吴海霞,苟志辉,杜尚嘉,等.基于全基因组重测序的五指山大叶茶种质资源多样性分析[J].山西农业大学学报(自然科学版),2025,45(2):1-8.
Wu H X, Gou Z H, Du S J,et al.Diversity analysis of germplasm resources of large-leaf tea in Wuzhishan mountains based on whole genome resequencing[J].Journal of Shanxi Agricultural University(Natural Science Edition), 2025,45(2):1-8.
doi: 10.13842/j.cnki.issn1671-8151.202407042

基于全基因组重测序的五指山大叶茶种质资源多样性分析

吴海霞[1,2],苟志辉[1,2],杜尚嘉[1],符溶[1],符生波[1],符明珠[1],吴挺聪[1],陈国德[1,2],郭霞[1*]

(1.海南省林业科学研究院/海南省红树林研究院,海南 海口 571199;
2.海南省热带林业资源监测与应用重点实验室,海南 海口 571199)

摘　要:[目的]野生大叶茶在海南岛五指山地区分布广泛,具有较高的利用价值,然而对其遗传背景及基因组多样性鲜有研究。本研究在已有的大叶茶基因组的基础上,对海南中部五指山区代表性区域野生大叶茶进行基因组重测序以及基于基因组重测序的种质资源多样性分析,为五指山野生大叶茶的保护及在茶树遗传育种中的利用提供理论基础。[方法]本研究选取了69份五指山野生大叶茶代表性样品,对其开展了基因组重测序,在此基础上开展PCA、系统发育分析以及遗传结构分析,并筛选了水满乡境内大叶茶的选择性清除位点。[结果]对野生大叶茶进行基因组重测序,共获得全长为60 001 469 742 bp的基因组重测序数据。将其比对到大叶茶参考基因组上检测变异位点,共获得41 381 015个变异位点,其中有39 970 726个SNP位点,141 029个Indel位点。考察核苷酸多态性的π值在基因组中的分布密度趋势,发现其与SNP、Indel在基因组中的分布高度一致。主成分分析、系统发育分析和遗传结构分析结果均表明,五指山野生大叶茶可分为5个群体,其中群体B、C、E遗传背景较纯,群体A和D的遗传背景较为复杂。水满乡境内大叶茶资源位于D和E群体,造成这2个群体间分化的部分变异位点分布于2273个蛋白编码基因中。[结论]本研究在获得基因组重测序数据基础上,多种遗传多样性分析均表明五指山野生大叶茶可分为5个群体,而水满乡境内大叶茶分布于其中2个群体,这为揭示五指山大叶茶资源的丰富性、制定保护策略及在育种上的应用提供了基础。

关键词:大叶茶; 五指山; 野生种质资源; 重测序; 遗传结构
中图分类号:S571.1　　**文献标识码**:A　　**文章编号**:1671-8151(2025)02-0001-08

Diversity analysis of germplasm resources of large-leaf tea in Wuzhishan mountains based on whole genome resequencing

Wu Haixia[1,2], Gou Zhihui[1,2], Du Shangjia[1], Fu Rong[1], Fu Shengbo[1], Fu Mingzhu[1], Wu Tingcong[1], Chen Guode[1,2], Guo Xia[1*]

(*1.Hainan Academy of Forestry/Hainan Academy of Mangrove, Haikou 571100, China; 2.Key Laboratory of Tropical Forestry Resources Monitoring and Application of Hainan, Haikou 571100, China*)

Abstract:[Objective]Wild large-leaf tea plants (*Camellia sinensis*) are widely distributed in the Wuzhishan region of Hainan Island and holds significant utilization value. However, there is limited research on its genetic background and genomic diversity. Based on the existing large-leaf tea plant genome, this study conducted the whole genome resequencing and germplasm resource diversity analysis of wild large-leaf tea from representative regions in the central Wuzhishan area of Hainan. The aim was to provide a theoretical foundation for the conservation of Wuzhishan wild large-leaf tea plants and its application in tea plant genetic breeding. [Methods]In this study, 69 representative samples of wild germplasms of tea plants in Wuzhishan were

收稿日期:2024-07-16　　**修回日期**:2024-12-04
基金项目:五指山市科技项目(WZSKJXM2021005);省属科研院所技术创新专项(SQKY2022-0037)
作者简介:吴海霞(ORCID:0009-0005-3799-287X),林业工程师,研究方向:资源保护及开发利用;苟志辉,林业工程师,研究方向:林业有害生物生态调控和林产品质量提升评价;吴海霞和苟志辉为同等贡献作者
*通信作者:郭霞,硕士研究生,高级工程师,E-mail:lengning77@126.com

collected for whole-genome resequencing. Based on the resequencing data, principal component analysis (PCA), phylogenetic analysis, and genetic structure analysis were performed. Additionally, selective sweep sites were identified for large-leaf tea within the Shuiman Township. [Result]Whole genome resequencing of wild large-leaf tea generated a total of 60 001 469 742 bp of resequencing data. By aligning this data to the reference genome of large-leaf tea, 41 381 015 variant sites were detected, including 39 970 726 SNP sites and 141 029 Indel sites. The distribution density trend of nucleotide polymorphism (π) values across the genome were found to be highly consistent with the distribution of SNPs and Indels. Principal component analysis, phylogenetic analysis, and genetic structure analysis collectively indicated that the wild large-leaf tea of Wuzhishan could be divided into five populations. Among these, populations B, C, and E exhibited relatively pure genetic backgrounds, while populations A and D had more complex genetic backgrounds. The large-leaf tea resources within Shuiman Township were located in populations D and E, and some of the variant sites contributing to the differentiation between these two populations were distributed across 2273 protein-coding genes. [Conclusion]Based on the whole genome re-sequencing data, multiple genetic diversity analyses revealed that Wuzhishan wild large-leaf tea can be classified into five populations, with the large-leaf tea from Shuiman Township distributed across two of these populations. This study provided a foundation for understanding the richness of Wuzhishan large-leaf tea resources, formulating conservation strategies, and applying these findings in breeding programs.

Keywords: *Camellia sinensis*, Wuzhishan, Wild germplasm, Resequencing, Genetic structure

茶是世界上最重要的饮料之一,而中国作为茶文化的起源国,在世界茶文化中拥有举足轻重的地位,其源远流长的茶艺传统深深影响着全球茶文化的发展和演变。研究表明,中国西南地区是茶树(*Camellia sinensis*)的起源地,而中国南方多样化的地理和气候特征,孕育了丰富的茶树种质资源,这是中国茶文化赖以繁荣的根本之一[1]。在茶学和茶经济发展中,野生茶树种质资源具有独特的价值。它不仅在生物多样性保护中具有重要的科学与实践意义,也在茶文化的传承与发展中发挥着不可替代的作用。野生茶树保留了大量原始遗传特性,它们是改良茶树品种的重要遗传基础[2]。研究这些野生茶树有助于了解茶树的进化历史、适应机制和生态功能,从而为茶树育种提供珍贵的遗传资源。这些野生茶树基因组中的抗病虫害基因、抗逆境基因等抗性基因,以及品质性状相关基因,对于改良现有茶树品种、提升茶叶的质量和产量具有重要意义。此外,野生茶树不仅是茶树育种和茶叶生产的源头,也是茶文化的载体。通过对野生茶树及其栽培技术的研究和保护,可以更好地传承和弘扬茶文化,促进茶文化相关研究和创新。

中国野生茶树分布区域分为横断山脉分布区、滇桂黔分布区、滇川黔分布区、南岭山脉分布区等[3]。然而,近年来多项调查表明,除了这些区域以外,在江西、福建、广东、江苏、湖北、陕西等地也广泛分布大量的野生茶树种质资源[4-9]。此外,基于基因组从头组装和基因组重测序的研究也表明,茶树丰富的种质资源造就了不同茶树资源间极度多样化的基因组成分[10-16]。然而,近年来,受栖息地丧失、过度采摘、农业扩张以及气候变化等多方面因素的影响,野生茶树栖息环境急剧缩小,其赖以生存的生态系统遭受严重破坏,使得野生茶树种群面临濒危甚至灭绝的风险[17]。这些栖息地的破坏,不仅削弱了茶树的自然繁殖能力,还限制了茶树遗传多样性的保留[18]。因此,发掘野生茶树资源,对于茶树种质资源保护利用以及茶学发展具有重要价值。

鉴于以上原因,采用分子方式对茶树种质资源进行梳理,对植物种质资源保护利用具有重要价值。近年来,多个团队采用SSR、CAPS等标记方式,对茶树种质资源进行了分类和梳理,获得了大量茶树种质资源分类信息[18-25]。而高质量茶树基因组的公布,为采用基因组重测序的方式开展茶树种质资源研究提供了新的机遇[10-11, 13-14, 16],其价值也在最近的茶树重测序研究中得以展现[12, 15]。

五指山位于海南岛中部,具有高温高湿、四季差异不明显等气候特征,这种气候孕育了独特且丰富的茶树种质资源。本团队前期调研表明,五指山的茶树种类较多,遗传多样性高,包括野生茶树和人工栽培品种。本团队观察表明,五指山地区这些独特的茶树资源具有抗病虫害、耐旱、耐荫等特性。此外,本团队前期基于代谢组的研究表明,相对于人工茶树,五指山野生茶树的茶叶中富含代谢物,这些代谢物在糖酵解、丙酮酸合成、氨基糖代谢和核苷糖代谢等糖代谢通路、以及酚类

物质合成、核黄素代谢、鞘磷脂代谢和黄酮合成等代谢通路中显著富集[25]。因此,五指山特殊的热带雨林气候和丰富的茶树种质资源,为茶叶生产和科学研究提供了得天独厚的自然条件,使得该地区在中国的茶产业和茶文化中占有重要地位。本研究采用基因组重测序,对69份五指山野生大叶茶的遗传多样性开展分析,以期为五指山大叶茶的保护和利用提供科学依据。

1　材料和方法

1.1　试验材料

本研究中野外采集的样本分别来自五指山市、乐东县和琼中县(附表1)。采集方法:每份材料,分别采集幼嫩叶片(一芽三叶),采集后立即用无菌水冲洗,将水吸干后液氮速冻,保存于干冰中。

1.2　试验方法及统计分析

1.2.1　基因组DNA提取、基因组重测序和测序数据质控

在野外采集69份五指山野生大叶茶样品叶片,采用改良的CTAB法提取叶片基因组DNA,琼脂糖凝胶电泳检测DNA质量,NanoDrop检测DNA浓度和纯度,送深圳华大基因进行基因组重测序(DNBSEQ-T7测序仪),获得重测序数据。去接头后,采用fastp进行质控、过滤、校正(去重复,去接头,过滤碱基质量值低于Q20且比例达到20以上的read,切除序列上游10个碱基,去除N碱基大于5的一对reads)[26]。

1.2.2　变异位点检测和筛选

将reads比对到大叶茶基因组上[27],采用deepvariant进行变异检测[28]。采用fastp对低质量的变异位点进行过滤,保留变异位点标准为:缺失率小于20、DP为5-100、QD大于40、位点质量值Qual值大于40,接着过滤MAF小于0.05、位点缺失率大于20、哈-温平衡检验p值小于0.001的位点,只保留二等位位点。CIRCOS图由CIRCOS v 0.69-8绘制[29]。

1.2.3　群体结构分析

对筛选得到的标记位点进行连锁不平衡(linkage disequilibrium,LD)分析,依据LD分析结果筛选vcf所有标记,以50个标记的窗口大小,10为步长,R2为0.2作为阈值进行筛选。利用筛选后的位点,基于样本的SNP信息进行PCA分析[30]。采用vcf2phylip(https://github.com/edgardomortiz/vcf2phylip)将SNP转化为Phylip格式;采用MEGA 11.0.13绘制系统发育树(Bootstrap Replications 1000;Site Coverage Cutoff 50)[31]。采用Admixture进行群体结构分析,分别运行了K从3到6的运算,最终对每个K值分析中的交叉验证误差(cross-validation error)进行分析[32]。

1.2.4　群体选择信号检测

采用复合似然比检验(CLR,composite likelihood ratio),基于多位点频率的检测方法(Site Frequency Spectrum),使用SweeD软件(4.0.0)计算CLR(composite likelihood ratio)[32]。以20 kbp为窗口划分窗格,计算CLR,筛选前5的选择信号。采用VCFtools软件,以200 kbp为窗口大小,20 kbp为步长进行滑窗计算核苷酸多态性(π)[33]。采用CMplot 4.4.1绘制曼哈顿图(https://github.com/YinLiLin/Cmplot)。

2　结果和分析

2.1　重测序数据的获取和变异检测统计

对69份五指山大叶茶样本进行基因组重测序;数据质控、过滤后,共获得440 060 267条reads,全长60 001 469 742 bp。测序数据平均GC含量为39.7。将质控后的reads比对到大叶茶基因组上,平均比对率为99.4(附表1)。对检测得到的变异位点进行过滤,共获得41 381 015个可信的变异位点,其中有39 970 726个SNP位点,141 029个Indel位点。这些变异位点大部分(79.3)位于基因间区域。在基因区域中,有7.0分布于内含子,1.0分布于外显子,6.3分布于上游区域,6.5分布于下游区域。其中,有597个变异造成了起始密码子的丢失,548个变异造成了终止密码子的丢失;此外,6043个变异产生了新的终止密码子,202 030个变异造成了编码蛋白氨基酸残基的改变。变异位点在大叶茶各染色体的分布如表1所示,在第1条染色体分布的变异位点最多,而在第15条染色体变异位点最少。考察变异位点频次,平均变异频次为74,其中在第5条染色体变异频次最高(82),在第15条染色体变异频次最低(65)。

2.2　核苷酸多态性在基因组中的分布

π值是考察种群内核苷酸多态性的重要指标。本研究计算了基因组各区域的π值,并将其与

表1 变异位点在大叶茶各染色体的分布

Table 1 Distribution of mutation sites on each chromosome of *Camellia sinensis*

染色体 Chromosome	染色体长度/bp Chromosome length	变异位点 Mutation sites	变异位点频次[a] Frequency of mutation sites
1	336 926 639	4 275 464	78
2	282 168 575	3 684 859	76
3	273 085 858	3 768 803	72
4	223 221 884	2 738 767	81
5	217 491 895	2 646 442	82
6	217 078 568	2 917 065	74
7	204 210 641	3 026 852	67
8	193 972 626	2 674 342	72
9	183 443 147	2 545 064	72
10	178 569 637	2 277 200	78
11	177 334 939	2 291 247	77
12	171 534 477	2 180 380	78
13	168 803 194	2 309 223	73
14	148 273 249	2 154 637	68
15	122 945 380	1 890 670	65

注:[a]为变异位点频次,表示多少个碱基出现1个变异位点。

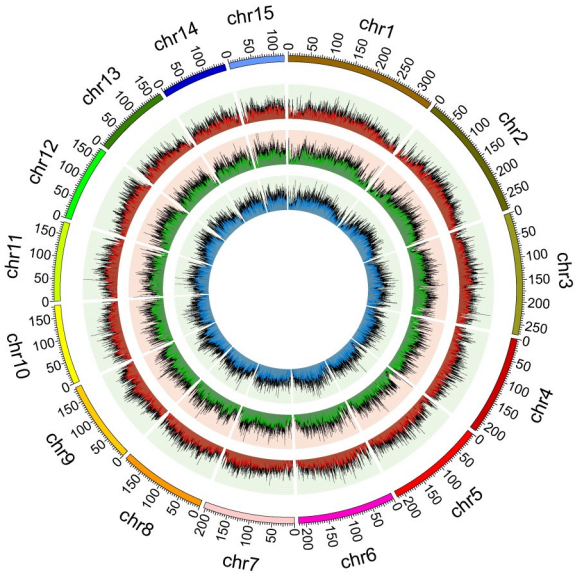
Note:[a] is the frequency of the variant site, which indicates how many bases have a variation site.

SNP、Indel的分布密度一起绘制了圈图,如图1所示。π值在基因组中的分布密度趋势与SNP、Indel高度一致,即π值高的区域SNP、Indel分布密度也高,表明SNP、Indel分布较密的区域,核苷酸多态性也较高。

2.3 五指山野生大叶茶的群体结构

利用变异位点数据,对五指山野生大叶茶进行主成分分析(PCA),结果如图2所示,69份五指山野生大叶茶样本可分为5个群体。根据SNP位点构建系统发育树,发现五指山大叶茶也分为5个显著群体,与PCA结果一致(图3)。

交叉验证误差分析表明,当k=5时,交叉验证误差最低,表明五指山大叶茶样本分为5个群体最为合适。采用STRUCTURE构建遗传群体,当k=5时,群体分类结果与系统发育分析结果一致(图3)。其中,群体B、C、E遗传背景较纯,而群体A和D的遗传背景较为复杂。其中,部分群体A样本和部分群体D样本均混有较多群体E的遗传物质。此外,少量群体B和群体E样本也携带有少许其它群体的遗传成分。群体C样本基本未发现其它群体的遗传成分,表明群体C的遗传背景最纯(图3)。



最外圈为染色体,其次为SNP分布,以及Indel分布,最内圈为π值分布

The outermost ring is the chromosome schematic, followed by the SNP distribution, and the Indel distribution, and the innermost circle is the π distribution

图1 SNP、Indel和π值在大叶茶基因组中的分布

Fig.1 Distribution of SNP, Indel and π values in the genome of *Camellia sinensis*

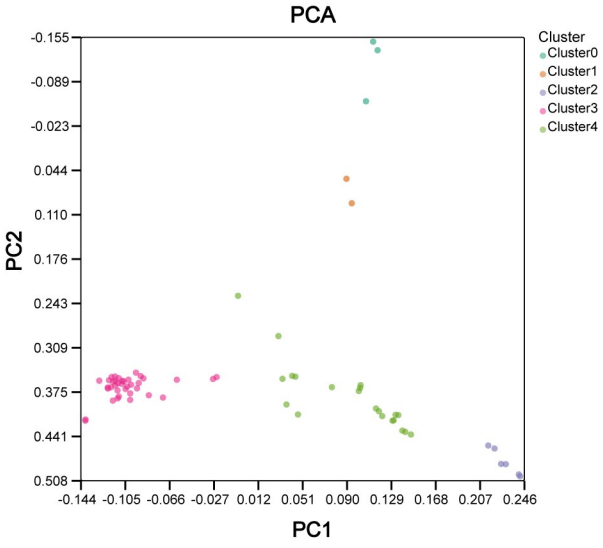


图2 五指山野生大叶茶的主成分分析

Fig.2 Principal component analysis of Wuzhishan wild *Camellia sinensis*

2.4 五指山市水满乡大叶茶样本的选择消除位点

近年来,五指山大叶茶深受欢迎,尤其是水满乡的茶叶最为知名。本研究中,水满乡大叶茶样本集中分布于D群体和E群体,其中D群体样本来

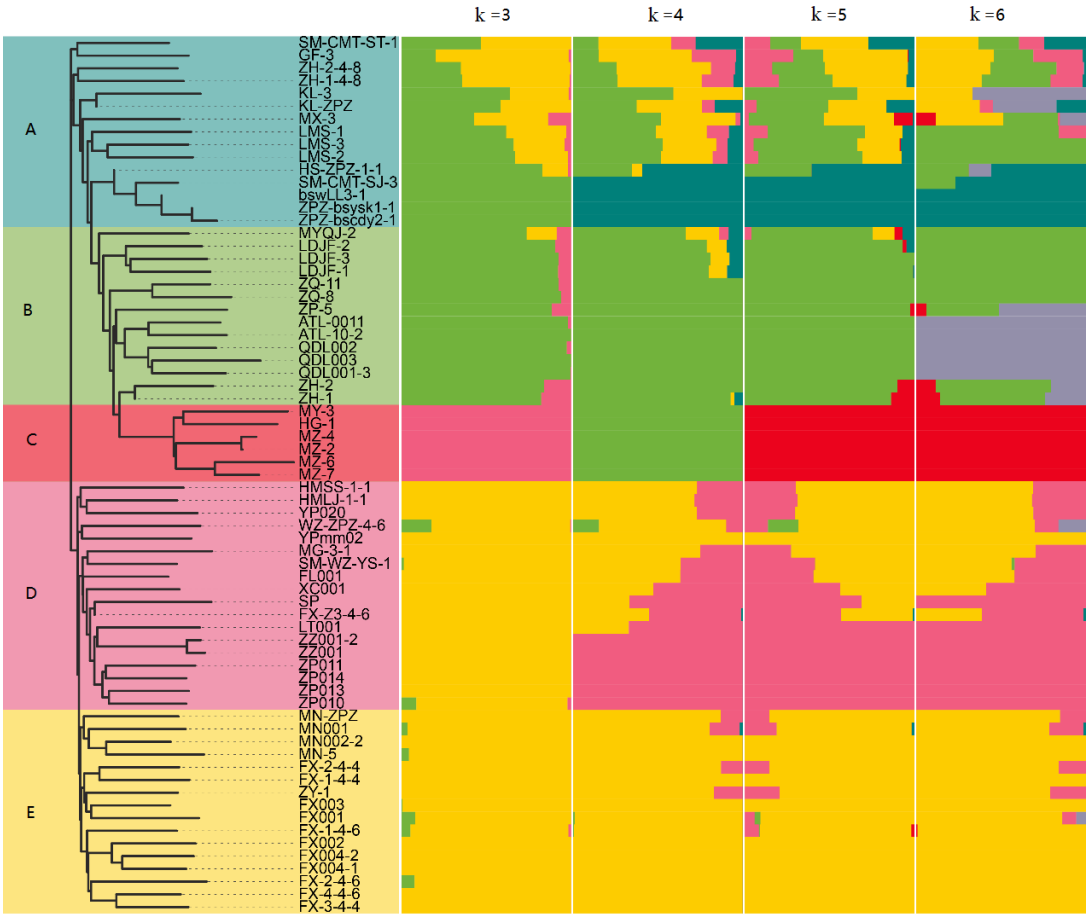


图3　五指山大叶茶的遗传结构

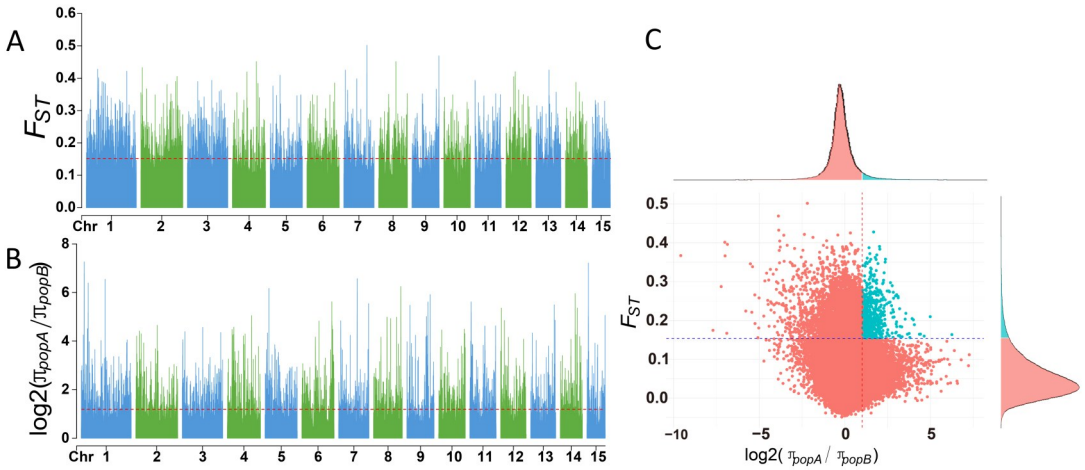
Fig.3　Genetic structure of Wuzhishan *Camellia sinensis*

自什再、什蒲、方龙、牙排等地,而E样本来自毛纳、方响、什应等地。其中,E群体较纯,而D群体部分样本混有E群体的背景(图3)。针对D群体、E群体,本研究开展了选择消除分析,以考察D群体和E群体之间遗传背景上的差异。为此,分别计算了D和E群体间的分化指数(Fst)和核苷酸多态性(π),它们在大叶茶基因组中的分布分别如图4A和图4B所示。将所有窗口的Fst值和log2(π)值绘制到散点图中,其受选择的前5受选择区域如图4C蓝色部分所示。这些受选择区域分布于2273个蛋白编码基因中,表明D群体和E群体的分化有可能影响部分基因的功能。

3　讨论

已有研究表明,茶叶具有非常丰富且极其复杂的遗传多样性,而研究特定区域的大叶茶的多样性有利于当地茶资源的保护和利用[34]。近年来,基于基因组重测序,开展植物遗传背景及结构的研究已逐渐成为常用技术[35]。本研究以海南五指山野生大叶茶为研究对象,通过基因组重测序的方式,从多个角度对其遗传背景进行了研究。研究结果表明,五指山大叶茶拥有丰富的遗传变异,本研究采集的69份样品可分为5个群体,其中水满乡大叶茶样本可分为两个群体,这两个群体间的基因组多态性位点部分位于蛋白编码基因中。这是首次采用高通量测序技术研究五指山大叶茶的遗传多样性。前期,本团队采用代谢组技术,研究了五指山野生大叶茶的代谢组分,表明在该野生茶中含有一系列特异的代谢组分[25]。本研究中,五指山大叶茶呈现了高度的遗传多样性。因此,采用基因组学技术,研究不同群体五指山大叶茶的代谢组分的异同,这将为五指山大叶茶的研究、保护和利用提供更加丰富的基础信息。

近年来,多份茶树基因组得以测序,且基因组质量均较高[3,10,16,34,36]。本研究采用云南野生茶树基因组作为参考基因组,进行变异检测[28],主要是



A:Fst在大叶茶基因组中的分布(红色虚线以上为前5的Fst信号);B:log2(π)在大叶茶基因组中的分布(红色虚线以上为前5的log2(π)信号);C:Fst和log2(π)均为前5的信号

A:Distribution of Fst in the genome of *Camellia sinensis* (signals above the red dotted line is the first 5 of log2(π) signals); B:Distribution of log2(π) in the genome of *Camellia sinensis* (Above the red dashed line are the top 5 signals);C: Fst and log2 (π) are both the top 5 of signals

图4 水满乡D群体和E群体样本选择消除分析

Fig.4 Sample selection and elimination analysis of population D and population E in Shuiman Township

基于以下三点考虑:第一,该基因组质量是所有茶树中质量最高的基因组之一;第二,该基因组依赖的材料为野生材料,与本研究主题较为契合;第三,已有研究表明,云南的野生茶树与海南的茶树遗传背景较为接近[16]。本研究结果表明,采用该基因组作为参考基因组序列,获得了质量较高的变异检测结果,且依赖该检测结果进行的后续分析较为可靠,进一步证明了采用该材料基因组的合理性。

在本研究中检测到五指山大叶茶的5个群体,其中2个群体遗传背景较纯,而其它4个群体的植株混有其它群体的遗传成分,有可能与其它群体植株间产生了杂交。在茶树其它群体材料的重测序研究中,也发现类似现象[36],凸显了杂交在茶树群体形成中的重要性。

4 结论

PCA分析、系统发育分析和遗传结构分析均表明,五指山野生大叶茶可分为五个群体,这为五指山野生大叶茶资源保护提供了理论基础。遗传结构分析表明,B、E群体遗传背景较纯,而A、C、D群体混有其它群体的基因组成分。本研究对水满乡茶树材料遗传背景进行了进一步的研究,发现该地茶树可分为两个群体,这两个群体间存在的部分多态性位点位于蛋白编码基因中,这为该地茶树材料标记的筛选以及其在育种中的应用提供了基础。

附加材料:附表1(高通量测序及比对样品统计)见中国知网增强出版内容。

参 考 文 献

[1]牛素贞,安红卫,刘文昌,等. 贵州古茶树种质资源的生育期及春梢萌发特性[J]. 贵州农业科学,2020,48(3):10-15.
Niu S Z, An H W, Liu W C, et al. Growth stage and spring shoot germination characteristics of ancient tea germplasm resources in Guizhou [J]. Guizhou Agricultural Sciences, 2020, 48(3):10-15.

[2]张文驹,戎俊,韦朝领,等. 栽培茶树的驯化起源与传播[J]. 生物多样性,2018,26(4):357-372.
Zhang W J, Rong J, Wei C L, et al. Domestication origin and spread of cultivated tea plants [J]. Biodiversity Science, 2018, 26 (4):357-372.

[3]王平盛,虞富莲. 中国野生大茶树的地理分布、多样性及其利用价值[J]. 茶叶科学,2002,22(2):105-108,134.
Wang P S, Yu F L. The geographic distribution, diversity and utilization of wild tea camellias in China [J]. Journal of Tea Science, 2002, 22(2):105-108, 134.

[4]朱艳宇,于文涛,高水练,等. 福建安溪茶树种质资源遗传多样性与铁观音衍生品种遗传关系[J]. 浙江农业学报,2024,36(7):1591-1601.
Zhu Y Y, Yu W T, Gao S L, et al. The diversity of tea germplasm resources and genetic relationship of 'Tieguanyin'-derived varieties in Anxi, Fujian, China[J]. Acta Agriculturae Zhejiangensis, 2024, 36(7):1591-1601.

[5]崔清梅,梁金波,马会杰,等. 湖北巴东县野生茶树种质资源遗

传多样性及种群结构亲缘关系分析[J]. 茶叶科学,2024,44(2):193-206.
Cui Q M, Liang J B, Ma H J, et al. Genetic diversity and population structure relationship analysis of wild tea germplasm resources in Badong County, Hubei Province [J]. Journal of Tea Science,2024,44(2):193-206.

[6]王泽涵. 闽粤相邻地区茶树种质资源遗传多样性及秃房野生茶花器官研究[D]. 福州:福建农林大学,2022.
Wang Z H. Study on genetic diversity of tea germplasm resources and wild *Camellia* organs in bare house in adjacent areas of Fujian and guangdong [D]. Fuzhou: Fujian Agriculture and Forestry University,2022.

[7]李琛,江新凤,张贱根,等. 基于不同海拔分布的庐山茶树资源品质成分变化研究[J]. 食品安全质量检测学报,2022,13(9):2886-2891.
Li C, Jiang X F, Zhang J G, et al. Research on quality component changes of tea resource based on different altitude ranges in Lushan[J]. Journal of Food Safety & Quality,2022,13(9):2886-2891.

[8]周高新,王朝阳,刘运华,等. 安康市野生大茶树资源调查报告[J]. 中国茶叶,2022,44(3):66-68.
Zhou G X, Wang Z Y, Liu Y H, et al. Investigation report on ancient tea plant resources in Ankang City[J]. China Tea, 2022,44(3):66-68.

[9]王梦荷,杨彬,田茂荣,等. 南京市栖霞山野生茶树种质资源调查与品质性状遗传多样性分析[J]. 江苏农业科学,2022,50(1):126-132.
Wang M H, Yang B, Tian M R, et al. Investigation of wild tea germplasm resources and genetic diversity analysis of quality traits in Qixia Mountain, Nanjing [J]. Jiangsu Agricultural Sciences,2022,50(1):126-132.

[10]Chen J D, Zheng C, Ma J Q, et al. The chromosome-scale genome reveals the evolution and diversification after the recent tetraploidization event in tea plant[J]. Horticulture Research, 2020,7:63.

[11]Samarina L S, Matskiv A O, Shkhalakhova R M, et al. Genetic diversity and genome size variability in the Russian genebank collection of tea plant [*Camellia sinensis* (L). O. kuntze][J]. Frontiers in Plant Science,2022,12:800141.

[12]Wang Y, Chen F, Ma Y C, et al. An ancient whole-genome duplication event and its contribution to flavor compounds in the tea plant (*Camellia sinensis*)[J]. Horticulture Research, 2021,8(1):176.

[13]Xia E H, Li F D, Tong W, et al. The tea plant reference genome and improved gene annotation using long-read and paired-end sequencing data[J]. Scientific Data,2019,6:122.

[14]Xia E H, Tong W, Hou Y, et al. The reference genome of tea plant and resequencing of 81 diverse accessions provide insights into its genome evolution and adaptation[J]. Molecular Plant, 2020,13(7):1013-1026.

[15]Zhang X T, Chen S, Shi L Q, et al. Haplotype-resolved genome assembly provides insights into evolutionary history of the tea plant *Camellia sinensis*[J]. Nature Genetics, 2021, 53(8):1250-1259.

[16]童韵琳,童航.生物多样性与古茶树资源保护——以《云南省古茶树保护条例》为例[J]. 福建茶叶,2023,45(9):190-192.
Tong Y L, Tong H. Biodiversity and the protection of ancient tea tree resources—taking the regulations on the protection of ancient tea trees in Yunnan Province as an example[J]. Tea in Fujian,2023,45(9):190-192.

[17]陈杰丹,马春雷,陈亮. 我国茶树种质资源研究40年[J]. 中国茶叶,2019,41(6):1-5,46.
Chen J D, Ma C L, Chen L. Research on tea germplasm resources in China for 40 years[J]. China Tea, 2019, 41(6): 1-5,46.

[18]蔡一鸣,蒋双丰,毛光志,等. 基于SSR分子标记的41份茶树种质资源遗传多样性及亲缘关系分析[J]. 中国农学通报,2023,39(30):54-60.
Cai Y M, Jiang S F, Mao G Z, et al. 41 tea germplasm resources: genetic diversity and relationship analysis based on SSR molecular markers [J]. Chinese Agricultural Science Bulletin,2023,39(30):54-60.

[19]林开勤,李悦欣,魏杰,等. 基于SSR分子标记分析石阡茶树资源遗传多样性[J]. 西南农业学报,2024,37(7):1435-1441.
Lin K Q, Li Y X, Wei J, et al. Genetic diversity of tea plant resources in Shiqian based on SSR molecular markers [J]. Southwest China Journal of Agricultural Sciences, 2024, 37(7):1435-1441.

[20]刘娇,杨雄伟,何胜聪,等.云南南涧县古茶树资源遗传多样性的EST-SSR分析[J/OL].热带亚热带植物学报,1-9[2024-11-26].http://kns.cnki.net/kcms/detail/44.1374.q.20240517.1459.002.html.
Liu J, Yang X W, He S C, et al. EST-SSR analysis of genetic diversity of ancient tea tree resources in Nanjian County, Yunnan Province [J/OL]. Journal of Tropical and Subtropical Botany, 1-9 [2024-11-26]. http://kns. cnki. net/kcms/detail/44.1374.q.20240517.1459.002.html.

[21]彭磊,赵懿琛,陆莹霞,等.41份古茶树种质资源遗传多样性SSR标记分析[J/OL].分子植物育种,1-15[2024-11-26].http://kns.cnki.net/kcms/detail/46.1068.S.20240315.1058.006.html.
Peng L, Zhao Y C, Lu Y X, et al. Genetic diversity analysis of 41 ancient tea germplasm resources by SSR markers[J/OL]. Molecular Plant Breeding, 1-15 [2024-11-26]. http://kns.cnki.net/kcms/detail/46.1068.S.20240315.1058.006.html.

[22]苏芹,邵陈禹,黄梦迪,等.耐寒茶树新品种CAPS分子标记辅助育种技术体系构建[J/OL].分子植物育种,1-14[2024-11-26].http://kns. cnki. net/kcms/detail/46.1068. S. 20240531.1609.016.html.
Su Q, Shao C Y, Huang M D, et al. Construction of CAPS molecular marker-assisted breeding technology system for new cold-tolerant tea varieties[J/OL]. Molecular Plant Breeding, 1-14 [2024-11-26]. http://kns. cnki. net/kcms/detail/46.1068. S.20240531.1609.016.html.

[23]田悦,张垚,赵卫国,等. 基于SSR标记对茶树紫阳群体种种

质资源的遗传多样性研究[J]. 江西农业学报,2023,35(12):69-75.
Tian Y, Zhang Y, Zhao W G, et al. Genetic diversity of *Camellia sinensis* Ziyang population germplasm resources based on SSR markers[J]. Acta Agriculturae Jiangxi,2023,35(12):69-75.

[24]王攀,于文涛,吕水源,等. 福建永泰茶树种质资源遗传多样性分析及DNA指纹图谱构建[J]. 福建农业学报,2024,39(1):66-74.
Wang P, Yu W T, Lü S Y, et al. Genetic diversity and DNA fingerprints of tea germplasms in Yongtai, Fujian [J]. Fujian Journal of Agricultural Sciences,2024,39(1):66-74.

[25]苟志辉,杜尚嘉,符生波,等. 五指山野生大叶茶叶片代谢物鉴定和特征分析[J]. 热带生物学报,2024,15(4):400-406.
Gou Z H, Du S J, Fu S B, et al. Identification and characteristic analysis of metabolites from the leaves of wild Assam tea (C. sinensis var. assamica) in Wuzhishan [J]. Journal of Tropical Biology,2024,15(4):400-406.

[26]Chen S F, Zhou Y Q, Chen Y R, et al. Fastp: an ultra-fast all-in-one FASTQ preprocessor [J]. Bioinformatics, 2018, 34(17):i884-i890.

[27]Zhang W Y, Zhang Y J, Qiu H J, et al. Genome assembly of wild tea tree DASZ reveals pedigree and selection history of tea varieties[J]. Nature Communications,2020,11(1):3719.

[28]Poplin R, Chang P C, Alexander D, et al. A universal SNP and small-indel variant caller using deep neural networks[J]. Nature Biotechnology,2018,36(10):983-987.

[29]Krzywinski M, Schein J, Birol I, et al. Circos: an information aesthetic for comparative genomics [J]. Genome Research, 2009,19(9):1639-1645.

[30]He W M, Xu L, Wang J X, et al. VCF$_2$PCACluster: a simple, fast and memory-efficient tool for principal component analysis of tens of millions of SNPs[J]. BMC Bioinformatics,2024,25(1):173.

[31]Kumar S, Stecher G, Peterson D, et al. MEGA-CC: computing core of molecular evolutionary genetics analysis program for automated and iterative data analysis [J]. Bioinformatics,2012,28(20):2685-2686.

[32]Pavlidis P, Živkovic D, Stamatakis A, et al. SweeD: likelihood-based detection of selective sweeps in thousands of genomes[J]. Molecular Biology and Evolution,2013,30(9):2224-2234.

[33]Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools[J]. Bioinformatics,2011,27(15):2156-2158.

[34]Wang F, Zhang B H, Wen D, et al. Chromosome-scale genome assembly of *Camellia sinensis* combined with multi-omics provides insights into its responses to infestation with green leafhoppers[J]. Frontiers in Plant Science, 2022, 13:1004387.

[35]Wang X C, Feng H, Chang Y X, et al. Population sequencing enhances understanding of tea plant evolution [J]. Nature Communications,2020,11(1):4447.

[36]Wei C L, Yang H, Wang S B, et al. Draft genome sequence of *Camellia sinensis* var. *sinensis* provides insights into the evolution of the tea genome and tea quality[J]. Proceedings of the National Academy of Sciences of the United States of America,2018,115(18):E4151-E4158.

(编辑:吕俊俐)