

利用 MGI 平台对大豆进行全基因组重测序分析

许亚茹

(中国农业科学院 作物科学研究所 重大平台中心, 北京 100089)

摘要: 本研究建立了 MGI 平台全基因重测序的方法。MGI 平台对大豆的全基因进行重测序结果显示, 测序数据质量良好, 且与参考基因组比对率较高, 符合后续分析要求, 对其进行 SNP 和 Indel 的变异检测和注释, 此结果说明今后可利用 MGI 平台对其它样品进行全基因重测序分析。

关键词: MGI 平台; 全基因重测序

Whole genome resequencing analysis of soybeans using the MGI platform

Xu yaru

Abstract: In this study, a method for whole gene resequencing on the MGI platform was established. The results of resequencing the whole genes of soybean by MGI platform showed that the sequencing data was of good quality and had a high comparison rate with the reference genome, which met the requirements of subsequent analysis, and the variation detection and annotation of SNP and Indel were carried out, which indicated that the MGI platform could be used to perform whole gene resequencing analysis on other samples in the future.

Keywords: MGI platform; Whole gene resequencing

1 研究背景

大豆是重要的粮食作物和油料作物, 也是人类最主要的植物蛋白来源[1]。我国是野生大豆的发源地, 有着极其丰富的大豆种质资源基础, 但是育种和产量较其他大豆主产国显得略有不足, 究其原因我国对大豆的研究和发掘力度存在不足, 因此, 对大豆育成品种的改良势在必行。自 2010 年起, 大豆群体水平的重测序也全面开展, 在大豆的全基因组变异图谱上也得到了一定的研究进展[2]。本研究利用 MGI 平台对大豆全基因组进行重测序分析, 挖掘全基因组水平上的突变。

2 实验仪器

主要实验仪器：MGISP-960、MGIDL-T7、DNBSEQ-T7

3 实验结果

3.1 测序数据质量

根据 MGI 平台的测序特点，使用双端测序的数据，要求 Q30 平均比例在 85% 以上，可以看出大豆重测序数据 Q30 平均比例在 94.72% 以上，说明大豆测序数据质量良好，满足分析要求。

表 1 测序数据统计表

Samples	ID	Clean reads	Clean bases	GC Content	%≥Q20	%≥Q30
P117	P117	169494922	25424238300	36.18%	98.49%	95.27%
P118	P118	166483906	24972585900	36.47%	98.61%	95.70%
P119	P119	186127112	27919066800	35.89%	98.57%	95.61%
P120	P120	192397276	28859591400	36.46%	98.22%	94.72%
P198	P198	141636468	21245470200	37.11%	98.67%	95.84%
P199	P199	169468714	25420307100	36.55%	98.60%	95.66%
P200	P200	155078286	23261742900	37.90%	98.77%	96.14%

样品原始数据碱基质量值可由图 1 看出不存在异常碱基，6 个大豆碱基测序错误率分布均如图 1。

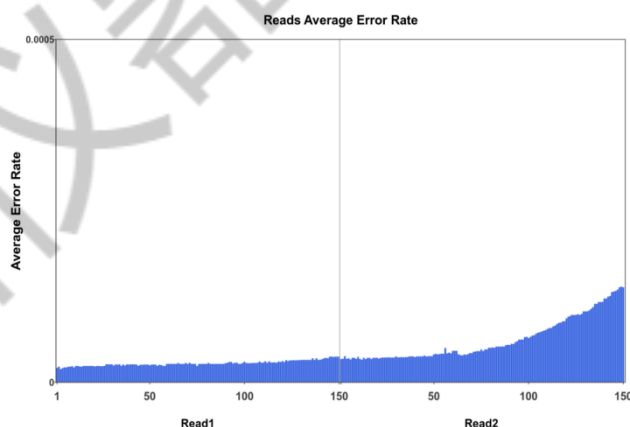


图 1 碱基测序错误率分布图

碱基类型分布检查可用于检测有无 AT、GC 分离现象，若有碱基分离现象可能是测序或