

Population Genomics of US Peanut Mini Core Collection using Genome-Wide SNP Genotyping

B. S. F. MÜLLER*, Y. CHU, and P. OZIAS-AKINS, Institute of Plant Breeding, Genetics & Genomics, Department of Horticulture, The University of Georgia, Tifton, GA 31793; C. CHEN, Auburn University, Auburn, AL 36849; and C. C. HOLBROOK, USDA-Agricultural Research Service, Crop Genetics and Breeding Research Unit, Tifton, GA 31793.

The US mini core has been a useful genetic resource for peanut breeding. In order to study the population genetics and structure of peanut germplasm, 108 accessions of the mini core, 16 cultivars and 12 accessions from two new botanical types were genotyped using the version 2 of Axiom SNP Array, resulting in 17,298 polymorphic markers. Linkage disequilibrium was calculated for all pairwise physical distances (7,840,819) among all the polymorphic SNPs on each chromosome separately. The average of genome-wide LD for pairs of SNPs (r^2) was 0.18, being 0.16 and 0.20 for A and B subgenomes, respectively. The genome-wide LD decayed to an r^2 below 0.2 within 23 Mb for the whole genome, and the LD of A subgenome (14 Mb) decayed faster than of the B subgenome (38 Mb). Population structure subdivided the accessions into two groups ($K = 2$), which largely conforms to the two subspecies: *hypogaea* (83 accessions) and *fastigiata* (53 accessions). The genetic divergence between these two subpopulations was high ($F_{ST}=0.45$) and the first two principal components explained 22% and 15.5% of the genetic variance through principal component analysis (PCA). The observed heterozygosity ($H_O=0.09$) was lower than expected heterozygosity ($H_E=0.26$) for the whole population, slightly higher for *fastigiata* ($H_O=0.11$, $H_E=0.19$) than *hypogaea* ($H_O=0.07$, $H_E=0.18$) subpopulation. The population genomics analyses were performed using non-overlapping 100 Kb sliding windows to estimate the F_{ST} , diversity from Nei (π), nucleotide diversity, Watterson's θ (θ_w) and Tajima's D within the two-subspecies for each window of the genome. F_{ST} was higher than 0.3 between the subspecies for the majority of the chromosomes, with the highest level at chromosomes 8 and 17. On average, the genetic diversity statistics (π , nucleotide diversity and θ_w) were higher for the *hypogaea* than *fastigiata* subpopulation on chromosomes 1 to 6 and 11 to 16, with the opposite for 9 to 10 and 18 to 20. In chromosomes 7 and 8, the diversity between the subspecies had more variation, perhaps due to genome rearrangements. The estimates of diversity between the two-subspecies varied along chromosome 17 mostly in the centromeric region, which had the highest F_{ST} in the whole genome. Negative Tajima's D estimates were observed for *hypogaea* subgroup mainly on chromosomes 18 to 20, which could indicate that a positive selection is driving divergence between these subspecies. The *hypogaea* subgroup may have passed through a recent selective sweep and possible population expansion after a recent bottleneck. The population genomics analyses identified genomic regions under selection putatively involved in the process of adaptation, providing good potential regions for further validation to detect candidate genes through genome-wide scans that can be useful to peanut breeding programs.