III International Symposium on Molecular Markers in Horticulture

Riva del Garda (Trento)

Italy,

FONDAZIONE EDMUND MACH

POSTER - Accurate SNP-calling with the Illumina Infinium 8K apple SNP-chip

Corresponding Author: Stijn Vanderzande (stijn.vanderzande@biw.kuleuven.be)

Name	Surname	Insititution	Address	Country
Vanderzande	Stijn	KULeuven	Willem de Croylaan 42, Heverlee	Belgium
Van de weg	Eric	Wageningen University and Research Centre	Droevendaalsesteeg 1, Wageningen	The Netherlands
Troggio	Michela	Foundation Edmund Mach	Via E. Mach 1, 38010 - S. Michele all'Adige	Italy
Keulemans	Wannes	KULeuven	Willem de Croylaan 42, Heverlee	Belgium
Davey	Mark	KULeuven	Willem de Croylaan 42, Heverlee	Belgium

Abstract: Introduction: Recently an Illumina Infinium 8K apple SNP-chip was developed by the International RosBREED SNP Consortium for high-throughput genotyping and association studies (Chagne et al, 2012). We have used this SNP-chip to genotype 144 individuals of a *Malus* x *domestica* germplasm collection consisting of old Belgian and commercial cultivars, as well as 128 individual progeny of a mapping population.
Aims: The genotype data generated was used i) to evaluate the overall performance of the chip and the robustness of the SNP calling both within a group of related individuals, as well as across a wider germplasm and ii) to develop a set of filter parameters to allow accurate automatic calling of the correct SNP genotype.
Results: Overall only 9% of the SNPs failed completely. However, when comparing the genotypes of cultivars common to two different datasets, it was clear that differences in population structure between the datasets strongly influences SNP-calling, due to difficulties in clustering fluorescent intensities into the different genotype classes. This in turn is due to the presence of homologous sequences that also hybridize to the chip probes used for the SNP genotyping. This leads to shifts in the observed fluorescence used for genotype calling (AA, AB, BB) and shifts in the allele frequencies depending on the number of homologous loci and the degree of sequence identity. In order to remove those SNPs that show differences between datasets, a bootstrap analysis was done and cut-off values for a range of quality parameters commonly used for filtering SNP-chip data, were established using the variability of the callings over the bootstrap populations.
Secondly a visual inspection of the intensity clustering used to genotype each SNP was carried out. This showed that even after automatic filtering, some SNPs showed deviations from the expected distributions due to the presence of the paralogous sequences. The fluorescence plot of these SNPs showed shifted clusters, the presence of additional genotype groups in the plot or a combination of both making it often hard to determine the true genotypes for the different individuals. In order to remove SNPs for which the genotype calling was influenced by the presence of paralogous sequences in the genome, two additional parameters were introduced: one for the cluster position and one for the cluster width since the additional groups were combined into the existing clusters leading to wider clusters.
Depending on the stringency of the applied cut-off values for the filter parameters, the number of SNPs that are retained varies between 2,000and 3,500 SNPs. With manual inspection, the proportion of SNPs that were considered to be accurately called rose from 45% without filtering to 85-90.6% depending on the stringency of the filter parameters.
Reference: Chagné, D., *et al*. (2012). Genome-wide SNP detection, validation, and development of an 8K SNP array for apple. PLoS ONE 7(2): e31745.

Keywords: *Malus* x *domestica,* SNP-genotyping, Germplasm, SNP-calling, SNP-chip

Type of presentation: Poster

Topics:

Germplasm collections and molecular characterisation
Young researchers mini-symposium (phD student or under 32 years old)