WHOLE GENOME SEQUENCING OF SUGARBEET AND DEVELOPMENT OF PUBLIC SNP GENOTYPING PLATFORM

Imad A. Eujayl* and Carl A. Strausbaugh
USDA-ARS, Northwest Irrigation and Soils Research Laboratory,
3793 N. 3600 E. Kimberly, ID 83341

ABSTRACT

A simplified approach has been implemented to generate public sugarbeet (*Beta vulgaris* subsp. *vulgaris*) whole genome sequence data (WGS). Next generation sequencing (NGS) systems based on generating large number of short sequence data is suitable to achieve this approach. The objectives of this study, were to; identify a large number of single nucleotide polymorphic (SNP) markers to be shared among public and private sugar beet researchers and specifically target polymorphic SNP between beet curly top resistant and susceptible genotypes, in addition to enriching sugar beet public genomic sequence data

The genome of KDH13 doubled haploid line has been sequenced using Illumina HiSeq2000 NGS platform. This line (PI663862) was released by USDA-ARS as a genetic stock resistant to beet curly top. Sequencing of a standard paired end and a 2kb-insert mate-pair genomic libraries, constructed from a leaf sample yielded 82.9 Gb of sequence data. Less than 2% of the raw sequence data was filtered out when eliminating mitochondria and chloroplast nucleotides. Nuclear sequence data totaling 51.5 Gb with a 35.7% GC content was assembled, using ABySS software, into contigs that achieved 68X genome coverage. The assembly of KDH13, named BvvSeq-1, was generated from a total of 3,850,207 contigs of which 95,923 were 1.15 Kb in length at N50. The contigs were assembled into 126,533 scaffolds in 1.8 Kb lengths at N50. These scaffolds cover 238 Mb which spans approximately 31% of the 758 Mb sugarbeet genome. BvvSeq-1 was aligned to the draft genome sequence of sugarbeet (RefBeet-0.9), which was generated from the KWS2320 doubled haploid line (susceptible to beet curly top) using MUMmer software. A total of 404 Mb of BvvSeq-1 were mapped to RefBeet-0.9 which accounts for 69%. A comprehensive array of single nucleotide variants were identified between the two genomes using Genome Analysis TK software. A thousand polymorphic single nucleotide variants between KDHL13 and KWS2320 were selected, based on the depth of sequencing (coverage) and variation in the flanking regions.

To establish a SNP panel for public genotyping and genetic mapping we used the Fluidigm® SNPtype assay coupled to the IFC and EP1 reader systems. This platform offers a flexible small scale genotyping system available in the following chip formats: 24.192, 48.48, and 96.96. We deployed the 96.96 format that allows for genotyping 96 individuals using 96 SNP markers and generated 9,216 data points.