

欧洲云杉液相芯片开发及应用

杨晓伟^{1,2} 张苗苗¹ 欧阳芳群³ 李朋乐¹ 安三平⁴ 王瑞辉² 王军辉^{1*}

(1. 中国林业科学研究院林业研究所 北京 10091; 2. 中南林业科技大学林学院 长沙 410000; 3. 国家植物园 北京 10091; 4. 小陇山林科所 天水 741000)

摘要:【目的】为开发适用于欧洲云杉高效稳定且价格低廉的育种芯片, 本研究结合欧洲云杉固相芯片“50 K array”和已发表的 SNP 位点, 开发了欧洲云杉 GenoBaits® *Picea abies* 40K GBTS panel。并利用该芯片开展了群体遗传结构、全基因组关联分析和基因组选择研究。【方法】利用该芯片对收集在甘肃天水的 465 份欧洲云杉无性系开展了群体遗传结构分析, 同时挖掘了控制树高、胸径、材积、MOE 值、Pilo、AV 等重要性状的 SNP 位点; 另外, 本研究也利用该芯片和 GBLUP、BayesA、LightGBM 以及 RandomForest 4 个模型进行了全基因组选择分析。【结果】1) 开发的 40K 液相芯片平均检出率为 99.27%, 一致率为 99.10%, 采用 GBTS 技术, 该芯片实际可提供高达 74687 个 mSNP 位点; 40K GBTS 芯片和 mSNP 大部分位于基因间区域; 最常见的 SNP 类型为[C/G]、[C/T]和[A/G], 占总 40K 芯片的 61.65%; mSNP 和 40K GBTS 中, 12.64%和 22.43%的 SNP 的 MAF 值大于 0.05, 其中分别有 58142 (77.85%) 和 29761 (74.40%) 个 SNP 属于中等多态性位点 (PIC 值>0.20)。2) 从 PCA、Admixture 和 NJ-tree 分析可以看出很明显的人工选育引起的群体结构差异, 均将待试群体划分为较为清晰的 3 个亚群。3) 通过 GWAS 分析共鉴定出 MA_136235_10537、MA_20503_36941、MA_10434168_1395 等 29 个与生长性状高度相关的 SNP 位点, MA_282528_2264、MA_18312_64895、MA_27456_53319 等 17 个与木材材性相关的 SNP 位点, 1 个与干型相关的位点 (MA_458872_5043)。4) 4 个模型中 BayesA 模型的预测精度最高, 除 SS 性状外, 其余几个性状均随着标记数量的增加, 预测精度逐渐升高, 当标记数量达到 10000-15000 时, 预测精度最高, 随后逐渐趋于稳定; 尽管在 GBLUP 中我们采用了 3 种策略试图去提高预测精度, 但从结果来看提升的幅度不大; BayesA、GBM 和 RF 中却可通过结合 GWAS 对标记进行预选, 使用较少的标记取得较好预测精度 (SS 性状除外), H、DBH、Pilo、MOE、AV 最高精度分别可达 0.38, 0.60, 0.58, 0.44 和 0.65。【结论】综上, 本研究设计了一款欧洲云杉 40K GBTS 液相芯片, 具有功能位点丰富、分型准确、成本低廉等优点, 欧洲云杉 GWAS 和 GS 结果也证明了利用该芯片进行相关研究的可靠性。相信随着芯片位点的不断优化升级, 将使得相关研究更加准确和高效, 进而更好地服务于欧洲云杉分子标记辅助育种工作。

关键词: 欧洲云杉; 液相芯片; 群体遗传; 全基因组关联分析; 全基因组选择;