

大麻 EST-SSR 遗传结构分析及指纹图谱构建

边 境, 王晓楠, 曹 焜, 赵 越, 王云云, 张晓艳, 孙宇峰

(黑龙江省科学院大庆分院, 大庆 163319)

摘要: 大麻是一年生草本植物, 一种多用途、可持续的作物。迄今为止, 关于大麻遗传结构的研究还很少。本研究通过 EST-SSR 分子标记分析大麻的遗传多样性和种群结构。结果表明, 20 对引物共扩增出 113 个清晰条带, 其中 113 个 (100%) 是多态性的; 共检测到 232 个等位基因, 平均每对引物检测到 4.0176 个等位基因; 观测杂合度 (H_o) 平均为 0.7102, 期望杂合度 (H_e) 平均为 0.6935; 200 个个体香农信息指数介于 0.7204~2.4625 之间, 平均值为 1.5368; 多态信息含量 (PIC) 变化范围为 0.3519~0.8801, 平均为 0.6558; 平均基因流 (N_m) 平均值为 13.6525。基于种群遗传结构、主成分分析和未加权的算术平均对组法 (UPGMA) 分析, 将大麻材料聚类为 3 组。不同聚类方法之间结果相似, 但 3 种模型的少数个体植株分布不同。聚类结果、基因多样性和遗传相似系数表明, 大麻个体总体亲缘关系较为密切。同时用 5 对核心引物能够区分参试种质, 并为每份种质构建了指纹图谱。研究结果为今后的大麻育种、遗传改良和核心种质资源收集提供了参考。

关键词: 大麻; EST-SSR; 遗传多样性; 种群结构

Genetic Structure Analysis and Fingerprint Construction of Hemp by EST-SSR Analysis

BIAN Jing, WANG Xiaonan, CAO Kun, ZHAO Yue, WANG Yunyun, ZHANG Xiaoyan, SUN Yufeng

(Daqing Branch of Heilongjiang Academy of Sciences, Daqing 163319)

Abstract: Cannabis is an annual herb and a versatile, sustainable crop, while uncovering the population diversity remain poorly investigated. In this study, the EST-SSR molecular markers were used to analyze the genetic diversity and population structure of cannabis. The results showed that a total of 113 score bands were amplified using 20 pairs of primers, of which all (100%) were polymorphic. A total of 232 alleles were detected, with an average of 4.0176 alleles per primer pair. The average observed heterozygosity (H_o) was 0.7102, and the average expected heterozygosity (H_e) was 0.6935. The Shannon information index of 200 individuals ranged from 0.7204 - 2.4625, with an average value of 1.5368. Polymorphism information content (PIC) ranged from 0.3519 to 0.8801, with an average of 0.6558. The mean gene flow (N_m) was 13.6525. Based on population genetic structure, principal component analysis and unweighted group analysis (UPGMA) with arithmetic mean, cannabis materials were grouped into 3 groups. The results were similar between the clustering methods, but the distribution of minority individual plants was different among the three models. The results of clustering, genetic diversity and genetic similarity coefficient showed that cannabis individuals were closely related to each other. In addition, five pairs of core primers, which were able to distinguish the test germplasm, were deployed for the fingerprint construction. Collectively, these results provided a reference for the future

收稿日期: 2023-05-31 修回日期: 2023-06-20 网络出版日期: 2023-07-07

URL: <https://doi.org/10.13430/j.cnki.jpgr.20230531001>

第一作者研究方向为分子植物育种, E-mail: 18645993347@163.com

通信作者: 张晓艳, 研究方向为作物栽培育种, E-mail: zhangxylibin@163.com

孙宇峰, 研究方向为作物栽培育种, E-mail: sunyf888@163.com

基金项目: 黑龙江省科学院对外合作项目 (DWHZ2023DQ01); 黑龙江省科学院青年创新基金项目 (CXMS2023DQ01); 黑龙江省省属科研院所科研业务费项目 (CZKYF2021-2-A004, CZKYF2022-1-A004)

Foundation projects: Foreign Cooperation Project of Heilongjiang Academy of Sciences (DWHZ2023DQ01); Youth Innovation Fund of Heilongjiang Academy of Sciences (CXMS2023DQ01); Research Expenses of Provincial Research Institutes of Heilongjiang Province (CZKYF2021-2-A004, CZKYF2022-1-A004)

breeding, genetic improvement and collection of core germplasm resources of cannabis.

Key words: hemp; EST-SSR; genetic diversity; population structure

大麻(*Cannabis sativa* L.), 俗称汉麻, 是一种草本植物, 属于大麻科。它被认为是最古老的栽培植物之一, 可用于多个应用领域, 从农业和植物修复到食品、饲料、化妆品、建筑和制药行业。根据不同的应用领域, 可以从大麻中获得各种具有工业价值的产品, 例如纤维和碎屑, 生物建筑和隔热材料, 具有重要营养和功能特性的种子、面粉、油和具有药理学意义的生物活性化合物^[1]。

遗传多样性是生物多样性的的重要组成部分, 是生态系统和物种多样性的基础^[2-4]。了解种质资源的遗传多样性和结构有助于高效、合理地开发、保护和利用种质资源^[5-6]。目前, 研究人员将多种作物的遗传多样性和种群结构研究作为重要的基础研究, 并且开展了许多的相关研究。使用形态学和农艺学特征检测遗传变异, 这些特征通常表现出受环境因素强烈影响的多基因遗传。除了环境差异外, 地理隔离、系统地理学模式、基因流动和种群动态也会导致选择压力, 从而导致空间结构遗传变异。由于基因组中基因的适应性变化, 物种将在表型和物候方面进行适应性进化^[7-8]。然而, 由于缺乏大量非模式物种的基因组信息, 无法确定与逆境相关的多个候选基因。同时由于这些候选基因在其基因组的局部适应中起着重要作用, 因此不容忽视^[9]。分子标记分析有效地替代了形态学和农艺学特征分析。目前, 最常见的遗传多样性分子标记包括 SNP、RFLP、RAPD、ISSR、简单序列重复序列(SSR, simple sequence repeat)和 AFLP^[10]。

使用每种技术的选择受到诸如应用的难易程度、基因组覆盖率、成本和自动化兼容性等因素的影响。简单序列重复分子标记是共显性的, 符合孟德尔定律。它操作简单, 具有高度的重现性和可靠性, 能够揭示子代和亲本之间不受基因表达、培养条件或环境条件影响的遗传差异, 并且可以显示出大量的多态性^[11-12]。SSR 分子标记广泛应用于植物种质鉴定、遗传多样性、遗传连锁图谱构建、基因定位和克隆以及数量性状位点(QTLs)分析^[13-16]。由于表达序列标签(EST, expressed sequence tag)-SSR 标记来自转录区域, 因此它们具有很高的成功扩增率和相关基因注释^[17]。使用 SSR 分子标记的遗传多样性分析已广泛应用于多种作物^[15, 18-21], 如多年生黑麦草、紫花苜蓿和小麦。信朋飞等^[22]利用 SSR

标记对大麻种质资源进行指纹图谱的构建。大量文献表明, SSR 分子标记结果可以揭示近缘物种的亲缘关系, 鉴定品种^[23]。

表达序列标签是剖析复杂性状以及估计分子多样性和种群结构的基础^[24], SSR 标记价格低廉且易于通过聚合酶链反应检测, 因此作为有价值的分子标记被广泛应用于群体筛查。本研究基于前期开发的大麻 EST-SSR 分子标记^[22], 筛选适用于大麻纤维用、籽用和花叶用类型及其种群遗传多样性和亲缘关系分析的核心引物, 并构建指纹图谱, 以期为大麻品种鉴定及种质创新提供依据。

1 材料与方法

1.1 试验材料与 DNA 提取

试验材料为黑龙江省科学院大庆分院审定品种及搜集保存的 200 份大麻种质资源(详见 <https://doi.org/10.13430/j.cnki.jpgr.20230531001>, 附表 1), 包括纤维用类型 78 份、籽用类型 63 份、花叶用类型 59 份。

选取大麻叶片 0.5 g, 运用柱式植物组织基因组 DNA 抽提试剂盒(上海生工生物技术有限公司)提取 DNA。-20℃保存备用。

1.2 EST-SSR 引物设计及 PCR 扩增

从前期开发的 EST-SSR 引物中随机筛选出 40 对, 由上海生工生物技术有限公司合成。随机选取 8 个品种的 DNA 混合样品, 对 SSR 引物进行多态性筛选, 筛选出条带清晰、稳定性好的 20 对引物用于所有材料的 PCR 扩增, 引物序列详见 <https://doi.org/10.13430/j.cnki.jpgr.20230531001>, 附表 2。试验所用试剂均购自上海生工生物技术有限公司。

PCR 反应体系为 25 μ L, 其中 DNA (20~50 ng/ μ L) 1 μ L, 10 \times Taq Buffer (with MgCl₂) 2.5 μ L, 引物 (10 μ mol/L) 各 0.5 μ L, Taq 酶 (5 U/ μ L) 0.2 μ L, dNTP (mix) (5 μ mol/L) 0.5 μ L, 用 dd H₂O 补足。PCR 扩增程序为 95 $^{\circ}$ C 5.0 min; 94 $^{\circ}$ C 30 s, 60 $^{\circ}$ C 30 s, 72 $^{\circ}$ C 30 s, 30 个循环; 72 $^{\circ}$ C 延伸 10 min。PCR 扩增产物的片段大小采用 QIAxcel 高级毛细管电泳仪检测。

1.3 数据分析

利用 BioCalculator 软件通过计算各峰的峰数、峰高、峰宽、峰面积等特征, 对扩增产物的单次数据

进行准确分析。选用的多态性较好的5对引物(E-20、E-24、E-26、E-31、E-17)扩增出的DNA片段按从小到大的顺序排列,将分辨率高,条带清晰记为“1”,无条带或模糊条带记为“0”。

根据电泳结果,在凝胶相同迁移率位置上,有条带的记为“1”,无条带的记为“0”,构成“0,1”序列数据阵。计算多态位点数及多态率,使用Excel 2021 计算多态性条带数(NPB, the number of polymorphic bands)和多态性条带百分比(PPB, the percentage of polymorphic bands)。GeneAEx 6.51b2^[25]软件转换各种文件格式进行不同分析,计算遗传多样性参数,包括等位基因数(N_a , the number of alleles)、有效等位基因数(N_e , the effective number of alleles)、香农信息指数(I , Shannon's information index)、基因流(N_m , gene flow)和遗传分化系数(F_{st} , the F-Statistic)。利用PIC Calc软件计算各引物多态性信息量(PIC, the polymorphic information content)。使用GeneAEx 6.51b2软件进行遗传距离分析、主成分分析(PCoA),基于Nei的无偏遗传距离

矩阵与MEGA 5.1进行未加权的算术平均对组法(UPGMA, unweighted pair-group method with arithmetic means)聚类分析^[26]。

使用STRUCTURE 2.3.4^[27]软件分析种群遗传结构,使用基于模型的聚类算法实现贝叶斯框架和马尔可夫链蒙特卡罗(MCMC, markov chain monte carlo)算法。为了确定最佳的亚群数量(K),对范围从2到10的每个K值进行了5次独立运行^[28]。每次运行都包含10,000步的老化期,然后是100,000次MCMC迭代。根据Evanno等^[29]开发的模型,估计 ΔK 参数基于连续K值之间数据对数概率的变化率,以确定最佳K值。

2 结果与分析

2.1 简单序列重复标记的多态性

在PCR反应中,20对SSR引物对200个个体共扩增出113个位点,全部表现出多态性。每个引物组合的多态位点数量从2到9不等,平均为5.65个位点(表1)。20对引物的扩增片段大小不一,在

表1 20条大麻EST-SSR引物的多态性分析

Table 1 Polymorphism analysis of 20 EST-SSR primers of *Cannabis*

引物 Primer	总条带 数 TNB	多态性 条带数 NPB	多态性			香农信息 指数 I	观测 杂合度 H_o	期望 杂合度 H_e	多态性 信息含量 PIC	遗传分化 系数 F_{st}	基因流 N_m
			条带百 分比 (%) PPB	等位 基因数 N_a	有效等位 基因数 N_e						
E11	5	5	100	5	1.6361	0.7204	0.3015	0.3898	0.3519	0.0087	28.4476
E17	5	5	100	16	5.4226	1.8902	0.8200	0.8176	0.7901	0.0105	23.5450
E18	9	9	100	9	2.8520	1.2180	0.7700	0.6510	0.5842	0.0146	16.8654
E20	5	5	100	20	7.8965	2.3713	0.7739	0.8756	0.8620	0.0112	22.0531
E22	6	6	100	14	4.4984	1.8273	0.6550	0.7796	0.7530	0.0088	28.0691
E23	7	7	100	8	3.0620	1.2248	0.8400	0.6751	0.6083	0.0139	17.6978
E24	3	3	100	19	9.0166	2.4625	0.7889	0.8913	0.8801	0.0120	20.5975
E25	6	6	100	15	2.7707	1.5314	0.7300	0.6407	0.6133	0.0088	28.0020
E26	5	5	100	12	6.0606	2.0438	0.6900	0.8371	0.8170	0.0279	8.7112
E27	4	4	100	11	3.2549	1.4694	0.8000	0.6945	0.6560	0.0109	22.7555
E3	6	6	100	9	4.2465	1.5819	0.7300	0.7664	0.7254	0.0611	3.8401
E31	4	4	100	15	6.0146	1.9361	0.9050	0.8358	0.8120	0.0154	16.0308
E32	7	7	100	9	3.5470	1.4486	0.8342	0.7199	0.6751	0.0121	20.4881
E37	8	8	100	10	3.4583	1.4709	0.4900	0.7126	0.6683	0.0418	5.7290
E4	5	5	100	10	2.4153	1.1295	0.7424	0.5875	0.5131	0.0064	38.6339
E40	3	3	100	8	3.9149	1.4935	0.8550	0.7464	0.7015	0.0094	26.2232
E43	2	2	100	18	4.2055	1.9170	0.7839	0.7641	0.7414	0.0096	25.8443
E48	7	7	100	9	1.8701	0.9827	0.5350	0.4664	0.4374	0.0386	6.2266
E6	3	3	100	9	2.4977	1.2097	0.6784	0.6011	0.5493	0.0295	8.2317
E7	5	5	100	6	1.7120	0.8076	0.4800	0.4169	0.3757	0.0061	40.9830
均值 Mean	5.65	5.65	100	11.6	4.0176	1.5368	0.7102	0.6935	0.6558	0.0180	13.6525

TNB: Total number of bands; NPB: Number of polymorphic bands; PPB: Percentage of polymorphic bands

100~300 bp之间变化。所有引物对均具有较高的基因多样性值,鉴定出较高水平的多态性。平均有效等位基因数(N_e)为4.0176个。同时共检测到232个等位基因,每个引物的平均等位基因数(N_a)为11.6。引物还表现出较高的香农信息指数(I),200份大麻种质的香农信息指数介于0.7204~2.4625之间,平均值为1.5368。 PIC 及香农指数最高值均为引物E24。观测杂合度(H_o , observed heterozygosity)的变化范围为0.3015~0.9050,平均值为0.7102,期望杂合度(H_e , expected heterozygosity)在0.3898~0.8913之间,平均值为0.6935。基因流(Nm)平均值为

13.6525,表明存在基因交流($Nm>1$)。

2.2 不同种群遗传多样性分析

对3种类型参试群体进行遗传多样性指数分析,结果显示(表2),3个种群平均观测杂合度(H_o)为0.7063,期望杂合度(H_e)为0.6840,期望杂合度均小于观测杂合度,说明各种群内遗传多样性较高。籽用型群体的等位基因数(N_a)为9.65,香农信息指数(I)为1.5545,期望杂合度(H_e)为0.7015,观测杂合度(H_o)为0.7017, PIC 值为0.6602,均高于其他两个群体,说明籽用型群体遗传多样性与其他两个类型相比较高。

表2 3个大麻种质群的遗传多样性比较分析

Table 2 Comparative analysis of genetic variation for the three groups of Cannabis

群体 Population	样本数量 Sample size	等位基因数 N_a	有效等位 基因数 N_e	香农信息 指数 I	观测杂合度 H_o	期望杂合度 H_e	多态性信息 含量 PIC
纤用Fiber	78	8.90	3.9575	1.5014	0.7556	0.6930	0.6503
花叶用Floral leaf	63	7.60	3.5174	1.3756	0.6616	0.6580	0.6134
籽用Seed	59	9.65	4.0388	1.5545	0.7017	0.7015	0.6602
均值Mean		8.72	3.8379	1.4772	0.7063	0.6840	0.6413

2.3 群体间的遗传距离和遗传相似度

将200份大麻材料按用途划分为3个种群进行遗传距离比较,如表3所示,3个种群间的遗传距离在0.0314~0.0805范围内,Nei's遗传一致性在0.9227~0.9691范围。籽用型种群与纤用型种群遗传距离最小,基于遗传距离构建的UPGMA聚类树也验证了籽用型与纤用型种群亲缘关系较近(图1)。同时分析3个种群的基因流和遗传分化系数(表4)观察到花叶用种群与其他种群的基因流相对较小,遗传分化系数较大,这可以解释为什么花叶用种群被单独归为一组。

表3 3个种群 Nei's 遗传一致度和遗传距离的无偏估计

Table 3 Unbiased estimation of Nei's genetic identity and genetic distance among three populations

群体 Population	纤用 Fiber	花叶用 Floral leaf	籽用 Seed
纤用Fiber		0.9227	0.9691
花叶用Floral leaf	0.0805		0.9392
籽用Seed	0.0314	0.0627	

对角线上方为Nei's遗传一致度,对角线下方为Nei's遗传距离
Nei's genetic identity is above diagonal and genetic distance is below diagonal

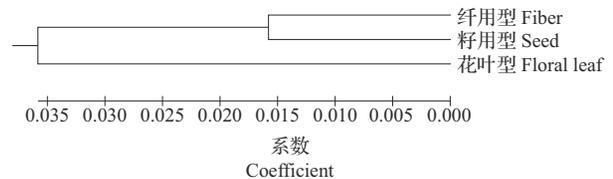


图1 基于遗传距离构建的UPGMA聚类树
Fig. 1 UPGMA clustering tree based on genetic distance

表4 三个种群之间的基因流遗传分化系数

Table 4 Gene flow and genetic differentiation coefficient between the three populations

群体 Population	纤用 Fiber	花叶用 Floral leaf	籽用 Seed
纤用Fiber		8.036	40.137
花叶用Floral leaf	0.030		11.448
籽用Seed	0.011	0.021	

对角线上方为基因流(Nm),对角线下方为遗传分化系数(F_{st})

Above the diagonal is gene flow (Nm), below the diagonal is the F_{st} Statistics (F_{st})

2.4 种群结构分析

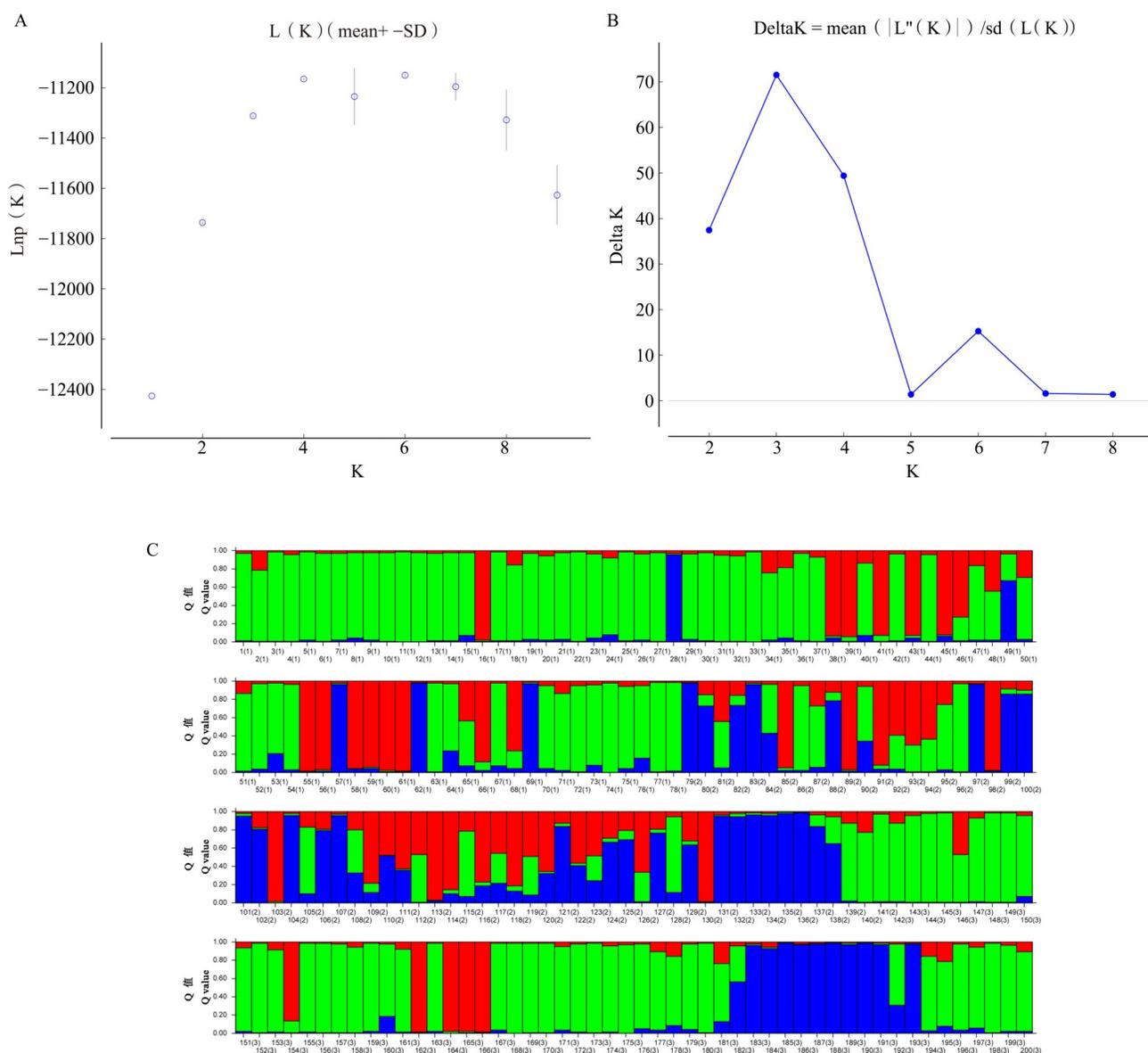
通过UPGMA聚类分析、PCoA分析和遗传结构分析,进一步探讨了基于遗传距离的不同类群和亚类种质之间的关系。使用Structure 2.3.4软件分析

了从200个大麻个体扫描的20个SSR引物对的标记。 $\ln P(K)$ 随着K的增加而不断变化,没有最大值,无法确定最佳亚群数(图2A)。根据Evanno等^[29]的方法确定最佳分类数。当 ΔK 值随K值变化时,一个明显的峰值可以确定为最佳分类数(图2B)。当K=3时,划分为3个类群(图2C)。其中类群1(绿色)113份,主要为纤维型资源,类群2(红色)44份,花叶型资源居多,类群3(蓝色)43份,多为籽用型品种资源。

UPGMA 树状图表明,种质可以分为3个组群

(图3)。200份大麻的聚类个体材料与种群遗传结构分析结果基本相符。基于UPGMA 树状图,组群I主要为纤维型大麻资源,组群II花叶型资源居多,组群III主要为籽用型资源。

对200份大麻个体进行主成分分析(图4),图中位置的距离代表亲缘关系的距离。PCoA 结果与UPGMA 树状图和种群结构分析结果基本一致。3组大麻个体分布相对集中,表明它们之间的亲缘关系密切。



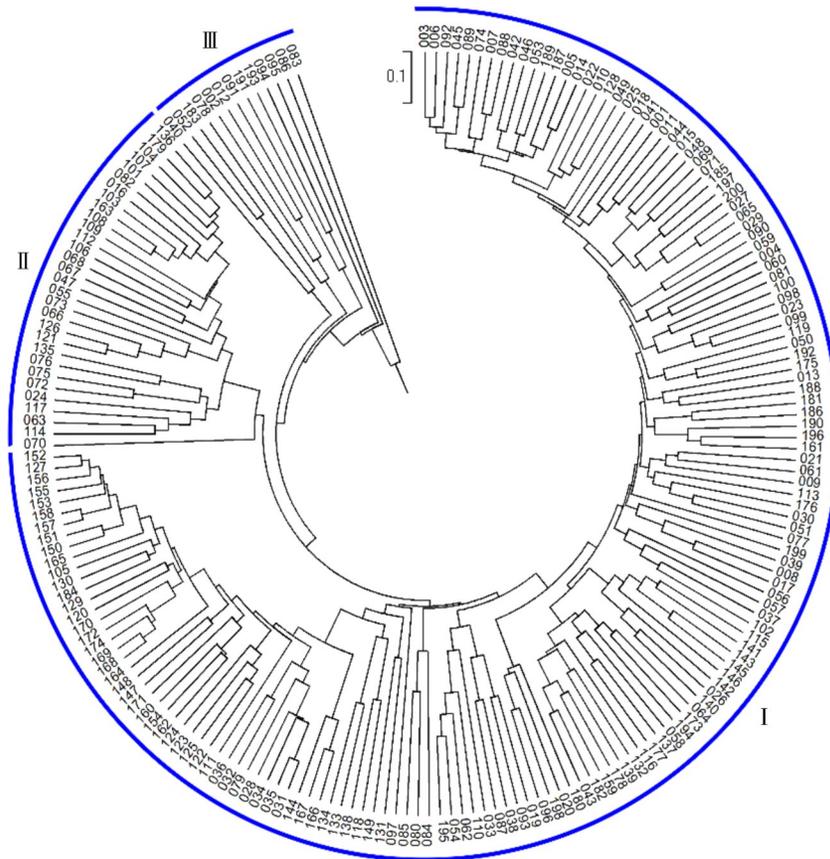
A: K值对K数的平均对数;B: K和 ΔK 的数量之间的关系;C: 使用STRUCTURE 2.3.4软件(K=3)确定的大麻种群结构。

横坐标数字代表资源序号和用途类型,1,绿色:纤维型;2,红色:花叶型;3,蓝色:籽用型

A: Average logarithm of K value to K number; B: The relationship between the quantities of K and ΔK ; C: Population structure of industrial hemp determined using STRUCTURE 2.3.4 software (K=3). The horizontal number represents the resource serial number and type of use. 1, green: Fiber type; 2, red: Mosaic type; 3, blue: Seed type

图2 200份大麻种群结构分析

Fig. 2 Population structure analysis of 200 industrial hemp samples



图中编号代表200份大麻个体(与附表1相对应)

The numbers in the figure represent 200 cannabis individuals (corresponding to schedule 1)

图3 基于UPGMA的200份大麻资源聚类分析

Fig. 3 Cluster analysis of 200 industrial hemp resources based on UPGMA

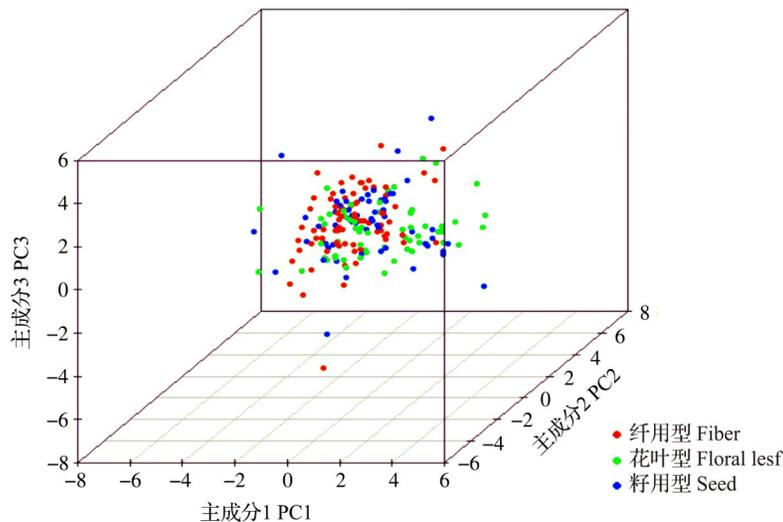


图4 200份大麻资源遗传多样性的主成分分析

Fig. 4 Principal component analysis of genetic diversity of 200 industrial hemp resources

2.5 指纹图谱的构建

经统计分析,5对引物在200份材料中共检测到25条清晰条带。将筛选出的5对核心引物E-20、

E-24、E-26、E-31、E-17依次编号为A~E,依据5对引物扩增的多态性位点,按照指纹图谱构建方法为每一份材料建立SSR指纹图谱代码(表5)。

表 5 200 份大麻种质指纹图谱

Table 5 Germplasm fingerprints of 200 industrial hemp samples

编号 Code	指纹图谱 Fingerprints	编号 Code	指纹图谱 Fingerprints	编号 Code	指纹图谱 Fingerprints
1	A0000B00010C01000D00101E00100	34	A00001B00000C10000D10010E10000	67	A01010B01000C10000D00000E00010
2	A01000B11000C00100D00001E00110	35	A00001B00000C10000D10010E10001	68	A00000B01100C11000D10100E00011
3	A00000B10010C00010D00101E00101	36	A10010B00000C10000D00110E10100	69	A01000B00000C01100D00001E00011
4	A00100B01000C01000D00101D00101	37	A01001B10000C01000D10100E00000	70	A00101B01000C11000D00000E00010
5	A00100B01000C00001D01100E00110	38	A10000B00000C00000D10100E01100	71	A01000B01000C01000D00001E00001
6	A00000B01000C00010D00100E00001	39	A10000B10000C00000D00101E00001	72	A00000B00010C11000D10100E00010
7	A10000B01100C01000D01100E00101	40	A01000B00100C11000D10100E00100	73	A00000B10010C01000D01100E00011
8	A01001B00100C00000D10100E00110	41	A10000B01000C01000D01100E00011	74	A00010B01000C10000D10100E00110
9	A00000B00100C00000D01100E00101	42	A00100B01000C00100D10001E00110	75	A00001B00000C11000D10100E00110
10	A10001B10000C10000D10000E00011	43	A00011B00100C11000D10000E00101	76	A01100B01000C10000D10000E00110
11	A10000B10000C01000D00101E00100	44	A01010B00010C01000D00101E00101	77	A01000B00100C01001D10000E00101
12	A00100B01000C00010D10100E00010	45	A00010B11000C01000D00101E00110	78	A01000B01000C10000D00001E00001
13	A00001B01000C01000D10000E00100	46	A00010B01000C10001D00101E00110	79	A00000B00000C10000D00110E10010
14	A00000B01000C10010D01100E00100	47	A10100B01000C10000D00101E00101	80	A01000B01100C10100D10001E00101
15	A01000B00000C01100D00001E00011	48	A01001B11000C01010D00101E00001	81	A00001B00001C01001D00100E00110
16	A01010B01000C01100D10000E00011	49	A00000B01100C10000D00010E00110	82	A01100B01001C01000D10000E01010
17	A00100B00100C00000D00100E00110	50	A10001B00000C00000D10001E01010	83	A00100B11000C01000D00010E10000
18	A10010B01000C01000D00010E00010	51	A10000B01000C00000D10100E11000	84	A00100B00010C00001D10000E00000
19	A00001B00000C01000D00100E00011	52	A01000B00001C01000D00110E00101	85	A01000B10000C01000D00101E00101
20	A00001B00100C00100D11000E01001	53	A01000B10000C00001D10100E00110	86	A00000B00000C00100D00100E00001
21	A01100B00000C00010D01100E01100	54	A01000B01010C00000D00110E00110	87	A01000B01000C00010D10100E01000
22	A00000B00100C11000D00100E00010	55	A00110B01000C10000D01000E01000	88	A10001B01100C00000D00100E00101
23	A00001B00000C10000D01100E	56	A01000B00000C00000D10000E01000	89	A00100B10000C01001D00101E00110
24	A10000B00000C00100D10000E00100	57	A01000B10000C10000D11000E00000	90	A00000B01000C00100D10010E00101
25	A00000B00100C00010D00101E10001	58	A00000B01000C10000D01000E00101	91	A00100B10000C10001D01000E01010
26	A00001B10000C01000D00101E00110	59	A00101B10000C00011D00001E00001	92	A00000B01010C01000D00101E11000
27	A01000B01000C10001D00001E00101	60	A00000B00000C00010D00101E00001	93	A01000B00000C11000D10100E00010
28	A10000B00000C10000D0100E10001	61	A01000B00000C10000D10100E01100	94	A00001B00000C10010D01100E01000
29	A01001B01000C00000D00100E01100	62	A01000B00000C00000D00110E00110	95	A01000B00000C10000D01010E00011
30	A01001B00010C00100D10100E01000	63	A00010B10000C10000D11000E00011	96	A00000B01010C10000D01000E01001
31	A00001B00000C10000D10010E10000	64	A00010B00000C10000D10100E00100	97	A00010B00010C10000D00010E00001
32	A10000B00000C10000D00110E00110	65	A01000B01000C01010D00100E00001	98	A00100B00000C10000D00100E00011
33	A00001B00001C10000D10010E00010	66	A00110B01000C11000D01000E01001	99	A00000B00000C10000D01010E00110

表5(续)

编号 Code	指纹图谱 Fingerprints	编号 Code	指纹图谱 Fingerprints	编号 Code	指纹图谱 Fingerprints
100	A00101B10000C01000D01010E00110	134	A01100B11000C01000D01100E10100	168	A01000B00000C01000D00010E10000
101	A01000B01000C01000D01000E00011	135	A00010B01000C01000D11000E00110	169	A01000B00000C01000D00010E10000
102	A01000B00100C10000D10100E00000	136	A00100B01000C01000D11000E00011	170	A01000B01000C10000D00110E10000
103	A00000B01001C01000D11000E01010	137	A01001B01000C01000D11000E00101	171	A10000B00100C01000D00110E10010
104	A00100B11000C01000D11000E00001	138	A00000B00010C01001D01000E00101	172	A01000B00000C00000D00010E10000
105	A10000B00000C01000D00110E10100	139	A00110B00000C11000D11000E01010	173	A01000B10000C10000D00011E10100
106	A00110B10001C01000D01000E00011	140	A00010B00000C01000D01000E00100	174	A01000B00000C01000D00010E10000
107	A01100B11000C01000D10000E00001	141	A00101B10000C01000D01100E00100	175	A00011B01000C00000D00001E00100
108	A10100B01010C01000D10000E01010	142	A00101B00000C01000D10100E00100	176	A01010B11000C10000D10010E01100
109	A00100B01001C11000D10000E00010	143	A01001B00100C11000D00100E01110	177	A01000B00001C10010D11000E00101
110	A01000B00001C00000D00110E00010	144	A00001B00100C01000D10010E00001	178	A00010B01000C11000D10100E00110
111	A01001B00010C01000D01010E10100	145	A00001B00000C00100D00100E00100	179	A00100B10100C10000D11000E00011
112	A00100B10000C01000D10000E00001	146	A00100B00000C01000D00100E01001	180	A00100B00100C00100D01010E00011
113	A01001B00000C00000D00100E10100	147	A11000B01000C01000D00010E01001	181	A00100B00100C10000D00100E10001
114	A01010B00000C01000D01000E01001	148	A01010B01000C11000D11000E01000	182	A00011B00000C11000D00100E00011
115	A00010B00000C01000D00100E00100	149	A00000B00000C01000D01000E00010	183	A10000B10000C01000D10100E10100
116	A00100B01000C11000D10100E00110	150	A01000B01000C11000D01010E10010	184	A00000B00000C00100D01010E10100
117	A00010B00000C11000D10000E00001	151	A01000B01000C10000D00101E10010	185	A01000B01000C00010D00001E00011
118	A01000B00100C11000D01000E00001	152	A01000B01000C10000D01010E10000	186	A10001B01000C00000D10000E00101
119	A00010B00100C00100D10000E00010	153	A01000B00000C01000D01010E10100	187	A10000B01000C00010D01100E00110
120	A01000B00000C10000D00110E10100	154	A01000B00000C10000D00010E10001	188	A00001B01100C11000D01000E00100
121	A00100B11000C10000D01100E00100	155	A01000B00000C10000D01010E10000	189	A00101B10000C00100D10100E00100
122	A00001B10000C10000D01010E10100	156	A01000B01000C10000D01010E10100	190	A00100B00000C10100D11000E01100
123	A00000B00000C10000D00010E10001	157	A01000B00000C11000D01010E10100	191	A00101B00000C11000D00101E00010
124	A01001B00000C10000D01000E10000	158	A01000B01000C11000D01010E00110	192	A00001B00001C01000D10010E01010
125	A01001B00000C10000D11000E00011	159	A00100B01010C10000D10000E01100	193	A11000B01000C01010D11000E01010
126	A00100B01000C01000D01000E01001	160	A10001B00000C01000D00010E00101	194	A00110B10000C10000D11000E10100
127	A01000B01000C10000D01010E11000	161	A01000B11000C01000D00101E00101	195	A00100B00000C00000D00110E00101
128	A00001B01000C00001D00010E00110	162	A01001B00000C01000D00010E01001	196	A00001B01000C01010D00001E00101
129	A01000B11000C10000D10100E1000	163	A10100B00011C01000D10000E00010	197	A01000B01000C10010D00001E00011
130	A01000B01000C10000D01010E00011	164	A01000B01000C01000D01010E10001	198	A01000B00100C00000D11000E01001
131	A01000B00001C10000D01100E00010	165	A01000B01000C01000D01010E10100	199	A00101B00000C01001D10100E01010
132	A00100B00001C11000D10010E01010	166	A01000B00000C00000D00110E00101	200	A01000B11000C00101D00100E00110
133	A00001B00000C01000D01010E10001	167	A00000B01000C00000D00110E00001		

3 讨论

3.1 分子标记多态性分析

EST-SSR是利用现有的EST序列进行电子筛查,然后进行PCR检测^[30]。使用EST开发SSR,避免了开发SSR引物过程中的克隆和测序步骤,充分利用现有数据,降低了开发成本^[31]。EST-SSR保守性好,在不同物种间具有良好的通用性,可以区分亲缘关系较近的材料。

探索性条件下基因型间的多态性率被认为是衡量DNA标记多样性分析效率的关键因素。大量研究指出,标记的多态性影响植物的遗传多样性水平。一般而言,使用多态性高的引物比使用多态性差的引物,供试材料的遗传参数更可靠。信朋飞等^[22]基于大麻EST信息建立SSR标记,为大麻遗传多样性研究提供理论依据。而本研究首次利用EST-SSR标记对大麻种群结构进行分析,从40对EST-SSR引物中筛选出20对引物对大麻种群进行标记。本研究中检测到的高遗传多样性可能是由于使用了为大麻基因组开发的EST-SSR分子标记,可以更好地区分基因位点。香农信息指数(I)介于0.7204~2.4625之间,平均值为1.5368。这一发现表明所选引物能够客观地揭示大麻种质资源的遗传多样性。总体而言,20对EST-SSR引物可以充分分析大麻材料之间的遗传差异。

3.2 大麻材料遗传变异及种群间遗传多样性

本研究根据大麻种质资源用途类型,将其分为3个亚群进行遗传结构分析。各种群香农信息指数介于1.3756~1.5545之间,遗传多样性水平较高。大麻种群的高度遗传多样性可能与该物种的异花授粉有关。有研究表明 $PIC \geq 0.5$,为高度多态位点; $0.25 < PIC < 0.5$,为中度多态位点; $PIC \leq 0.25$,为低度多态位点。在本研究中,20对引物 PIC 值平均为0.6558,说明大麻EST-SSR标记均表现为高度多态性,适合大麻遗传多样性分析。

可以中和种间分化和种内遗传漂变的高度基因流动在异花授粉植物中极为常见,可导致个体或种群之间的遗传多样性较低^[17]。本研究中,花叶用种群与其他种群之间的大麻基因流(Nm)较小,其他种群之间的 Nm 较高。Norman等^[32]提出, $Nm < 1$ 表明植物种群间遗传分化程度高,而 $Nm > 1$ 表明植物种群间遗传分化程度低。George等^[33]在异花授粉的白三叶草中证实了Norman等^[32]的理论。由于种群间基因交换频繁,种群间的亲缘关系得以维持,

因此种群间的遗传分化不显著。在进化过程中,大麻发生了巨大的遗传变异,与其他种群的遗传物质频繁交换。因此,在遗传育种层面形成了多态性丰富、遗传变异度高的种质资源。不同类型的种质资源收集和引进历史也可能影响基因流动并塑造新的种群遗传结构。

物种的遗传结构受多种因素的相互作用影响,例如种子和花粉的传播模式、种群统计历史、地质事件、地理或生态障碍以及环境因素的发散选择^[34]。根据3个群体两两之间的Nei's遗传距离,这3个种群被分为三组。籽用型和纤用型亲缘关系最近,同时在田间测评过程中发现籽用型和纤用型品种其外部形态特征较为相似,这可能是由于在育种过程中,将来源相同的品种资源按照不同用途而划分为不同的遗传分支。

3.3 大麻种质资源种群结构

本研究根据种群遗传结构分析,将200份材料分成3个类群,并将来自同一种群的个体也聚类到不同的类群中。在UPGMA聚类图中,花叶用类型的一些单株与籽用型和纤用型单株聚在一起,但在种群结构分析中,花叶用材料独立聚类一组。聚类可能是育种和驯化的结果,对多样性结构影响很大。选择和育种倾向于使植物保持具有经济价值的性状^[35]。此外,不同的环境也会引起遗传变化,从而影响种群结构的划分。也可能是大麻材料含有不同植物个体的遗传物质,由于所有个体的遗传信息由种群结构整合而成,因此会影响种群的结构。UPGMA聚类是基于大麻材料的遗传距离进行的,将密切相关的材料聚为一组,这可能具有不同遗传结构的材料聚集。Shen等^[36]分析了64份燕麦种质资源的群体遗传结构,也发现同一种族的种质资源聚类成不同的类群。

根据PCoA、UPGMA和STRUCTURE分析,本研究中的大麻种质基本被分为3组。200份大麻单株材料的分类基本符合群体遗传结构分析,但也存在一定差异。这可能是由于不同方法应用了不同的统计原理^[15]。PCoA可以根据原始数据的相异矩阵提供更有效的分类,这并不严格符合Hardy-Weinberg平衡假设。种群结构分析可以更好地了解遗传多样性,估计种质资源的变异情况,有利于对其进行有效利用。种群遗传结构通过贝叶斯聚类方法按概率将种质分配给亚群,并用于自然异交种群的细分。使用UPGMA分析的种质聚类是基于遗传距离实现的,它显示了种质之间更详细的关

系。总的来说,这3种方法对全面了解大麻种群遗传结构可以提供有效的帮助。此外,本研究所揭示的大麻差异性可以评价其在选择亲本组合时具有较高的育种和杂交优势,为选择具有强遗传差异的大麻杂交组合提供依据。

3.4 SSR标记构建指纹图谱

SSRs 作为一类具有等位变异高、共显性、检测简单快速、稳定性好等的分子标记,已在遗传多样性分析、指纹构建、性状标记和遗传连锁图谱构建等多个领域得到成熟应用^[37]。许多农学家和遗传学家对 SSRs 进行了广泛的研究和应用。本研究选择的5对核心引物,其 PIC 值在 0.7901~0.8801 之间,均属于高多态性引物,有利于品种资源鉴定。

4 结论

本研究利用 EST-SSR 标记对大麻的遗传结构进行分析,并且对 200 份大麻种质资源进行了指纹图谱的构建。研究表明,籽用型种群与纤维型种群遗传距离最小,基于遗传距离构建的 UPGMA 聚类树也验证了籽用型与纤维型种群亲缘关系较近;同时通过 UPGMA 聚类分析、PCoA 分析和遗传结构分析,进一步确定了 200 份大麻的聚类个体材料与种群遗传结构分析结果基本相符。分类结果、基因多样性和遗传相似系数表明,大麻个体总体亲缘关系较为密切。同时本研究选取 5 对核心引物对参试种质构建大麻指纹图谱,利用组合构成了大麻特有的 DNA 指纹,能够将这些材料逐一地区分开来。结果证实,大麻种质具有足够的遗传多样性。研究结果将为大麻杂交组合、标记辅助改良、种质资源保护和核心种质收集提供分子依据。

参考文献

- [1] Irakli M, Tsaliki E, Kalivas A, Kleisariis F, Sarrou E, Cook C M. Effect of genotype and growing year on the nutritional, phytochemical, and antioxidant properties of Industrial Hemp (*Cannabis sativa* L.) Seeds. *Antioxidants* (Basel, Switzerland), 2019, 8(10):491
- [2] Bailey J K, Schweitzer J A, Ubeda F, Koricheva J, Leroy C J, Madritch M D, Rehill B J, Bangert R K, Fischer D G, Allan G J, Whitham T G. From genes to ecosystems: A synthesis of the effects of plant genetic factors across levels of organization. *Biological Sciences*, 2009, 364(1523): 1607-1616
- [3] Haddad N M, Crutsinger G M, Gross K, Haarstad J, Tilman D. Plant diversity and the stability of foodwebs. *Ecology Letters*, 2011, 14(1): 42-46
- [4] Zhang C, Vornam B, Volmer K, Prinz K, Kleemann F, Köhler L, Polle A, Finkeldey R. Genetic diversity in aspen and its relation to arthropod abundance. *Frontiers in Plant Science*, 2014, 5: 806
- [5] Costa R, Pereira G, Garrido I, Tavares-De-Sousa M M, Espinosa F. Comparison of RAPD, ISSR, and AFLP molecular markers to reveal and classify orchardgrass (*Dactylis glomerata* L.) germplasm variations. *PLoS ONE*, 2016, 11(4): e0152972
- [6] Sork V L, Aitken S N, Dyer R J, Eckert A J, Legendre P, Neale D B J T G. Putting the landscape into the genomics of trees: Approaches for understanding local adaptation and population responses to changing climate. *Genomes*, 2013, 9(4): 901-911
- [7] Feng X J, Jiang G F, Fan Z. Identification of outliers in a genomic scan for selection along environmental gradients in the bamboo locust, *Ceracris kiangsu*. *Scientific Reports*, 2015, 5: 13758
- [8] Rellstab C, Gugerli F, Eckert A J, Hancock A M, Holderegger R. A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, 2015, 24(17): 4348-4370
- [9] Li Y, Zhang X X, Mao R L, Yang J, Miao C Y, Li Z, Qiu Y X. Ten years of landscape genomics: Challenges and opportunities. *Frontiers in Plant Science*, 2017, 8:2136
- [10] Li H, Ma Y, Pei F, Zhang H, Jiang M. Large-scale advances in SSR markers with high-throughput sequencing in *Euphorbia fischeriana* Steud. *Electronic Journal of Biotechnology*, 2021, 49:50-55
- [11] Wang K, Lin Z, Wang L, Wang K, Shi Q, Du L, Ye X. Development of a set of PCR markers specific to *Aegilops longissima* chromosome arms and application in breeding a translocation line. *Theoretical and Applied Genetics*, 2018, 131(1): 13-25
- [12] Daudi H, Shimelis H, Mathew I, Oteng-Frimpong R, Ojiewo C, Varshney R K. Genetic diversity and population structure of groundnut (*Arachis hypogaea* L.) accessions using phenotypic traits and SSR markers: Implications for rust resistance breeding. *Genetic Resources and Crop Evolution*, 2021, 68(2): 581-604
- [13] Ren R, Xu J, Zhang M, Liu G, Yao X, Zhu L, Hou Q. Identification and molecular mapping of a gummy stem blight resistance gene in wild watermelon (*Citrullus amarus*) germplasm PI 189225. *Plant Disease*, 2020, 104(1): 16-24
- [14] Chen C, Chang J, Wang S, Lu J, Liu Y, Si H, Sun G, Ma C. Cloning, expression analysis and molecular marker development of cinnamyl alcohol dehydrogenase gene in common wheat. *Protoplasma*, 2021, 258(4): 881-889
- [15] Wu F, Ma S, Zhou J, Han C, Hu R, Yang X, Nie G, Zhang X. Genetic diversity and population structure analysis in a large collection of white clover (*Trifolium repens* L.) germplasm worldwide. *PeerJ*, 2021, 9:e11325
- [16] Jiang W Z, Yao F J, Lu L X, Fang M, Wang P, Zhang Y M, Meng J J, Lu J, Ma X X, He Q, Shao K S. Genetic linkage map construction and quantitative trait loci mapping of agronomic traits in *Gloeostereum incarnatum*. *Journal of*

- Microbiology (Seoul, Korea), 2021, 59(1): 41-50
- [17] Sun M, Dong Z, Yang J, Wu W, Zhang C, Zhang J, Zhao J, Xiong Y, Jia S, Ma X. Transcriptomic resources for prairie grass (*Bromus catharticus*): Expressed transcripts, tissue-specific genes, and identification and validation of EST-SSR markers. *BMC Plant Biology*, 2021, 21(1): 264
- [18] 徐照龙, 易金鑫, 余桂红, 张大勇, 何晓兰, 王秀娥, 马鸿翔. 藜科6种耐盐植物遗传多样性的EST-SSR分析. *植物遗传资源学报*, 2011, 12(1): 113-120
- Xu Z L, Yi J X, Yu G H, Zhang D Y, He X L, Wang X E, Ma H X. EST-SSR based genetic diversity analysis on salt tolerant plants from six species in Chenopodiaceae. *Journal of Plant Genetic Resources*, 2011, 12(1): 113-120
- [19] 张金渝, 杨维泽, 崔秀明, 金航, 虞泓, 陈中坚, 沈涛, 杨涛. 三七栽培居群遗传多样性的EST-SSR分析. *植物遗传资源学报*, 2011, 12(2): 249-254
- Zhang J H, Yang W Z, Cui X M, Jin H, Yu H, Chen Z J, Shen T, Yang T. A rapid method for DNA extraction from plant tissue. *Journal of Plant Genetic Resources*, 2011, 12(2): 249-254
- [20] Zhang F, Wang C, Li M, Cui Y, Shi Y, Wu Z, Hu Z, Wang W, Xu J, Li Z. The landscape of gene-CDS-haplotype diversity in rice: Properties, population organization, footprints of domestication and breeding, and implications for genetic improvement. *Molecular Plant*, 2021, 14(5): 787-804
- [21] 张水明, 陈程, 陈芳芳, 汪天. 16个蝴蝶兰品种EST-SSR遗传多样性分析. *植物遗传资源学报*, 2013, 14(3): 560-564
- Zhang S M, Chen C, Chen F F, Wang T. Analysis of genetic diversity of 16 phalaenopsis cultivars using EST-SSR markers. *Journal of Plant Genetic Resources*, 2013, 14(3): 560-564
- [22] 信朋飞, 臧巩固, 赵立宁, 高春生, 程超华. 大麻SSR标记的开发及指纹图谱的构建. *中国麻业科学*, 2014, 36(4): 174-182
- Xin P F, Zang G G, Zhao L N, Gao C S, Cheng C H. Development of SSR markers and construction of fingerprint for *Cannabis* (*Cannabis sativa* L.). *Plant Fiber Sciences in China*, 2014, 36(4): 174-182
- [23] Wu X B, Duan L Z, Chen Q, Zhang D Q. Genetic diversity, population structure, and evolutionary relationships within a taxonomically complex group revealed by AFLP markers: A case study on *Fritillaria cirrhosa* D. Don and closely related species. *Global Ecology and Conservation*, 2020, 24: e01323
- [24] Stavridou E, Lagiatis G, Kalaitzidou P, Grigoriadis I, Bosmalis I, Tsaliki E, Tsiotsiou S, Kalivas A, Ganopoulos I, Madesis P. Characterization of the genetic diversity present in a diverse sesame landrace collection based on phenotypic traits and EST-SSR markers coupled with an HRM analysis. *Plants* (Basel, Switzerland), 2021, 10(4): 656
- [25] Peakall R, Smouse P E. GenAlEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research—an update. *Bioinformatics* (Oxford, England), 2012, 28(19): 2537-2539
- [26] Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Molecular Biology and Evolution*, 2018, 35(6): 1547-1549
- [27] Pritchard J K, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*, 2000, 155(2): 945-959
- [28] Wang M L, Zhu C, Barkley N A, Chen Z, Erpelding J E, Murray S C, Tuinstra M R, Tesso T, Pederson G A, Yu J. Genetic diversity and population structure analysis of accessions in the US historic sweet sorghum collection. *Theoretical and Applied Genetics*, 2009, 120(1): 13-23
- [29] Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 2005, 14(8): 2611-2620
- [30] Zheng Y, Zhang Z, Wan Y, Tian J, Xie W. Development of EST-SSR markers linked to flowering candidate genes in *Elymus Sibiricus* L. based on RNA Sequencing. *Plants* (Basel, Switzerland), 2020, 9(10): 1371
- [31] Fan M, Gao Y, Wu Z, Zhang Q. Linkage map development by EST-SSR markers and QTL analysis for inflorescence and leaf traits in *Chrysanthemum morifolium* Ramat.). *Plants* (Basel, Switzerland), 2020, 9(10): 1342
- [32] Norman C E, Diane R E. Population genetic consequences of small population size: Implications for plant conservation. *Annual Review of Ecology and Systematics*, 1993, 24: 217-242
- [33] George J, Dobrowolski M P, Van Zijl De Jong E, Cogan N O, Smith K F, Forster J W. Assessment of genetic diversity in cultivars of white clover (*Trifolium repens* L.) detected by SSR polymorphisms. *Genome*, 2006, 49(8): 919-930
- [34] Smith A L, Hodkinson T R, Villellas J, Catford J A, Csörgő A M, Blomberg S P, Crone E E, Ehrlén J, Garcia M B, Laine A L, Roach D A, Salguero-Gómez R, Wardle G M, Childs D Z, Elder D D, Finn A, Munné-Bosch S, Baudraz M E A, Bódis J, Brearley F Q, Bucharova A, Caruso C M, Duncan R P, Dwyer J M, Gooden B, Groenteman R, Hamre L N, Helm A, Kelly R, Laanisto L, Lonati M, Moore J L, Morales M, Olsen S L, Pärtel M, Petry W K, Ramula S, Rasmussen P U, Enri S R, Roeder A, Roscher C, Saastamoinen M, Tack A J M, Töpper J P, Vose G E, Wandrag E M, Wingler A, Buckley Y M. Global gene flow releases invasive plants from environmental constraints on genetic diversity. *Proceedings of the National Academy of Sciences of the United States of America*, 2020, 117(8): 4218-4227
- [35] Li Z, Yun L, Gao Z, Wang T, Ren X, Zhao Y. EST-SSR primer development and genetic structure analysis of *Psathyrostachys juncea* Nevski. *Frontiers in Plant Science*, 2022, 13: 837787
- [36] Shen G W, Li J S, Ren C Z, Hu Y G. Analysis of genetic diversity and population structure of oat germplasm from China and Canada. *Journal of Triticeae Crops*, 2010, 30: 617-624
- [37] Zheng X, Cheng T, Yang L, Xu J, Tang J, Xie K, Huang X, Bao Z, Zheng X, Diao Y, You Y, Hu Z. Genetic diversity and DNA fingerprints of three important aquatic vegetables by EST-SSR markers. *Scientific Reports*, 2019, 9(1): 14074

附表 1 200 份大麻材料明细表

Table S1 Information of 200 Cannabis millet accessions

编号 Code	名称 Name	类型 Type	来源 Source												
1	云麻 4 号	纤维用	中国云南	51	陵川	籽用	中国陵川	101	中龙红麻	药用	中国福建	151	2020-3	籽用	中国黑龙江
2	公主岭-1	纤维用	中国公主岭	52	包头	籽用	中国包头	102	T6-3	药用	中国黑龙江	152	2020-4	籽用	中国黑龙江
3	五常 40	纤维用	中国公主岭	53	H2018-1	籽用	中国黑龙江	103	HA2-2	药用	加拿大	153	2020-5	籽用	中国黑龙江
4	沈阳 1	纤维用	中国龙江县头站	54	M-1	籽用	中国黑龙江	104	HA2-4	药用	加拿大	154	2020-6	籽用	中国黑龙江
5	大庆-1	纤维用	中国黑龙江	55	五台-1	籽用	中国黑龙江	105	HA2-5	纤维用	加拿大	155	2019-1	籽用	中国黑龙江
6	大庆-2	纤维用	中国黑龙江	56	晋麻 1 号	籽用	中国山西	106	LSA2-2	药用	加拿大	156	2021-1	籽用	中国黑龙江
7	肇东农家	纤维用	中国肇东	57	甘肃-1	籽用	中国甘肃	107	LSA2-3	药用	加拿大	157	2021-2	籽用	中国黑龙江
8	龙江县农-1	纤维用	中国龙江县	58	吉引-1	籽用	中国吉林	108	WA2-1	纤维用	加拿大	158	2022-1	籽用	中国黑龙江
9	龙江县农-2	纤维用	中国龙江县	59	吉引-2	籽用	中国吉林	109	WA2-2	纤维用	加拿大	159	2021-4	药用	中国黑龙江
10	哈市道外区	纤维用	中国哈尔滨道外区	60	吉引-4	籽用	中国吉林	110	云麻 7 号	药用	中国云南	160	2021-5	药用	中国黑龙江
11	榆树	纤维用	中国海伦共和	61	吉引-5	籽用	中国吉林	111	龙大麻 5 号	药用	中国黑龙江	161	汉麻 22	药用	中国黑龙江
12	绥化-1	纤维用	中国绥化	62	2020-7	籽用	中国云南	112	龙大麻 6 号	药用	中国黑龙江	162	梦兰	药用	加拿大
13	绥化-2	纤维用	中国绥化	63	龙麻 5 号	籽用	中国哈尔滨	113	龙大麻 10 号	药用	中国黑龙江	163	北大荒	药用	中国黑龙江
14	公主岭-3	纤维用	中国公主岭	64	吉引-6	籽用	中国吉林	114	牡麻 2 号	药用	黑龙江	164	美引 1 号	药用	美国
15	白城 1	纤维用	中国白城	65	火麻 1 号	纤维用	中国黑龙江	115	方华 1 号	药用	黑龙江	165	汉麻 7 号	药用	中国黑龙江
16	白城 2	纤维用	中国白城	66	龙麻 1 号	籽用	中国黑龙江	116	方科 1 号	药用	黑龙江	166	赛麻 1 号	药用	中国黑龙江
17	大同	纤维用	中国佳木斯	67	龙麻 2 号	纤维用	中国黑龙江	117	金昶 1 号	药用	黑龙江	167	汉麻 11 号	药用	中国黑龙江
18	松原	纤维用	中国松原	68	汉麻 1 号	纤维用	中国黑龙江	118	Canada1 号	药用	加拿大	168	2017-158	纤维用	中国黑龙江
19	云南	纤维用	中国昆明	69	汉麻 2 号	纤维用	中国黑龙江	119	美生 2 号	药用	美国	169	2019-182	纤维用	中国黑龙江
20	通榆县 1	纤维用	中国通榆县	70	汉麻 4 号	纤维用	中国黑龙江	120	美生 3 号	药用	美国	170	2019-162	纤维用	中国黑龙江
21	通榆县 2	纤维用	中国通榆县	71	汉麻 5 号	纤维用	中国黑龙江	121	美生 4 号	药用	美国	171	2019-23	纤维用	中国黑龙江

22	通榆县 3	纤用	中国通榆县	72	汉麻 6 号	纤用	中国黑龙江	122	加美 3 号	药用	加拿大	172	新麻 1 号	纤用	中国黑龙江
23	内蒙古	纤用	中国内蒙	73	汉麻 9 号	籽用	中国黑龙江	123	以色列 1 号	药用	以色列	173	山西榆社	籽用	中国山西榆社
24	甘肃 1	纤用	中国甘肃	74	华夏汉麻 1 号	纤用	中国黑龙江	124	以色列 2 号	药用	以色列	174	甘肃大粒	籽用	中国甘肃
25	阳曲麻	纤用	中国公主岭	75	格雷西亚	纤用	俄罗斯	125	澳大利亚 1 号	药用	澳大利亚	175	望奎农家	纤用	中国望奎
26	绿化 1 号	纤用	中国龙江县头站	76	汉麻 12 号	纤用	中国黑龙江	126	加美 15	药用	加拿大	176	林甸小粒	纤用	中国林甸
27	延寿农家	纤用	中国延寿县	77	汉麻 13 号	纤用	中国黑龙江	127	中升 9	药用	加拿大	177	Mar-22	籽用	中国哈市
28	云南	纤用	中国昆明	78	汉麻 17 号	纤用	中国黑龙江	128	中升 10	药用	加拿大	178	XD45	纤用	中国吉林
29	甘南农家 1	纤用	中国甘南县	79	汉麻 19 号	纤用	中国黑龙江	129	八一农大 1	药用	中国黑龙江	179	XD39	纤用	中国吉林
30	甘肃 2	纤用	中国甘南县	80	汉麻 20 号	纤用	中国黑龙江	130	八一农大 2	药用	中国黑龙江	180	陕西定州	籽用	陕西
31	云南-1	药用	中国昆明	81	汉麻 21 号	纤用	中国黑龙江	131	八一农大 3	药用	中国黑龙江	181	俄罗斯	纤用	俄罗斯
32	云南-2	药用	中国昆明	82	A2	籽用	中国黑龙江	132	仁和药业 1	药用	中国黑龙江	182	大庆肇源	纤用	黑龙江肇源
33	云南-3	药用	中国昆明	83	二龙	籽用	中国龙江县	133	仁和药业 2	药用	中国黑龙江	183	吉林农家 1	籽用	中国吉林
34	云南-4	药用	中国昆明	84	东柳	籽用	中国龙江县	134	仁和药业 3	药用	中国黑龙江	184	吉林农家 2	纤用	中国吉林
35	云南-5	药用	中国昆明	85	农家品种 1	籽用	中国安达市	135	仁和药业 4	药用	中国黑龙江	185	吉林榆树农家籽	纤用	吉林
36	云南-6	药用	中国昆明	86	农家品种 2	籽用	中国安达市	136	仁和药业 6	药用	中国黑龙江	186	吉林农安农家籽	籽用	中国吉林
37	内蒙	纤用	中国内蒙	87	甘肃 1	籽用	中国甘肃	137	仁和药业 7	药用	中国黑龙江	187	内蒙赤峰农家籽	纤用	中国内蒙赤峰
38	内蒙	纤用	中国内蒙	88	沈阳 2	籽用	中国沈阳	138	仁和药业 8	药用	中国黑龙江	188	内蒙和林格尔农家籽	纤用	中国内蒙赤峰
39	延寿农家	纤用	中国延寿县	89	松原	籽用	中国吉林	139	仁和药业 9	药用	中国黑龙江	189	陕西火麻籽	籽用	中国陕西
40	白山	纤用	中国哈尔滨	90	太古晋农	籽用	中国黑龙江	140	仁和药业 10	药用	中国黑龙江	190	甘肃火麻籽	纤用	中国甘肃
41	云麻 2 号	纤用	中国林甸县红旗乡	91	2016-1	籽用	中国黑龙江	141	京麻 1 号	药用	中国黑龙江	191	山西火麻籽	纤用	中国山西

42	2013-4	纤维用	中国公主岭	92	2016-2	籽用	中国黑龙江	142	美国 1 号	药用	美国	192	云南火麻籽	籽用	中国云南
43	山西	籽用	中国公主岭	93	云麻 5 号	籽用	中国云南	143	2020Q01	药用	中国黑龙江	193	绿花 2 号	纤维用	中国黑龙江
44	2013 三连	籽用	中国伊春南岔	94	大麻 1 号	籽用	中国云南	144	2020Q08	药用	中国黑龙江	194	庆麻 1 号	纤维用	中国黑龙江
45	道外 2016	籽用	中国公主岭	95	蒙麻 1 号	籽用	中国内蒙古	145	2020Q36	药用	美国	195	绿花 2 号	籽用	中国黑龙江
46	海伦农场	籽用	中国云南	96	汾 3	籽用	中国山西	146	2020S02	药用	中国黑龙江	196	大庆大同农 家籽	纤维用	中国黑龙江
47	云麻-3	籽用	中国云南	97	汾 4	籽用	中国山西	147	同株材料 1	药用	中国黑龙江	197	紫花	纤维用	中国黑龙江
48	吉林范家屯	籽用	中国公主岭	98	龙大麻 1 号	纤维用	中国黑龙江	148	同株材料 2	药用	中国黑龙江	198	南非-1	药用	南非
49	吉林松原	籽用	中国松原	99	龙大麻 2 号	纤维用	中国黑龙江	149	2020-1	籽用	中国黑龙江	199	南非-3	药用	南非
50	晋麻 2	籽用	中国陵川	100	龙大麻 3 号	纤维用	中国黑龙江	150	2020-2	籽用	中国黑龙江	200	南非-4	药用	南非

附表 2 多态性 SSR 引物序列

Table S2 Polymorphism SSR primers sequences

引物名称 Primer name	正向引物序列(5'-3') Forward primer sequences(5'-3')	退火温度 (°C) Tm	反向引物序列(5'-3') Reverse primer sequences(5'-3')	退火温度 (°C) Tm	片段长度 (bp) Length	重复基序 Repeated motif
E-3	TCGATCCAAAAGGAAGCAAC	60.19	AGGTCAGAACCCAGCTCAGA	59.99	216	(AAT)7
E-4	CAAAAGTAGCAACTTGGTTTGG	58.85	TTGTTTGGCACCAAGTTTCA	60.13	214	(TCC)8
E-6	TGAATGGGATGATTTTGTGTG	59.24	TTGTTTCTCGTAATCCCTTTTCA	59.99	229	(TTG)8
E-7	CCGGAACAAGGTTGTCTGT	60.01	CATTGCTCGAGTCACCAGA	59.98	152	(CTG)7
E-11	TGTCATCATCCCAATTCTTGAG	59.93	TGACAAAGCACAAGGGATCA	60.24	189	(GAA)7
E-17	TCATCCTATCCCCTCTCTCTCTC	59.47	ATGGCTGCTGCATCATATCG	62.11	166	(TC)25
E-18	AGAAGCTCCTCCTCCTCCTC	59.13	ACAAAAGCGATCTGGTCCAT	59.56	150	(CT)11
E-20	CAACAAACAAGAAAGAAGAAGA	59.55	AGTATCCTCTTCAGGTGACCA	56.27	163	(TC)22
E-22	GCCAATTGTCGAGGTCCATA	60.68	ACAAAGAACACATAAACTAAACG	53.24	197	(AT)11
E-23	CATGAACAACCCATCTCACG	59.96	ATGGCGAACGATTCTCAAAG	60.21	159	(CTT)7
E-24	ACCTATTAACAGGAGAATCTTT	52.57	TGTGGATGCCATTTTGTATG	60.32	169	(AT)10
E-25	GATCCCAAGAAACCAAGGT	60.17	AACCACGAACCAAGGATGAG	59.97	198	(CT)10
E-26	ACCAACGTTTGACCTTGCAT	60.42	GGCTAGTCAAGTGGCCATA	60.37	152	(TA)10
E-27	TAGCCACGGAGGAAGAAGAA	59.95	CGTCCAATATCTTGCCGAAT	59.92	176	(GAA)8
E-31	CCCCCAAATTCCAATCTAT	59.85	GCCTCGTGATCCATCTCTCT	59.36	237	(CAT)7
E-32	GCTAGGGCTAATTAATAAGATGAAGAG	58.69	AACTGCCGAAGCTTCTCCTT	60.52	236	(GTT)11
E-37	TTTCTCTTTCTACCATACACCAA	60.04	CGATTAATCACACCATCAAACG	60.24	156	(AG)10
E-40	CGACAACAACGACGTTTCAG	60.34	AAATAACCTCGCCGGTATGA	59.43	195	(ACA)11
E-43	TGGCTCTCATCTCCTACACA	56.86	AGCACAAAAGCTGAGACACA	57.17	218	(TC)12
E-48	CGGTATGCAGTGTGGTTA	57.17	GTCAGTTCTCTCAAATGC	52.34	285	(GCCAA)4