

DOI:10.13275/j.cnki.lykxyj.2020.01.019

西伯利亚白刺基因组信息初探

朱礼明^{1,2}, 黎梦娟^{1,2}, 张景波³, 杨秀艳^{3,4}, 成铁龙^{1,2*}

- (1. 南京林业大学, 林木遗传与生物技术省部共建教育部重点实验室, 江苏 南京 210037;
2. 南京林业大学南方现代林业协同创新中心, 江苏 南京 210037;
3. 中国林业科学研究院沙漠林业实验中心, 内蒙古 磴口 015200;
4. 中国林业科学研究院国家林业和草原局盐碱地研究中心, 北京 100091)

摘要: [目的] 为了解西伯利亚白刺 (*Nitraria sibirica* Pall) 基本基因组信息, 如基因组大小、复杂程度等特征。[方法] (1) 以西伯利亚白刺的幼苗嫩叶为材料, 以番茄为标定, 用解离液 mG 解离叶片后, 经流式细胞仪检测细胞悬液, 收集细胞并计算西伯利亚白刺的 C 值。(2) 通过构建小片段文库, 进行 Illumina HiSeq 测序并分析测序结果, 预估其基因组大小、GC 含量、杂合率、重复序列比例等, 使用 MISA 脚本对过滤后数据进行 SSR 位点分析。[结果] 西伯利亚白刺预估全基因组大小为 526.30 Mbp, 基因组 GC 含量为 36.78%, 基因组杂合率为 0.90%, 重复序列比例为 55.39%。对测序数据进行 SSR 位点标记, 共得到 521 125 个 SSR 位点, 其中, 单核苷酸重复为 342 883 个, 占总 SSR 位点的 65.80%。[结论] 由各项分析指标推测西伯利亚白刺为复杂基因组, 可以采用二代 Illumina 与三代 PacBio 测序辅以 Hi-C 技术联合分析, 有利于西伯利亚白刺基因组的组装及基因图谱的获取。

关键词: 西伯利亚白刺; 流式细胞术; 基因组大小; GC 含量; SSR 标记

中图分类号: S722

文献标志码: A

文章编号: 1001-1498(2020)01-0144-08

西伯利亚白刺 (*Nitraria sibirica* Pall) 系蒺藜科白刺属植物, 为第三纪孑遗植物, 分布于蒙古、中亚以及我国西北、华北、东北的沙地、盐碱地地区^[1]。西伯利亚白刺具耐盐碱、抗风沙等特性, 能在沙漠盐碱等恶劣环境下生存, 是一种优良的沙地、盐碱地改良物种, 其果实富含多种氨基酸、糖类、黄酮等物质^[2-4], 营养价值丰富, 其地上部分也可作为牲畜饲料。因此, 西伯利亚白刺兼有生态和经济价值, 有较好的开发利用前景。

目前, 关于西伯利亚白刺的研究主要集中在繁殖技术优化^[5-6]、果实成分测定^[7-8]及生理生化测定^[9-11]等方面, 有关西伯利亚白刺的分子生物学方面的研究较少^[12], 基因组学方面的研究也尚未见报道。宏观的研究只能从表层揭示西伯利亚白刺抗逆适应现象, 并不能从内部机制、进化等层面解释西伯利亚白刺抗逆机理, 而全基因组测序可以获取

典型基因组特征并获得大量基因序列, 对于剖析其生长、发育、抗逆等机理, 发掘西伯利亚白刺的生态和经济价值有积极意义^[13-14]。

全基因组调查通过了解待测生物基因组的基本特征, 可以对全基因组测序组装难度、组装时间和成本等作出大致的评估并作出相应的测序策略调整, 是基因组测序前必不可少的步骤之一。

流式细胞术是一种快速预测基因组大小的技术, 它通过比较待测植物和标定植物细胞悬液荧光吸收峰相对比值, 再根据标定植物的基因组大小来计算待测植物基因组大小^[15]。而随着基因组测序技术的成熟及成本的下降, 通过全基因组 survey 来探究待测植物的基因组基本特征不失为一种有效的方法, 作为近年来发展较快的基因组预测技术, 全基因组 survey 可以对生物的基因组基本特征测定评估^[16-17], 相比于流式细胞术等基因组大小预测方

收稿日期: 2019-03-02 修回日期: 2019-10-22

基金项目: 唐古特白刺 SOS 信号通路在耐受高盐胁迫中的作用及其表观遗传分析 (31770715)

* 通讯作者: 成铁龙, 男, 正高级工程师.E-mail: ctielong@126.com

法, 不仅可以精准预测基因组大小, 还可以对基因组复杂程度、杂合率、重复序列比例等有相应的评估, 更能切合生物的基因组特征, 因而有更好的参考价值。

SSR 分子标记以其高重复性、高多态性、共显性遗传、丰度高等优良特性成为了研究群体遗传学、遗传变异和标记辅助选择的有力工具, 对于了解西伯利亚白刺的进化有积极的作用。

本研究基于流式细胞术和全基因组 survey 测序的方法对西伯利亚白刺基因组大小、复杂程度、杂合率等基因组特征有一个较为详细的评估, 同时也对其测序方案的制定提出建议, 为后续西伯利亚白刺基因组组学研究奠定了良好的基础。

1 材料和方法

1.1 材料

将取自内蒙古磴口的野生西伯利亚白刺种子置于 4℃ 下沙藏 30 d, 置于萌发盒上进行萌发, 再将发芽的种子定植于 7 cm×7 cm 的塑料花盆中 (基质配方为河沙: 营养土=1:1, 并在其中掺入少量珍珠岩和蛭石), 幼苗生长 2 个月后取嫩叶备用。流式标定植物为 Jaroslav Doležal 博士惠赠的番茄 ‘Stupické polní rane’ 32 品种。

1.2 方法

1.2.1 流式细胞分析 使用 BD 公司 influx 型号流式细胞仪对西伯利亚白刺基因组大小进行分析, 选用 mG 解离液对植物叶片进行解离, 使用碘化丙啶 (PI) 溶液为荧光染料, 采用本番茄作为内标, 使用 Influx 自带分析软件 FACSTM 分析基因组大小。

操作步骤: 于塑料皿上滴加 1.5 mL mG 解离液, 分别取 0.5 g 西伯利亚白刺、番茄新鲜叶片用刀片迅速切碎后过 400 目滤网, 将收集的滤液 1 500 rpm, 离心 6 min, 吸除上清液后重新加入 500 μL 预冷的 mG 解离液, 加入 PI 染色液, 最后加入 10 μg·mL⁻¹ 的 Rnase, 避光 4℃ 孵育 5 min 后低速上机检测。

C 值计算公式: $C_{\text{待测样本}} = C_{\text{标定}} \times (G_0/G_1)_{\text{待测样本}} / (G_0/G_1)_{\text{标定}}$

式中: G_0/G_1 为流式荧光吸收强度。

mG 解离液配方:

45 mmol·L⁻¹ MgCl₂, 20 mmol·L⁻¹ MOPS, 30 mmol·L⁻¹ Na₃C₆H₅O₇·2H₂O, 1% (w/v) PVP-40, 0.2% (v/v) TritonX-100, 10 mmol·L⁻¹ Na₂EDTA,

20 μL·mL⁻¹ β-巯基乙醇, 调节 pH 至 7.0, -20℃ 下保存。PI 为碘化丙啶, 使用时至终浓度为 50 μg·μL⁻¹, 4℃ 保存。

1.2.2 DNA 的提取以及质量检测 采用 CTAB 法对西伯利亚白刺的新鲜叶片进行 DNA 提取, 得到的 DNA 样品用紫外分光光度计检测其浓度、 OD_{260}/OD_{280} , 再经 1% 琼脂糖凝胶电泳检测其完整性 (电泳条件为: 电压 180 V, 电泳时间: 30 min)。

1.2.3 文库制备及测序方法 检测合格的 DNA 样品通过 Covaris 超声波破碎仪打断成片段, 并进行末端修复, 加 poly-A 尾, 加测序接头, 纯化, PCR 扩增等步骤后, 构建出 350 bp 双端 PE150 待测序文库。文库通过 Illumina Hiseq 平台进行双端 PE 测序。

1.2.4 K-mer 分析 采用 K-mer 分析策略, 若每条序列的长度为 L, K-mer 长度为 K, 可以得到 L-K+1 个 K-mer, 再通过这些数据来对基因组大小进行预估, 通过 Lander-waterman 算法对西伯利亚白刺基因组大小进行估计, 满足公式:

$$C_{\text{base}} = C_{\text{K-mer}} \times L(L - K + 1);$$

$$G = N_{\text{K-mer}} / C_{\text{K-mer}} = N_{\text{base}} / C_{\text{base}}$$

式中: N_{base} 和 $N_{\text{K-mer}}$ 为序列的碱基总数和 K-mer 数, C_{base} 和 $C_{\text{K-mer}}$ 为覆盖碱基的期望深度和 K-mer 期望覆盖深度。

对预估的基因组大小进行修正, 将 K-mer 深度为 1 的情况认为是错误情况, 计算错误率, 并用于修正基因组大小, 修正公式为

$$G_{\text{revised}} = \text{Genome} \times (1 - E)$$

式中: G_{revised} 为修正后的基因组大小, E 为测序错误率。

通过 K-mer 数学分析模型, 基因组杂合率公式为:

$$X = \frac{a_{1/2} \times n_K / (2 \times K)}{n_K - a_{1/2} \times n_K / 2} = \frac{a_{1/2}}{K(2 - a_{1/2})}$$

式中: $a_{1/2}$ 为杂合 K-mer 种类数的百分比, n_K 为所有 K-mer 的种类数。

另外, 计算标准泊松分布和实际数据曲线峰值后的面积差值, 可得到重复序列百分比, 在这里我们计算纯合峰深度 1.8 倍后面的 K-mer 个数所占的比例来估计重复序列比例。

1.2.5 基因组组装 由于西伯利亚白刺基因组重复序列较多, 我们选择 K-mer=41 将打断的 DNA 序

列拼接组装到 Scaffold, 通过 reads 之间的 overlap 关系构建 de Bruijn 图并对其简化, 在重复区域边界位置进行剪切, 得到 contig 序列, 再根据大片段数据的 Pair-end 关系, 构建 Scaffold 序列, 最后用 reads 对 Scaffold 的 gap 区域进行填补, 完成组装过程, 具体配置参数为

pregraph : -K 41 -R -d 1

-K kmer: K value in kmer

-R (optional): unsolve repeats by reads (default no)

-d KmerFreqCutoff(optional): delete kmers with frequency no larger than (default 0)

contig : -D 1 -M 1 -R

-D EdgeCovCutoff(optional): delete edges with coverage no largert than (default 1)

-M mergeLevel (default 1,min 0, max 3): the strength of merging similar sequences during contiging

-R solve_repeats (optional): solve repeats by read paths(default: no)

map : -K 41

-K kmer (default: the same as in pregraph): k value in kmer

scaff : -F 1 -L 43

-F (optional) fill gaps in scaffold. (default 0; 1:normally; -1:only fill nonrepeat gap; 2:radically)

-L minLen : shortest contig (minus K value) for scaffolding

再根据组装结果统计其 contig 分布情况, 统计测序长度大于 500 bp 的测序深度和 GC 含量并做 GC 含量分布图。

1.2.6 SSR 分布特征分析 运行 MISA 脚本 (pgrc.ipk-gatersleben.de/misa) 对过滤后数据 SSR 位点鉴定并统计其类型、数量。筛选标准为单核苷酸 SSR 位点 ≥ 16 次, 双核苷酸 SSR 位点 ≥ 6 次, 三四核苷酸 SSR 位点 ≥ 5 次。

2 结果与分析

2.1 流式细胞基因组大小分析

将西伯利亚白刺和番茄的叶片混合解离液放入流式细胞运行并在 480 nm 波长下检测其荧光吸收强度 (图 1), 其中, P0 为西伯利亚白刺的吸收峰, P1 为番茄的吸收峰, 番茄参考 2C 值为 1.96 pg, 实验重复 3 次。将平均值代入 C 值计算公式得出: $2C_{\text{西伯利亚白刺}} = 2C_{\text{番茄}} \times (G_0/G_1_{\text{西伯利亚白刺}}) /$

$(G_0/G_1_{\text{番茄}}) = 1.96 \text{ pg} \times 0.534$, 得西伯利亚白刺 C 值大小为 523.4 Mbp。

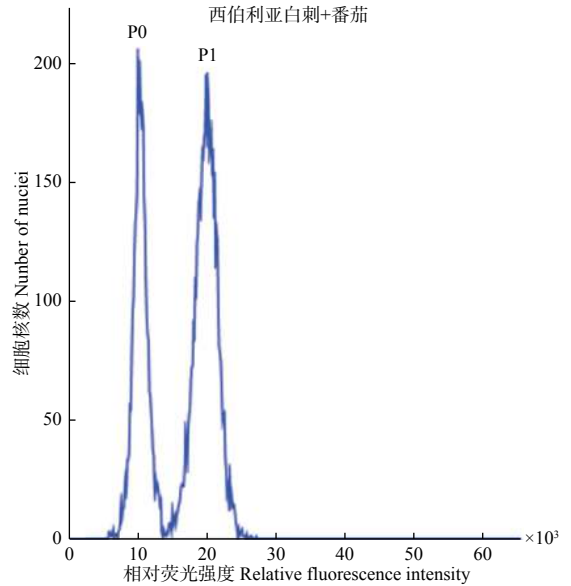


图 1 流式细胞测定结果

Fig. 1 Flow cytometry results

2.2 DNA 的提取以及质量检测

取 1 μL DNA 样品于分光光度计的检测, 结果显示 OD_{260}/OD_{280} 为 1.89, 浓度为 $206.9 \text{ ng} \cdot \mu\text{L}^{-1}$ 。再利用 1% 琼脂糖凝胶电泳检测其条带完整性, 图 2 表明: 电泳条带单一, 无明显杂带。综合二者推测, 此 DNA 完整度较高, 可用于下游实验。

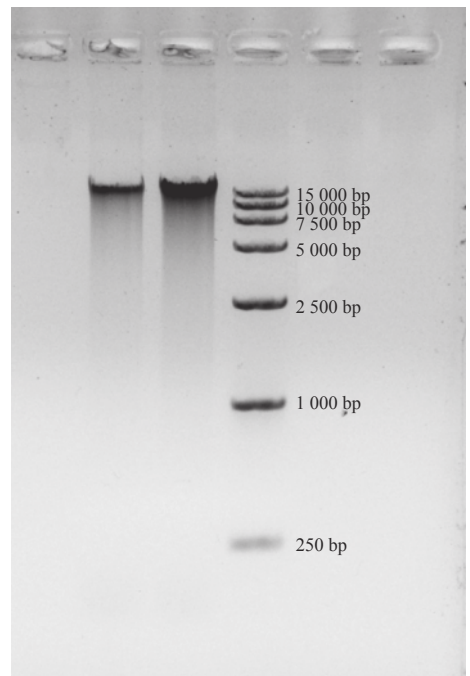


图 2 DNA 琼脂糖凝胶电泳图

Fig. 2 DNA agarose gel electrophoresis

2.3 测序数据产出及质控

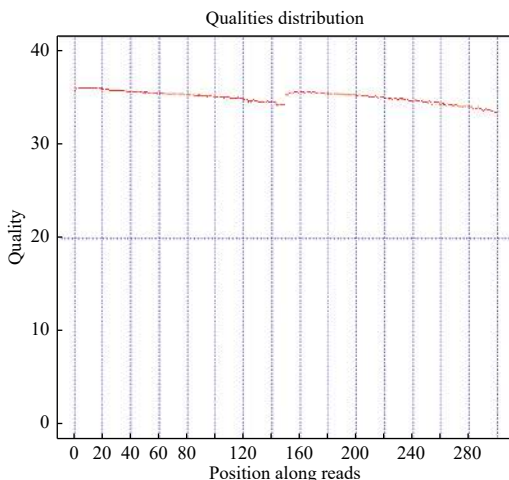
2.3.1 测序数据统计 过滤掉无效或低质量的 reads 数据, 再经图像识别、去污染等步骤, 得出最终的测序结果 (表 1): 其中, 测序的总 reads 数为 212 852 294 个, 测序的总数据大小为 63 855.69 Mbp, 按照 536.16 Mbp 的预估基因组大小得出本次测序深度为 119.09 \times , 测序的错误率为 0.04%, Q20 的含量为 95.59%, Q30 的含量为 89.33%, GC 含量为 36.78%。

表 1 测序结果统计

Table 1 Sequencing results statistics

reads数量 Raw paired reads	数据产量 Raw base/ Mbp	测序错误率 Error rate/%	Q20含量 Q20 Content/%	Q30含量 Q30 Content/%	GC含量 GC Content/%
212 852 294	63 855.69	0.04	95.59	89.33	36.78

2.3.2 测序质量检测 测序数据的质量主要分布在 Q30 ($\geq 80\%$) 以上, 这样才能保证后续分析的正常进行, 如图 3 所示, 实验 Q30 含量为 89.33% 满足后续分析要求。



注: 左为 read-1 数据质量分布, 右为 read-2 数据质量分布
Notes: left: read-1; right: read-2

图 3 数据质量分布

Fig. 3 Data quality distribution

此外, 测序错误率也影响测序结果的准确性, 对于下游分析至关重要, 本实验 2 个 reads 的测序错误率均低于 1% (图 4), 表明本次测序错误率控制良好。为进一步保证测序结果的可信性, 还需对本次测序的碱基含量分布进行分析。GC 含量分布检查用于检测有无 AT、GC 分离现象, 理论上 G 和 C 含量以及 A 和 T 含量在每个测序循环上应

分别相等, 且整个测序过程中稳定不变, 呈水平线。由于 DNA 模板扩增偏差等原因使测序前几个碱基测序质量值较低, 发生小幅度波动, 属于正常情况。本实验中 (图 5) 测序的 G 和 C 的含量和 A 和 T 的含量接近也保证了测序的可信度。

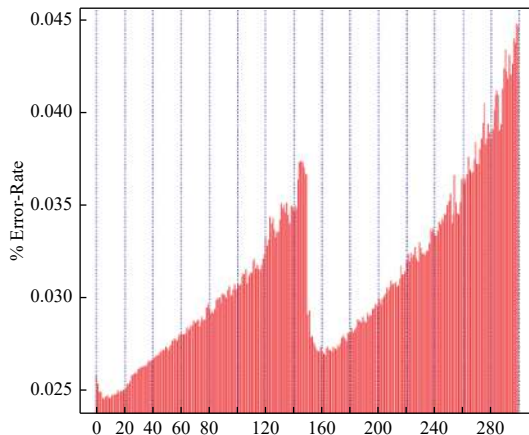


图 4 测序错误率分布

Fig. 4 Sequencing error rate distribution

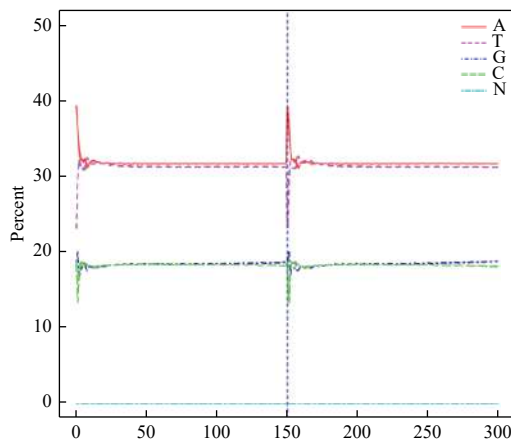


图 5 GC 含量分布图

Fig. 5 GC content distribution map

2.4 K-mer 分析

利用 K-mer 分析法对西伯利亚白刺基因组大小进行估计, 根据测序结果 (表 2、图 6) 发现: 当 K-mer 深度为 89 \times 时存在明显的主峰, 由 K-mer 相关公式计算得到的基因组大小为 536.16 Mbp, 并通过后续基因修正得修正后基因组大小为 526.30 Mbp; 而在主峰前横坐标二分之一处出现次峰。一般当目标序列存在杂合现象时, 存在杂合位点的 K-mer 被分成 2 份, 频率变成原频率的 1/2, 因此, 此峰为杂合峰, 并统计得出西伯利亚白刺基因组杂合率为 0.90%, 杂合率较高, 属于复

杂基因组。此外，在约为主峰 2 倍 depth 的地方存在次峰，并有明显的拖带现象，该片段出现的期望值是大部分的 2 倍，这些片段为重复片段，由相关统计结果得重复序列数占总序列数的 55.39%。

表 2 K-mer=17 分析所得各项数据
Table 2 K-mer=17 analysis of the data

数据名称 Data name	数据大小 Data size
K-mer 数/bp	47 718 560 459
修正前基因组大小/Mbp	536.16
修正基因组大小/Mbp	526.30
杂合率/%	0.90
重复率/%	55.39

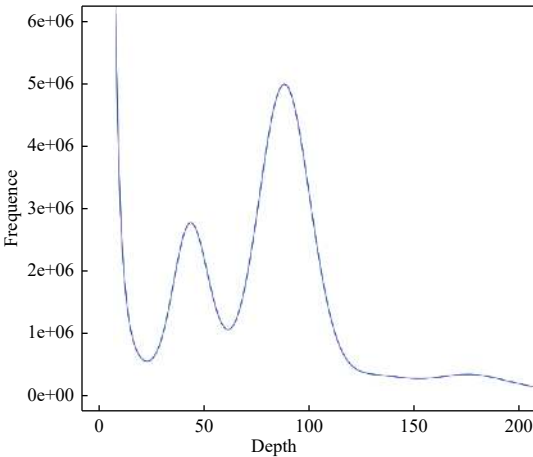


图 6 K-mer=17 Depth 和 K-mer 种类数频率分布图
Fig. 6 K-mer=17 Depth and K-mer species frequency distribution

2.5 基因组组装

2.5.1 数据组装结果 运用 Soapdenovo 软件拼接上述测序数据，并对数据进行纠错，构建 contig、scaffold 等优化过程，得到初步的基因组组装信息 (表 3)：针对组装好的长度大于等于 100 bp 的 scaffold 内部 contig 进行统计，得 N50 长度为 1 076 bp，N90 为 147 bp，组装得到最长的序列长度为 45 660 bp，组装的 contig 总数量为 917 423 个，总长度为 424 458 883 bp。进一步将所有文库测序得到的 reads 比对回初步得到的 contigs，利用 reads 之间的连接关系和插入片段大小信息，过滤掉长度 <100 bp 的 contig 序列，最终将 contigs 组装成 scaffolds，结果显示：N50 的长度的 1 889 bp，N90 为 189 bp，最长序列长度为 89 063 bp，组装总量为 717 232 个，总长度为 443 258 576 bp。

表 3 基因组组装结果统计

Table 3 Genomic assembly results statistics

数据类型 Types of data	拼接序列 Spliced sequence	
	Contig	Scaffold
Total_length/bp	424 458 553	443 258 576
N90_length/bp	147	189
Max_length/bp	45 660	89 063
N50_length/bp	1 076	1 889
Total_number/个	917 423	717 232

2.5.2 GC 含量分布分析 GC 含量是反映植物基因组组成的重要指标之一，GC 含量深度分析图用于检测测序是否存在 GC 分布偏向，样品是否存在细菌的污染等。由图 7 可得：西伯利亚白刺基因组测序没有明显的 GC 偏向。图中有 2 处 GC 聚集处，为了确认低测序深度区域是否为细菌污染造成，将低测序深度序列比对到 NCBI 核苷酸数据库，并没有细菌序列被比对上，说明样品没有被细菌污染，推测这是由于西伯利亚白刺基因组高杂合度所造成的。由于在组装过程中同源染色体上杂合部位只能被识别出一半，导致此部位的 GC 含量分布在低测序深度区域。

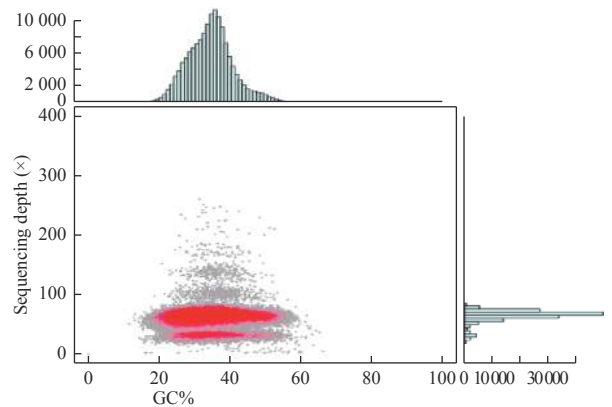


图 7 GC 含量与测序深度关联分析统计图
Fig. 7 GC content and sequencing depth correlation analysis

2.6 SSR 位点分析

由 MISA 脚本分析西伯利亚白刺基因组数据并统计 (表 4)，共搜寻到 521 125 个 SSR 位点，其中，单核苷酸位点出现比例最高，达 342 883 个，占总 SSR 位点的 65.80%；二核苷酸位点 146 312 个，占比 28.06%；三核苷酸位点 26 133 个，

占 5.02%; 四个及以上核苷酸位点 8 678 个, 占 1.67%。所以, 单核苷酸重复是西伯利亚白刺主要的 SSR 重复位点, 同时单核苷酸重复中 A/T 占比最多, 达到了 63.94%。

表 4 西伯利亚白刺 SSR 位点统计
Table 4 SSR locus statistics of *N. sibirica*

类型Type	结构structure	数目Number	所占比例Proportion/%
单碱基	A/T	333 188	63.94
	G/C	9 695	1.86
双碱基	AC/GT	10 978	2.10
	AG/CT	14 095	2.70
	AT/AT	118 038	22.65
	GG/CG	3 201	0.61
	AAC/GTT	946	0.18
三碱基	AAG/CTT	4 311	0.83
	AAT/ATT	16 521	3.17
	ACC/GGT	1 601	0.31
	ACG/CGT	87	0.02
	ACT/AGT	121	0.02
	AGC/CTG	452	0.09
	AGG/CCT	653	0.13
	ATC/ATG	1 259	0.24
	CCG/CGG	182	0.03
≥四碱基		8 678	1.67
合计		521 125	100

3 讨论

基因组大小是指生物单倍体染色体中 DNA 的含量, 也称为 C 值^[18]。目前为止已有数千种动植物的 C 值被检测并收录入相应的动植物 C 值库^[19-20]。DNA 的 C 值是生物体重要的基因特征, 是种群分类的证据之一, 也是开展各项基因工作的基础。了解基因组大小对于推测物种的演化趋势、进化地位、种属间进化关系、生物进化分类等具有深远的意义。

基因组大小预测常使用流式细胞术^[21]、Feulgen 图像分析法^[22]、全基因组 survey 调查^[23]等方法。流式细胞术通过比较待测植物和内标植物细胞悬液荧光吸收峰比值, 根据公式由内标植物的基因组来计算待测植物基因组的大小, 是一种快速、便捷的基因组预估的方法, 在测定动植物的基因组大小

方面均有较广的应用。

全基因组 survey 测序是基于小片段文库的低深度从头测序, 通过对原始数据进行图像识别, 去污染、去接头等步骤, 再进行 K-mer 分析, Soapdenovo 软件组装继而完成整个分析过程, 可对基因组的大小、GC 含量、杂合率以及重复序列的含量等重要的基因组特征信息进行分析, 相比于流式细胞仪、Feulgen 图像分析法等基因组大小预测方法更能切合所测生物体基因组特征, 是一种更精确的分析未知基因组特征的途径^[24-26]。

西伯利亚白刺基因组 GC 含量为 36.78%, 没有明显的过高或过低的情况^[27], 对 NGS 测序准确性影响较小; 而其杂合率为 0.9%, 基因组重复序列比例达 55.39%, 属于高杂合基因组。推测可能是由于西伯利亚白刺在地理分布上较广, 生态条件悬殊、植物形态变化也较大有关^[28]。

一般来说, 基因组杂合度越大, 重复片段越多, 该物种的组装难度就越大。西伯利亚白刺属于高杂合基因组植物, 而同为高杂合基因组的胡杨利用全基因组鸟枪法结合 Fosmid 拼装策略获得了精度较高的基因组图谱^[29]具有一定参考意义, 如果使用二代测序 Platanus 组装软件^[30]可能更适合于西伯利亚白刺基因组的拼装。随着近年来测序成本的下降和 3 代测序技术的普及, 二代 Illumina 搭配三代 Pacbio 辅以 Hi-C 技术的方案将会是西伯利亚白刺全基因组测序更好的选择, 更有利于获得高质量的全基因组图谱。

4 结论

本实验测得西伯利亚白刺基因组大小为 536.16 Mbp, 修正后为 526.30 Mbp, 杂合率为 0.90%, 重复序列比例为 55.39%; 西伯利亚白刺 Contig N50 为 1 076 bp, 总长为 424 458 553 bp, Scaffold N50 为 1 889 bp, 总长为 443 258 576 bp。西伯利亚白刺有 521 125 个 SSR 位点, 其中单核苷酸位点有 342 883 个, 二核苷酸位点有 146 312 个, 三核苷酸位点有 26 133 个, 四个及以上为 8 678 个, 单核苷酸为其主要的 SSR 特征。

参考文献:

- [1] 赵可夫, 范海, 江行玉, 等. 盐生植物在盐渍土壤改良中的作用[J]. 应用与环境生物学报, 2002, 8(1): 31-35.
- [2] 玉苏甫·买买提, 阿布都克里木·热依木, 阿丝叶·阿不都力米提. 西伯利亚白刺不同部位的氨基酸测定分析[J]. 中国农学通报, 2016,

- 32(22): 37-41.
- [3] 耿庆祯, 都恕兴, 尚艳霞, 等. 西伯利亚白刺与唐古特白刺果实营养成分比较分析[J]. 食品科技, 2008, 33(3): 101-104.
- [4] 高 航, 索有瑞. 柴达木盆地西伯利亚白刺和唐古特白刺的氨基酸含量及其营养评价[J]. 氨基酸和生物资源, 2002, 24(4): 4-7.
- [5] 王晨霞, 陈贵林. 西伯利亚白刺的组织培养与快速繁殖[J]. 植物生理学报, 2007, 43(6): 1143-1144.
- [6] 玉苏甫·买买提, 阿丝叶·阿不都力米提, 尤努斯·居玛. 于田县西伯利亚白刺播种育苗技术研究[J]. 北方园艺, 2016(19): 71-74.
- [7] Banaev E V, Vysochina G I, Kukushkina T A, *et al.* Variability in the content of biologically active substances in the leaves of *Nitraria sibirica* Pall. (Nitrariaceae)[J]. Contemporary Problems of Ecology, 2014, 7(1): 90-96.
- [8] Mamat Y, Reyim A, Abdulimit A. Detection and analysis of amino acids in different parts of *Nitraria sibirica*[J]. Chinese Agricultural Science Bulletin, 2016.
- [9] 李焕勇, 唐晓倩, 杨秀艳, 等. NaCl处理对西伯利亚白刺幼苗中矿物质元素含量的影响[J]. 植物生理学报, 2017, 53(12): 2125-2136.
- [10] 陈志强, 李庆贱, 时瑞亭, 等. 苏打盐碱胁迫对西伯利亚白刺光合和生长的影响[J]. 北京林业大学学报, 2011, 33(3): 31-37.
- [11] Wang L, Ma Y K, Li N N, *et al.* Isolation and characterization of a tonoplast Na⁺/H⁺ antiporter from the halophyte *Nitraria sibirica*[J]. Biologia Plantarum, 2016, 60(1): 113-122.
- [12] Li H, Tang X, Yang X, *et al.* De novo transcriptome characterization, gene expression profiling and ionic responses of *Nitraria sibirica* Pall. under salt stress[J]. FORESTS, 2017, 8(2116): 211.
- [13] 陈 勇, 柳亦松, 曾建国. 植物基因组测序的研究进展[J]. 生命科学研究, 2014, 18(1): 66-74.
- [14] Goff S A, Ricke D, Lan T H, *et al.* A draft sequence of the rice genome (*Oryza sativa* L. ssp. Japonica)[J]. Science, 2002, 296(5565): 92-100.
- [15] Gregory T R. Animal genome size database[DB]. Noncoding DNA, 2001.
- [16] 唐 其, 马小军, 莫长明, 等. 罗汉果全基因组Survey分析[J]. 广西植物, 2015, 35(6): 786-791.
- [17] 张 伟, 王 瑞, 杨晓野, 等. 捕食性真菌 *Duddingtonia Flagrans* 基因组DNA的提取及基因组Survey分析[J]. 中国兽医学报, 2017(11): 2090-2094.
- [18] Shi M J, Cheng Y Y, Zhang W T, *et al.* The evolutionary mechanism of genome size[J]. Chinese Science Bulletin, 2016.
- [19] Ryan G T, Nicol J A, Heidi T, *et al.* Eukaryotic genome size databases[J]. Nucleic Acids Research, 2007, 35(Database issue): D332-D338.
- [20] Sònia G, Leitch I J, Alba A R, *et al.* Recent updates and developments to plant genome size databases[J]. Nucleic Acids Research, 2014, 42(Database issue): D1159-D1166.
- [21] Doležel J, Greilhuber J, Suda J. Estimation of nuclear DNA content in plants using flow cytometry[J]. Nature Protocols, 2007, 2(9): 2233-2244.
- [22] Jeffery N W, Gregory T R. Genome Size Estimates for Crustaceans Using Feulgen Image Analysis Densitometry of Ethanol-Preserved Tissues[J]. Cytometry Part A the Journal of the International Society for Analytical Cytology, 2015, 85(10): 862-868.
- [23] 王 雪, 周佳熠, 孙会改, 等. 新疆沙冬青基因组调查测序与基因组大小预测[J]. 植物遗传资源学报, 2018, 19(1): 143-149.
- [24] 孟金涛, 苑建蕊, 魏彦杰, 等. 基于DeBruijn图的DeNovo序列组装软件性能分析[J]. 科研信息化技术与应用, 2013, 29(5): 39-41.
- [25] Li R, Fan W, Tian G, *et al.* The sequence and de novo assembly of the giant panda genome[J]. Nature, 2010.
- [26] Miller J R, Koren S, Sutton G. Assembly algorithms for next-generation sequencing data[J]. Genomics, 2010, 95(6): 315-327.
- [27] Aird D, Ross M G, Chen W S, *et al.* Analyzing and minimizing PCR amplification bias in illumina sequencing libraries[J]. Genome Biology, 2011, 12(2): R18.
- [28] Zhang M, Temirbayeva K, Sanderson S C, *et al.* Young Dispersal of Xerophil Nitraria Lineages in Intercontinental Disjunctions of the Old World[J]. Scientific reports, 2015, 5(1): 13840.
- [29] Ma T, Wang J, Zhou G, *et al.* Genomic insights into salt adaptation in a desert poplar[J]. Nature Communications, 2013, 4(4): 2797.
- [30] Cong Q, Borek D, Otwinowski Z, *et al.* Tiger swallowtail genome reveals mechanisms for speciation and caterpillar chemical defense[J]. Cell Reports, 2015, 10(6): 910-919.

A study on *Nitraria sibirica* Pall Genome

ZHU Li-ming^{1,2}, LI Meng-juan^{1,2}, ZHANG Jing-bo³, YANG Xiu-yan^{3,4}, CHENG Tie-long^{1,2}

(1. Key Laboratory of Forest Genetics and Biotechnology, Nanjing Forestry University, Nanjing 210037, Jiangsu, China; 2. Co-Innovation Center for Sustainable Forestry in Southern China, Nanjing 210037, Jiangsu, China; 3. Experimental Centre of Desert Forestry, Chinese Academy of Forestry, Dengkou 015200, Inner Mongolia, Mongolia, China; 4. Chinese Academy of Forestry, Beijing 100091, China)

Abstract: [Objective] This study aims at estimating the size, complexity and other characteristics of *Nitraria sibirica* Pall genome. [Method] Taking *Solanum lycopersicum* as DNA reference standard, the young leaves of *N. sibirica* were used as test material and mG as separate buffer to establish the *N. sibirica* genomic C-value by flow cytometry (FCM) method. A small fragment library was established for Illumina HiSeq sequencing. The genome size, heterozygosity, GC content and repetitive sequence ratio were studied and MISA script was used to analyze the SSR data. [Result] The size of *N. sibirica* genome was 526.30 Mbp, the heterozygosity rate was 0.90%, the GC content ratio was 36.78% and the repeat sequence ratio was 55.39%. Using the assembled genome data for SSR analysis obtained 521 125 loci, of which mononucleotide repeat were 342 883 loci, accounting for 65.80% of the total SSRs. [Conclusion] It is predicted that the *N. sibirica* genome has a complex nature and can be analyzed via Illumina combined to PacBio and Hi-C technologies in order to facilitate the assemble of *N. sibirica* genome and the acquisition of genetic maps.

Keywords: *Nitraria sibirica* Pall; Flow cytometry; genome size; GC content; SSR molecular marker

(责任编辑: 张 研)