READ ME

Supplemental data for:

Zhao, D., Sapkota, M., Glaubitz, J., Bassil, N., Mengist, M. F., Iorizzo, M., Heller-Uszynska, K., Mollinari, M., Beil, C. T., Sheehan, M. J. (2024). A public mid-density genotyping platform for cultivated blueberry (*Vaccinium* spp.). *Genetic Resources* 5 (9), 36–44. doi: 10.46265/genresj.WQZS1824.

Contents

- Zhao et al_Supplemental Figure 1, 2, 3.pdf
- Zhao et al_Supplemental File 1.xlsx
- Zhao et al_Supplemental File 2.csv
- Zhao et al_Supplemental File 3.xlsx
- Zhao et al_Supplemental Table 1.xlsx

Details

Supplemental Figure 1. Principle Component Analysis (PCA) plots of the 'Draper' x 'Jewel' F_1 population. *[See Zhao et al_Supplemental Figures 1, 2, 3.pdf]*

Supplemental Figure 2. Missing data rates for different grouped subsets of genetic material. *[See Zhao et al_Supplemental Figures 1, 2, 3.pdf]*

Supplemental Figure 3. Blueberry Genetic map construction for the F_1 population. *[See Zhao et al_Supplemental Figures 1, 2, 3.pdf]*

Supplemental File 1: Genomic information of the blueberry 3K DArTag marker panel.
- 1a. **Blueberry 3K DArTag** tab: List of SNP markers targeted in the Vaccinium corymbosum genome in the 3K DArTag panel for blueberry. *Marker_name*=Scaffold number and physical position based on the Draper reference genome assembly in Colle *et al*, 2019, *Chromosome*=Scaffold number, *Position*=physical position, *Strand*=indicates if the target locus is on the plus or minus strand of DNA, *In gene*=Indicates if the locus is in a genic region or nonGenic region, *GeneID*=gene name as annotated in Colle *et al*, 2019.
- 1b. **Gene number vs markers targeted** tab: Number of markers selected per chromosome is positively correlated with the number of genes per chromosome. *Chromosome*=linkage group assignment, *Scaffold*=scaffold assignment based on the Draper reference genome assembly v1.0 in Colle *et al*, 2019, *length of chromosome (bp)*=Total length of scaffold, *Count of marker loci*=distribution of the 3000 markers by linkage group, *% of genome*=percentage of basepair residing on each linkage group, *1 marker per genome bp*=the average frequency of selected markers on a chromosome in physical distance,

Number of genes=number of annotated genes per chromosome based on and gene model annotation v1.0 in Colle *et al*, 2019, *Rank order by number of loci (low to high)*=Assigned rank of the number of markers on the panel per chromosome where 1 is fewest and 12 is greatest, *Rank order by number of genes (low to high)*=Assigned rank of the number of genes per chromosome where 1 is fewest and 12 is greatest.

- o Correlation plot of the number of genes per chromosome to the number of targeted loci per chromosome (using rank order).
  - Draper genome: http://gigadb.org/dataset/100537
  - Draper assembly: https://ftp.cngb.org/pub/gigadb/pub/10.5524/100001_101000/100537/V_corymbosum_genome_v1.0.fasta
  - Gene annotation: https://ftp.cngb.org/pub/gigadb/pub/10.5524/100001_101000/100537/V_corymbosum_v1.0_geneModels.gff

[See Zhao et al_Supplemental File 1.xlsx]

Supplemental File 2: MADC report (as CSV) of the 375 samples used to validate the 3K DArTag panel. Rows 1-6 contain sample metadata with plate and well assignments. Row 7 contains headers in columns A-P and sample names in columns Q-OA. Rows 8-22920 contain all the detected microhaplotypes for the 3K DArTag loci where "|RefMatch" and "|AltMatch" refer to empirically discovered haplotypes that match either Ref or Alt at the target locus but have additional off-target SNPs within the 54bp sequence. Data in matrix Q8-OA22920 are read counts of all detected microhaplotypes per individual. Headers in Row 7: *AlleleID*=Microhaplotype name, *CloneID*=Locus name, *AlleleSequence*=Detected microhaplotype sequence, *ClusterConsensusSequence*=Reference genome sequence expected, *CallRate*=frequency of each microhaplotype observed in the dataset, *OneRatioRef, OneRatioSnp, FreqHomRef, FreqHomSnp, FreqHets, PICRef, PICSnp, AvgPIC, AvgCountRef, AvgCountSnp, RatioAvgCountRefAvgCountSnp* are all additional frequency statistic for each microhaplotype.

[See Zhao et al_Supplemental File 2.csv]

Supplemental File 3: Linkage group with their marker order, positions in cM, and parental phasing information where p1 represents 'Draper' and p2 represents 'Jewel', and the decimal place following p1 and p2 indicates which of the four available homologs the marker resides on in the parent (eg., p1.3 indicates Draper parent allele on homolog 3). Loci_Ct=running count of mapped loci, Marker Name=scaffold number and physical position based on the Draper reference genome assembly in Colle *et al*, 2019, LG=Linkage group, Ref Chrom=chromosome assignment based on the reference genome, Ref Position=physical position of the SNP based on the reference genome (in bp), Map Position=Emperical genetic map location of each locus from MAPpoly, ref=Reference nucleotide, alt=Alternative nucleotide, Dose in Draper (p1)=dosage count for each mapped locus in 'Draper' parent, Dose in Jewel (p2)=dosage count for each mapped locus in 'Jewel' parent.

[See Zhao et al_Supplemental File 3.xlsx]

Supplemental Table 1. Accessions used in the construction and testing of the blueberry 3K DArTag panel. Germplasm that was used in whole genome sequencing and SNP database construction are indicated by an asterisk (*). Samples not tested with the validation set are indicated by the pound sign (#). Where known, species, ploidy, and classification are indicated.

[See Zhao et al_Supplemental Table 1.xlsx]