植物遗传资源学报 2019, 20(5): 1309-1316
Journal of Plant Genetic Resources DOI: 10.13430/j.cnki.jpgr.20181218002

大麻状罗布麻的全基因组分析和 SSR 标记开发

宋立肖, 李国旗, 靳长青, 龚诗佩
(宁夏大学西北土地退化与生态恢复国家重点实验室培育基地 / 宁夏大学西北退化生态系统恢复与重建教育部重点实验室, 银川 750021)

摘要: 大麻状罗布麻是重要的经济和生态作物, 由于基因组数据匮乏和分子标记少而限制其遗传研究工作的开展。本研究利用 Illumina 测序平台对大麻状罗布麻的基因组大小进行测定, 通过生物信息学方法对其基因组杂合度和重复序列等基本信息进行预估, 并做了基因组的初步组装, 在此基础上对其基因组序列进行了 SSR 查找。研究结果表明, 总测序量为 31.94 Gb, 测序质量正常(Q20≥90%, Q30≥85%), 与 NCBI 核苷酸数据库(NT)比对显示样本不存在外源污染; K-mer 分析(K=17)结果显示, 大麻状罗布麻基因组大小为 239.02 Mbp, 杂合率为 0.56%, 重复序列占全基因组比例为 36.72%, 初步预估大麻状罗布麻基因组为复杂基因组; 采用 K-mer=41 进行基因组初步组装, 共获得 273336 条 contigs, N50 为 3838 bp, 总长为 222723253 bp, 进一步将 contigs 进行连接、延长, 组装得到 224587 条 scaffolds, N50 为 6421 bp, 总长为 226378236 bp; 此外, 对基因组数据进行 SSR 分子遗传标记分析, 共鉴定出 117511 个 SSR, 不同类型核苷酸重复差异较大, 单核苷酸重复最多, 六核苷酸重复最少。该研究为后续全基因组 *de novo* 测序及组装策略提供依据。

关键词: 大麻状罗布麻; 基因组测序; K-mer 分析; SSR 分子标记

Whole Genome Sequencing and Development of SSR Markers in *Apocynum cannabinum*

SONG Li-xiao, LI Guo-qi, JIN Chang-qing, GONG Shi-pei
(*Breeding Base for State Key Laboratory of Land Degradation and Ecological Restoration in Northwest China, Ningxia University / Key Laboratory for Recovery and Restoration of Degraded Ecosystem in North-western China of Ministry of Education, Ningxia University, Yinchuan 750021*)

Abstract: *Apocynum cannabinum* L. is an important economic and ecological crop.Genetic studies on this crop remains largely behind due to the un-availability of genome reference and limited amount of molecular markers.In this study, we performed a whole genome sequencing of *A.cannabinum* L. by sequencing technology (Illumina), and analyzed the ratios of heterogeneity and repetitive elements, followed by exploration of SSR markers.A total of 31.94 Gb high-quality sequences (Q20 ≥ 90 and Q30 ≥ 85%) have been generated.By searches with NCBI nucleotide database (NT), no exogenous contamination in the sample was detected.*De novo* assembly and *K-mer* analysis revealed a genome size of 239.02 Mbp, with 0.56% heterozygosity and 36.72% of repetitive sequences.273336 contigs with a N50 of 3838 bp in length have been detected, and the total length was 222723253 bp.224587 scaffolds were further assembled, with a N50 of 6421 bp in length, and the total length was 226378236 bp.By predicting for the simple sequence repeats (SSR), 117511 have been detected potentially for exploring SSR markers.Of all SSR categories, the mononucleotide and hexanucleotide repeat units formed the largest and the least categories, respectively.Thus, this work generated a sequence dataset that might be useful source for future *de novo* assembling of *A.cannabinum* L. genome.

Key words: *Apocynum cannabinum* L.; genome sequencin; K-mer analysis; SSR molecular marker

收稿日期: 2018-12-18 **修回日期:** 2019-01-07 **网络出版日期:** 2019-01-24
URL: http://doi.org/10.13430/j.cnki.jpgr.20181218002
第一作者主要从事植物资源生态学方向的研究, E-mail: 1186720040@qq.com
通信作者: 李国旗, 主要从事植物生态学等研究工作, E-mail: guoqilee@163.com
基金项目: 国家重点研发计划(2017YFC0504406); 宁夏大学研究生创新项目资助(GIP2019045)
Fundation project: National Key R&D Program of China (2017YFC0504406); Postgraduate Innovation Project of Ningxia University (GIP2019045)

罗布麻(*Apocynum venetum* L.)是一种抗逆性极强的生态经济型植物,属于夹竹桃科(Apocynaceae)罗布麻属(*Apocynum* L.)多年生宿根草本[1]。国内从20世纪50年代董正钧[2]先生发现并命名罗布麻开始,就对罗布麻的多种经济价值[3]进行了研究,尤其在纤维的纺织利用方面[4]。据北美植物志记载[5],罗布麻属有20多种植物,但中国只有1种,即罗布麻(*A.venetum* L.)。为了丰富罗布麻属植物遗传资源,充分利用其生态和经济价值,在国家林业局"948"项目的支持下,实验室于2004年自美国引进了大麻状罗布麻(*Apocynum cannabinum* L.)并栽培成功,初步研究证明,大麻状罗布麻比罗布麻耐旱性更强,茎秆更粗直、分叉更少、叶片更大[6-7]。大麻状罗布麻(*A.cannabinum* L.),又名加拿大麻、印第安麻,多产自北美温带和亚热带地区,除了传统的纤维用[8]以外,因含有强心苷、树脂、挥发油、橡胶、鞣质和淀粉等物质,美洲土著民族医学中,用其根做药用,治疗心脏病和腹腔积液等病症[9-10]。因大麻状罗布麻具有很强的适应性,往往成为演替的早期物种[11-12]。

近年来,生物信息学的发展日新月异,随着二代高通量测序技术的飞速发展以及三代单分子测序技术的成熟,测序成本降低,时间缩短。2000年,采用传统的Sanger法顺利完成拟南芥(*Arabidopsis thaliana* (L.)Heynh.)基因组全序列测定,这是人类首次全部破译出高等植物的全基因组序列。2002年,完成了水稻(*Oryza sativa* L.)全基因组序列解析[13],第一次可以从基因组水平上对植物的生长、发育、进化、起源等问题进行研究,同时加快了功能基因的挖掘和植物改良的速度,也为其他植物的基因组测序研究奠定了基础。目前,植物基因组学与生物信息学已经进入快速发展阶段。至2018年3月,已完成向日葵(*Helianthus annuus* L.)[14]、紫芝(*Ganoderma sinense* Zhao.Hsu & Zhang)[15]、丹参(*Salvia miltiorrhiza* Bunge)[16]和杜仲(*Eucommia ulmoides* Oliv.)[17]等193种植物的基因组测序工作。通过从基因组水平分析植物的生长、发育等问题,不仅加深了我们对植物的认识,同时也加快了植物新基因的发现与植物品种的改良。开展大麻状罗布麻全基因组测序工作,将能够从分子水平揭示其抗逆性机理,并为科学合理利用其经济价值提供理论支撑;而本研究开展大麻状罗布麻低覆盖度的基因组调查分析以及SSR分子遗传标记分析,旨在全面了解该基因组特征,为后续全基因组 *de novo* 测序及组装策略提供依据。

1 材料与方法

1.1 试验材料

供试材料为大麻状罗布麻,2004年引自美国华盛顿州艾伦斯堡(Washington Ellensburg),于2018年3月下旬,采自宁夏石嘴山市平罗县罗布麻实验基地,取其地下根段带回实验室,进行盆栽培养。5月初挑选长势良好健康的植株取其顶端幼嫩的叶和茎,液氮速冻后,置于–80 ℃超低温冰箱中保存备用。

1.2 试验方法

1.2.1 样品提取及检测 采用改良的CTAB法提取罗布麻基因组DNA[18],采用琼脂糖凝胶电泳和NanoDrop 2000(Thermo, Inc.)分别检测模板的完整性和浓度,要求A260/280 ≥ 1.80。

1.2.2 测序数据产出及质控 经检验合格的大麻状罗布麻DNA样品通过Covaris超声波破碎仪随机打断成长度为350 bp的片段,电泳回收所需长度的DNA片段(0.2~5 kb),再进行末端修复、加A尾和测序接头,得到的小片段纯化后进行电子扩增E-PCR,完成文库的制备。通过Illumina NovaSep平台进行高通量双末端(Paired-End)测序,为了确保信息分析质量的准确性,过滤掉Illumina测序产生的原始序列(raw reads)中含有接头污染的、N(N表示无法确定碱基信息)比例大于10%的、低质量的reads,再进一步比对、过滤叶绿体序列,从而得到有效序列(clean reads)。整个基因组测序由北京诺禾致源生物信息科技有限公司完成。

1.2.3 17-mer分析预测基因组大小、杂合率和重复序列 在基因组组装前,通过K-mer分析,利用有效序列(clean reads),对基因组大小、杂合率和重复序列进行初步预估。取K为17进行分析,将clean reads连续分割,得到长度为17的碱基序列,然后统计K-mer频数分布,做频率分布图,获得K-mer分布曲线。从曲线中得出K-mer深度估计值,估计基因组大小[19]。通过计算纯合峰深度1.8倍后面的K-mer个数所占的比例,可以得到重复序列比例。杂合率通过公式(1)计算得到,假设对于每一个杂合位点,有2×K个杂合K-mer覆盖,因此杂合K-mer的期望深度为1/2。所以,可通过$a_{1/2}$(杂合K-mer种类数的百分比)和$n_{Kspecies}$(所有K-mer的种类数)估算杂合率。

$$\phi = \frac{a_{1/2} \times n_{Kspecies} / (2 \times K)}{n_{Kspecies} - a_{1/2} \times n_{Kspecies} / 2} = \frac{a_{1/2}}{K(2 - a_{1/2})} \quad (1)$$

其中$a_{1/2}$为杂合K-mer种类数的百分比,$n_{Kspecies}$为所有K-mer的种类数。

1.2.4　基因组初步组装　利用 SOAP denovo 软件进行基因组初步组装[20]，首先对频数低的 K-mer 进行纠错，然后采用 K-mer=41 将纠错后的小片段库 reads 截成更小的序列片段，构建 Contig 序列并进一步组装为 Scaffold，并进行 Scaffold 上的 gap 补洞处理。以 10 kb 为窗口，在碱基序列上无重复前进，计算每个窗口的平均深度与 GC 含量，获得 GC_depth 点图。从图中可以看出测序是否有明显的 GC 偏向、细菌污染等情况。同时，根据 GC 聚成块的分层情况可以判断基因组的重复率和杂合率[21]。

1.2.5　SSR 分子遗传标记分析　过滤掉序列长度小于 1000 bp 的 Scaffold 后，利用 MISA（Micro Satellite identification too）软件进行基因组 SSR 分子遗传标记分析，搜索基因组序列中所有的核苷酸重复单元，统计出 SSR 位于 Scaffold 的位置、起始位点、终止位点、数量、长度。进行 SSR 查找前按如下标准设置参数：单核苷酸、二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸各重复单元重复次数最小值分别为：10、6、5、5、5、5。

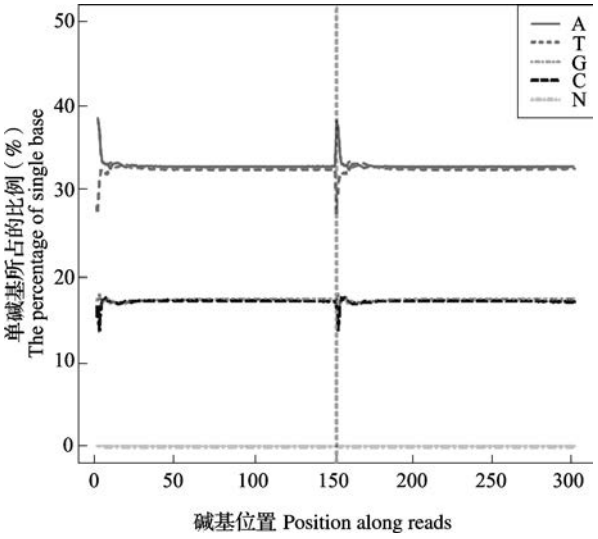
2　结果与分析

2.1　测序数据统计及质量评估

采用 Illumina NovaSep 平台进行高通量双末端（Paired-End）测序，得到大麻状罗布麻 raw reads 31.95 Gb，过滤掉低质量数据后得到 clean reads 31.94 Gb，占原始序列的 99.96%。评估 Q20 和 Q30，测序质量显示，Q20、Q30 分别是 95.97% 和 89.78%，表明大麻状罗布麻高通量测序准确度高。图 1 是碱基类型分布图，检测 AT、GC 是否存在分离现象，被测的序列为随机打断的 DNA 片段，理论上，按照随机性打断及碱基互补配对原则 G 和 C、A 和 T 的含量在每个测序循环中是分别相等的，但是长链非编码测序多采用链特异建库，所以 GC、AT 会存在一定的波动。由图 1 可以看出 A 和 G、C 和 T 的含量接近，而 N 含量接近零，表明测序质量较好。

2.2　17-mer 分析预测基因组大小、杂合率和重复序列结果

以每个 K-mer 深度为横坐标，K-mer 出现的频度为纵坐标绘制 K-mer 深度频度分布图（图 2），对大麻状罗布麻测序有效数据进行 K-mer 分析，从图 2 中可以观察到，depth=99 附近是主峰值，即 K-mer 的期望深度。利用 SOAP denovo 软件统计得到 K-mer 总数约是 24.18 Gb，通过公式（基因组大小 =K-mer 总数 /K-mer 深度）估算基因组大小约为



虚线的左半部分为 read-1 GC 含量分布，右半部分为 read-2 GC 含量分布，从上到下依次代表 A、T、G、C 碱基类型，用于检测 AT、GC 是否存在分离现象

The left half of the dotted line in figure 1 is the read-1 GC content distribution, and the right half is the read-2 GC content distribution, The A, T, G, and C base types are represented from top to bottom, which is used to detect whether AT, GC separation is present

图 1　GC 含量分布图

Fig.1　Distribution figure of GC content

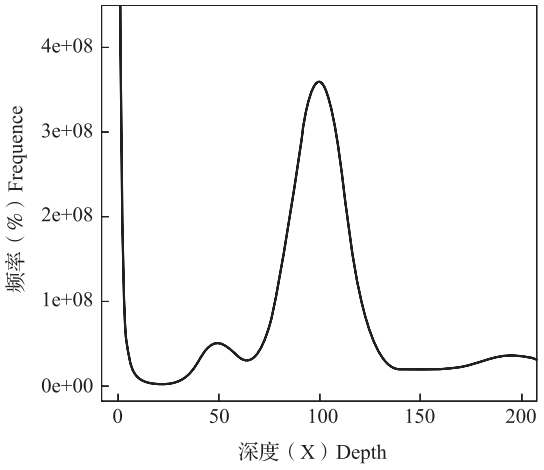


图 2　K-mer 分布曲线

Fig.2　Distribution curve of K-mer

244.26 Mbp，修正后的基因组大小为 239.02 Mbp。K-mer 曲线存在拖尾现象，说明大麻状罗布麻基因组含有重复序列，根据主峰后 1.8 倍的 K-mer 总数占所有 K-mer 数的百分比，计算得到重复序列比例为 36.72%。利用公式（1）计算出大麻状罗布麻基因组杂合率为 0.56%。

2.3　基因初步组装结果

利用 K-mer=41，对 31.94 Gb 的 clean reads 进行 Contig 和 Scaffold 组装，得到了最优的组装效

果，结果如表 1 所示。总共得到 273336 条 contigs，序列的总长为 222723253 bp，Contig N50 长度为 3838 bp，最长序列长度为 150728 bp；进一步组装后得到 224587 条 scaffolds，总长 226378236 bp，组装得到的最长序列长度为 235065 bp，构建 Scaffold 的 N50 为 6421 bp。从图 3 和图 4 中可以看出有明显的峰，主峰前的 1/2 的峰为杂合峰，主峰后 2 倍的位置的峰为重复峰能够判断出峰值在 77x 左右的是纯合峰，纯合峰前 1/2 的峰是杂合峰，初步判断大麻状基因组有一定的杂合率，是复杂基因组。

表 1 组装结果统计表

Table 1 Statistics of the assembled genome sequences in *A.cannabinum* L.

项目 Item	重接群 Contig		拼接群 Scaffold	
	长度（bp）Length	数目 Number	长度（bp）Length	数目 Number
N50	3838	11393	6421	7028
N60	2221	19064	3682	11678
N70	1173	33030	1764	20633
N80	598	60074	770	40717
N90	268	116082	322	87046
总长度（bp）Total length	222723253		226378236	
总数 Total number	273336		224587	
最长序列（bp）Max length	150728		235065	
GC 含量（%）GC content	33.60			

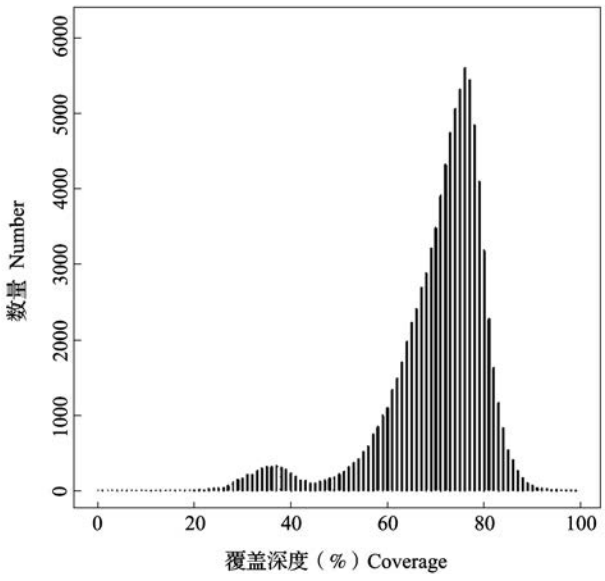


图 3 Contig 覆盖深度和数量分布图

Fig.3 Distribution figure of Contig coverage depth and number

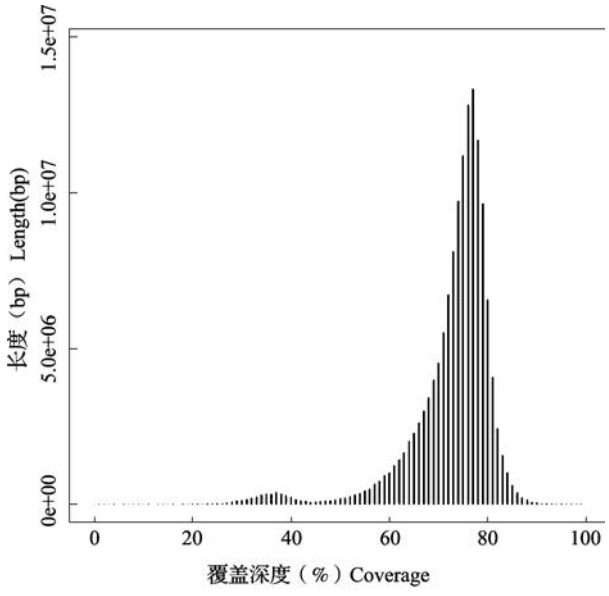


图 4 Contig 覆盖深度和长度分布图

Fig.4 Distribution figure of Contig coverage depth and length

2.4 GC 含量及分布情况

根据基因组测序序列 GC_depth 分布图可以看出测序是否有明显的 GC 偏向，一般高 GC 或低 GC 含量区域测序深度与正常区域有较大差异，覆盖度较低，这些区域也可能是由样品被细菌污染造成的。以 10 kb 无重叠区域为窗口计算 GC 含量和统计其测序深度，发现几乎所有的窗口 GC 含量都在 20%~60% 且深度大于 20 倍（图 5），大麻状罗布麻样品无明显异常，GC 含量无明显偏向，GC_depth 深度分布分为 2 层，低深度处分布着 1 个区域。将低深度分布区的序列提取出来，通过 Blast 软件比对 NCBI 核苷酸数据库（NT 库），结果显示该样品不

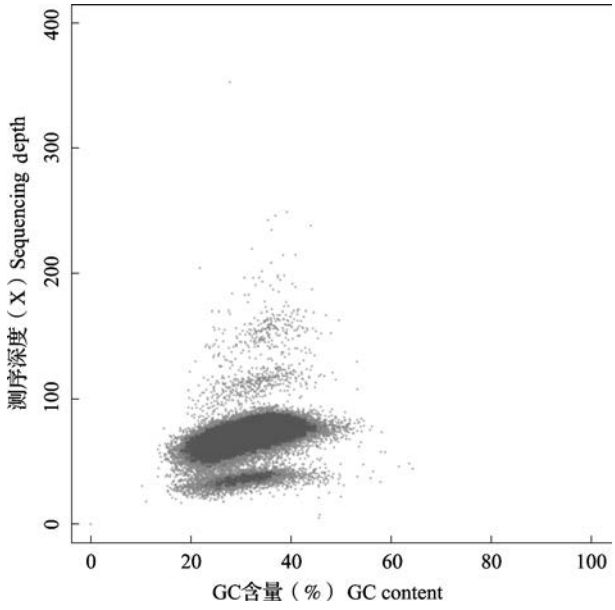


图 5　GC_depth 分布图
Fig.5　Distribution figure of GC_depth

存在外源污染。GC 聚成的块分成了较为明显的 2 层，可能是由杂合引起的。因为杂合会使 2 条同源染色体杂合的部位只装出了 1 条，或 2 条都有装出，同时该部位以上的 read 乘数是整个基因组乘数的一半，导致 GC 含量图中出现较低的一层。

2.5　大麻状罗布麻基因组 SSR 位点查找及分布情况

利用 MISA 软件对大麻状罗布麻基因组中 1~6 核苷酸重复完整性 SSR 进行查找分析，从组装的 Scaffold 大麻状罗布麻基因组中查找，序列总长度为 173723507 bp，序列总数为 32271 个，其中 23050 个序列中共鉴定出 117511 个 SSR，平均每 1478 bp 出现 1 个 SSR，有 16600 个序列含有的 SSR 数大于等于 2，部分大麻状罗布麻基因组 SSR 数据信息见表 2。

表 2　大麻状罗布麻基因组 SSR 数据库的部分结果
Table 2　Part of the *A.cannabinum* L. genomic SSR database

Scaffold ID	SSR 类型 SSR type	重复单元 Repeat unit	SSR 长度（bp） SSR length
Scaffold 5	单核苷酸	（A）12	1278
Scaffold 35	单核苷酸	（T）12	1810
Scaffold 317	二核苷酸	（AT）16	30814
Scaffold 451	二核苷酸	（TG）6	10687
Scaffold 627	三核苷酸	（AAT）5	13710
Scaffold 121023	三核苷酸	（CAT）5	16071
Scaffold 94989	四核苷酸	（ATTT）6	9554
Scaffold 99271	四核苷酸	（TTAT）5	1645
Scaffold 93953	五核苷酸	（AAGAA）5	2649
Scaffold 120373	五核苷酸	（CACAT）5	3599
Scaffold 6441	六核苷酸	（TCTGTC）5	9795
Scaffold 9255	六核苷酸	（GGTTTA）5	18802

2.6　大麻状罗布麻基因组 SSR 序列分析

对大麻状罗布麻的 117511 个 SSR，按照不同类型重复单元的核苷酸数量进行分类统计：单核苷酸重复单元最多，占总重复单元的 67.79%，其次为二核苷酸重复单元，占到总重复单元的 25.30%，六核苷酸重复单元最少，占总重复单元的 0.11%。三核苷酸、四核苷酸、五核苷酸重复单元分别占总重复单元的 6.09%、0.54%、0.18%。每种核苷酸重复单元包含了不同种类的核苷酸，其中单核苷酸重复单元由 4 种不同的核苷酸组成，二核苷酸、三核苷酸、四核苷酸、五核苷酸和六核苷酸重复单元的组成类型分别为 12 种、60 种、78 种、77 种、99 种。不同类型核苷酸重复单元的重复核苷酸组成数据库部分结果见表 3。

表 3 不同重复单元类型 SSR 的重复核苷酸组成数据库部分结果

Table 3 Part of the repeated nucleotide composition database of SSR with different repeat unit types

项目 Item	单核苷酸 Mononucleotide	二核苷酸 Dinucleotide	三核苷酸 Trinucleotide	四核苷酸 Tetranucleotide	五核苷酸 Pentanucleotide	六核苷酸 Hexanucleotide
SSR 组成类型 SSR type	A/T C/G	AT/TA AG/TC	ATC/TAG AAC/TTG	ACAT/ATGT AGAT/ATCT	ACACG/CGTGT AAAAG/CTTTT	AGAGGG/CCCTCT AATCAC/ATTGTG
SSR 类型数量 Type quality of SSR	4	12	60	78	77	99
SSR 总数 Number of SSR	79665	29727	7151	632	209	127
占重复单元比例(%) Rate of repeat type	67.79	25.30	6.09	0.54	0.18	0.11

3 讨论

1997 年 5 月美国国家植物基因组计划(NPGI)首次提出,通过对植物全基因组的研究,绘制基因组图谱,分析碱基序列,以确定基因的功能。从基因组获得信息和知识将通过生物遗传工程和新育种的方法,用来改良植物的有用特性。2010 年,我国学者陈士林提出了“本草基因组计划(HerbGP)”,具有重要经济价值和典型次生代谢途径的药用植物进行全基因组测序和后基因组学研究的系列计划[22]。对于探索未知物种的基因组大小及特征,高通量测序 K-mer 估计基因组特征法[23]方法简便,结果准确。樟树(*Cinnamomum camphora*(L.)J.Presl)[24]、罗汉果(*Siraitia grosvenorii*(Swingle)Jeffrey ex Lu & Zhang)[21]、鹅掌楸(*Liriodendron chinense*(Hemsl.)Sarg.)[25]、沙冬青(*Ammopiptanthus mongolicus*(Maxim.ex Kom.)S.H.Cheng)[26]等多种物种的全基因组大小预测都成功运用了这种方法。本研究通过高通量测序 K-mer 估计法首次估算出了基因组大小、杂合率、GC 含量、重复序列等结果,为进一步研究大麻状罗布麻基因组奠定了基础。首次评估出大麻状罗布麻基因组大小为 239.02 Mbp,与鳗草(*Zostera marina* L.)(237.5 Mb)和野生草莓(*Fragaria vesca* L.)(245 Mb)比较接近。Shangguan 等[27]通过多数已释放的植物基因组数据总结发现,GC 含量大多在 30% ~ 47% 之间。而 Aird 等[28]的研究也表明 GC 含量过高(>65%)或过低(<25%),都会导致高通量测序出现误差,影响拼接数据的准确性。本研究中大麻状罗布麻基因组 GC 含量为 33.60%。植物基因组杂合受到品系、繁殖方式等的影响,甜橙通过珠心胚无性繁殖,并以嫁接的方式培育,其基因组具有高度的杂合性[29],青稞自花授粉的繁育方式致使其杂合度水平低于异花授粉的植物[30]。而罗布麻的授粉和繁殖方式均比较复杂,繁育系统是兼性自交为主、异交为辅的方式,有性繁殖和无性繁殖并存,形成了独特的“生殖补偿机制”[7],罗布麻基因组杂合率为 0.56%,其杂合的原因还有待进一步研究。采用 K-mer=41 进行基因组初步组装,共获得 273366 条 contigs,N50 为 3838 bp,进一步将 contigs 进行连接、延长,组装得到 224587 条 scaffolds,N50 长为 6421 bp。根据对大麻状罗布麻基因组的调查分析,建议在后续的大规模测序时可以采用“2+3”(Illumina+PacBio)测序技术相结合的策略,辅以 Hi-C 技术和重测序技术进行大麻状罗布麻全基因组的研究。

大麻状罗布麻基因组学发展较慢,分子标记数量少,仅有几个(如 ALFP、ISSR 和 RAPD)能够用来进行种属鉴定的分子标记[7],因此,需要加强大麻状罗布麻基因组学研究,筛选和开发多种分子标记并建立统一的方法和标准,用于种属鉴定以及种质资源区分。本研究基于高通量基因组测序数据,利用 MISA 软件进行 SSR 分子遗传标记,在总长为 173723507 bp 的大麻状罗布麻基因组序列中共检测到 117511 个 SSR,重复类型种类丰富,从单核苷酸重复到六核苷酸重复均有分布。其中重复类型最多的是单核苷酸重复,占总重复单元的 67.79%,这与杜仲(*Eucommia ulmoides* Oliv.)[31]和一串红(*Salvia splendens* Sellow ex Wied-Neuw.)[32]的 SSR 遗传标记分析结果相似,单核苷酸重复单元占主导,可能与目前大麻状罗布麻的基因组数据不完全有很大关系。对于不同类型的 SSR 重复单元,随着重复单元数量增加,其基因组 SSR 分布频率呈现逐步降低的趋势[33],大麻状罗布麻基因组 SSR 重复单元与分布频率的关系与这一发现一致。Morgante 等[34]研究发现,基因组大小与 SSR 数量或 SSR 密度之间呈

现负相关关系，大麻状罗布麻和其他一些物种基因组 SSR 分布特征印证了这一点，具有较大基因组（739 Mb）的高粱，SSR 密度最低[35]。国内外一些学者提出，基因组中低级重复单元较多则表示该物种进化水平较高，相反，高级重复单元比例较多的物种其进化时间段或变异频率低[36-37]，可以推测大麻状罗布麻的物种进化水平较高。大多数植物的 SSR 分子遗传标记分析也显示，二核苷酸重复单元占主导，例如黄芩（*Scutellaria baicalensis* Georgi）、沙冬青基因组中二核苷酸重复最多，分别占总重复单元的比例为 68.32% 和 56.39%[38,32]。另外，大麻状罗布麻基因组的 SSR 分析结果显示，不同核苷酸重复单元类型中，不同重复单元的比例差异较大。其中单核苷酸重复以 A/T 重复单元为主，占单核苷酸重复的 92.2%，二核苷酸重复中 AT/TA 最多，占二核苷酸重复的 67.2%；而一串红基因组的 SSR 主要重复类型以 A/T 和 AG/CT 为主[32]，黄芩基因组的 SSR 主要重复类型是 CT/GA 和 TTC/GAA[38]。因此，SSR 位点的重复单元类型在不同物种间的分布有较大差异。

参考文献

[1] 中国科学院植物志编辑委员会 . 中国植物志 . 北京：科学出版社，1977：157–161
Editorial Committee of Botany of the Chinese Academy of Sciences. Flora of China.Beijing：Scientific and Technical Publishers，1977：157–161

[2] 董正钧 . 我国新发现的高级纺织纤维植物 - 罗布麻 . 科学通报，1957，2（19）：607-608
Dong Z J，A new advanced textile fiber plant in China-Apocynum.Chinese Science Bulletin，1957，2（19）：607-608

[3] 本书编写组 . 罗布麻的综合利用 . 北京：科学出版社，1978：52-58
Compilation of this book.Comprehensive Utilization of Apocynum. Beijing：Scientific and Technical Publishers，1978：52-58

[4] 刘士侠 . 高级纤维作物，罗布麻 . 上海：上海科学技术出版社，1959：96-118
Liu S X.High fiber crop，*Apocynum*.Shanghai：Shanghai Scientific and Technical Publi shers，1959：96-118

[5] George W M，Harold W R .North American Flora.The New York City Botanical Garden，1949，29：188-192

[6] 王东清，李国旗，王磊 . 干旱胁迫下红麻和大麻状罗布麻水分生理及光合作用特征研究 . 西北植物学报，2012，32（6）：1198-1205
Wang D Q，Li G Q，Wang L.Daily dynamics of photosynthesis and water physiological characteristics of *Apocynum venetum* and *A.cannabinum* under drought stress.Acta Botanica Boreali-Occidentalia Sinica，2012，32（6）：1198-1205

[7] 李国旗，陈彦云 . 罗布麻生理生态学研究 . 北京：科学出版社，2012：69-93
Li G Q，Chen Y Y.Physioecology and of *Apocynum*.Beijing：Scientific and Technical Publishers，2012：69-93

[8] Indian Use of Apocynum cannabinum as a Textile Fibre. Proceedings of the Academy of Natural Sciences of Philadelphia，1884，36：30

[9] Duprey A J B，Eng M R C S，Lond L R C P.A case of mitral incompetency and ascites treated with *Apocynum cannabinum*. Lancet 1905，166（4283）：955-956

[10] 肖正春，袁昌齐，束成杰，张广伦，张卫明 . 罗布麻类植物资源开发利用及展望 . 中国野生植物资源，2018，37（1）：1-4
Xiao Z C，Yuan C Q，Shu C J，Zhang G L，Zhang W M.Development，utilization and outlook of *Apocynum* resources.Chinese Wild Plant Resources，2018，37（1）：1-4

[11] Keever C.Mechanisms of plant succession on old fields of Lancaster County，Pennsylvania.Bulletin of the Torrey Botanical Club，1979，106（4）：299-308

[12] Mulhouse J M，Galatowitsch S M.Revegetation of prairie pothole wetlands in the mid-Continental US：twelve years post-reflooding.Plant Ecology，2003，169（1）：143-159

[13] Goff S A，Ricke D，Lan T H，Presting G，Wang R L，Dunn M，Glazebrook J，Sessions A，Oeller P，Varma H，Hadley D，Hutchison D，Martin C，Katagiri F，Lange B M，Moug H T，Xia Y，Budworth P，Zhong J P，Miguel T，Paszkowski U，Zhang S P，Colbert M，Sun W L，Chen L L，Cooper B，Park S，Wood T C，Mao L，Quail P，Wing R，Dean R，Yu Y，Zharkikh A，Shen R，Sahasrabudhe S，Thomas A，Cannings R，Gutin A，Pruss D，Reid J，Tavtigian S，Mitchell J，Eldredge G，Scholl T，Miller R M，Bhatnagar S，Adey N，Rubano T，Tusneem N，Robinson R，Feldhaus J，MacalmaT，Oliphant A，Briggs S.A draft sequence of the rice genome（*Oryza sativa* L.ssp. japonica）.Science，2002，296（5565）：92-100

[14] Badouin H，Gouzy J，Grassa C J，Murat F，Staton S E，Cottret L，Lelandais-Brière C，Owens G L，Carrère S，Mayjonade B，Legrand L，Gill N，Kane N C，Bowers J E，Hubner S，Bellec A，Bérard A，Bergès H，Blanchet N，Boniface M C，Brunel D，Catrice O，Chaidir N，Claudel C，Donnadieu C，Faraut T，Fievet G，Helmstetter N，King M，Knapp S J，Lai Z，Le Paslier M C，Lippi Y，Lorenzon L，Mandel J R，Marage G，Marchand G，Marquand E，Bret-Mestries E，Morien E，Nambeesan S，Nguyen T，Pegot-Espagnet P，Pouilly N，Raftis F，Sallet E，Schiex T，Thomas J，Vandecasteele C，Varès D，Vear F，Vautrin S，Crespi M，Mangin B，Burke J M，Salse J，Muños S，Vincourt P，Rieseberg L H，LangladeN B.The sunflower genome provides insights into oil metabolism，flowering and Asterid evolution.Nature，2017，546（7656）：148-152

[15] Zhu Y J，Xu J，Sun C，Zhou S G，Xu H B，Nelson D R，Qian J，Song J Y，Luo H M，Xiang L，Li Y，Xu Z C，Ji A J，Wang L Z，Lu S F，Hayward A，Sun W，Li X W，Schwartz D C，Wang Y T，Chen S L.Chromosome-level genome map provides insights into diverse defense mechanisms in the medicinal fungus *Ganoderma sinense*.Scientific Reports，2015，5：11087

[16] Xu H B，Song J Y，Luo H M，Zhang Y J，Li Q S，Zhu Y J，Xu J，Li Y，Song C，Wang B，Sun W，Shen G A，Zhang X，Qian J，Ji A J，Xu Z C，Luo X，He L，Li CY，Sun C，Yan H X，Cui G H，Li X W，Li X E，Wei J H，Liu J Y，Wang Y T，Hayward A，Nelson D，Ning Z M，Peters R J，Qi X Q，Chen S L.Analysis of the genome sequence of the medicinal plant *Salvia miltiorrhiza*. Molecular Plant，2016，9（6）：949-952

[17] Wang W C，Chen S Y，Zhang X Z.Whole-genome comparison

reveals heterogeneous divergence and mutation hotspots in chloroplast genome of *Eucommia ulmoides* Oliver.International Journal of Molecular Sciences, 2018, 19(4): 1037-1047

[18] Xu M, Sun Y, Li H .EST-SSRs development and paternity analysis for *Liriodendron* spp.. New Forest, 2010, 40(3): 361-382

[19] Huang S S, Li R Q, Zhang Z H, Li L, Gu X F, Fan W, Lucas W J, Wang X W, Xie B Y, Ni P X, Ren Y Y, Zhu H M, Li J, Lin K, Jin W W, Fei Z J, Li G C, Staub J, Kilian A, Vossen E A G, Wu Y, Guo J, He J, Jia Z Q, Ren Y, Tian G, Lu Y, Ruan J, Qian W B, Wang M W, Huang Q F, Li B, Xuan Z L, Cao J J, Asan, Wu Z G, Zhang J B, Cai Q L, Bai Y Q, Zhao B W, Han Y H, Li Y, Li X F, Wang S H, Shi Q X, Liu S Q, Cho W K, Kim J Y, Xu Y, Heller-Uszynska K, Miao H, Cheng Z H, Zhang S P, Wu J, Yang Y H, Kang H X, Li M, Liang H Q, Ren X L, Shi Z B, Wen M, Jian M, Yang H L, Zhang G J, Yang Z T, Chen R, Liu S F, Li J W, Ma L J, Liu H, Zhou Y, Zhao J, Fang X D, Li G Q, Fang L, Li Y R, Liu D Y, Zheng H K, Zhang Y, Qin N, Li Z, Yang G H, Yang S, Bolund L, Kristiansen k, Zheng H C, Li S C, Zhang X Q, Yang H M, Wang J, Sun R F, Zhang B X, Jiang S Z, Wang J, Du Y C, Li S G.The genome of the cucumber, *Cucumis sativus* L..Nature Genetics, 2009, 41(12): 1275-1281

[20] Li R Q, Fan W, Tian G, Zhu H M, He L, Cai J, Huang Q F, Cai Q L, Li B, Bai Y Q, Zhang Z H, Zhang Y P, Wang W, Li J, Wei F W, Li H, Jian M, Li J W, Zhang Z L, Nielsen R, Li D W, Gu W J, Yang Z T, Xuan Z L, Ryder O A, Leung F C, Zhou Y, Cao J J, Sun X, Fu YG, Fang X D, Guo X S, Wang B, Hou R, Shen F J, Mu B, Ni P X, Lin R M, Qian W B, Wang G D, Yu C.The sequence and *de novo* assembly of the giant panda genome.Nature, 2010, 463(7279): 311-317

[21] 唐其,马小军,莫长明,潘丽梅,韦荣昌,赵欢 . 罗汉果全基因组 Survey 分析 . 广西植物, 2015, 35(6): 786-791
Tang Q, Ma X J, Mo C M, Pan L M, Wei R C, Zhao H.Genome survey analysis in *Siraitia grosvenorii*.Guihaia, 2015, 35(6): 786-791

[22] 陈士林,孙永珍,徐江,罗红梅,孙超,何柳,程翔林,张伯礼,肖培根 . 本草基因组计划研究策略 . 药学学报, 2010, 45(7): 807-812
Chen S L, Sun Y Z, Xu J, Luo H M, Sun C, He L, Chen X L, Zhang B L, Xiao P G.Strategies of the study on Herb Genome Program.Acta Pharmaceutica Sinica, 2010, 45(7): 807-812

[23] Chen W B, Hasegawa D K, Arumuganathan K, Simmons A M, Wintermantel W M, Fei Z J, Ling K S.Estimation of the whitefly *Bemisia tabaci* genome size based on K-mer and flow cytometric analyses.Insects, 2015, 6(3): 704-715

[24] 伍艳芳,肖复明,徐海宁,章挺,江香梅 . 樟树全基因组调查 . 植物遗传资源学报, 2014, 15(1): 149-152
Wu Y F, Xiao F M, Xu H N, Zhang T, Jiang X M.Genome Survey in *Cinnamomum camphora* L.Presl.Journal of Plant Genetic Resources, 2014, 15(1): 149-152

[25] 钟永达,张新,李彦强,刘立盘,余发新 . 鹅掌楸全基因组调查 . 分子植物育种, 2017, 15(2): 507-512
Zhong Y D, Zhang X, Li Y Q, Liu L P, Yu F X.Genome Survey of *Liriodendron* chinense (Hemsl.) Sarg.Molecular Plant Breeding, 2017, 15(2): 507-512

[26] 王雪,周佳熠,孙会改,禹瑞敏,高飞,周宜君 . 新疆沙冬青基因组调查测序与基因组大小预测 . 植物遗传资源学报, 2018, 19(1): 143-149
Wang X, Zhou J Y, Sun H G, Yu R M, Gao F, Zhou Y J.Genomic survey sequencing and estimation of genome size of *Ammopiptanthus mongolicus*.Journal of Plant Genetic Resources, 2018, 19(1): 143-149

[27] Shangguan L F, Han J, Kayesh E, Sun X, Zhang C Q, Pervaiz T, Wen X C, Fang J G.Evaluation of genome sequencing in selected plant species using expressed sequence tags.PLoS One, 2013, 8(7): e69890

[28] Aird D, Ross M G, Chen W S, Danielsson M, Fennell T, Russ C, Jaffe D B, Nusbaum C, Gnirke A.Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries.Genome Biology, 2011, 12(2): R18

[29] 李亚莉 . 云南迪庆藏区藏族传统文化影响下的青稞传统种质资源遗传多样性研究 . 昆明: 中国科学院昆明植物研究所, 2008: 57-63
Li Y L.Impact of traditional tibetan culture on genetic diversity of *Hulless Barley* (*Hordeum vulgare* var.*nudum*) landraces from Diqing Prefecture, Yunnan.Kunming: Kunming Institute of Botanty, Chinese Academy of Sciences, 2008: 57-63

[30] 焦文标 . 基于二代测序技术的甜橙基因组杂合度与起源研究 . 武汉: 华中农业大学, 2013
Jiao W B.Genome heterozygosity and origin study of sweet orange basedon second generation sequencing technology. Wahan: Huazhong Agricultural University, 2013

[31] 吴敏,杜红岩,乌云塔娜,刘攀峰,荆腾 . 杜仲基因组微卫星特征及 SSR 标记开发 . 林业科学研究, 2015, 28(3): 387-393
Wu M, Du H Y, Wu yun T N, Liu P F, Jing T.Characterization of genomic microsatellites and development of SSR markers of *Eucommia ulmoides*.Forest Research, 2015, 28(3): 387-393

[32] 王硕,葛秀秀,孔维一,陈洪伟,刘克锋,王红利 . 一串红全基因组调研及 SSR 特征分析 . 北京农学院学报, 2018, 33(2): 15-22
Wang S, Ge X X, Kong W Y, Chen H W, Liu K F, Wang H L.Genome survey and characteristic analysis of SSR in *Salvia splendens*.Journal of Beijing University of Agriculture, 2018, 33(2): 15-22

[33] Sonnet H, Carpendale S, Strothotte T.Long microsatellite alleles in *Drosophila melanogaster* have a downward mutation bias and short.Genetics, 2000, 155(3): 1213

[34] Morgante M, Hanafey M, Powell W.Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes.Natuer Genetics, 2002, 30(2): 194-200

[35] Huo N, Lazo G, Vogel J, You F M, Ma Y, Hayden D M, Coleman-Derr D, Hill A, Dvorak J, Anderson O D, Luo M C, Gu Y Q.The nuclear genome of *Brachypodium distachyon*: analysis of BAC end sequences.Functtional Integrative Genomics, 2008, 8(2): 135-147

[36] Tóth G, Gáspári Z, Jurka J.Microsatellites in different eukaryotic genomes: survey and analysis.Genome Research, 2000, 10(7): 967

[37] 高亚梅,韩毅强,汤辉,孙东梅,王彦杰,王伟东 根瘤菌基因组内简单重复序列的分析 . 中国农业科学, 2008, 41(10): 2992-2998
Gao Y M, Han Y Q, Tang H, Sun D M, Wang Y J, Wang W D.Analysis of simple sequence repeats in *Rhizobium* genomes. Scientia Agricultura Sinica, 2008, 41(10): 2992-2998

[38] 齐琳洁,龙平,蒋超,袁媛,黄璐琦 . 黄芩基因组 SSR 分子标记的开发及遗传多样性分析 . 药学学报, 2015, 50(4): 500-505
Qi L J, Long P, Jiang C, Yuan Y, Huang L Q.Development of microsatellites and genetic diversity analysis of *Scutellaria baicalensis* Georgi using genomic-SSR markers.Acta Pharmaceutica Sinica, 2015, 50(4): 500-505