Fast, Accurate, Low Coverage Sequencing for Genome Wide Genotyping Large Populations for Genetic Improvement

J. CLEVENGER*; W. KORANI. Hudson Alpha Institute for Biotechnology, 601 Genome Way Huntsville, AL 35806

Genetic improvement using genomics relies on the quality, quantity, and cost of genome wide molecular markers. Marker types have evolved dramatically since the advent of the peanut genome initiative and subsequent release of a high-quality reference genome for peanut. Currently, there are two large scale single nucleotide polymorphism (SNP) arrays that provide a cost effective and accurate set of markers for researchers. However, these SNP arrays were designed using only a small population of peanut genotypes with available sequence and as such the arrays suffer from strong ascertainment bias. The most effective marker system uses genome-wide sequencing to theoretically assay all DNA sequence variation within a population, but sequence cost and informatics has been a limiting factor. We have recently developed an informatics pipeline, Khufu, that is highly accurate calling alleles in peanut with very low coverage sequencing (1X coverage). Using published datasets and sets of known true SNPs, we show that Khufu exhibits higher than 99% allele call accuracy in peanut using as low as 0.5X coverage whole genome sequencing. Additionally, more than 90% of the possible SNPs are recovered with 3X coverage sequencing. The cost of sequencing and accuracy with low coverage reduces per sample cost to that of fixed arrays. The combination of population-specific, *de novo* SNP sets, with low cost and high throughput makes genotyping with Khufu the best option for genetic mapping, diversity studies, and other genomics analyses.