PgtSNP chip: A high-throughput SNP genotyping array

Les J. Szabo[1], Pablo Olivera[2], Maria Newcomb[2], Dave Hodson[3] & Yue Jin[1]

[1]USDA ARS Cereal Disease Laboratory, St. Paul, MN 55108, USA
[2]Department of Plant Pathology, University of Minnesota, St. Paul, MN 55108, USA
[3]CIMMYT-Ethiopia, Addis Ababa, Ethiopia

Understanding the genetic diversity and population structure of *Puccinia graminis* f. sp. *tritici* (*Pgt*) has been hampered by the obligate, dikaryotic nature of this pathogen and the lack of robust high-throughput genotyping tools. A custom 1536-SNP Illumina GoldenGate array (PgtSNP 1.5k chip) was developed based on *Pgt* reference genome (U.S. isolate 75-36-700) and NGS sequence data from four selected isolates (78-21-BB463, U.S.; 04KEN156-4, Kenya; 06YEM34-1, Yemen; ANZ21-0, Australia). A balanced selection scheme was use to identify candidate SNP loci evenly distributed along the largest supercontigs, covering 50% of the *Pgt* genome. A final set of 1524 SNPs was selected with loci that are evenly spaced (average spacing of 27 kb) and spanning supercontigs 1-26. These loci are partitioned between genic (82.9%) and intergenic (17.1%). The genic loci are further partitioned between coding (66.5%), intronic (14.5%) and untranslated (2.9%) regions. Twenty-two additional loci previously identified for genotyping members of the "Ug99" race group were added. To test the performance of PgtSNP 1.5k chip a representative sample set of 50 *Pgt* isolates, spanning both geographical (15 countries) and temporal (56 years) diversity, was used. Replicates of each sample were included. Data was filtered using the following criteria: GenTrain score > 0.6; 10% GC score > 0.6; call frequency > 95%; and no replicate errors. The resulting SNP set containing 1236 loci was analyzed by Principal Component analysis with the software package 'Poppr'. The 50 *Pgt* isolates were divided into six genetic clusters. In general, the isolates divided geographically, with the North American samples distributed between two of the clusters and the remaining four clusters containing samples from Africa, Asia and Europe. In a second study, 41 *Pgt* isolates collected in Ethiopia during the 2013/14 main wheat-growing season was analyzed. Phylogenetic analysis divided these isolates into four well-supported clades, based on 918 SNP loci. Each of these clades represented a distinct *Pgt* race phenotype or race group: clade I, Ug99 race group; clade II, JRCQC; clade III, TRTTF/RRTTF; clade IV, TKTTF. Clade IV, was further subdivided into two distinct sub-clades and represents the *Pgt* race that was responsible for the 2013 wheat stem rust epidemic in Ethiopia. Analysis of Ethiopian *Pgt* collections made during 2014 wheat season will be discussed.