Genomics and Phenomics to Identify Yield and Drought Tolerance Alleles for Improvement of Camelina as a Biofuel Crop

Hussein Abdel-Haleem[1*] (Hussein Abdel-Haleem@usda.gov), Zinan (Lily) Luo[1], Aaron E. Szczepanek[1], Matthew Conley[1], Hao Zhang[2], Yufeng Ge[2], Daniel Schachtman[2], Megan M. Augustin[3], Jordan R. Brock[3], Noah Fahlgren[3], Toni Kutchan[3], Sheeja George[4], Russ Gesch[5], and **John M. Dyer[1]**

[1]USDA-ARS, US Arid-Land Agricultural Research Center, Maricopa, AZ; [2]University of Nebraska-Lincoln, Lincoln, NE; [3]Donald Danforth Plant Science Center, St. Louis, MO; [4]University of Florida, Quincy, FL; and [5]USDA-ARS, North Central Soil Conservation Research Laboratory, Morris, MN

Project Goals: Plant oils represent renewable sources of energy-dense hydrocarbons that can be used for biofuel, but a major challenge is to produce these oils in high-yielding non-food crops that can grow under marginal conditions. Our goal is to improve the suitability of camelina as a bioenergy crop. The major objectives are to: 1) Develop and apply automated, non-destructive high-throughput phenotyping (HTP) protocols to evaluate the phenotypic diversity of a camelina panel consisting of 250 accessions, grown under well-watered and water-limited conditions; 2) Discover alleles/genes controlling morphological, physiological, seed, and oil yield traits using genome-wide association studies; and 3) Identify, test, and validate useful germplasm under diverse environments and marginal production areas. These studies will enable the discovery of new genes associated with crop yield and stress tolerance and identify high-yielding cultivars that are suitable for certain geographical regions.

In recent years, *Camelina sativa* has received considerable attention as a potential non-food biofuel crop, but significant challenges remain to develop stable, high-yielding, geographically adapted germplasm suitable for biofuels production. Here, we are using advanced high-throughput phenotyping and genomics-based approaches to discover useful gene/alleles controlling seed yield and oil content and quality in camelina under water-limited conditions and identify high-yielding cultivars suitable for production in different geographical regions. Our experimental design includes complimentary field- and greenhouse-based HTP experiments conducted on a spring panel of 250 camelina accessions grown under well-watered and water-limited conditions.

To develop genomic tools and information for underpinning genome-wide association studies (GWAS), the camelina panel was genotyped using genome-by-sequencing technology (GBS), resulting in identification of 6,192 high-quality SNPs distributed throughout the genome. Population genetics analysis revealed two distinct populations of camelina corresponding to main geographical regions of cultivar collection. As an example of the utility of using these SNP markers to find genes associated with agronomically important traits, we conducted a GWAS of cuticular wax traits in the camelina diversity panel. Cuticular wax content and composition are important for regulating non-stomatal water loss in aerial organs of plants, and identification of markers and genes associated with these traits might be useful for developing more drought-tolerant camelina varieties. Wax content and composition were first determined for select varieties of camelina and closely related species, resulting in identification of 49 distinct lipid molecular species. These compounds were then characterized for the entire camelina diversity panel, revealing variations in wax content and composition. Broad-sense heritabilities further revealed a strong genetic component for certain lipid classes and lipid molecular species. GWAS studies were subsequently conducted, resulting in identification of SNP markers associated with certain wax components. Several of these markers were associated with specific genes, revealing candidates (and markers) that might be used for rational manipulation of wax content using molecular breeding approaches.

To identify genes regulating camelina oil content, composition, and yield, the entire camelina panel was cultivated and analyzed using a greenhouse-based high throughput phenotyping system (LemnaTech) available at the Danforth Center, with two replicates of each accession cultivated under well-watered and water-limited conditions, from 7 to 35 days after planting. Phenotypic data for each plant were collected and included approximately 200,000 images produced during the course of the experiment. Images were processed using PlantCV software to extract trait data including plant height and plant area. Algorithms are currently being developed to detect plant flowering. Seeds were also harvested from each plant and analyzed for oil content, composition, nitrogen and glucosinolate content, as well as total seed yield. Data analysis showed significant phenotypic variations among genotypes as well stress levels. Data are being prepared for subsequent GWAS analyses.

The camelina panel was also cultivated under well-watered and water-limited conditions in fields located in Nebraska and Arizona. Field high-throughput phenotypic data including plant height, canopy temperature, and spectral reflectances (e.g., NDVI) were collected using ground-based phenotyping vehicles equipped with various sensors and data logging equipment. More than 20 phenotypic traits related to yield components and fatty acids were collected. These data from well-watered and water-limited trails were analyzed and indicated significant phenotypic diversity in the camelina population. Genome-wide association analysis using GAPIT was conducted to identify putative single nucleotide polymorphism (SNP) markers that are significantly associated with these traits. Together, seven and nine significant SNP markers on different chromosomes were associated with plant height, seed weight and seed yield under water-limited and well-watered conditions, respectively, meaning that some genes might be actively triggered under either of the two conditions. No significant markers overlapped between environments. BLASTx annotation identified gene functions for 8 out of 16 genes associated with significant SNP markers. These functions were mainly related to general maintenance of normal cell function and development. These SNP could be good candidates for marker assisted selection. Data from more environments, years and locations will be combined for a more comprehensive GWAS analysis. This GWAS study will lay a foundation for future molecular breeding programs to help accelerate the selection of *C. sativa* varieties with superior traits of interests. Field data will be compared to data obtained from the greenhouse-based studies to help identify select cultivars that consistently perform better under water-limited conditions, and also provide a means to determine how well greenhouse-based studies mimic conditions observed in the field.

This material is based upon work that is supported by the National Institute of Food and Agriculture, U.S. Department of Agriculture, under award number 2016-67009-25639.