Transcriptomic diversity among maize inbreds

Xiao Li1, Wei Huang1, Sanyang Liu1,2, Yinzhu Huang3,4, Jinhua Yang1, Wei Wu1,5, Kay Ying1, Cheng-Ting “Eddy” Yeh1, Dan Nettleton1, Patrick S. Schnable1,2

1Iowa State University; Ames, IA 50011-3650 2China Agricultural University; Beijing, China 100193
3Current Address: Department of Plant Pathology, Kansas State University, Manhattan, KS 66506-5502
4Current Address: DATA Biotechnology Beijing Co. Ltd., Beijing, China, 102206
5Current Address: Crop Genetics Research and Development; DuPont, Pioneer; Johnston, IA 50131

Current Address: Genome Technology Branch, the National Human Genome Research Institute, NIH, Rockville, MD 20892

Summary

1. Genotype-wise expression variable genes are enriched in regulatory functions.
2. ~5M SNPs and InDels are discovered among NAM founders, 65% of which are genetic variants.
3. 1,319 high confident gene set genes were affected by large effect SNPs.
4. Large effect SNPs with high allele frequencies were mainly splice site SNPs.
5. 16% of FGS genes have more than 10 haplotypes.
6. Gene haplotypes and CNV changes correlate with gene expression levels; CNV-by-tissue interactions were also discovered.
7. 14,441 nonFGS WGS genes and 5,597 novel transcribed regions were defined with RNAseq data.

Introduction

Zea mays has extensive genetic diversity. Many CNVs and PAVs were detected through array Comparative Genomic Hybridization (Springer et al 2009, Swanson-Wagner et al 2010). RNA sequencing technology enables simultaneous characterization of genomic variation, expressed genomic regions, gene expression and alternative splicing. To globally identify the expressed PAV genes as well as studying diversity of transcriptomes, we deeply sequenced the transcriptomes of 5 tissues (ear, shoot, root, tassel and shoot apex) from 27 NAM founders and conducted comprehensive genomic and functional genomic analyses of this dataset.

Variant calling via RNA-seq

(A) Concordance rate of SNPs/InDels among HapMap1, HapMap2 and RNA-seq datasets (B) Reference allele frequency distribution (C) Definition of large effect SNP (LES) and LES genes (D) Large effect allele frequency (E) Proportions of SNPs effects with large effect allele frequency categories

We want to thank Dr. Nathan Springer for helpful discussions and suggestions. This work is supported by NSF grants DBI- 1027527 (P. Schnable, PI) and DBI-0820610 (M. Scanlon, PI).