【ART001-009】eQTL分析解析早期玉米籽粒发

文章链接：https://onlinelibrary.wiley.com/doi/full/10.1111/tpj.14193

发表期刊：The Plant Journal

第一作者：Junling Pang

通讯作者：赵军 （范云六、王国英）

1. 实验材料

使用了前人研究中的材料（Yang et al., 2010; Fu et al., 2013），从368份材料中选出282份材料调查了授粉后5天的玉米籽粒，268份在授粉后15天的。以及另外14份玉米温带材料。

RNA建库，PE125测序。

2. 序列比对

过滤掉低质量的数据后，使用 tophat进行比对。Picard 去重。

3. 表达定量

使用GenomicFeatures 和 GenomicAlignments 软件进行表达定量， 使用 DESeq2 软件包进行表达量归一化。保留群体中值表达水平 >0 的基因用于关联分析。

为了避免异常值以及关联分析时基因表达的非正态性对结果的影响，使用R软件qqnorm 包对表达量进行转化。

4. SNP 检测及缺失基因型插补

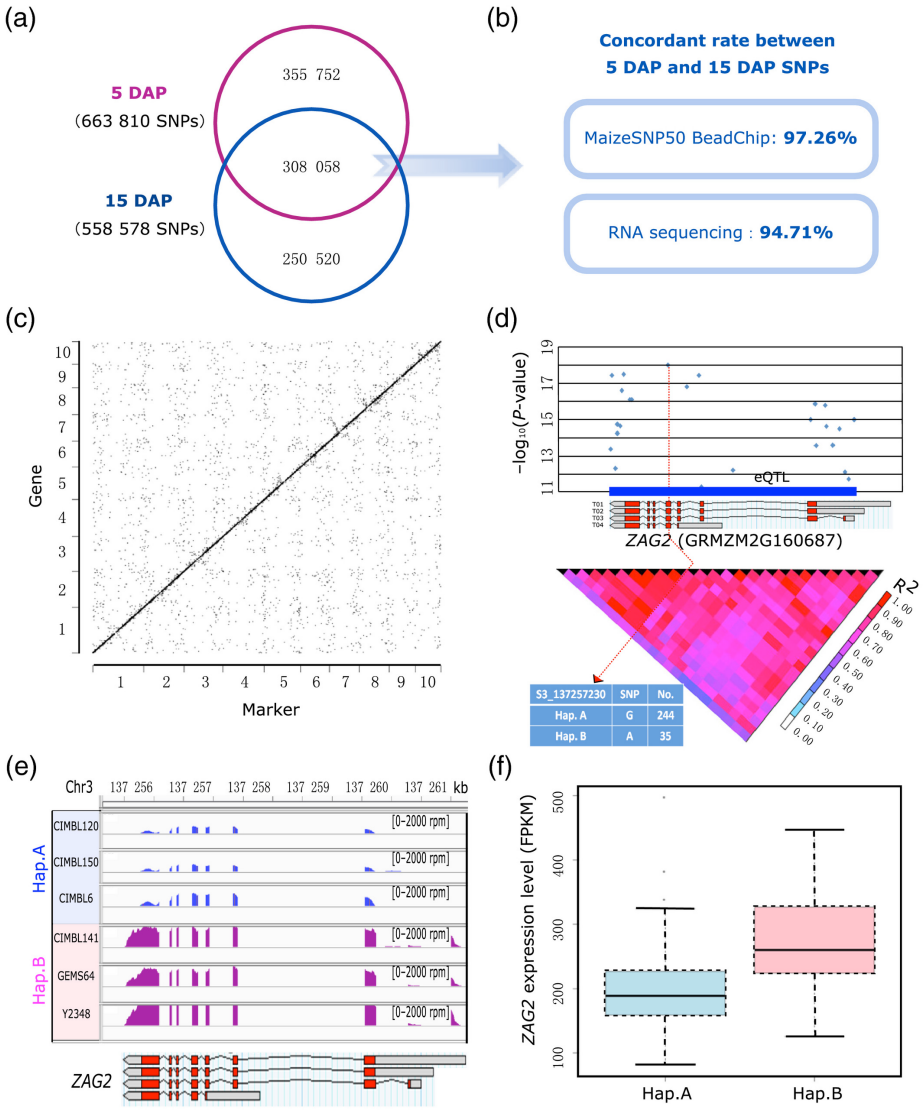
使用 GATK 3.6 软件基于RNAseq best practice 对282份玉米自交系进行SNP 检测。过滤杂合率 >10% 或者缺失率 > 60% 的 SNP 后将杂合位点定义为缺失，同时过滤 Maf < 0.02 的位点后使用 Beagle4.1 进行插补。获得 663810 个高质量 SNP。

作者首先使用无参方法对全部位点进行插补，然后使用有参方法进行插补，然后将两种方法进行整合：如果存在有参方法的位点，则将无参方法中的位点使用其替换。插补后的杂合位点使用缺失代替。

5. 基因型比较验证

将之前授粉后 15 天RNA测序得到的基因型与本研究得到的基因型进行比较，计算相同材料的SNP一致性，发现 308058 个SNP在两组数据中一致（46.41% of 5 DAP and 55.15% of 15 DAP）。而前期中 15 DAP 的基因型主要来自 Maize SNP50K 芯片（49256个）以及 RNASeq（509322）两种方法。

作为另外的SNP质量控制方法，作者将这些一致性位点分别与来自这两种方法的位点相比，发现5DAP的基因型一致性分别为97.26%及94.71%（5DAP-vs-50K, 5DAP-vs-RNASeq）。



6. 基因型整合

为了增大标记密度，作者将两次得到的基因型进行整合，本研究中得到了 914330 个SNP 标记，覆盖了 28280 个基因，相比之前的研究多了 3105 个基因。本研究中，89.73% 的SNP落在了基因区域内。每个基因平均有 32 个SNP。对于重叠的SNP，作者保留了本研究中得到的基因型。

6. SNP 注释

作者使用 VariantAnnotation 软件对得到的SNP进行注释，并根据在基因组上的位置对其进行分类。

7. 群体结构，亲缘关系以及隐藏因子分析

作者使用 TASSEL v5 软件，基于得到的 236205 个 Maf > 0.05 的 SNP 以及27103 个基因进行主成分及亲缘关系计算。选取了前三个主成分做为群体结构。使用 Tassel 中Normalized_IBS 方法计算其亲缘关系。

使用 PEER 软件（基于贝叶斯算法解释复杂的非遗传因素）计算其隐藏因素。将前三个主成分（群体结构）做为协变量带入模型，使用先验参数（aα = 10−7 G and bα = 10−1 G）进行计算，最终得到10个PEER 隐藏因素以进行后续分析

算。最终得到10个PEER 隐藏因素以进行后续分析。

> We used the prior parameters aα = 10−7 G and bα = 10−1 G in the noise precision distribution, where G is the total number of genes.

8. 关联分析及eQTL鉴定

使用 TASSEL 软件 MLM 模型进行关联分析，前三个主成分及10个隐藏因子做为协变量，亲缘关系矩阵做为随机效应的方差-协方差矩阵。使用 BH 方法进行假阳性控制（α = 0.05，P < 2.6E-6）。

eQTL 的鉴定方法与前人研究（Fu et al., 2013）一样，距离小于 5KB 的相邻SNP被分为一簇，至少包含三个显著SNP的簇被定义为一个 eQTL。如果相同基因的两个候选 eQTL 低于LD（R^2>0.1）,则保留包含更显著SNP的eQTL，如果两个eQTL具有相同的 p 值，则过滤效应较小的 eQTL。

最终，共鉴定得到 22966 个与 18377 个基因相关的 eQTL。这些 eQTL 的平均长度为 4086 bp，平均每个 eQTL 由 13 个 SNP 构成。

9. 多目标 eQTL 鉴定

作者比较了在两个发育时期（5dap，15dap）中鉴定得到的 eQTL，将在两个时期中关联大于 3 个基因的 eQTL 定义为了多目标 eQTL。并使用 Circos 软件对其进行展示。

10. 籽粒长度的关联分析

同样使用 Tassel v5 软件 MLM 及 GLM 模型进行关联分析。

11. 数量性状转录本

12. GO 富集分析

使用工具 agriGO 进行富集分析。

13. Motif 富集分析

14. 数据

The RNA sequencing data for the 282 maize inbred lines in 5-DAP kernels have been submitted to the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra/) with the BioProject ID PRJNA413629 or study accession ID SRP119998. The processed gene expression abundance data, including reads counts as well as FPKM values, could be accessed through NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo/) with the series entry GSE110315. The SNP dataset was submitted to the European Variation Archive (EVA; https://www.ebi.ac.uk/eva) with the accession number as PRJEB24974. All the procedures and scripts for reads mapping

the accession number as PRJEB24974. All the procedures and scripts for reads mapping, gene expression analysis and SNP calling can be found in the following website: https://github.com/Pang-Junling/maize-5DAP-eQTL.