

SNP calling for oat biparental genotyping arrays

Malachy Campbell

March 2020

1 Summary

Briefly, the experiment consists of 912 samples (one is a blank) genotyped with the Oat_GenMills_6k Array. Initial clustering was performed by Jason Fiedler (Jason.Fiedler@usda.gov) using the Polyploid module for GenomeStudio. I'm not sure what parameters were used by Jason. I (Malachy) cleaned up the clusters following the protocol described by the Merriman Lab at the University of Otago (<https://github.com/MerrimanLab/merrimanlab.github.io/wiki/Genome-Studio-Genotype-Quality-Control-Protocol>). It is basically a summary of Gao et al (2014; <https://doi.org/10.1038/nprot.2014.174>) with some additional steps.

First, I sorted markers by cluster separation and looked at the clustering results for each marker to determine a threshold where the clusters were reliably separated. A cluster separation value of 0.1 seemed to provide decent separation, thus this was used a lower bound. Next, markers exceeding this value were sorted based on the call frequency and the clustering for each marker was inspected manually with priority being placed on markers with low call frequencies. In cases where the clusters were well-defined but many individuals had missing calls the clusters were adjusted manually to include missing individuals and increase the call frequency. Markers were set as missing in cases where there were more than three clusters (AA, AB, BB) or individuals were not well-clustered. In total, I removed 1,158 of the 4,852 markers on the array leaving 3,654 markers. Since I'm not sure of everyone's goal for these data, I did not do any filtering based on MAF for missingness.

2 File Description

In this folder you should find:

1. Cornell2019JanninkMetabolomics-Oat6K: Raw data files for GenomeStudio provided by Jason
2. OatBiparentalGenotypingArrays_filteredMC_3-14-20: GenomeStudio files after filtering.

3. CleanedMrks.txt: Matrix of marker genotypes in the Illumina format (AA, AB, BB). This file lists *all* markers, missing or not.

Let me know if you have any questions or comments campbell.malachy@gmail.com.