## 1.1 DNA extraction and sequencing

Samples were sent to Diversity Arrays Technology Pty. Ltd. (DArT; Canberra, Australia - http://www.diversityarrays.com) for DNA extraction and sequencing. DNA was extracted using the GeneCatchTM Blood and Tissue Genomic Mini Prep Kit (Epoch Life Science, Inc) following the manufacturer guidelines. SNP discovery was performed for each sub-sample using the standard DArTseq protocol (DArT 2018). DArTseq is a genotype-by-sequencing method (Sansaloni et al. 2011; Kilian et al. 2012) able to perform genome-wide marker discovery using the Illumina next-generation sequencing (NGS) platform (Andrews et al. 2016; Zhang et al. 2018). The DArTseq protocol is explained in brief below.

To ensure the quality of all genomic DNA, template DNA was incubated in a 1X solution of Multi-CoreTM restriction enzyme buffer (Promega) for 2 hours at 37°C. Approximately 100ng per µL of each DNA sample was then digested with a combination of the two restriction enzymes *Pstl* and *Sxphl.* Each individual sample was ligated to unique barcodes and adapters specific to these enzymes. PCR amplification of each sample followed using primers specific to the barcode and adaptor sequences used. PCR conditions consisted of 1 min initial denaturation at 94 °C, followed by 30 cycles of 20 s denaturation (94 °C), 30 s annealing (58 °C) and 45 s extension (72 °C), and a final extension of 7 min at 72 °C. To prepare for hybridization to the flow cell, approximately 10 µL of each sample were pooled, diluted and denatured using NaOH. The subsequent library was sequenced on an Illumina HiSeq®2500 platform for 77 cycles, resulting in 77 base pair (bp) long fragments (single read). 20% of the samples were processed a second time following the preceding protocol to create a set of technical replicates that were used to later asses the reproducibility of SNP calls.

Raw genetic sequences were converted to fastq format using the Illumina HiSeq2500 platform. Individual samples were then demultiplexed based on their unique ligated barcode. All remaining reads were checked for contamination using GenBank viral and bacterial example sequences alongside a database curated by DArT to ensure for quality control of reads.

## 1.2 Generation of SNP dataset

Single nucleotide polymorphism (SNPs) were identified and called using in DArT proprietary pipeline DArTSoft14TM (Diversity Arrays Technology) standard procedures. Monomorphic clusters were removed retaining only homozygous and heterozygous forms. Filtering for SNPs with average ratio of read depths between alleles of 0.75, reproducibility average of >90%, and a minimum read depth of 5.

**References**

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA. 2016. Harnessing the power of RADseq for ecological and evolutionary genomics. Nature Reviews Genetics **17**:81.

DArT. 2018. DNA Extraction Protocol for DArT, Available from <https://ordering.diversityarrays.com/files/DArT_DNA_isolation.pdf> (accessed 8.11.18 2018).

Kilian A, Wenzl P, Huttner E, Carling J, Xia L, Blois H, Caig V, Heller-Uszynska K, Jaccoud D, Hopper C. 2012. Diversity arrays technology: a generic genome profiling technology on open platforms. Pages 67-89. Data production and analysis in population genomics. Springer.

Sansaloni C, Petroli C, Jaccoud D, Carling J, Detering F, Grattapaglia D, Kilian A. 2011. Diversity Arrays Technology (DArT) and next-generation sequencing combined: genome-wide, high throughput, highly informative genotyping for molecular breeding of Eucalyptus. Page P54. BMC proceedings. BioMed Central.

Zhang L, Yang X, Qi X, Guo C, Jing Z. 2018. Characterizing the transcriptome and microsatellite markers for almond (Amygdalus communis L.) using the Illumina sequencing platform. Hereditas **155**:14.