植物遗传资源学报 2019，20（1）：26-36
Journal of Plant Genetic Resources
DOI：10.13430/j.cnki.jpgr.20180529001

基于全基因组 SNP 的贵州久安古茶树遗传关系分析

郭　燕，乔大河，杨　春，李　燕，陈正武，陈　娟
（贵州省农业科学院茶叶研究所，贵阳 550006）

摘要：贵州久安分布着大量的古茶树资源。为了明晰这些茶树资源间的遗传关系，本研究以分布于久安 5 个不同区域的 100 份古茶树为材料，首先利用 GBS（genotyping by sequencing）技术对它们的全基因组 SNP 进行了鉴定，然后基于鉴定到的 SNP 进行了系统进化树构建、主成分分析以及遗传结构分析。100 份古茶树材料共获得 548597 个高质量 SNP，并对这些 SNP 进行了变异类型注释。系统进化树、主成分分析以及遗传结构分析的结果高度一致，结果表明同一区域内的材料间亲缘关系较近；100 份资源可以分为 3 个类群；古茶园（G）的材料与其他 4 个区域的材料遗传背景差异较大，这 4 个区域的资源可能有相同的亲本来源。

关键词：古茶树；SNP；遗传结构；进化关系

Genetic Diversity of Old Tea Plant Resources in Jiuan City of Guizhou Province, Using Genome-Wide SNP

GUO Yan, QIAO Da-he, YANG Chun, LI Yan, CHEN Zheng-wu, CHEN Juan
（*Institute of Tea, Guizhou Academy of Agricultural Science, Guiyang 550006*）

Abstract: A large number of old tea plant resource is widely distributed in Jiuan city of Guizhou province, China.The genetic relationship of these tea plant resources still remain unclear.In this study, one-hundred old tea plant samples, which were collected from five different regions of Jiuan city, were subjected for diversity analysis by genotyping by sequencing（GBS）.A total of 548597 high-quality SNP were identified in this collection, and the variation of these SNP were annotated.By using these genome-wide SNP, the phylogenetic tree, the principal component analysis（PCA）and population structure were conducted.A similar pattern on genetic relationship was revealed by three approaches.These samples were assigned into three groups and the samples collected in the same/similar area were clustered in each group.Thus, our results suggested a genetic difference of samples in the Guchayuan（G）, relative to the samples from other four regions, which likely shared the same and unique parental origin.

Key words: ancient tea plant; SNP; genetic structure; evolutionary relationship

茶树［*Camellia sinensis*（L.）Kuntze］是一种重要的多年生木本经济作物（2n=30），目前已在全球超过 52 个国家和地区广泛种植。茶树主要起源于我国的西南地区，处于茶树起源地核心位置的贵州拥有丰富的茶树种质资源，包括长期栽培驯化的地方群体品种、野生种、半野生种等[2-3]。尤其是近年来在全省各地陆续发现的野生和半野生茶树资源极大地丰富了贵州的茶树种质基因库。按照 2005 年 3 月召开的古茶山国际研讨会上提出的将“分布于天然林中的野生古茶树及其群落，半驯化的人工栽培的野生茶树和人工栽培的百年以上的古茶园（林）”统一定义为“古茶树”，那么贵州各地的大量

收稿日期：2018-05-29　**修回日期：**2018-06-04　**网络出版日期：**2018-08-15
URL：http://kns.cnki.net/kcms/detail/11.4996.S.20180814.1651.001.html
第一作者研究方向为茶树资源与育种，E-mail：710191785@qq.com
通信作者：陈娟，研究方向为茶树资源与育种，E-mail：chenjuan309@163.com
基金项目：贵阳市 - 贵州省农科院院地合作项目（院地农科合字［2014］7 号）；贵州省 - 农科院省院联合基金项目（黔科合 LH 字［2015］7075 号）；贵州省科技支撑计划项目（黔科合支撑［2017］2557 号）
Foundation project: The Guiyang City-Guizhou Academy of Agricultural Sciences Cooperation Project（20147）, The Science and Technology Cooperation Project of Guizhou Province（LH20157075）, The Science and Technology Support Project of Guizhou Province（20172557）

野生和半野生茶树资源都可以划分为古茶树资源。对这些有着深厚文化底蕴以及丰富遗传变异的古茶树资源的深入研究对探索茶树的起源以及挖掘优异茶树种质资源具有重要意义。

茶树种质资源是种质创新与茶树新品种培育的物质基础，对茶树种质资源的鉴定与评价是进行茶树资源分类、育种应用以及茶树起源研究的前提[4]。传统的鉴定方法大多依靠形态学和生化组分等表型性状，易受人为的主观性以及环境条件的影响，鉴定周期长，极大地限制了种质资源鉴定与评价的准确性和效率[5]。而基于 DNA 序列多态性的分子标记因其稳定性、准确性和高效性已广泛应用于种质资源鉴定、遗传图谱构建以及分子标记辅助育种等各个方面[6]。目前，基于 DNA 序列多态性开发并利用的分子标记主要包括随机扩增多态性标记（RAPD，random amplified polymorphic DNA）、限制性片段长度多态性标记（RFLP，restriction fragment length polymorphism）、扩增片段长度多态性标记（AFLP，amplified fragment length polymorphism）、简单序列重复（SSR，simple sequence repeat）、简单序列重复区间（ISSR，inter simple sequence repeat）以及单核苷酸多态性标记（SNP，single nucleotide polymorphism）等[6-8]，此外，基于叶绿体 DNA 特定位点的标记也广泛应用于种质资源的鉴定[9]。虽然上述分子标记在包括茶树在内的不同物种中都一定程度上得到了很好的应用，但是除了 SNP 标记外，其余的分子标记都会受限于标记的数量以及应用的成本。而 SNP 变异作为植物基因组中最丰富的变异类型，除了数量巨大外，其获得对 DNA 片段大小也没有特殊要求，尤其是随着高通量测序（NGS）技术的发展，为高效、低廉、大量的 SNP 鉴定与标记开发提供了极大便利[10-11]。目前在水稻[12]、玉米[13]、棉花[14]等作物中都已开发出了高通量的 SNP 芯片。在茶树的研究中，王丽鸳等[15]、Fang 等[16]先后利用公共数据库中茶树的 EST 序列信息开发了 EST-SNP 标记并应用于种质鉴定；Ma 等[17]基于 SLAF-seq 技术、利用一个含有 148 个 F_1 个体的作图群体，构建了一个包含 6042 个 SNP 标记的茶树遗传图谱；Yang 等[11]利用 RAD-Seq 技术鉴定到了 15444 个 SNP，从全基因组水平对 18 份栽培和野生茶树的系统进化关系进行了研究。这些研究表明 SNP 标记在茶树育种研究中也有着巨大潜力。

GBS（genotyping by sequencing）作为一项基于 NGS 的简化基因组测序技术，目前已在不同作物的全基因组 SNP 鉴定、种质遗传多样性分析、全基因组关联分析、遗传图谱构建与基因定位等方面广泛应用[18-21]。如 Verma 等[20]利用 GBS 技术从包含 177 个个体的鹰嘴豆 RIL 群体中鉴定到了 119672 个 SNP，并构建了一个包含 3228 个高质量 SNP 的连锁图谱，对籽粒性状进行了 QTL 定位分析；Hamon 等[22]利用 GBS 技术从 86 个咖啡个体中获得了 118317 个 SNP，并将其应用于咖啡的起源研究以及咖啡因性状的进化研究中。由此可见，基于 GBS 的全基因组 SNP 挖掘不仅适用于一般的粮食作物，同样也适用于多年生的经济作物，这为利用 GBS 技术挖掘茶树基因组 SNP 提供了新的方法。

贵州久安分布着大量的野生和半野生茶树群，是贵州古茶树资源集中分布的代表地区之一。虽然这些茶树资源的表型性状存在着广泛的多样性，但是对于它们的遗传背景和它们之间的亲缘关系还不清楚。在本研究中，利用 GBS 技术对分布于贵州久安 5 个不同区域的 100 份古茶树资源的全基因组 SNP 进行了系统鉴定，基于鉴定到的 SNP 构建了这 100 份茶树资源的系统进化树并分析它们的遗传结构关系，以期探索它们的迁移历史和亲缘关系特征，为该地区茶树种质资源的保护和利用提供参考；同时本研究鉴定到的 SNP 也将有助于后续茶树分子生物学研究的应用。

1　材料与方法

1.1　试验材料

本研究利用的 100 份茶树资源分布于贵州久安 5 个不同区域，长期缺乏人为干涉，属于半野生资源，且区域彼此之间存在地理隔离。其中久安村 10 份（JA1~10），古茶园 22 份（G1~22），打通村 25 份（DT1，DT3~26），打通路 23 份（DTL1~4，DTL6~17，DTL19~25），樱桃树 20 份（YT1~6，YT8~15，YT18~23）。于 2017 年 4 月采集 1 芽 2 叶，液氮速冻后在 –80 ℃超低温冰箱保存备用。

1.2　试验方法

1.2.1　茶树基因组 DNA 提取与 GBS 文库构建　100 份茶树叶片 DNA 提取采用改良的 CTAB 法。提取的 DNA 经 1% 琼脂糖凝胶电泳和紫外分光光度计检测质量和浓度后送北京诺禾致源科技股份有限公司按照 Elshire 等[23]的方法进行 GBS 文库构建和测序。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。

1.2.2 茶树基因组SNP挖掘 测序获得的原始图像数据文件经碱基识别分析转化为原始序列数据（raw read），对原始序列数据去除含接头序列的reads、单端测序序列中未测出的碱基超过该序列长度10%的序列以及低质量（≤5）碱基数超过该序列长度50%的序列，最终获得高质量序列数据（clean read）。有效的高质量序列数据经BWA（Burrows-Wheeler Aligner）软件[24]比对到茶树基因组（http://www.plantkingdomgdb.com/tea_tree/）[25]。利用SAMTOOLS软件[26]进行SNP检测，获得的SNP经过测序深度3×（dp3）、Miss0.3、次要等位基因频率（MAF）>0.01的过滤后得到高质量的SNP用于后续分析。另外，过滤后的高质量SNP通过ANNOVAR软件[27]工具进行注释。

1.2.3 群体进化树构建 利用个体的SNP计算群体之间的遗传距离（p-距离），两个个体 *i* 和 *j* 的之间的p-距离计算公式为：

$$D_{ij} = \frac{1}{L}\sum_{l=1}^{L} d_{ij}$$

式中L为高质量SNP区域长度，假设位置1的等位基因为A/C，如果两个个体基因型为AA和AA，则 d_{ij}=0；如果两个个体的基因型是AA和AC或者AC和AC，则 d_{ij}=0.5；如果两个个体基因型为AA和CC，则 d_{ij}=1。运用TreeBest（http://treesoft.sourceforge.net/treebest.shtml）软件计算距离矩阵，在距离矩阵基础上，利用邻接法（neighbor-joining method）构建系统进化树。引导值（bootstrap values）经过达1000次计算获得。

1.2.4 主成分分析 利用获得的高质量SNP，基于个体间的SNP差异，通过GCTA（http://cnsgenomics.com/software/gcta/pca.html）软件计算特征向量以及特征值，然后运用R语言绘制PCA分布图。

1.2.5 群体遗传结构分析 根据100份古茶树资源的地理分布将其分为5个亚群，利用获得的高质量SNP，通过PLINK（http://pngu.mgh.harvard.edu/~purcell/plink/）进行群体结构分析，同时运用admixture软件构建群体遗传结构。

2 结果与分析

2.1 测序质量

100份古茶树资源共获得62.95G高质量测序数据，平均每个样品获得0.63G数据，碱基错误率在1%以下（Q20）的平均占比96.44%，而错误率低于0.1%（Q30）的平均占比91.13%，表明测序质量较高。100份资源平均获得高质量序列条数4.69百万条，平均有4.54百万的序列条数可以比对到茶树参考基因组，平均比对率为96.68%，平均测序深度11.72。平均比对率和平均测序深度满足重测序分析要求，可以进行后续分析。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（Accession number：CRA000946）。每个样品的数据概况如表1所示。

2.2 SNP类型与分布

SAMTOOLS软件用于SNP检测，获得的SNP经过滤后最终得到548597个高质量SNP用于后续分析。对这些SNP的变异类型统计发现6种变异类型的SNP中A/G和C/T变异类型最多，分别占比37.3%和37.4%，即409860个SNP发生了碱基转换、138737个发生了碱基颠换，转换与颠换的比率为2.95（图1A）。进一步对它们在基因组的分布情况统计发现有89%的SNP位于基因间区，5%的位于内含子区，外显子区、5′UTR、3′UTR区则分别占比2%（图1B）。对于基因外显子区的SNP，共涉及2958个基因，平均每个基因的编码区有4个SNP；这些变异中共有6544个SNP（涉及2225个基因）导致编码氨基酸的改变，剩余5443个SNP（涉及2390个基因）变异没有造成氨基酸序列的改变；另外共有211个SNP导致178个基因提前获得终止密码子、13个SNP导致13个基因终止密码子丧失（图1C）。

2.3 群体进化树与主成分分析

利用鉴定到的高质量SNP对这100份茶树资源进行系统进化分析。100份茶树资源的系统进化树如图2A所示，从图中可以看出，来自同一个取样地点的材料间亲缘关系较近，而不同地点的材料间有明显的差异，预示着这5个取样地点内部的材料可能有相同的亲本来源。另外从图中也可以看出，来自久安村（JA）的材料与来自古茶园（G）的材料亲缘关系相对较近，而来自打通村（DT）和打通路（DTL）的材料间亲缘关系较近，因此进一步将这100份材料分为3个类群。通过主成分分析方法再次分析材料间的亲缘关系，如图2B所示，主成分分析的结果与系统进化树的结果高度一致，根据主成分1（PC1）和主成分2（PC2）同样可以将这100份材料分为3个类群。

表 1　试验材料及测序数据概况

Table 1　The materials used in this study and overview of NGS dataset

区域 Area	样品 Sample	原始数据量（bp） Raw base	有效数据量（bp） Clean base	有效数据率（%） Effective rate	Q20（%）	Q30（%）	有效片段数 Clean reads	可比对片段数 Mapped reads	比对效率（%） Mapping rate	测序深度 Average depth
DT	DT1	627318720	627297696	100.00	96.67	91.44	4356234	4215478	96.77	11.31
	DT3	631841472	631812960	100.00	96.95	92.16	4387590	4189839	95.49	11.24
	DT4	549190368	549166176	100.00	96.98	92.31	3813654	3675039	96.37	10.27
	DT5	662796864	662774400	100.00	96.91	91.96	4602600	4457742	96.85	11.75
	DT6	733737024	733709376	100.00	96.22	90.30	5095204	4934567	96.85	12.32
	DT7	759168864	759143808	100.00	97.17	92.69	5271832	5117343	97.07	12.57
	DT8	723569472	723544992	100.00	96.88	91.91	5024618	4857693	96.68	12.06
	DT9	768081024	768053376	100.00	96.86	92.07	5333704	5164499	96.83	12.41
	DT10	715951872	715930848	100.00	97.00	92.23	4971742	4817145	96.89	11.95
	DT11	772914240	772885440	100.00	96.64	91.31	5367260	5174267	96.40	12.18
	DT12	457022880	457005888	100.00	96.60	91.19	3173652	3063334	96.52	9.73
	DT13	580735008	580712832	100.00	96.56	91.06	4032728	3907462	96.89	11.03
	DT14	705542976	705513888	100.00	96.59	91.16	4899402	4750994	96.97	11.73
	DT15	783071712	783037440	100.00	96.88	91.91	5437760	5275910	97.02	12.49
	DT16	636805440	636784704	100.00	97.15	92.62	4422116	4280187	96.79	11.30
	DT17	512423712	512400960	100.00	96.91	92.05	3558340	3443674	96.78	10.15
	DT18	751020768	750988512	100.00	96.80	91.88	5215198	5054610	96.92	12.62
	DT19	800876448	800838432	100.00	97.13	92.62	5561378	5385793	96.84	12.48
	DT20	664009920	663987456	100.00	97.08	92.44	4611024	4465507	96.84	11.34
	DT21	815474880	815436576	100.00	96.19	90.23	5662754	5472823	96.65	12.78
	DT22	630630144	630608832	100.00	96.50	90.99	4379228	4234563	96.70	11.31
	DT23	853943616	853910496	100.00	97.01	92.24	5929934	5745896	96.90	13.01
	DT24	445597056	445513536	99.98	95.97	90.25	3093844	2969910	95.99	9.31
	DT25	731523456	731395008	99.98	96.18	90.67	5079132	4905858	96.59	11.02
	DT26	764896032	764751744	99.98	96.11	90.41	5310776	5131079	96.62	11.34
DTL	DTL1	748040544	747827136	99.97	96.54	91.28	5193244	5022732	96.72	12.48
	DTL2	724904928	724709952	99.97	95.10	87.95	5032708	4874688	96.86	12.66
	DTL3	759611808	759376224	99.97	96.77	91.96	5273446	5092934	96.58	12.68
	DTL4	703745568	703524096	99.97	96.73	91.80	4885584	4729333	96.80	12.18
	DTL6	428991552	428898528	99.98	96.03	90.34	2978462	2867314	96.27	9.97
	DTL7	684342432	684204192	99.98	96.28	90.83	4751418	4611612	97.06	11.95
	DTL8	694332000	694179072	99.98	96.17	90.50	4820688	4678293	97.05	11.73
	DTL9	724587264	724442688	99.98	96.30	90.94	5030852	4852915	96.46	12.19
	DTL10	675326592	675179136	99.98	96.38	91.14	4688744	4528991	96.59	11.73

表 1（续）

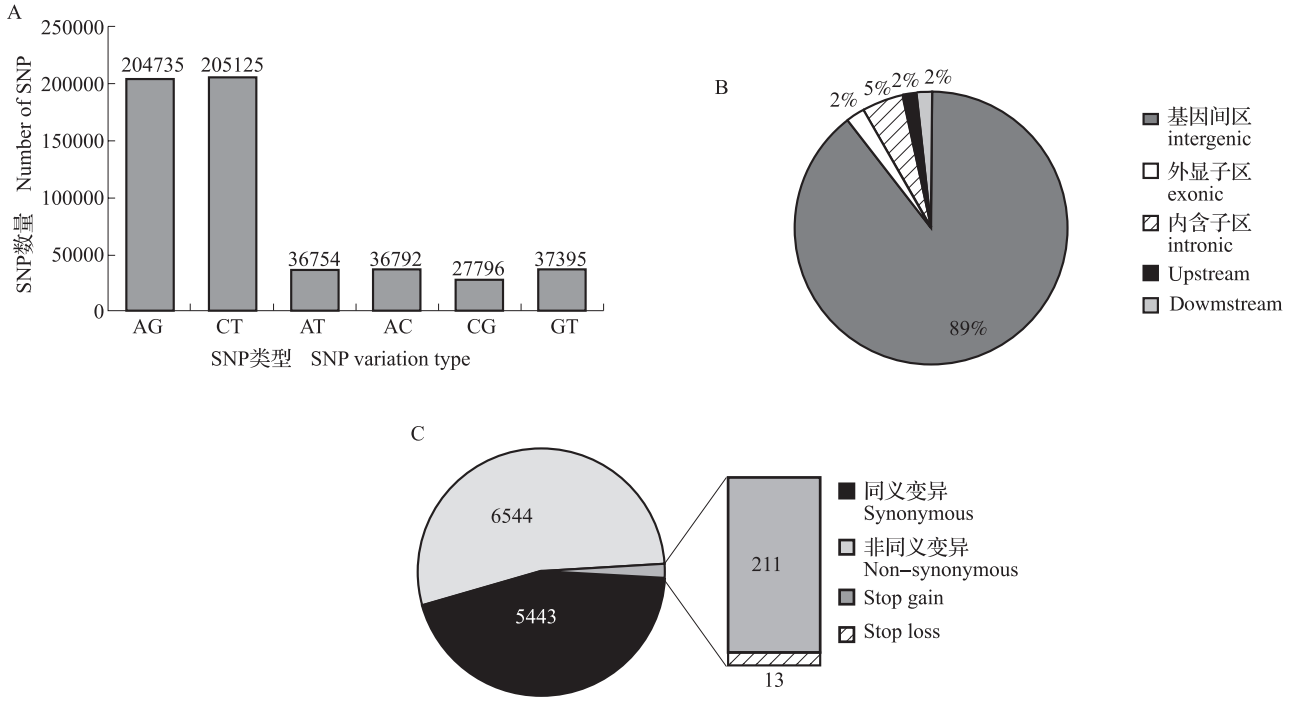
区域 Area	样品 Sample	原始数据量（bp） Raw base	有效数据量（bp） Clean base	有效数据率（%） Effective rate	Q20（%）	Q30（%）	有效片段数 Clean reads	可比对片段数 Mapped reads	比对效率（%） Mapping rate	测序深度 Average depth
DTL	DTL11	646662528	646526016	99.98	96.26	90.75	4489764	4346682	96.81	11.66
	DTL12	614294208	614173248	99.98	95.65	89.29	4265092	4134286	96.93	11.55
	DTL13	650655648	650507040	99.98	96.41	91.25	4517410	4379863	96.96	11.63
	DTL14	703304352	703155744	99.98	96.09	90.46	4883026	4729298	96.85	11.72
	DTL15	665030592	664878528	99.98	96.30	90.91	4617212	4475234	96.93	11.81
	DTL16	621510336	621366912	99.98	96.31	90.92	4315048	4184579	96.98	11.40
	DTL17	727229376	727076736	99.98	96.00	90.17	5049144	4872258	96.50	12.60
	DTL19	440431488	440345376	99.98	95.98	90.01	3057954	2960531	96.81	9.80
	DTL20	541311552	541203264	99.98	95.95	89.96	3758356	3644863	96.98	10.83
	DTL21	573969312	573850944	99.98	95.93	89.94	3985076	3866069	97.01	11.20
	DTL22	602921088	602775648	99.98	96.23	90.70	4185942	4053129	96.83	11.32
	DTL23	641363328	641221632	99.98	96.35	91.19	4452928	4285924	96.25	11.14
	DTL24	682570080	682405344	99.98	96.35	90.97	4738926	4597875	97.02	11.89
	DTL25	695816928	695662560	99.98	96.31	90.91	4830990	4685166	96.98	12.19
G	G1	788943456	788716800	99.97	96.84	92.07	5477200	5286158	96.51	12.41
	G2	671025888	670831776	99.97	96.36	91.05	4658554	4461019	95.76	11.17
	G3	850013568	849778560	99.97	96.48	91.13	5901240	5717488	96.89	13.56
	G4	722960064	722745792	99.97	96.50	91.17	5019068	4863637	96.90	12.47
	G5	604064448	603871200	99.97	96.50	91.17	4193550	4072461	97.11	11.23
	G6	659705472	659509056	99.97	96.76	91.87	4579924	4451693	97.20	11.76
	G7	726962976	726748992	99.97	96.98	92.46	5046868	4892397	96.94	12.15
	G8	699071904	698849280	99.97	96.85	92.09	4853120	4697926	96.80	11.97
	G9	759846528	759622752	99.97	96.84	92.08	5275158	5116839	97.00	12.44
	G10	819231264	818972064	99.97	96.94	92.38	5687306	5499062	96.69	12.41
	G11	709008768	708798528	99.97	96.92	92.28	4922212	4768842	96.88	12.07
	G12	692174880	691976160	99.97	96.16	90.43	4805390	4626717	96.28	11.44
	G13	786121920	785895552	99.97	96.58	91.36	5457608	5277931	96.71	12.54
	G14	782468064	782238816	99.97	96.88	92.19	5432214	5235828	96.38	12.75
	G15	503559648	503417376	99.97	96.35	90.86	3495954	3390489	96.98	10.69
	G16	621409824	621206784	99.97	96.91	92.41	4313936	4161000	96.45	11.39
	G17	636069312	635873472	99.97	96.95	92.41	4415788	4262545	96.53	11.47
	G18	672317856	672122880	99.97	96.78	91.94	4667520	4520531	96.85	11.63
	G19	733072896	732853728	99.97	96.81	92.11	5089262	4917052	96.62	12.25
	G20	621443808	621261792	99.97	96.44	91.06	4314318	4166310	96.57	11.49
	G21	814036320	813799584	99.97	96.23	90.50	5651386	5481482	96.99	13.10
	G22	734716512	734499936	99.97	96.66	91.69	5100694	4922709	96.51	12.40

表 1（续）

区域 Area	样品 Sample	原始数据量（bp） Raw base	有效数据量（bp） Clean base	有效数据率（%） Effective rate	Q20（%）	Q30（%）	有效片段数 Clean reads	可比对片段数 Mapped reads	比对效率（%） Mapping rate	测序深度 Average depth
JA	JA1	498978432	498836736	99.97	96.48	91.37	3464144	3307798	95.49	10.75
	JA2	680395392	680197536	99.97	96.76	91.90	4723594	4554272	96.42	11.56
	JA3	702226944	702015264	99.97	96.59	91.52	4875106	4701005	96.43	12.06
	JA4	676901664	676702080	99.97	96.66	91.82	4699320	4508416	95.94	11.64
	JA5	478743264	478602144	99.97	96.78	92.06	3323626	3192990	96.07	9.73
	JA6	765494784	765256608	99.97	96.69	91.77	5314282	5127539	96.49	11.53
	JA7	777868128	777659616	99.97	96.24	90.58	5400414	5219790	96.66	12.06
	JA8	542302848	542132928	99.97	96.69	92.00	3764812	3620089	96.16	10.26
	JA9	613301472	613125792	99.97	96.52	91.44	4257818	4103977	96.39	10.48
	JA10	844850880	844595424	99.97	96.79	92.01	5865246	5648791	96.31	12.46
YT	YT1	716275584	716122944	99.98	96.43	91.27	4973076	4816143	96.84	12.03
	YT2	730432512	730274400	99.98	96.46	91.24	5071350	4918691	96.99	12.26
	YT3	739911168	739760256	99.98	95.58	89.19	5137224	4952394	96.40	12.32
	YT4	750746016	750586176	99.98	96.09	90.25	5212404	5050798	96.90	12.63
	YT5	812485728	812308608	99.98	96.38	91.08	5641032	5459956	96.79	13.20
	YT6	579582144	579466080	99.98	95.88	89.82	4024070	3896535	96.83	11.07
	YT8	654657696	654502464	99.98	96.36	91.23	4545156	4402812	96.87	11.71
	YT9	684880704	684719136	99.98	96.40	91.22	4754994	4594339	96.62	11.71
	YT10	696526560	696375936	99.98	96.19	90.69	4835944	4674168	96.65	12.16
	YT11	751574880	751403232	99.98	96.31	91.00	5218078	5055628	96.89	12.49
	YT12	621890208	621758880	99.98	95.78	89.68	4317770	4188093	97.00	11.74
	YT13	654716736	654588288	99.98	95.69	89.33	4545752	4383088	96.42	11.70
	YT14	717240960	717091200	99.98	96.09	90.48	4979800	4815247	96.70	12.69
	YT15	702972000	702833184	99.98	96.00	90.10	4880786	4733156	96.98	12.40
	YT18	553820544	553728672	99.98	94.48	86.71	3845338	3725676	96.89	11.04
	YT19	680384160	680233536	99.98	96.27	90.87	4723844	4570606	96.76	11.94
	YT20	718569504	718392960	99.98	96.14	90.57	4988840	4817724	96.57	12.46
	YT21	510267744	510251040	100.00	96.61	91.42	3543410	3401595	96.00	10.40
	YT22	563819904	563800896	100.00	96.92	92.03	3915284	3779168	96.52	10.62
	YT23	542857536	542792448	99.99	94.78	86.88	3769392	3658261	97.05	9.85
平均 Average		676019989	675884007	99.98	96.44	91.13	4693639	4538346	96.68	11.72

Q20：错误率在 1%以下的碱基所占的百分比；Q30：错误率在 0.1%以下的碱基所占的百分比

Q20：the percentage of the bases with error rates below 1%，Q30：the percentage of the bases with error rates below 0.1%



A：6 种变异类型 SNP 数目统计；B：不同基因组位置的 SNP 比例统计；C：外显子区 SNP 变异类型注释。Upstream：基因上游 1 kb 区域；Downstream：基因下游 1 kb 区域；Stop gain：使基因获得终止密码子的变异；Stop loss：使基因失去终止密码子的变异

A: Six types of SNP and the number of SNP in each type, B: The position of the SNP in the gene structure, C: The annotation of the SNP in exon. Upstream: the SNP located in 1 kb upstream of a gene, Downstream: the SNP located in 1 kb downstream of a gene, Stop gain: the variation causes the gene to be terminated, Stop loss: the variation causes the gene to lose the terminator codon

图 1 SNP 的变异类型统计与注释

Fig. 1 Statistics of SNP

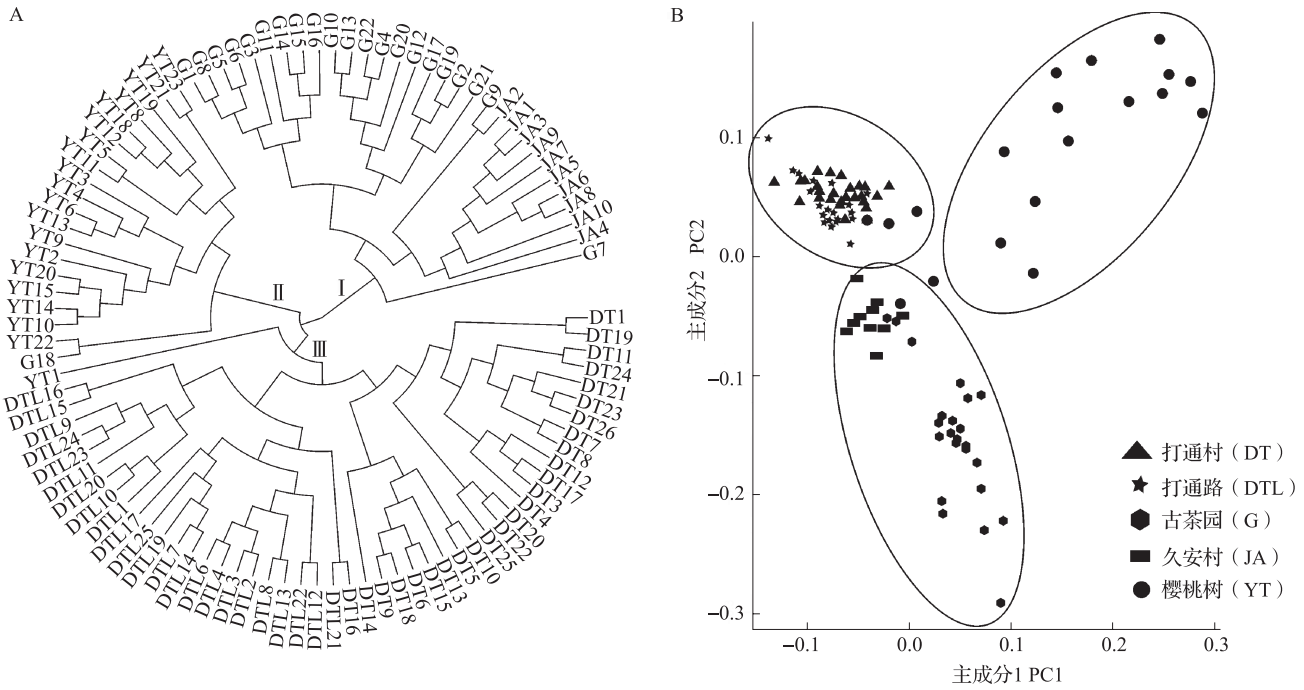


图 2 100 份古茶树资源的系统进化树（A）与主成分分析（B）

Fig. 2 The phylogenetic tree（A）and principal component analysis（B）of the 100 old tea plant samples

2.4　群体遗传结构分析

为了进一步明晰 100 份材料个体间的亲缘关系，对它们进行了遗传结构解析（图 3）。在 K=2 时，并不能明确这些材料间的遗传背景，而当 K=3 时，材料的遗传背景与系统进化树和主成分分析的结果呈现出高度一致，同时也能清楚地看出 5 个取

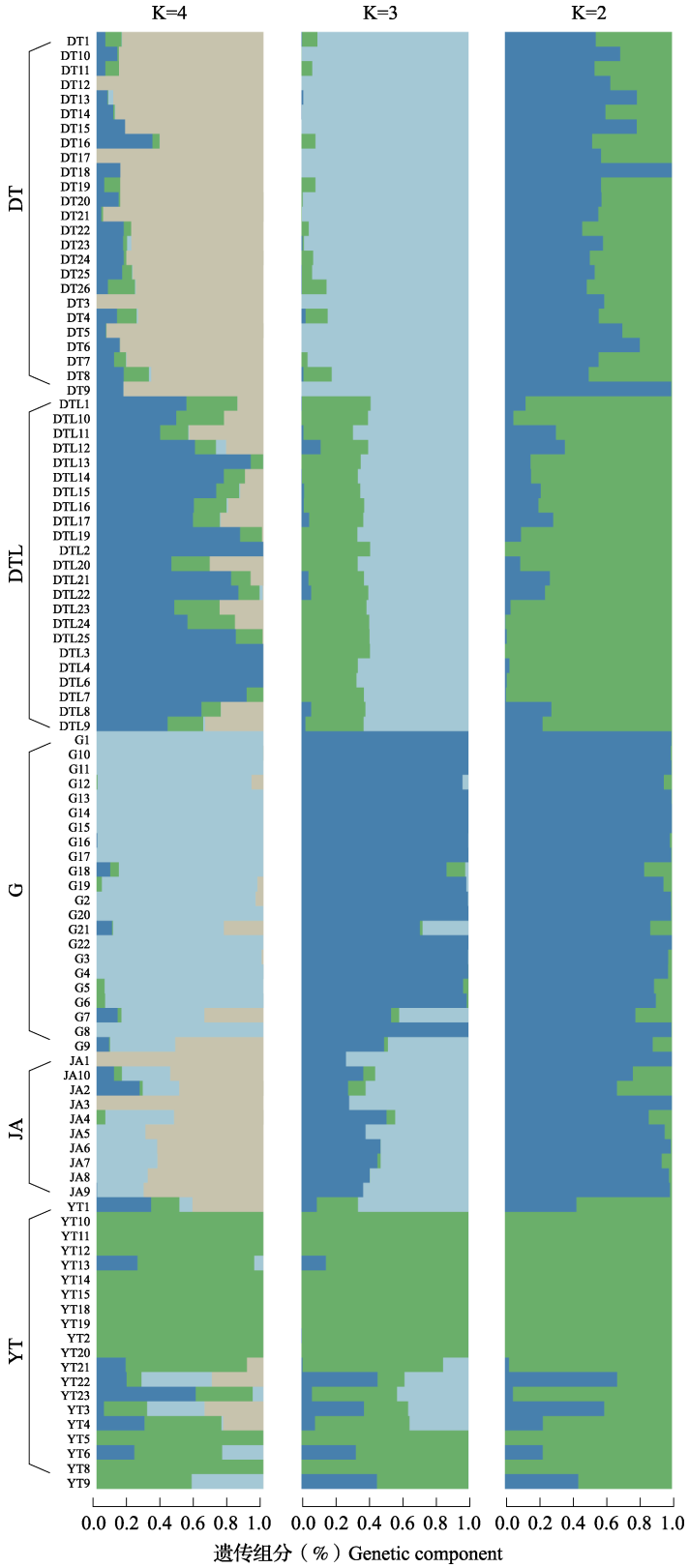


图 3　100 份古茶树资源的遗传结构关系

Fig. 3　The population structure of the 100 old tea plant samples

样地点材料的遗传背景差异。在 K=3 时，除来自古茶园（G）的材料外，其余 4 个地点的材料间都有相同部分的遗传背景，其中来自打通村（DT）的材料与来自打通路（DTL）的材料很可能具有相同的双亲，而打通村（DT）的材料与久安村（JA）的材料以及打通路（DTL）与樱桃树（YT）的材料则可能有相同的父本或者母本。而来自古茶园（G）的材料仅与久安村（JA）的材料有部分相同遗传背景，表明古茶园（G）的材料还保存着特异的遗传背景。进一步当 K=4 时，此时的材料间遗传结构关系同样支持上述猜测。

3 讨论

3.1 茶树基因组 SNP 的鉴定研究

SNP 变异是植物基因组中最主要的、覆盖度最广的序列变异类型，很容易通过序列比对的方式进行鉴定[16]。另外，得益于其廉价、高效和全基因组覆盖特性，SNP 标记已迅速发展为基因分型、种质鉴定、遗传图谱构建以及 QTL 定位等研究最受广泛应用的分子标记类型[10-11]。而茶树作为多年生木本植物，高度自交不亲和、人工杂交结实率低、遗传背景复杂，SNP 标记在茶树遗传研究中的应用还很局限。

在本研究中，利用 GBS 技术对 100 份茶树资源进行简化基因组测序分析，平均每个样品获得 0.63G 数据，约占茶树基因组的 21%。获得的 SNP 中，A/G 和 C/T 变异类型的 SNP 最多，这与前人对茶树中 SNP 鉴定的结果一致[11, 15, 28-29]，表明碱基转换是茶树中主要的 SNP 变异类型。本研究鉴定到的 SNP 的转换与颠换比率为 2.95，这要高于前人利用 RAD-Seq 鉴定的茶树 SNP 中碱基转换 / 颠换比 1.48 以及 EST-SNP 的 2.0[11, 19]，这可能与本研究参考的是茶树基因组序列而后两者参考的是茶树转录组序列有关。另外，本研究在 2958 个茶树基因的外显子区鉴定到了 12211 个 SNP，这些 SNP 导致的非同义变异和同义变异比率为 1.2，这略高于 Yang 等[11]基于 RAD-Seq 的 1.1。此外，178 个基因中的 SNP 变异导致基因编码提前终止、13 个基因的终止密码子丧失。这为后续开发功能基因 SNP 标记提供了可能，同时这些非同义突变的 SNP 也可以为后续的基因功能研究提供参考。

3.2 基于 SNP 的久安古茶树遗传结构关系

久安古茶树大部分为明朝时期人工栽培的中小叶种茶树，已发现有 3 万余丛，是目前国内最大的灌木型古茶树类群[30]，且已入选第三批中国重要农业文化遗产名录（http://jiuban.moa.gov.cn/zwllm/tzgg/tz/201510/t20151012_4861519.htm）。这些古茶树资源不仅为茶树种质创新提供了丰富的变异来源，同时也承载着厚重的历史文化内涵。前期的研究者从立地生态环境[31]、基于农艺性状的资源分类[32]、生化组分[33]等方面对久安古茶树资源已经做了初步研究，并筛选出了一些表型特异的古茶树资源。然而为了便于后续对这些资源的育种应用，很有必要从 DNA 水平了解资源间的亲缘关系以及遗传结构关系。在本研究中，从贵州久安 5 个不同古茶树分布区域选取了 100 份古茶树资源（每个区域选取材料数不少于 10 份）。由于贵州独特的喀斯特地形，这 5 个区域存在着地理隔离，彼此区域间的材料不存在自由传粉情况；此外，由于长期缺乏人为管理，这些茶树资源基本处于半野生状态。因此，以此为材料探索它们的遗传结构与亲缘关系不仅可以明晰久安古茶树的迁移历史，也有助于后续针对性的育种开发利用。

本研究显示不同区域的材料具有明显的分支关系。这 5 个区域的古茶树资源可以分为 3 个类群：久安村（JA）和古茶园（G）的材料分为一个类群；打通村（DT）和打通路（DTL）的材料分为一个类群；樱桃树（YT）的材料单独为一个类群。这与它们彼此间的地理距离呈正相关。进一步对这些资源进行遗传结构解析，在 K=3 时，遗传结构分析的结果与系统进化树和主成分分析的结果高度一致。同一区域材料的遗传背景非常相似，且不同植株的遗传背景差异很小，很有可能是通过无性繁殖获得的，因此可以推测在上百年前甚至几百年前当地已通过无性繁殖进行茶苗的繁育。此外，从遗传结构也可以看出，古茶园（G）的材料的遗传背景与其余 4 个区域的具有明显的差异，这预示着古茶园的种质可能属于外来种质，因此在后续可以考虑利用古茶园（G）的材料与其余 4 个区域的材料进行人工杂交创制新种质，而同一区域内部的材料间要谨慎进行种质杂交利用。

总之，本研究利用基于高通量测序的 GBS 技术对 100 份古茶树资源进行了全基因组的 SNP 挖掘，共鉴定到近 55 万个高质量 SNP，并利用鉴定到的 SNP 对这些资源的亲缘关系和遗传结构进行了系统分析。这些覆盖全基因组的 SNP 不仅有助于准确地进行茶树资源间遗传关系的划分和起源分化探

索，同时增加了茶树中可利用的 SNP 标记。另外对于位于基因内部的 SNP，在后续研究中可考虑结合基因功能注释、材料农艺性状以及生化组分、抗逆性等表型数据，利用基于连锁不平衡（LD）的关联分析进行功能基因位点定位，进而有助于茶树分子标记辅助育种的开展。

参考文献

[1] Chen L, Apostolides Z, Chen Z M.Global tea breeding: achievements, challenges and perspectives.Hangzhou: Zhejiang University Press, 2012: 1

[2] 陈亮，虞富莲，童启庆 . 关于茶组植物分类与演化的讨论 . 茶叶科学，2000，20（2）：89-94
Chen L, Yu F L, Tong Q Q.Discussions on phylogenetic classfication and evolution of Sect.*Thea*.Journal of Tea Science, 2000, 20(2): 89-94

[3] 鄢东海 . 贵州茶树种质资源研究进展及野生茶树资源调查 . 贵州农业科学，2009，37（7）：184-187
Yan D H.Research progress on tea germplasm resources and investigation of wild tea resource in Guizhou.Guizhou Agricultural Sciences, 2009, 37(7): 184-187

[4] 陈亮，杨亚军，虞富莲 . 中国茶树种质资源研究的主要进展和展望 . 植物遗传资源学报，2004，5（4）：389-392
Chen L, Yang Y J, Yu F L.Tea germplasm research in China: recent progresses and prospects.Journal of Plant Genetic Resources, 2004, 5(4): 389-392

[5] 马建强，姚明哲，陈亮 . 茶树种质资源研究进展 . 茶叶科学，2015（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 (1): 11-16

[6] 郭宁，高怀杰，韩硕，宗梅，王桂香，张月云，刘凡 . 观赏羽衣甘蓝 SSR 标记分型与亲缘关系研究 . 植物遗传资源学报，2017，18（2）：349-357，366
Guo N, Gao H J, Han S, Zong M, Wang G X, Zhang Y Y, Liu F.Genotypic and genetic relationship analysis of ornamental kale (*Brassica oleracea* var.*acephala*) by SSR markers.Journal of Plant Genetic Resources, 2017, 18(2): 349-357, 366

[7] 张成才，刘园，姜燕华，吴立赟，王丽鸳，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

[8] 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

[9] 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

[10] Th B E, Thornsberry J M.Plant molecular diversity and applications to genomics.Current Opinion in Plant Biology, 2002, 5(2): 107-111

[11] 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 SNP from RAD sequencing. PLoS One, 2016, 11(3): e0151424

[12] 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

[13] 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

[14] 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

[15] 王丽鸳，张成才，成浩，韦康 . 茶树 EST-SNP 分布特征及标记开发 . 茶叶科学，2012，32（4）：369-376
Wang L Y, Zhang C C, Cheng H, Wei K.Characterization of EST-derived SNP and development of SNP-markers in tea (*Camellia sinensis*).Journal of Tea Science, 2012, 32(4): 369-376

[16] 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

[17] 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

[18] Huang Y F, Poland J A, Wight C P, Jackson E W, Tinker N A.Using genotyping-by-sequencing (GBS) for genomic discovery in cultivated oat.PLoS One, 2014, 9(7): e102448

[19] He J, Zhao X, Laroche A, Lu Z X, Liu H, Li Z.Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding.Frontiers in Plant Science, 2014, 5: 484

[20] 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

[21] Uncu A O, Frary 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

[22] Hamon P, Grover C E, Davis A P, Rakotomalala J J, Raharimalala N E, Albert V A, Sreenath H L, Stoffelen P, Mitchell S E, Couturon E, Hamon S, de Kochko A, Crouzillat D, Rigoreau M, Sumirat U, Akaffou S, Guyot R.Genotyping-by-sequencing provides the first well-resolved phylogeny for coffee (Coffea) and insights into the evolution of caffeine

content in its species: GBS coffee phylogeny and the evolution of caffeine content.Molecular Phylogenetics & Evolution, 2017, 109: 351-361

[23] 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

[24] Li H, Durbin R.Fast and accurate short read alignment with burrows-wheeler transform.Bioinformatics, 2009, 25(14): 1754-1760

[25] 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

[26] 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

[27] Wang K, Li M, Hakonarson H.ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.Nucleic Acids Research, 2010, 38(16): e164

[28] 张成才,王丽鸳,韦康,成浩 . 基于茶树 EST-SNP 的 CAPS 标记开发 . 分子植物育种, 2013, 11(6): 817-824
Zhang C C, Wang L Y, Wei K, Cheng H.Conversion of EST-SNP into CAPS markers in tea plant (*Camellia sinensis*). Molecular Plant Breeding, 2013, 11(6): 817-824

[29] Zhang C C, Wang L Y, Wei K, Cheng H.Development and characterization of single nucleotide polymorphism markers in *Camellia sinensis* (Theaceae).Genetics & Molecular Research GMR, 2014, 13(3): 5822-5831

[30] 黄富贵 . 贵阳花溪久安古茶树概况及保护利用 . 贵州茶叶, 2017, 45(2): 37-38
Huang F G.General situation, protection and utilization of Jiuan ancient tea plants in Huaxi, Guiyang.Journal of Guizhou Tea, 2017, 45(2): 37-38

[31] 宋勤飞,牛素贞,陈正武,尹杰,周绍均,岑春娇 . 基于主成分分析的花溪古茶树立地土壤养分评价 . 浙江农业学报, 2017, 29(11): 1844-1853
Song Q F, Niu S Z, Chen Z W, Yin J, Zhou S J, Cen C J.Evaluation of nutrient status in site soil of ancient tea trees in Huaxi on principal component analysis.Acta Agriculturae Zhejiangensis, 2017, 29(11): 1844-1853

[32] 龚雪,杨春,郭燕,周顺珍,郑文佳 . 贵州久安古茶树资源的分类研究 . 种子, 2015, 34(4): 56-58, 63
Gong X, Yang C, Guo Y, Zhou S Z, Zheng W J.The classification study of the ancient tea plant resources in Jiuan Guizhou.Seed, 2015, 34(4): 56-58, 63

[33] 宋勤飞,牛素贞,何嵩涛,尹杰,奉红琼 . 久安古茶树春梢芽叶性状及生化成分分析 . 西南农业学报, 2013, 26(2): 505-509
Song Q F, Niu S Z, He S T, Yin J, Feng H Q.Analysis on leaf bud characters and biochemical components in spring shoots of Jiuan ancient tea Tree.Southwest China Journal of Agricultural Sciences, 2013, 26(2): 505-509