Abstract

A major goal of maize genomic research is to identify sequence polymorphisms responsible for phenotypic variation in traits of economic importance. Large-scale detection of sequence variation is critical for linking genes, or genomic regions, to phenotypes. However, due to its size and complexity, it remains expensive to generate whole genome sequences of sufficient coverage for divergent maize lines, even with access to next generation sequencing (NGS) technology. Because methods involving reduction of genome complexity, such as genotyping-by-sequencing (GBS), assess only a limited fraction of sequence variation, targeted sequencing of selected genomic loci offers an attractive alternative. We therefore designed a sequence capture assay to target 29 Mb genomic regions and surveyed a total of 4,648 genes possibly affecting biomass production in 21 diverse inbred maize lines (7 flints, 14 dents). Captured and enriched genomic DNA was sequenced using the 454 NGS platform to 19.6-fold average depth coverage, and a broad evaluation of read alignment and variant calling methods was performed to select optimal procedures for variant discovery. Sequence alignment with the B73 reference and *de novo* assembly identified 383,145 putative single nucleotide polymorphisms (SNPs), of which 42,685 were non-synonymous alterations and 7,139 caused frameshifts. Presence/absence variation (PAV) of genes was also detected. We found that substantial sequence variation exists among genomic regions targeted in this study, which was particularly evident within coding regions. This diversification has the potential to broaden functional diversity and generate phenotypic variation that may lead to new adaptations and the modification of important agronomic traits. Further, annotated SNPs identified here will serve as useful genetic tools and as candidates in searches for phenotype-altering DNA variation. In summary, we demonstrated that sequencing of captured DNA is a powerful approach for variant discovery in maize genes.