 引物对浙江省的 36 个育成品种进行了鉴定研究,并筛选了 10 对核心引物构建了指纹图谱,同时利用核心引物对未知茶苗进行了遗传鉴定; Tan 等^[14]利用 30 对 SSR 引物对 128 份中国无性系良种进行鉴定分析,从分子层面证实其中 29 份材料的亲本之一均为福鼎大白茶; Meegahakumbura 等^[15]利用 23 对 SSR 引物对来自中国和印度的 402 份茶树材料进行了驯化和育种历史研究,表明云南有两个阿萨姆种茶树的起源中心,并认为对中国阿萨姆种茶树种质及地方种质的收集具有重要育种利用价值。然而由于 SSR 标记数目可应用的有限性,致使其面对高分辨率种质遗传背景筛选、高密度遗传图谱构建等研究需求时,原有的优势已不再明显。而以单核苷酸多态性(SNP, single nucleotide polymorphism) 为基础的分子标记则显示出其独特的优势:数量巨大、全基因覆盖、应用成本低廉等。尤其是随着高通量测序技术的发展,为从复杂基因组中挖掘全基因组的 SNP 提供了便利^[16]。在玉米^[17]、水稻^[18]等作物中都已开发出包含数以万计 SNP 标记位点的高通量 SNP 芯片,但是目前对于茶树中 SNP 标记的开发和利用还很有限^[19-20],茶树中已报道的基于 SNP 标记的遗传图谱也仅包含 6042 个 SNP^[21]。Genotyping by sequencing(GBS) 是一项基于高通量测序技术的挖掘基因组 SNP 变异进行基因分型的方法,目前已在多种作物的系统进化研究、全基因关联分析、高密度遗传图谱构建等方面广泛应用^[22-24]。

本研究首先利用 35 对 SSR 引物对 27 份贵州

主要栽培茶树品种和 27 份新选育的品系进行遗传多样性分析并筛选核心引物构建了指纹识别图谱;然后利用基于高通量测序技术的 GBS 方法对这 54 份材料的全基因组 SNP 进行鉴定,利用鉴定到的 SNPs 分析了它们的亲缘关系和遗传结构。旨在为贵州省内主要栽培茶树品种高效鉴定和后续茶树育种项目的开展提供理论和数据支撑,同时获得的 SNP 也有助于后续茶树 SNP 标记的开发与应用。

表 1 54 份茶树样品信息
Table 1 Detailed information of the 54 tea plant samples used in this study

代号 ID	名称 Name	种质类型 Germplasm type	代号 ID	名称 Name	种质类型 Germplasm type
BH1	保靖黄金茶 1 号	品种	C550	C550	改良品系
CQ1	楠叶齐	品种	DH1	DH1	改良品系
FD1	福鼎大白茶	品种	DH2	DH2	改良品系
GY8	贵茶育 8 号	品种	DH3	DH3	改良品系
JG1	金观音	品种	DH4	DH4	改良品系
MP1	苗圃 1 号	品种	DH5	DH5	改良品系
NW	鸟王种 1 号	品种	DH6	DH6	改良品系
QC1	黔茶 1 号	品种	DZ1	单株 0 号	改良品系
QC7	黔茶 7 号	品种	DZ2	单株 2 号	改良品系
QC8	黔茶 8 号	品种	DZ3	单株 3 号	改良品系
QM308	黔湄 308	品种	FD05	FD05	改良品系
QM502	黔湄 502	品种	FD06	FD06	改良品系
QM601	黔湄 601	品种	FX1	福选 1 号	改良品系
QM701	黔湄 701	品种	FX2	福选 2 号	改良品系
QM809	黔湄 809	品种	QZ1	QZ1	改良品系
TC12	台茶 12 号	品种	QZ2	QZ2	改良品系
TX1	苔选 0310	品种	TB0322	0322 变异	改良品系
TX2	苔选 0307	品种	J1	久安绿 1 号	古茶树
TX3	苔选 0314	品种	GYH1	GYH1	古茶树
TX4	苔选 0331	品种	GYH2	GYH2	古茶树
YG0843	云贵山 0843	品种	GYH4	GYH4	古茶树
ZJ1	紫娟	品种	GYH5	GYH5	古茶树
ZZ1	紫魁	品种	GYH7	GYH7	古茶树
LJ43	龙井 43	品种	GYH8	GYH8	古茶树
QF4	黔辐 4 号	品种	GYH9	GYH9	古茶树
QM419	黔湄 419	品种	GYH10	GYH10	古茶树
MS131	名山白毫 131	品种	GYH12	GYH12	古茶树

1 材料与方法

1.1 材料

本研究涉及的茶树材料共计 54 份,其中育成品种 27 份、改良品系 17 份、古茶树资源 10 份,材料的详细信息如表 1 所示。54 份材料全部种植于贵州省农业科学院茶叶研究所贵阳和湄潭的茶树种质资源圃中,于 2017 年 4 月对各个材料的一芽二叶取样,液氮速冻后放入 -80 ℃冰箱中保存备用。

1.2 茶树基因组 DNA 提取及 SSR 分型

茶树基因组 DNA 提取采用改良的 CTAB 法^[25]。提取基因组 DNA 经 1% 琼脂糖凝胶电泳和紫外分光光度计检测质量后用于 SSR 扩增。

本研究用到的 35 对 SSR 引物筛选自公开发表的文献^[26-27]，由上海生工生物工程有限公司合成，引物的详细信息如表 2 所示。20 mL PCR 扩增体系包含 10 mL 预混 Taq(TaKaRa)，1 mL 模板 DNA，正反向引物各 1 mL，7 mL 无菌水。PCR 扩增程序为 94 ℃ 预变性 4 min，94 ℃ 变性 10 s，

退火 30 s(退火温度根据引物设定)，72 ℃ 延伸 40 s，35 个循环后 72 ℃ 再延伸 10 min。PCR 产物经 Fragment Analyzer™ 96 全自动毛细管电泳系统进行 SSR 分型，毛细管电泳相关试剂采用系统配套的 dsDNA 905(DNF-905)试剂盒，反应体系配制按照操作说明：2 mL PCR 产物加入到 22 mL 的 1 × TE buffer 中，在 96 孔 PCR 板的第 H12 孔加入 24 mL DNA ladder。利用系统配套的 PROSize™ 2.0 软件进行片段大小的读取，其中 1 bp 的下游 marker 和 500 bp 的上游 marker 用于片段的矫正。

表 2 本研究使用的 SSR 标记
Table 2 The character of the SSR markers used in this study

引物名称 Primer name	原始编号 Original code	引物序列 Primer sequences (5'-3')	重复单元 Repeat motif	退火温度 (℃) Tm	片段大小 (bp) Size	引物来源 Reference
TP1	TM043	F: TAAATTCTGTCCACCCATCT R: GTAACATAATCTCCCTCCTTCT	(TTTC) 5	56	130~146	[26]
TP2	TM044	F: CACGAAGAATAACCCAAGAT R: GAAGATACATGCGAATACATAC	(TC) 10	56	123~147	[26]
TP3	TM046	F: AACACCATAAGCTCATCTAC R: ACTCCATTACCCGCTACTAT	(AG) 12	56	102~131	[26]
TP4	TM055	F: CAGGGTATGATTGATGAGGA R: CAAGCAACTTATTTCTGTGGT	(CAC) 6	56	266~295	[26]
TP6	TM080	F: AGATCCAAGCGAGTGTCTTA R: CAAATAGAGGCAATTCACCA	(CT) 11	56	100~120	[26]
TP7	TM089	F: CTATGATGCCTCCTTCAGTTT R: TGTACTTGATAACACGACCC	(AT) 10	56	236~260	[26]
TP8	TM099	F: CTGCCCCTGAAGTTAGTTTT R: ATCGAATCAACCATTAGAAAAGT	(TC) 10	56	236~268	[26]
TP9	TM103	F: GTCCCCATTGCTCTTAGTTT R: ATCATTGACCACCACATCAT	(TATGTG) 4	56	195~220	[26]
TP13	TM153	F: CACCCTTACCCTTCACA R: CCTTCCCACTACTTCCA	(ATTTT) 4	56	222~ 236	[27]
TP14	TM281	F: TTGGCAAACACCATTCAC R: GCTTGTGCTGGGAGGAG	(CT) 10	56	156~182	[27]
TP16	TM227	F: TGCGCTAATCCTTTGTCTTCC R: CGTTCATCGGCATTTCCA	(TTTT) 3	56	140~158	[27]
TP17	TM221	F: TCCGGACATCAAATAGTA R: AAGCCTCATGGACAGCC	(CACCAA) 3	56	90~108	[27]
TP19	TM212	F: GAAGCAGTCACAAGAGGC R: CCATCCCCAAATACCAAG	(TC) 9	56	241~265	[27]
TP20	TM213	F: CATTAACTGAAACAACCC R: GCAACAAGCCAGCCTCC	(TAGGGT) 3	56	205~ 230	[27]
TP21	TM205	F: CAAACCCCTGAAGCACAA R: GGACCCATCAACATTCC	(ACC) 6	56	123~160	[27]
TP22	TM207	F: TCCGCTTAGATCGCATA R: GGGCATCAGGTGAGGTG	(AAGAA) 3	56	243~265	[27]

表 2(续)

引物名称 Primer name	原始编号 Original code	引物序列 Primer sequences (5'-3')	重复单元 Repeat motif	退火温度 (°C) T _m	片段大小 (bp) Size	引物来源 Reference
TP23	TM208	F: CCTATTGTTTGCATCCC R: GACATTTATTGCTGCCACT	(TTTTGT) 3	56	128~158	[27]
TP25	TM188	F: TACACCGAAACAATCACC R: AAGTATAGCTTGGAGGAC	(AAGAA) 3	56	215~230	[27]
TP26	TM190	F: CACAATCAACCCCAATAA R: CAGTGGGTTTGAATCAT	(TTTTC) 5	56	105~135	[27]
TP27	TM048	F: CAGTTTGGTCTTTGTACTGT R: CTCAAATCAATCCCTTTCT	(AG) 9	58	139~174	[26]
TP28	TM050	F: CAACCCATCTCCATTCATCTC R: GGACTCCGCATCCATTACAG	(CT) 10	58	114~144	[26]
TP29	TM051	F: AATCATGCCCAAGGACATTC R: CAACCACTACCCATTTCAT	(GGT) 6	58	160~188	[26]
TP30	TM052	F: CCCAGCAATCTTGCTCAAT R: GTTGGCTTTCTGACCTATCC	(AG) 9	58	136~164	[26]
TP31	TM057	F: CTCGCTCCAATCATCACAA R: TGAAGGCCAGGAGAAAGAAA	(CTT) 9	58	320~352	[26]
TP32	TM058	F: CATTATCCCTTTCTTGTTCCA R: GGAGGGAGTAGGAGGTGGTCT	(TCA) 6	58	253~295	[26]
TP33	TM060	F: CTTGTGCTGGGAGGAGTTAT R: CTTTGGCAAACACCATTCAC	(CT) 8	58	158~178	[26]
TP34	TM062	F: TGAGACTCAGCCTCAGAACC R: ACCGATTGCTTTATACCCAG	(CT) 15	58	108~140	[26]
TP35	TM070	F: GCAGAAGGCGTTAGCAAATA R: GGAAGAGTGGGTGCGTGAAG	(CT) 11	58	150~186	[26]
TP36	TM075	F: ACCCCTGTTTCGTCTTTAC R: GTTGCTGTTGATTCGTCTGA	(CT) 16	58	157~179	[26]
TP37	TM083	F: CGAAGACGTCCTCACTAGAT R: TGTCTGTTGTTTCCAGGTAT	(TGAGT) 4	58	210~230	[26]
TP38	TM087	F: GGAACTCTTCTCCCTCCTTT R: CCTACTCAATCATAACAGCA	(CT) 17	58	203~253	[26]
TP39	TM107	F: TGTGGAATAAACAATCCTCCTG R: TCAAAGCAAAGCCCCTCAAT	(TTTGT) 4	58	135~160	[26]
TP40	TM114	F: AGACCTACCGAAGCGACCAA R: ACAGCCCAATCCATCCAAAT	(AT) 9	58	202~224	[26]
TP41	TM137	F: GGACTTGGAACCTCGCTCT R: ATCAGCAGCAACAGAACG	(AGAGA) 3	58	210~225	[27]
TP42	TM140	F: CCGGTGCATGATGTCTC R: CCGCAATCTCTCCAAT	(GCACCA) 3	58	264~300	[27]

1.3 SSR 数据分析

利用 PowerMarker 3.25 软件^[28] 进行 SSR 分型数据的遗传参数分析, 包括主要等位基因频率 (MAF, Major allele frequency)、等位基因数 (N_a , Number of alleles)、基因型数 (Genotype No.)、观察杂合度 (H_o , Observed heterozygosity)、多态信息含量 (PIC , Polymorphic information content) 等。

1.4 SSR 核心引物筛选与指纹图谱构建

利用 PI (Probability of identity) 和 $PIsibs$ 两个参数来评估 SSR 标记的鉴别力^[29], PI 表示两个随机个体拥有相同基因型的平均概率, 计算公式为:

$$PI = 2(\sum p_i^2)^2 - \sum p_i^4$$

$PIsibs$ 表示同胞群体 (Sibs) 中的 PI 值, 计算公式为:

$$PIsibs = 0.25 + (0.5\sum p_i^2) + [0.5(\sum p_i^2)^2] - (0.25\sum p_i^4)$$

两个公式中 p_i 均表示一个基因座上第 i 等位基因的频率。两个参数的计算利用 GenALEX 6.5 Excel 插件进行^[30]。核心引物的筛选参照 Tan 等^[14] 描述的原则进行, 筛选出的核心引物按照引物顺序和各引物扩增条带大小构建指纹图谱。

1.5 GBS 文库构建与测序

用于 GBS 文库构建的 DNA 样品经检测合格后送北京诺禾致源科技股份有限公司进行测序文库构建和测序。GBS 文库构建参照 Elshire 等^[31] 的方法进行, 即质检合格的基因组 DNA 用 Mse I 限制性内切酶进行酶切, 在酶切后的片段两端加上相应的适配接头, 利用 PCR 扩增两接头之间的片段, 对样品混合之后进行电泳回收纯化, 纯化产物质检合格后利用 Illumina HiSeq4000 平台进行双末端 PE150 测序。

1.6 全基因组 SNP 挖掘

测序获得的原始序列数据 (Raw reads) 进行过滤, 包括去除含有接头序列的 reads、单端序列中未知碱基 (N) 超过该序列长度 10% 的序列、低质量碱基数超过该序列长度 50% 的序列, 最终得到高质量序列数据 (Clean reads) 并存储为 Fastq 格式。利用 Burrows-Wheeler Aligner (BWA) 软件^[32] 将 Clean reads 比对到茶树基因组 (http://www.plantkingdomgdb.com/tea_tree/)^[1]; 利用 SAMTOOLS 软件^[33] 进行 SNP 挖掘, 获得的 SNP 经测序深度 $3\times$ 、Miss0.3 和次要等位基因频率 >0.01 的条件质检过滤后用于后续分析。利用 ANNOVAR 软件^[34] 工具对高质量的 SNPs 进行变异类型注释; 利用 OmicShare 平台的 GO 富集分析工具 (<http://www.omicshare.com/tools/Home/Soft/gogsea>) 对基因

SNPs 相关基因进行 GO 功能注释分析。

1.7 基于全基因组 SNP 的茶树遗传结构分析

利用鉴定到的 SNP 进行遗传距离 (p-distance) 计算, 两个个体间遗传距离的计算方法参照 Yang 等^[20] 的描述。利用 TreeBest (<http://treesoft.sourceforge.net/treebest.shtml>) 软件计算距离矩阵, 在距离矩阵基础上运用邻接法 (NJ-method) 构建系统进化树 (bootstrap values=1000); 利用 GCTA (<http://cns.genomics.com/software/gcta/pca.html>) 软件计算特征向量以及特征值进行主成分分析 (PCA); 利用 PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/>) 进行遗传结构分析, 并借助 Admixture 软件构建遗传结构。

2 结果与分析

2.1 基于 SSR 标记的遗传多样性分析

基于 35 对 SSR 引物, 共检测到 317 个等位基因位点, 每对引物检测到的等位基因位点数介于 3~16 之间, 平均每对引物扩增等位基因数为 9.06; 主要等位基因频率为 0.11~0.86, 平均 0.35; 每对引物检测的基因型数为 3~37, 平均 16.60; 观察杂合度介于 0~0.87 之间, 平均 0.46; 多态信息含量介于 0.22~0.92 之间, 平均 0.73 (表 3)。

2.2 核心 SSR 引物筛选及指纹图谱构建

PI 和 $PIsibs$ 用于评估 SSR 标记的指纹鉴别力, 对于每对引物, PI 介于 0.01~0.59 之间, $PIsibs$ 介于 0.29~0.78 之间; 对于两个随机个体在这 35 对 SSR 引物下具有相同基因型的概率为 4.5×10^{-42} , 而在这 35 对 SSR 引物下的 $PIsibs$ 为 3.6×10^{-15} (表 3)。对于仅有 1 对 SSR 引物来说, 此时 $PI=0.04$, 当有 2 对 SSR 引物时, $PI=0.001$ (图 1A), 而 PI 介于 0.0001~0.01 之间已足够鉴定大多数自然群体^[29], 也就意味着这 35 对 SSR 引物中的任意两对一同使用就能区分本研究中的茶树材料。然而考虑到位点缺失的存在, 最终根据 Tan 等^[14] 描述的原则, 本研究从中选取了 5 对 SSR 引物 (TP4、TP23、TP28、TP32、TP34) 作为核心引物, 它们的等位基因位点频率如图 1B 所示, $PI=9.01 \times 10^{-8}$, $PIsibs=4.76 \times 10^{-3}$, 预示着它们足够用于更大群体的指纹识别。按照这 5 对 SSR 引物的顺序 (TP4-TP23-TP28-TP32-TP34) 以及每对引物的等位基因大小顺序对本研究的茶树材料构建了指纹鉴定码 (表 4)。

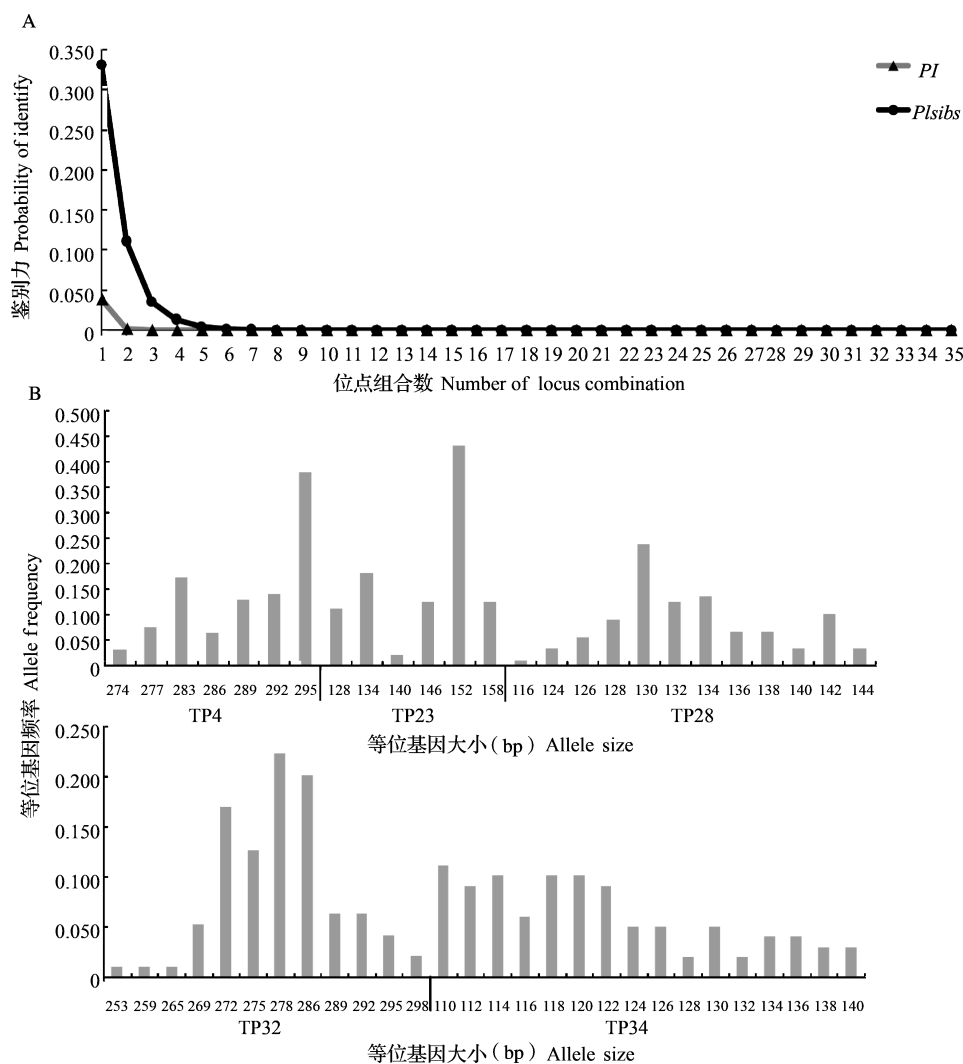
表 3 35 对 SSR 标记的遗传参数统计

Table 3 Statistics of the 35 SSR markers

标记 Marker	主要等位基 因频率 MAF	基因型数 Genotype No.	等位基因数 <i>Na</i>	观察杂合度 <i>Ho</i>	多态信息含量 <i>PIC</i>	<i>PI</i>	<i>PIsibs</i>
TP1	0.19	22	9	0.53	0.84	0.04	0.33
TP2	0.20	19	11	0.28	0.83	0.04	0.33
TP3	0.24	26	14	0.52	0.86	0.03	0.32
TP4	0.36	15	7	0.30	0.76	0.07	0.38
TP6	0.24	19	11	0.66	0.84	0.04	0.33
TP7	0.14	26	14	0.42	0.89	0.02	0.31
TP8	0.20	26	14	0.37	0.89	0.02	0.31
TP9	0.28	15	6	0.76	0.75	0.08	0.38
TP13	0.71	3	3	0	0.36	0.39	0.63
TP14	0.34	18	10	0.35	0.80	0.05	0.35
TP16	0.56	9	4	0.39	0.55	0.21	0.50
TP17	0.49	7	4	0.87	0.59	0.18	0.47
TP19	0.42	10	8	0.71	0.71	0.10	0.40
TP20	0.62	5	4	0.04	0.47	0.28	0.55
TP22	0.59	6	4	0.05	0.51	0.25	0.53
TP21	0.51	14	10	0.39	0.67	0.12	0.43
TP23	0.44	12	6	0.74	0.70	0.11	0.41
TP25	0.52	9	5	0.40	0.59	0.18	0.47
TP26	0.35	12	5	0.41	0.71	0.10	0.40
TP27	0.20	19	13	0.52	0.87	0.02	0.31
TP28	0.23	22	11	0.83	0.86	0.03	0.32
TP29	0.20	23	10	0.60	0.84	0.04	0.33
TP30	0.16	26	12	0.60	0.88	0.02	0.31
TP31	0.27	25	12	0.66	0.82	0.04	0.34
TP32	0.20	21	12	0.32	0.84	0.04	0.33
TP33	0.35	18	9	0.33	0.80	0.05	0.36
TP34	0.11	37	16	0.87	0.92	0.01	0.29
TP35	0.18	29	16	0.71	0.88	0.02	0.31
TP36	0.21	23	12	0.54	0.86	0.03	0.32
TP37	0.86	3	3	0.23	0.22	0.59	0.78
TP38	0.13	22	16	0.33	0.91	0.01	0.30
TP39	0.36	12	5	0.64	0.68	0.12	0.42
TP40	0.41	16	12	0.35	0.71	0.10	0.40
TP41	0.54	7	4	0.45	0.55	0.21	0.50
TP42	0.50	5	5	0	0.57	0.20	0.48
平均 Mean [*]	0.35	16.60	9.06	0.46	0.73	4.5×10^{-42}	3.6×10^{-15}

^{*}: 对于 *PI* 和 *PIsibs*, 此处的值是所有标记位点组合使用时的可能性

^{*}: For the *PI* and *PIsibs*, the values are the combined probabilities, which are the products of the *PI* or *PIsibs* of all individual loci



A: SSR 标记的鉴别力评估; B: 5 对核心引物的基因型频率分布

A: Evaluation of the probability of identification of the SSR markers, B: Allele sizes and frequencies for the five core SSR markers.

图 1 SSR 标记在茶树样品中的指纹鉴别力评估

Fig.1 Evaluating the fingerprinting power of SSR markers in tea plant samples

2.3 测序质量

54 份茶树材料共获得 40.27 G 测序数据, 经过滤后高质量测序数据共计 40.26 G, 平均每个材料获得 0.76 G 数据, 大小约为茶树参考基因组的 25%。54 份样品的测序质量较高 ($Q20 \geq 96.2\%$, $Q30 \geq 90.5\%$), 平均 GC 含量 39.3%, 每个样品获得的 Clean reads 介于 3.6 百万~8.0 百万之间, 平均 5.6 百万条, 最终平均 5.4 百万条序列比对到茶树参考基因组, 平均比对率 96.7%。测序获得的原始数据 (Raw data) 已上传 BIG Data Center 的 Genome Sequence Archive (GSA) 数据库, 登录号为 CRA000946。

2.4 SNP 类型与分布

54 份茶树材料共获得 698117 个 SNP, 对 SNP

的 6 种变异类型统计显示 A/G 和 C/T 替换的频率远高于其他 4 种 (图 2A); SNP 大多数 (89%) 分布在基因间区, 另外有 2% 左右的 SNP 分布在基因外显子区 (图 2B); 对于分布于基因编码区的 SNP, 共涉及 3535 个茶树基因 (约占茶树参考基因组基因数量的 0.1%), 其中 8439 个 SNP 导致 2887 个基因发生同义突变, 6898 个 SNP 导致 2729 个基因发生非同义突变, 非同义突变与同义突变的比率为 0.82, 此外还分别有 271 个和 21 个 SNP 分别导致了 227 个和 21 个基因编码提前终止和编码延伸 (图 2C); 进一步分析显示每个基因的 SNP 数目介于 1~35 之间, 平均一个基因中含有 4 个 SNP (图 2D)。

表 4 基于 5 对 SSR 引物的茶树材料指纹识别码

Table 4 The fingerprinting codes of the tea plant samples based on five SSR markers

代号 ID	SSR 指纹识别码 SSR fingerprinting code [*]	代号 ID	SSR 指纹识别码 SSR fingerprinting code [*]
BH1	A0B152C130-138D275-295E110	J1	A0B152-134C130D278E120
C550	A283B152C128-132D286-298E118-132	JG1	A283-289B0C132-142D269E114-138
CQ1	A295B152-134C130-138D278-286E110-118	LJ43	A289-295B0C0D0E110
DH1	A277B152C130-140D289-292E120-140	MP1	A292-295B152-134C132D278-286E112-120
DH2	A292-295B152C130D259-275E110-118	MS131	A295B0C130-138D0E0
DH3	A286-292B0C128D278E112-122	NW	A283B152C130-136D0E114-120
DH4	A277B158-145C0D275-286E122-128	QC1	A295B146-128C128-138D289E112-134
DH5	A295B0C130D272-275E0	QC7	A295B152-134C126-140D272E112-118
DH6	A289B158-152C134-142D286E118-134	QC8	A294B146-126C128-137D292E120-129
DZ1	A292-295B152-134C0D286E0	QF4	A295B146-128C126-134D278-286E112
DZ2	A286B152-140C132-136D295E112-120	QM308	A283B158-134C0D272-272E114-136
DZ3	A292-295B158-134C0D275E124	QM419	A295B152-134C130-138D0E112
FD05	A277B152-128C126-136D278E110-122	QM502	A274B152-134C124-130D278E116-122
FD06	A295B0C134-144D275E120-126	QM601	A0B158C0D286E130-140
FD1	A292B0C128D0E114-126	QM701	A286-292B158-152C130-134D269-275E116-126
FX1	A295B0C132-136D272-289E114-122	QM809	A283-283B146-128C130-138D253-286E118-136
FX2	A283B152-128C126-134D286E112-124	QZ1	A286B152-134C130-134D272E0
GY8	A289B152-128C0D278E110-122	QZ2	A286B152-128C132-142D278E122-122
GYH1	A295B158-146C134-142D289E118-136	TB0322	A295B0C130-142D286-295E112-134
GYH10	A289B152C132-142D272E114	TC12	A0B152C134-144D272E114
GYH12	A283B152-146C132-136D286E118-130	TX1	A292B146-128C132-140D278E126-136
GYH2	A292B158-134C130D272E120-130	TX2	A295B152C0D278E110-128
GYH4	A277-295B152-134C132-138D286E118-124	TX3	A295B152-134C134-144D272-292E124-138
GYH5	A283-292B152-128C130D0E114-122	TX4	A289-295B152C130-142D272E110-116
GYH7	A283-295B152-134C0D0E0	YG0843	A295B146C134-142D292E118-138
GYH8	A289B158-134C124-134D275E120-132	ZJ1	A274-289B158-134C134-142D275E116-130
GYH9	A295B146-139C124-128D265-289E110-134	ZZ1	A283-295B152-134C130-136D269E116-126

*: A、B、C、D、E 分别表示引物 TP4、TP23、TP28、TP32 和 TP34; “0” 表明没有条带

*: A, B, C, D and E represent the primer TP4, TP23, TP28, TP32 and TP34, respectively, “0” indicate no amplification

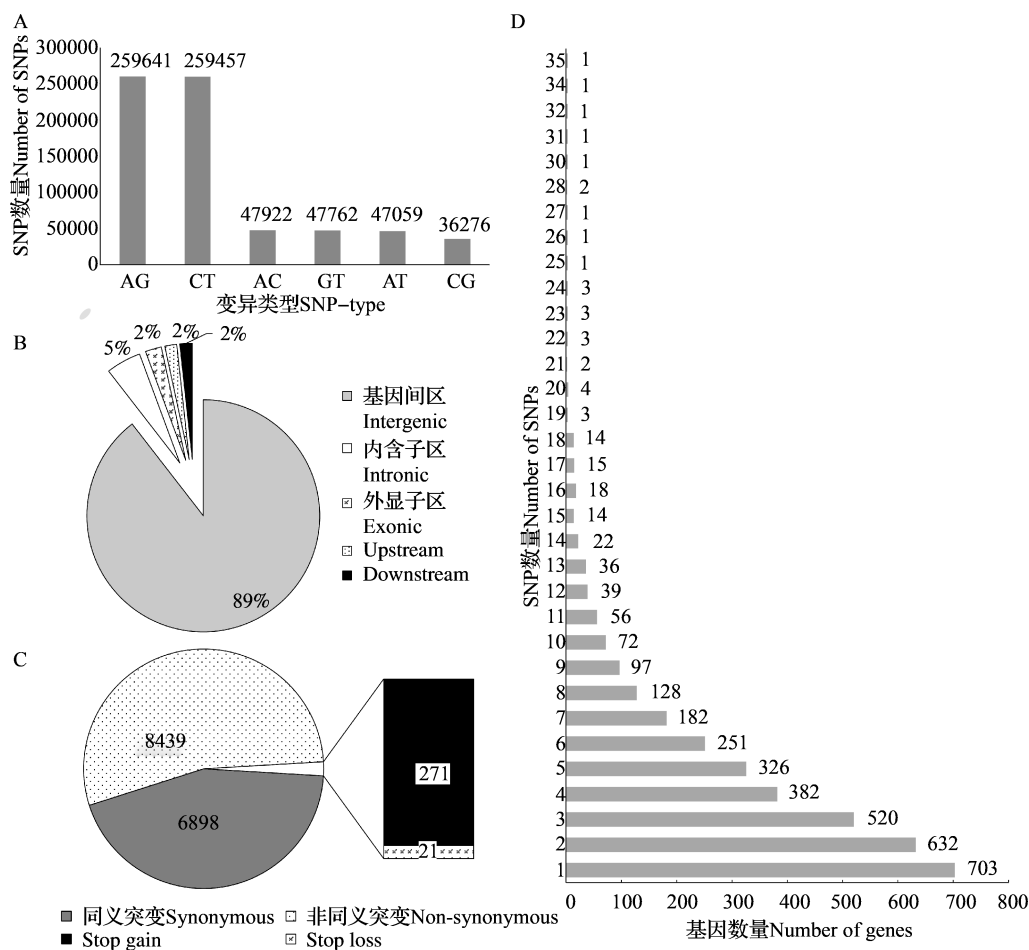
2.5 基因 SNP 相关基因的功能分析

对于上述的 3535 个茶树基因进一步比对茶树参考基因组的功能注释信息 (http://www.plantkingdomgdb.com/tea_tree/data/fun/), 其中 1275 个基因比对到 GO 类别的生物过程 Biological Process、487 个基因比对到 GO 类别的细胞组分 Cellular Component、1823 个基因比对到 GO 类别的分子功能 Molecular Function (图 3)。对于生物过程 Biological Process 类别下涉及代谢过程 metabolic process (GO: 0008152)、细胞过程 cellular process (GO: 0009987) 和单组织过程 single-organism process (GO: 0044699) 等亚类的基因数目最多; 对于细胞组分 Cellular Component 类别下涉及膜 membrane (GO: 0016020)、细胞 cell (GO: 0005623) 和细胞部分 cell part (GO: 0044464) 亚类的基因数目最多; 对于分子功能 Molecular Function 类别下涉及结合 binding (GO: 0005488) 和催化活

性 catalytic activity (GO: 0003824) 亚类的基因数目最多。

2.6 基于全基因组 SNP 的遗传结构分析

基于高质量的全基因 SNP 对 54 份茶树材料的遗传关系进行分析。根据遗传距离的系统进化显示 54 份材料可以分为 4 个类群 (图 4A); 根据已知的部分育种信息, 其中 FX1、FX2、FD05、FD06 和 QM809 都具有福鼎大白茶 (FD1) 遗传背景, QM308、QM502、QM601、QM701 以及 ZJ1 都具有云南大叶种的遗传背景, 在本研究中, 前者都聚在了第 I 类群、后者都聚在了第 II 类群; 此外, TB0322 是 QC1 的自然突变单株、QF4 为 QM419 的种子辐照后代, 它们也分别聚在了一起, 该结果不仅从分子层面证实了它们的遗传关系, 同时也反应出基于全基因 SNP 进行种质鉴定的可靠性。主成分分析的结果与系统进化分析结果高度一致 (图 4B), 同时也显示出了这些材料的遗传多样性。对比育成品种



A: 6 种 SNPs 变异类型统计; B: SNPs 在基因组不同位置的分布; Upstream/Downstream 分别表示 SNP 位于基因上游 / 下游 1kb 区域;

C: CDS 区域 SNPs 类型分布; Stop gain/Stop loss 分别表示获得 / 失去终止密码子; D: 不同基因内 SNP 数量统计

A: Summary of the six variation types SNPs identified in tea plant, B: Distribution of SNPs on the basis of their location in different genomic regions. Upstream, the SNP located in 1kb upstream of a gene, Downstream, the SNP located in 1kb downstream of a gene, C: The distribution statistics of SNPs in the CDS region. Stop gain, the variation causes the gene to be terminated, Stop loss, the variation causes the gene to lose the terminator codon, D: The statistics of SNPs in different genes

图 2 SNP 类型与分布统计

Fig.2 The distribution statistics of the identified SNPs

和新选育的新品系,结果发现二者间的遗传背景具有重叠性;而 GYH 系列为近年来从贵州沿河土家族自治县收集的古茶树资源中单株选育的新品系,这些资源内部遗传背景较为相近,而与目前已育成的品种具有明显的遗传差异。进一步分析这 54 份茶树材料的遗传结构,在 $K=2$ 时,这些资源可以粗略分为具有贵州本地茶树遗传背景的材料和不具有贵州本地茶树遗传背景的材料两大类;在 $K=3$ 时,又可以从具有贵州本地茶树遗传背景的材料中分出同时具有贵州和云南茶树资源遗传背景的材料;在 $K=4$ 时,遗传结构的结果与系统进化以及主成分分析的结果呈现出高度一致性(图 4C)。

在系统进化结果的第 I 类群中,除了已知的 FD1 及其上述 5 个后代材料外,剩余的 QC1 和 TX3

是从湄潭苔茶群体种中选育的材料、QM419 为贵州地方品种的自然杂交后代,MS131 虽然是四川选育的品种,但是有研究^[14]已证实其属于福鼎家族,因此可以推断 QC1、TX3 和 QM419 的遗传背景中应该存在福鼎的血缘,再结合它们遗传结构关系,进而可以将第 I 类群归为具有福鼎大白茶遗传背景的材料。对于第 II 类群,除了已知的以云南大叶种作为亲本之一的 5 份材料外, J1 为贵州久安古茶树群体单株选育材料、YG0843 为贵州云贵山自然群体单株选育材料、QZ1 为昆明中叶种自然杂交后代、DH3 遗传背景未知,根据本研究的结果将它们归为具有云南茶树资源背景的材料。对于第 III 类群,除了 DZ1~DZ3 为祁门群体种单株选育材料外,其余材料均为不同省市的育成品种,且根据记录并没有

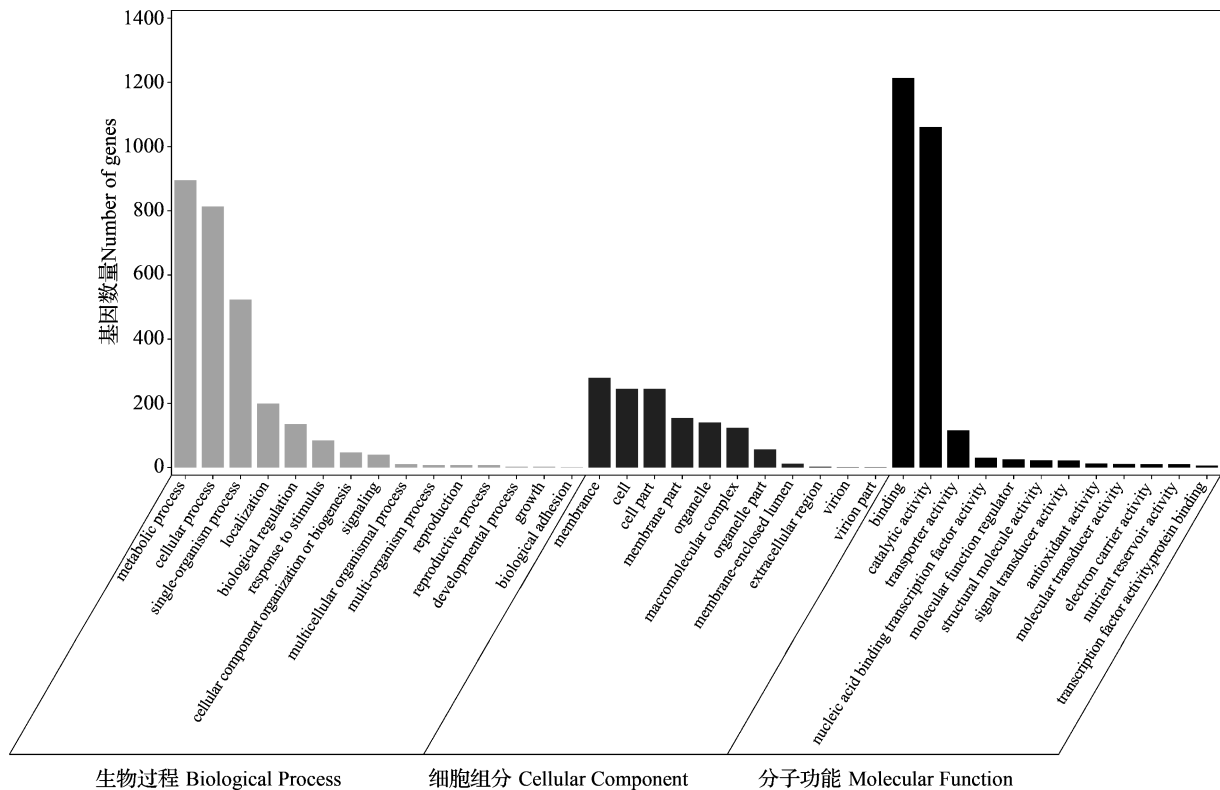


图3 基因 SNP 相关茶树基因的 GO 功能分类

Fig.3 Gene Ontology classifications of the identified genic SNP-associated tea unigenes

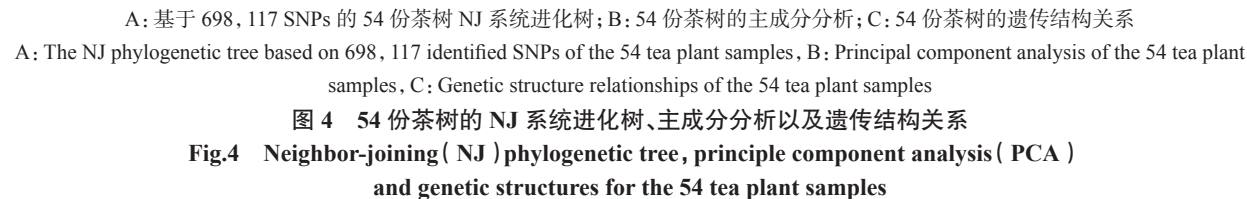
显示出它们具有相同的亲本或者相似的遗传背景,这可能与不同省市间资源交换有关,因此结合遗传结构关系将此类群归为其他类型材料。对于第IV类群, DH 系列遗传背景未知, GYH 系列为贵州沿河古茶树群体单株选育材料, TX 系列、C550 和 ZZ1 为湄潭苔茶群体单株选育材料, QC7 和 QC8 也均以湄潭苔茶作为亲本之一,因此将该类群归为贵州本地茶树材料,这同样可以通过遗传结构关系进一步得以确认。

3 讨论

基于 SSR 标记的分子指纹图谱是植物新品种或者新种质快速鉴定的有效方法,一方面可以进行品种真实性的鉴定来保护品种权所有者的合法权益,另一方面通过新种质与已知种质资源的对比鉴定能够进行未知种质资源类别划分便于育种应用^[10,13]。近年来,随着贵州茶园面积的激增,对优良茶树品种的需求也在增加。然而茶树生长周期长,因此能够在早期阶段对茶树品种真实性进行准确鉴定是有效剔除混杂、保证优质茶园建设的关键。本研究以贵州省目前主要的栽培茶树品种为材料,利用 SSR 标记进行指纹鉴定图谱构建,35 对 SSR 引物的平均 PIC 为 0.73,意味着从中挑选少量的引

物就可以区分出本研究的所有材料。为了评估引物组合的鉴别力,本研究进一步借助 *PI* 和 *PIsibs* 两个参数评估了每一对引物的鉴别力,任意两对 SSR 引物的 *PI* 平均为 0.001,而 *PI* 介于 0.0001~0.01 之间的标记已足够用于大多数自然群体的鉴定^[29],表明本研究中的任意两对引物就可以区分本研究的所有材料;另外考虑到本研究采用的是毛细管电泳,对等位基因的区分精度可以达到 1 bp,因此为了降低不同检测方法带来的误差,本研究参考前人的措施^[10,14,35]最终挑选 5 对 SSR 引物作为核心引物,构建了贵州省主要栽培茶树品种的指纹图谱。

对于茶树育种来说,从优良品种中进一步进行自然变异优良单株筛选或者进行优良品种间的杂交是新品种选育的重要策略^[5];而对于杂交育种,亲本间的遗传差异越大,杂交后代的遗传分离越广泛,因此对杂交亲本遗传背景的认识是开展杂交育种的重要前提。对于本研究的 54 份材料,其中一半为已育成品种,而另一半为新选育的品系,但是它们真实的遗传背景并不十分清晰。考虑到这些材料很可能继续作为亲本应用到后续的育种利用中,本研究进一步利用高通量测序技术对这些材料的全基因组 SNP 进行了挖掘,以期基于全基因组 SNP 深入明晰它们的遗传背景。54 份材料共鉴定到 698117 个高



传关系的分析结果以及已知品种的部分亲本来源信息,可以初步推测前期贵州茶树育种的一般模式:即以福鼎大白茶为亲本之一的遗传改良、以云南茶树资源为亲本的杂交选育、以其他外引品种为材料的单株选育以及对本地资源的单株选择,其中福鼎大白茶和云南茶树种质如云南大叶种等的应用占着主导地位,而对其他外引品种以及本地资源的利用还很有限。GYH 系列为从贵州沿河古茶树资源中

单株选育的新品系,它们之间的遗传背景相近而与育成品种间的亲缘关系较远,预示着本地的古茶树资源中蕴含着新的变异,具有潜在的育种利用价值。因此在后续的育种实践中可以考虑加强省外优良茶树品种以及贵州省优良古茶树资源的利用。

需要指出的是,本研究鉴定到的高质量 SNP 中有 2% 左右分布在基因 CDS 区域,共涉及 3535 个茶树功能基因,对这些功能基因进行 GO 功能注释表明大多数基因涉及茶树的代谢过程。类黄酮、茶氨酸、咖啡碱等次生代谢物的动态变化影响着茶叶的品质^[36],因此这些基因 SNP 可能在茶树次生代谢相关基因的功能调控中发挥着作用,这需要后续的研究。

致谢: 本文的毛细管电泳仪操作得到了贵州省农业科学院辣椒研究所宋拉拉的热情指导帮助,在此表示诚挚谢意!

参考文献

- [1] Xia E H, Zhang H B, Sheng J, Li K, Zhang Q J, Kim C, Zhang Y, Liu Y, Zhu T, Li W, Huang H, Tong Y, Nan H, Shi C, Shi C, Jiang J J, Mao S Y, Jiao J Y, Zhang D, Zhao Y, Zhao Y J, Zhang L P, Liu Y L, Liu B Y, Yu Y, Shao S F, Ni D J, Eichler E E, Gao L Z. The tea tree genome provides insights into tea flavor and independent evolution of caffeine biosynthesis. *Molecular Plant*, 2017, 10 (6): 866-877
- [2] Chen L, Apostolides Z, Chen Z M. Global tea breeding: achievements, challenges and perspectives. Hangzhou: Springer-Zhejiang University Press, 2012: 1
- [3] 乔小燕,陈栋,李家贤,吴华玲,黄华林,何玉媚,晏嫦好. 植物功能基因组研究主要方法在茶树上的应用. *茶叶科学*, 2012, 32 (4): 362-368
Qiao X Y, Chen D, Li J X, Wu H L, Huang H L, He Y M, Yan C Y. Methods of plant functional genomics and their application in tea plant (*Camellia sinensis*). *Journal of Tea Science*, 2012, 32 (4): 362-368
- [4] Mukhopadhyay M, Mondal T K, Chand P K. Biotechnological advances in tea (*Camellia sinensis* [L.] O. Kuntze): a review. *Plant Cell Reports*, 2016, 35 (2): 255-287
- [5] Chen L, Zhou Z X, Yang Y J. Genetic improvement and breeding of tea plant (*Camellia sinensis*) in China: from individual selection to hybridization and molecular breeding. *Euphytica*, 2007, 154: 239-248
- [6] 马建强,姚明哲,陈亮. 茶树种质资源研究进展. *茶叶科学*, 2015, 35 (1): 11-16
Ma J Q, Yao M Z, Chen L. Research progress on germplasms of tea plant (*Camellia sinensis*). *Journal of Tea Science*, 2015, 35 (1): 11-16
- [7] 巫桂芬,徐鲜均,徐建堂,陶爱芬,张立武,魏丽真,潘漠,方平平,林荔辉,祁建民. 利用 SRAP、ISSR、SSR 标记绘制黄麻基因源分子指纹图谱. *作物学报*, 2015, 41 (3): 367-377
Wu G F, Xu X J, Xu J T, Tao A F, Zhang L W, Wei L Z, Pan M, Fang P P, Lin L H, Qi J M. Construction of molecular fingerprinting map in gene pool of Jute with SRAP, ISSR, and SSR markers. *Acta Agronomica Sinica*, 2015, 41 (3): 367-377
- [8] 查美琴,赵玉玲,李疆,朱瑜,罗淑萍,胡月. 新疆枸杞种质资源遗传多样性分析及 DNA 指纹图谱构建. *植物遗传资源学报*, 2016, 17 (6): 1065-1071
Zha M Q, Zhao Y L, Li J, Zhu Y, Luo S P, Hu Y. Analysis of genetic diversity and construction of DNA fingerprint of germplasm resources of Wolfberry (*Lycium barbarum* L.) revealed by SCoT in Xinjiang of China. *Journal of Plant Genetic Resources*, 2016, 17 (6): 1065-1071
- [9] 赵久然,李春辉,宋伟,王元东,张如养,王继东,王凤格,田红丽,王蕊. 基于 SNP 芯片揭示中国玉米育种种质的遗传多样性与群体遗传结构. *中国农业科学*, 2018, 51 (4): 626-634
Zhao J R, Li C H, Song W, Wang Y D, Zhang R Y, Wang J D, Wang F G, Tian H L, Wang R. Genetic diversity and population structure of important Chinese maize breeding germplasm revealed by SNP-chips. *Scientia Agricultura Sinica*, 2018, 51 (4): 626-634
- [10] Liu S, Liu H, Wu A, Hou Y, An Y, Wei C. Construction of fingerprinting for tea plant (*Camellia sinensis*) accessions using new genomic SSR markers. *Molecular Breeding*, 2017, 37 (8): 93
- [11] 李莉,王俊峰,颜廷进,李娜娜,丁汉凤,樊守金. 基于 SSR 标记的山东省小麦 DNA 指纹图谱的构建. *植物遗传资源学报*, 2013, 14 (3): 537-541
Li L, Wang J F, Yan T J, Li N N, Ding H F, Fan S J. Establishment of DNA fingerprinting for wheat in Shandong province by SSR markers. *Journal of Plant Genetic Resources*, 2013, 14 (3): 537-541
- [12] 张云,龙云星,奎丽梅,徐高峰,涂建,辜琼瑶,李华慧,黄平,卢义宣. 我国水稻两用核不育系 SSR 遗传多样性分析. *分子植物育种*, 2017, 15 (7): 2836-2846
Zhang Y, Long Y X, Kui L M, Xu G F, Tu J, Gu Q Y, Li H H, Huang P, Lu Y X. SSR analysis on genetic diversity of rice dual-use nuclear male sterile lines in China. *Molecular Plant Breeding*, 2017, 15 (7): 2836-2846
- [13] 张成才,刘园,姜燕华,吴立赞,王丽鹭,韦康,成浩. SSR 标记鉴定浙江省主要无性系茶树品种的研究. *植物遗传资源学报*, 2014, 15 (5): 926-931
Zhang C C, Liu Y, Jiang Y H, Wu L Y, Wang L Y, Wei K, Cheng H. Application of SSR markers in cultivar identification of clonal tea plant in Zhejiang province, China. *Journal of Plant Genetic Resources*, 2014, 15 (5): 926-931
- [14] Tan L Q, Peng M, Xu L Y, Wang L Y, Chen S X, Zou Y, Qi G N, Cheng H. Fingerprinting 128 Chinese clonal tea cultivars using SSR markers provides new insights into their pedigree relationships. *Tree Genetics & Genomes*, 2015, 11 (5): 90
- [15] Meegahakumbura M K, Wambulwa M C, Li M M, Thapa K K, Sun Y S, Moller M, Xu J C, Yang J B, Liu J, Liu B Y, Li D Z, Gao L M. Domestication origin and breeding history of the tea plant (*Camellia sinensis*) in China and India based on nuclear microsatellites and cpDNA sequence data. *Frontiers in Plant Science*, 2018, 8: 2270
- [16] Cai C P, Zhu G Z, Zhang T Z, Guo W Z. High-density 80 K SNP array is a powerful tool for genotyping *G. hirsutum* accessions and genome analysis. *BMC Genomics*, 2017, 18 (1): 654
- [17] Cheng X, Ren Y H, Jian Y Q, Guo Z F, Zhang Y, Xie C X, Fu

- J J, Wang H W, Wang G Y, Xu Y B, Li P, Zou C. Development of a maize 55 K SNP array with improved genome coverage for molecular breeding. *Molecular Breeding*, 2017, 37(3): 20
- [18] Chen H, Xie W, He H, Yu H, Chen W, Li J, Yu R, Yao Y, Zhang W, He Y, Tang X, Zhou F, Deng X W, Zhang Q. A high-density SNP genotyping array for rice biology and molecular breeding. *Molecular Plant*, 2014, 7(3): 541-553
- [19] Fang W P, Meinhardt L W, Tan H W, Zhou L, Mischke S, Zhang D. Varietal identification of tea (*Camellia sinensis*) using nanofluidic array of single nucleotide polymorphism (SNP) markers. *Horticulture Research*, 2014, 1: 14035
- [20] Yang H, Wei C L, Liu H W, Wu J L, Li Z G, Zhang L, Jian J B, Li Y Y, Tai Y L, Zhang J, Zhang Z Z, Jiang C J, Xia T, Wan X C. Genetic divergence between *Camellia sinensis* and its wild relatives revealed via genome-wide SNPs from RAD sequencing. *PLoS One*, 2016, 11(3): e0151424
- [21] Ma J Q, Huang L, Ma C L, Jin J Q, Li C F, Wang R K, Zheng H K, Yao M Z, Chen L. Large-scale SNP discovery and genotyping for constructing a high-density genetic map of tea plant using specific-locus amplified fragment sequencing (SLAF-seq). *PLoS One*, 2015, 10(6): e0128798
- [22] 王小柯, 江东, 孙珍珠. 利用 GBS 技术研究 240 份宽皮柑橘的系统演化. *中国农业科学*, 2017, 50(9): 1666-1673
Wang X K, Jiang D, Sun Z Z. Study on phylogeny of 240 mandarin accessions with genotyping-by-sequencing technology. *Scientia Agricultura Sinica*, 2017, 50(9): 1666-1673
- [23] Uncu A O, Fray A, Karlovsky P, Doganlar S. High-throughput single nucleotide polymorphism (SNP) identification and mapping in the sesame (*Sesamum indicum* L.) genome with genotyping by sequencing (GBS) analysis. *Molecular Breeding*, 2016, 36(12): 173
- [24] Verma S, Gupta S, Bandhiwal N, Kumar T, Bharadwaj C, Bhatia S. High-density linkage map construction and mapping of seed trait QTLs in chickpea (*Cicer arietinum* L.) using genotyping-by-sequencing (GBS). *Scientific Reports*, 2015, 5: 17512
- [25] 唐玉海, 郭春芳, 张木清. 一种提取茶树基因组 DNA 的方法: 改良的 CTAB 法. *福建教育学院学报*, 2007(1): 99-101
Tang Y H, Guo C F, Zhang M Q. A modified method for extracting genome DNA of tea-modified CTAB method. *Journal of Fujian Institute of Education*, 2007(1): 99-101
- [26] Ma J Q, Zhou Y H, Ma C L, Yao M Z, Jin J Q, Wang X C, Chen L. Identification and characterization of 74 novel polymorphic EST-SSR markers in the tea plant, *Camellia sinensis* (Theaceae). *American Journal Botany*, 2010, 97(12): e153-e156
- [27] Yao M Z, Ma C L, Qiao T T, Jin J Q, Chen L. Diversity distribution and population structure of tea germplasms in China revealed by EST-SSR markers. *Tree Genetics & Genomes*, 2012, 8(1): 205-220
- [28] Liu K, Musel S V. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, 2005, 21: 2128-2129
- [29] Waits L P, Luikart G, Taberlet P. Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology*, 2001, 10: 249-256
- [30] Peakall R, Smouse P E. GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, 2012, 28(19): 2537-2539
- [31] Elshire R J, Glaubitz J C, Sun Q, Poland J A, Kawamoto K, Buckler E S, Mitchell S E. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*, 2011, 6(5): e19379
- [32] Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics*, 2009, 25(14): 1754-1760
- [33] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. The sequence alignment/map format and SAMtools. *Bioinformatics*, 2009, 25(16): 2078-2079
- [34] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*, 2010, 38(16): e164
- [35] 陈世军, 张明泽, 姚玉仙, 谢维斌. 基于 SSR 标记的黔南茶树种质资源 DNA 指纹图谱构建. *植物遗传资源学报*, 2017, 18(1): 106-111
Chen S J, Zhang M Z, Yao Y X, Xie W B. Establishment of DNA fingerprinting for tea germplasm from Qiannan prefecture by SSR markers. *Journal of Plant Genetic Resources*, 2017, 18(1): 106-111
- [36] Wei C L, Yang H, Wang S B, Zhao J, Liu C, Gao L P, Xia E H, Lu Y, Tai Y L, She G B, Sun J, Cao H S, Tong W, Gao Q, Li Y Y, Deng W W, Jiang X L, Wang W Z, Chen Q, Zhang S H, Li H J, Wu J L, Wang P, Li P H, Shi C Y, Zheng F Y, Jian J B, Huang B, Shan D, Shi M M, Fang C B, Yue Y, Li F D, Li D X, Wei S, Han B, Jiang C J, Yin Y, Xia T, Zhang Z Z, Bennetzen J L, Zhao S C, Wan X C. Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality. *Proceedings of the National Academy of Sciences of the United States of America*, 2018, 115(18): E4151-E4158