Performing an Internal Reference Genome Assembly, Whole-Genome Sequencing and *In Silico* Digestion for Improved Efficiencies in Marker Detection for Virginia-type Peanuts

C.S. NEWMAN*, A.T. OAKLEY, R.J. ANDRES, K.R. FRITZ, J.C. DUNNE Department of Crop and Soil Sciences, North Carolina State University, Raleigh, NC, 27695; A.M. HULSE-KEMP, USDA-ARS, Raleigh, NC, 27695

The steady decline in sequencing costs provides the opportunity for peanut breeding programs to utilize next-generation sequencing to identify single nucleotide polymorphisms (SNPs) across the genome for marker development. The resulting information helps guide the development of reduced representation (genotype-by-sequencing; GBS) sequencing protocols. GBS is a low-cost technique which can be used to rapidly genotype lines in a breeding program. Creating a modern genotyping approach involves: selecting or assembling a reference sequence most similar to the organism of interest; whole-genome sequencing (WGS) on a subset of diverse germplasm; aligning WGS data to the reference genome to identify SNPs; optimizing the selection of enzyme pairs to be used for GBS through *in silico* digestion of the reference genome and by maximizing SNP site quantity and sequencing read depth in vitro. To initiate this protocol development, an internal reference genome of cv. 'Bailey II' was assembled. Tissue from Bailey II was sequenced on the PacBio Sequel II system, yielding 253GB of raw data. The raw reads were then assembled, polished with Arrow and Pilon, and scaffolded with BioNano. Simultaneously, a subset of diverse lines from the North Carolina State University Peanut Breeding & Genetics program were submitted for WGS. The resulting data were cleaned and aligned to the reference genome to reveal novel SNPs and to validate extant SNP positions present on the current Affymetrix Arachis2 48k array. A novel Python script was written to digest the 'Tifrunner' genome using selected enzyme pairs and was then applied to the Bailey II internal reference genome. The fragments identified from the *in silico* digest of Bailey II will then be analyzed to maximize the recovery of SNPs using GBS. Validation of the optimal enzyme pairs for SNP recovery and read depth will be verified through the construction of GBS libraries based on enzyme pairs that 1. Maximize SNP recovery (potentially low depth), 2. Maximize depth (potentially low SNP count) 3. Intermediate SNP sites and depth. This high-throughput genotyping method would afford peanut breeding programs, based solely on phenotypic selection, the opportunity to develop genomic resources to implement marker-assisted or genomic selection for trait improvement.