

GENOTYPIC DIVERSITY IN AVOCADO (*Persea americana* Mill.) FROM THE NATIONAL GERMPLASM BANK AND THE NORTHWEST ANDES OF COLOMBIA

J. Berdugo-Cely¹, A.J. Cortés², P.H. Reyes-Herrera³, I. Cerón-Souza³, L.P. Delgadillo-Duran³, A. Navas², R. Yockteng³

¹Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA) - CI Turipaná, Km 13 vía Montería-Cereté (Córdoba), Colombia.

²Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA) - CI La Selva, Km 7 vía Llanogrande, Rionegro (Antioquia), Colombia.

³Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA) - CI Tibaitatá, Km 14 vía Mosquera, Mosquera, Colombia.

Avocado (*Persea americana* Mill.) production increased significantly in Colombia during the last years because of the fruit exportations to Europe, and in the following years the demand will increase as a result of the oncoming access to the United States market. The Colombian germplasm bank, guarded by Agrosavia, currently preserves 210 accessions of native avocado diversity. So far, genotypic characterization using single nucleotide polymorphism (SNP) of native Colombian avocado trees is not available to determine the genetic diversity within the germplasm bank and to discover SNPs associated with phenotypic variation of agronomic importance traits. To fill this gap, 144 accessions from the national germplasm bank (*i.e.* 129 natives and 15 commercial trees) and 240 materials from the Colombian North West Andes (*i.e.* 73 native trees and 167 rootstocks of grafted trees from 8 plantations) we selected and genotyped using a next-generation genotyping method, known as Genotyping-by-Sequencing (GBS), in order to discover SNP markers throughout the genome of the 384 selected samples. For genomic library construction, we used a double enzyme digestion (*ApeKI/PstI*) of DNA and reads were produced by a 150 bp paired-end sequencing on an Illumina HiSeq X system. A total of 3.132.393.704 reads with 472.991.449.304 total bp-length we obtained. The SNP discovery is currently running using: (1) an avocado reference genome assembly, with a genome draft of 446.76 Mb published in NCBI; and (2) a *de novo* strategy using the STACKS 2.2 pipeline. From previous results found using 13 microsatellite loci, we do not expect to detect clear population structure in the germplasm analyzed, supporting the hypothesis of high anthropic mobility of the native avocado materials in Colombia. Yet, the high-throughput sequencing technology used in this study will enhance the discovery of high-density SNPs through the native avocado genomes to further allow the association between genotypic and phenotypic variation using Genome-Wide Associations Studies (GWAS) and Genomic Selection (GS) strategies.

DIVERSIDAD GENOTÍPICA DEL AGUACATE (*Persea americana* Mill.) DEL BANCO DE GERMOPLASMA DE LA NACIÓN Y EL NORESTE DE LOS ANDES DE COLOMBIA

J. Berdugo-Cely¹, A.J. Cortés², P.H. Reyes-Herrera³, I. Cerón-Souza³, L.P. Delgadillo-Duran³, A. Navas², R. Yockteng³

¹Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA) - CI Turipaná, Km 13 vía Montería-Cereté (Córdoba), Colombia.

²Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA) - CI La Selva, Km 7 vía Llanogrande, Rionegro (Antioquia), Colombia.

³Corporación Colombiana de Investigación Agropecuaria (AGROSAVIA) - CI Tibaitatá, Km 14 vía Mosquera, Mosquera, Colombia.

La producción de aguacate (*Persea americana* Mill.) ha aumentado significativamente en Colombia durante los últimos años debido a las exportaciones de su fruta a Europa, y en los próximos años la demanda aumentará como resultado del acceso que se aproxima al mercado de los Estados Unidos. El banco de germoplasma colombiano, mantenido por Agrosavia, conserva actualmente 210 accesiones de la diversidad de aguacates nativos. Sin embargo, la caracterización genotípica a través de polimorfismos de un solo nucleótido (SNP) de los aguacates nativos de Colombia para determinar la diversidad genética dentro del banco de germoplasma y descubrir SNPs asociados con la variación fenotípica de características de importancia agronómica no está disponible. Para llenar este vacío, 144 accesiones del banco de germoplasma de la nación (129 nativos y 15 árboles comerciales) y 240 materiales del noreste de los Andes de Colombia (73 árboles nativos y 167 portainjertos de árboles injertados de 8 plantaciones) fueron seleccionados y genotipados usando un método de genotipificación de nueva generación, conocido como Genotipificación por Secuenciación (GBS, por sus siglas en inglés), para descubrir marcadores SNPs a lo largo del genoma de las 384 muestras seleccionadas. Para la construcción de las bibliotecas genómicas, fue utilizada una doble digestión de ADN (*ApeKI/PstI*) y las lecturas se produjeron mediante una secuenciación de 150 pares de bases (pb) *paired-end* en un secuenciador Illumina HiSeq X. Un total de 3.132.393.704 de lecturas con 472.991.449.304 bp de longitud total fueron obtenidas. El descubrimiento de SNPs se está ejecutando actualmente utilizando: (1) un ensamblaje con genoma de referencia usando el borrador del genoma de aguacate Hass de 446.76 Mb publicado en NCBI; y (2) una estrategia de ensamblaje *de novo* usando STACKS 2.2. A partir de resultados previos utilizando 13 marcadores microsatélite, se espera no detectar una estructura poblacional clara en el germoplasma analizado, apoyando la hipótesis de alta movilidad antrópica de los materiales de aguacate nativos en Colombia. Sin embargo, la tecnología de secuenciación de alto rendimiento utilizada en este estudio mejorará el descubrimiento de SNPs de alta densidad a través de los genomas nativos del aguacate para permitir aún más la asociación entre la variación genotípica y fenotípica mediante las estrategias de estudios de asociación de genoma completo (GWAS, por sus siglas en inglés) y de selección genómica (GS, por sus siglas en inglés).