Unlocking the Power of Haplotype-Based Molecular Breeding using Long Read DNA Sequencing

Alexander H.J. Wittenberg*, René Hogers, Koen Nijbroek, Esther Verstege, Rui Peng Wang, Sevgin Demirci, Harrie Schneiders, Anker Sørensen and Dick Roelofs
* presenting author (Alexander.Wittenberg@keygene.com)

Keygene N.V., Agro Business Park 90, 6708 PW, Wageningen, The Netherlands

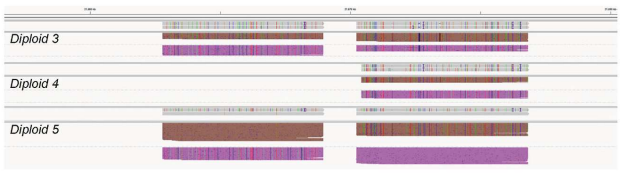
Introduction

The availability of high-quality, contiguous reference genomes is essential for effective marker development and gene discovery, associated with important crop breeding traits. The early adoption of long read sequencing technologies (e.g. Oxford Nanopore Technologies, ONT) has become an essential driver of innovation by KeyGene. Still, whole genome sequencing of large populations is not cost effective, especially when studying large and complex genomes. Therefore, KeyGene has developed and/or implemented innovations to reduce genome complexity in a random (e.g. Haplotype Based Sequencing[1], KeyGene® HaBSeq) or targeted manner (e.g. Adaptive Sequencing). Because these applications are long read sequencing-based, they can comprehensively detect all variation (from SNPs to SVs) in the context of genetic linkage, thereby opening possibilities to perform more precise haplotype-based molecular breeding contributing to crop improvement. Moreover, variation in DNA modifications, proven to be associated with important agricultural traits such as fruit ripening, can now be instantly revealed.

Adaptive sequencing

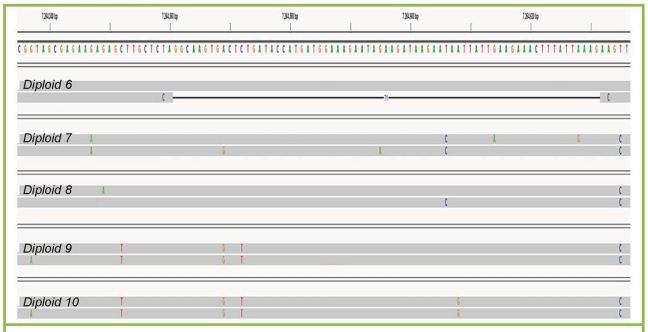
Adaptive sequencing applied in banana for enrichment of a QTL region. Haplotypes identified are indicated above the sorted reads.

KeyGene® HaBSeq



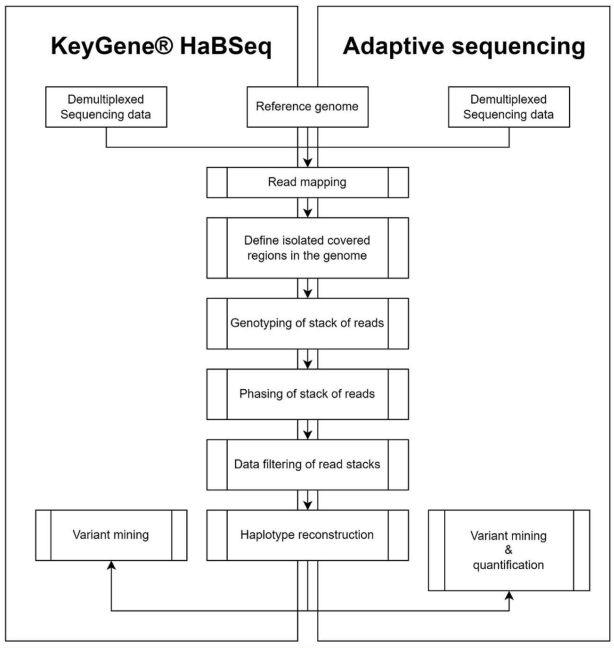
KeyGene® HaBSeq applied in three banana samples. Fragments of 5-7kb were selected, and haplotypes identified for each sample individually are indicated above the sorted reads.

Haplotype-phased consensus



HaBSeq applied in five banana samples and zoomed in at nucleotide level. Haplotypes with homozygous and heterozygous alleles are indicated.

Haplotype analysis workflow



Outline of the bioinformatics analysis workflow for haplotype generation

Impact of haplotype-based genotyping

- *Comprehensive variation detection: SNP, SV/CNV, methylation*
- *Identity by descent (more precise population structure prediction)*
- *Improved validation of variation due to haplotype context*
- *More informative in multi-parental mapping populations (e.g. MAGIC populations)*
- *More precise/unique mapping to reference genome locations*
- *KeyGene® HaBSeq enables better informed decisions of parent selection in breeding/selection schemes*

Conclusions

- *Analysis workflow developed for identification of random and/or targeted haplotypes ranging from ~ 1 kb up to several Mbs in size*
- *QTL regions can be resolved at haplotype-level, including elucidation of all variation in complex regions*