Specialty Wood Products Research Partnership

SCION FORESTS ▪ PRODUCTS ▪ INNOVATION

Douglas-fir breeding and genomics, Specialty Wood Products Research aligned SSIF

Authors: Dagmar Goeke, Natalie Graham, Emily Telfer, Jaroslav Klápště, Mari Suontama and Heidi Dungey

Date: June 2018

Publication No: SWP-T058

FOREST GROWERS RESEARCH

TABLE OF CONTENTS

supported by forestgrowers commodity levy

FOREST GROWERS RESEARCH

INTRODUCTION

Douglas-fir research in the SWP programme has taken a huge step towards testing genomics for implementation of genomic selection methodology in the breeding programme. Genomics research has been funded through the Scion Strategic Investment Funds (SSIF). The testing of genomic selection (GS) in a Douglas-fir breeding programme was initiated as part of the Specialty Wood Products Partnership Programme (SWP), to explore the potential for genomic tools to increase Douglas-fir competitiveness for the New Zealand forest industry. The project was initiated by a revision of the current breeding programme and the selection of suitable field experiments to create a robust genomic selection training population (Klápště et al. 2017).

The theoretical main advantage of genomic selection is to utilise genomic marker based relationship matrices in genetic evaluation which enhances gains by more accurate selections. Another advantage is that when adequately robust genomic based models are available, selection can be based solely on markers. This enables selection without phenotypes and the progeny testing phase can be skipped, leading to a shorter generation interval and increased genetic gains per unit of time. A robust genomic marker resource where the genotyping of breeding population candidates is based on is a prerequisite for the successful implementation of marker-based selection. We have obtained collaborative access to use the SNP resource (Single Nucleotide Polymorphism) developed by the Oregon State University in order for genotyping to take place.

As a further research for SNP quality assurance, we will be testing haploid material from Douglas-fir seed that can potentially provide additional knowledge of SNP markers. Conifers are large and complex organisms and can have large gene families. Therefore, when developing or utilising a SNP resource for the first time, inclusion of haploid material provides assurance that SNPs are biologically real and not a composites of pseudo markers from multiple highly homologous loci.

The first milestone in this task is to initiate genotyping of training population individuals. The second milestone in this task is to optimise DNA extraction of haploid megagametophyte tissue from Douglas-fir seed and the development of DNA markers suitable for distinguishing haploid and diploid material.

MATERIALS AND METHODS

Milestone 1: Genotyping training populations

Genomic marker data was generated by an external service provider, the commercial genotyping facility GeneSeek (a Neogen Corporation company). A previously developed Douglas-fir Axiom® SNP array was used for genotyping, with 58,350 SNP markers made available by the Oregon State University. The training population covers a total number of 2112 individuals from the 1996 breeding trials at Kaingaroa and Gowan Hill, for which cambial tissue was collected in 2017. DNA was extracted in batches of 96, as per Klápště et al. (2017), and quantified **fluorometrically using the Quant-iT™ PicoGreen® dsDNA assay kit (Invitrogen Ltd.) and a BMG POLARStar Galaxy microplate reader (BMG Lab-Technologies, Offenberg, GER).** All 2112 samples (22 plates of 96 samples each) were dried down and a minimum of 250 ng DNA per sample shipped to the GeneSeek laboratory (Nebraska, USA), with the first set for genotyping containing 1152 individuals (12 plates). The remaining 960 individuals will be genotyped during the financial year 2018/19.

Milestone 2: Extraction of haploid DNA from megagametohpytes

DNA extraction from Douglas-fir megagametophyte tissue
Megagametophyte (maternal haploid (1n)) tissue from Douglas-fir seeds was separated from the diploid (2n) embryo, the seed coat and the nucellus (Fig 1). Megagametophyte tissue was stored at -20°C until cell lysis and DNA extraction.

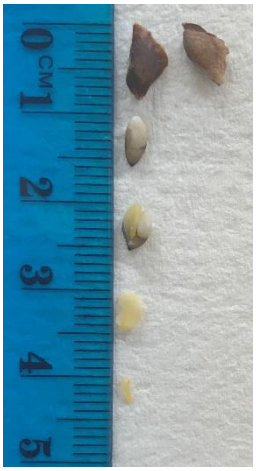

Figure 1: Stepwise separation of Douglas-fir megagametophyte tissue from other seed components

DNA was extracted from a single megagametophyte per reaction. Several DNA extraction methods were tested and optimised:

1) CTAB (cetyl trimethylammonium bromide) extraction protocol (DNA clean up via precipitation; Doyle and Doyle, 1990 and Telfer et al., 2013);
2) Macherey-Nagel (MN) NucleoSpin® Plant II 96 kit (96-well format) (PL1 lysis buffer or PL2 lysis buffer; DNA clean up through silica-membrane filter, vacuum based; Macherey-Nagel, Düren, Germany) following manufacturer's instructions;
3) MN NucleoSpin® Plant II kit (individual columns) (PL1 lysis buffer or PL2 lysis buffer; DNA clean up through silica membrane filter, centrifugation based) following manufacturer's instructions;
4) MN NucleoMag® Plant kit (DNA clean up magnetic bead based) following manufacturer's instructions.

DNA quantity was determined **fluorometrically using the Quant-iT™ PicoGreen® dsDNA assay kit (Invitrogen Ltd.) and the BMG POLARStar Galaxy microplate reader (BMG Lab-Technologies, Offenberg, GER), or the** *Quant-iT*™ Broad-Range *dsDNA Assay Kit and a Qubit. Testing of DNA purity was* performed with a NanoDrop™1000 (Thermo Scientific, Karlsruhe, Germany) by measuring absorbance ratios of 260 nm/280 nm (A260/280) and 260 nm /230 nm (A260/230). A ratio of ~1.8 for A260/280 is generally considered pure for DNA, while A260/230 values should be in the range of 2.0-2.2. Low A260/280 or A260/230 ratios indicate the presence of contaminants (ThermoSCIENTIFIC, T042-TECHNICAL BULLETIN).

Seed lots for testing:

- Plus tree 888_432 and 888_434 (Longmile Rd, Rotorua, New Zealand) stored at Scion's nursery – collected in 1990s (Kane Fleet, pers. comm.); seeds are small and most are dried out, some mouldy.
- Ernslaw One Ltd, Tapanui, Tramway, ESO 2013 – 200 seeds received 16 November 2017, parental origin unknown; seeds appear fresh and healthy.

DNA extraction method variations:

- Homogenisation methods (bead-beatingin lysis buffer at room temperature versus manual disruption in liquid nitrogen),
- CTAB method optimisation: lysis temperature, centrifugation speed, centrifugation temperature, centrifugation time, reconstitution procedure of DNA pellet.
- MN NucleoSpin® Plant II kit PL1 lysis buffer versus PL2/3 lysis buffer,
- MN NucleoSpin® Plant II kit (i) 96well vacuum filtration versus (ii) single column centrifugation versus (iii) magnetic bead isolation of DNA versus (iv) CTAB.

DNA extraction methods and results were captured on FreezerPro (Sample Management Software acquired by Brooks Automation) and unique IDs assigned.

Proof of haploid status for megagametophyte DNA

Three different categories of markers were tested for their suitability to confirm that the genomic DNA obtained from the haploid megagametophyte tissue was not contaminated by diploid embryo, seed coat or nucellus tissue. These included:

1) eight microsatellite marker gene regions (SSR, simple sequence repeats; Slavov et al., 2004),
2) the nuclear gene region ITS (two internal transcribed spacers - ITS1 and ITS2 - that flank the 5.8s nuclear ribosomal DNA region of plant genomes; White et al, 1990), and
3) nine SNP gene regions (single nucleotide polymorphism; Krutovsky and Neale, 2005)

Polymerase chain reactions (PCRs) for all gene regions tested were conducted and PCR products separated on agarose gels. PCR products of successfully amplified gene regions were sequenced in-house at Scion, Rotorua. Raw sequence data was cleaned and aligned with those obtained from Genbank (NCBI, *National Center for Biotechnology Information,* https://www.ncbi.nlm.nih.gov/).

RESULTS AND DISCUSSION

Milestone 1: Genotyping training population

The proportion of samples that passed all the DNA quality thresholds for genotyping was high, 98.6% (1136/1152 samples), with 16 failed samples. Genotyping data is stored in Q:\Forest Genetics\aaaClients\FOA Diverse Forests\SWP 17-18\A02050 D.-fir genomics\Genotyping. The total number of SNP's used in genotyping was 58,350.
The average call rate of SNPs in samples passing QC was 99.5%, out of which ~45% were polymorphic SNP markers (Conversion types of PolyHighResolution and NoMinorHom included in the polymorphic marker class) that can be used for genomic prediction of breeding values (Table 1). All default SNP calling metrics were applied.

Table 1. SNP metrics summary from Geneseek/Neogen.

ConversionType[1]	Count	Percentage
PolyHighResolution	20278	34.752
Other	13718	23.51
MonoHighResolution	12801	21.938
NoMinorHom	5929	10.161
CallRateBelowThreshold	5036	8.631
OTV	588	1.008

[1] https://biobank.ctsu.ox.ac.uk/crystal/docs/axiom_geno_analguide.pdf provides an explanation of the SNP metrics

The remaining 960 individuals will be genotyped during the financial year 2018/19.

Milestone 2: Extraction of haploid DNA from megagametohpytes

DNA extraction from megagametophyte tissue

In general, negligible amounts of DNA were recovered from apparently older and dehydrated megagametophyte tissue, in particular after homogenisation in liquid nitrogen or when using magnetic bead based isolation (MN NucleoMag® Plant kit).

Highest DNA yields were achieved by using the younger and fresher seeds (mostly available from seed lot received from Ernslaw One Ltd) combined with a modified protocol of the standard CTAB DNA extraction method.

In comparison, the modified CTAB method resulted in 3-4 times higher yields than the best performing MN NucleoSpin® Plant II protocol.

DNA purities (A260/280 and A260/230 ratios) obtained for DNA extracted using both the optimised CTAB protocol and the adapted MN NucleoSpin® Plant II kits are sufficient for downstream PCR based applications. However, DNA extractions obtained using the CTAB method were observed to be unstable through multiple freeze-thawing cycles, displaying rapid deterioration and becoming un-amplifiable, possibly due to the presence of residual contaminants. Dependent on the requirements of the downstream application, the MN NucleoSpin® Plant II 96 well kit can provide sufficient DNA yields and may be the kit of choice for high sample numbers and to provide more stable DNA extracts.

Proof of haploid status in megagametophyte DNA

Seven DNA samples were used to identify the most suitable method to prove that DNA extracts of megagametophyte tissue are not contaminated with diploid embryo, seed coat or nucellus DNA:

- one putatively diploid DNA sample obtained from cambium tissue,
- three putatively diploid DNA samples obtained from embryo tissue,
- three putatively haploid DNA samples obtained from megagametophyte tissue.

Sequencing analysis of the **Douglas-fir SSR marker** regions resulted in unclean, unreproducible chromatograms which were therefore not diagnostic for ploidy status. For successful SSR diagnostics, PCR products require generally fluorescent labelling using the M13F-tag on the 5' end of the forward primer and subsequent separation on a genetic analyser (Schuelke, 2000; Abdelkrim et al., 2009). The final step cannot be performed in-house at this stage. If other methods are not satisfactory to confirm the ploidy status of DNA extracted from megagametophyte tissue, genotyping can be performed externally but genotyping costs and turnaround time for genotyping have to be adjusted in the budget accordingly.

Amplification of the **nuclear ITS gene region** of the seven test samples inconsistently resulted in multiple PCR products. Sequencing reactions of isolated PCR products were not successful. Hence, this method is not suitable for determination of the ploidy of Douglas-fir DNA extracts.

Four **Douglas-fir SNP gene regions** were tested initially (lot I) and successfully amplified. For one SNP gene region, no ambiguities were identified for the seven test samples. For the second SNP gene region, cambium and embryo samples show several ambiguities and indel (insertion/deletion) variation between strands, while the three DNA extracts from megagametophyte tissue resulted in no ambiguities or indel variation as anticipated. For the remaining two SNP regions, sequencing reactions failed.

The second, potentially diagnostic SNP gene region was tested on further 42 DNA extracts obtained from cambium tissue, as well as repeated on the seven test samples to show reproducibility. All samples were successfully amplified and sequenced, repeat sequencing results were consistent, and only 7 out of the total 46 DNA extracts (from cambium and embryo tissue) showed no ambiguities and no indel variation between the two strands of the diploid plant genome. Based on this sample set, this indicates ~15% potentially false negative results could be expected when using this SNP region to confirm haploid status of megagametophyte DNA.

To improve the threshold for detecting potentially false negatives for haploid DNA extracts, another five SNP gene regions (lot II) were tested on the seven test samples. Amplification was successful for the cambium DNA extract but not for the other six test samples which were obtained from embryo and megagametophyte tissue by CTAB-based extraction methods. It has been observed before in-house that DNA extracts from megagametophyte tissue using the CTAB method produces less stable DNA extracts which deteriorate over time and are thus not suitable for long-term storage or multiple freeze-thaw cycles (Emily Telfer, Forest Genetics, Scion, pers. comm.) but the exact parameters are unclear. Hence, the DNA extracts test set was changed to DNA extracts solely obtained by using the Macherey-Nagel Nucleospin plant II kit.

The one diagnostic SNP gene region from test lot I and the five additional SNP gene regions (lot II) were successfully amplified and sequenced for twelve DNA extracts from cambium tissue (seven cambium samples that showed no ambiguities and indel variation for the diagnostic SNP gene region from test lot (I), another four cambium DNA extracts randomly picked and one repeat sample as control). For every SNP region tested, at least one ambiguity and/or indel variation was detected between the two DNA strands in at least a subset of the samples, the composition of which differed with different SNP regions. For the three most diagnostic SNP gene regions in combination, the test DNA extracts from cambium showed at least one ambiguity over a total of 72 known SNPs, 6 known indels, and a total 1877 bp to 1904 bp sequenced. This reduces the

likelihood of “potentially false negatives” when testing DNA extracts from megagametophyte tissue for contamination with diploid DNA to close to 0%.

16 DNA extracts obtained from megagametophyte tissue using the Macherey-Nagel Nucleospin plant II kit were tested for contamination with diploid DNA by sequencing the three most diagnostic SNP gene regions. Samples were successfully amplified and sequenced and showed no ambiguities or indel variation indicating that no contamination of diploid DNA is present.

For one SNP gene region from test lot I, sequencing resulted in 21 additional SNPs compared to GenBank sequence data. For the other two most diagnostic SNP gene regions from test lot II, sequencing resulted in three additional SNPs and nine additional SNPs, as well as one 1-bp insertion, respectively, in comparison to GenBank sequence data.

RECOMMENDATIONS AND CONCLUSIONS

Milestone 1: Genotyping training populations

Generally, the DNA quality for genotyping from cambium samples was good with only 16 samples out of 1152 failed. Genotyping results are promising with only a small number of missing data and the SNP information for polymorphic markers was good as well. The second set of genotyping will be undertaken next financial year. Analyses to fit quantitative genetic models to predicted genomic breeding values will be undertaken 2018/19 to the data sets and compared with pedigree-based models.

Milestone 2: Extraction of haploid DNA from megagametohpytes

Improved of DNA extraction yield using the Macherey-Nagel Nucleospin plant II kit:
We have successfully shown that DNA of suitable quantities and quality can be extracted from Douglas-fir megagametophytes, using the methods optimised in this study. This offers a means by which to obtain haploid DNA for this species, allowing for markers to be screened for their specificity to a single locus. This is highly beneficial in complex conifer species where genome duplication and a high degree of repetitive sequence content increase the risk for spurious SNP genotyping calls, and the detection and elimination of such SNP markers is critical.

Further improvements to the recommended DNA extraction method are still possible. For instance, in another study, DNA extractions using the Macherey-Nagel Nucleospin plant II kit from *Pinus radiata* megagametophyte tissue has shown increased yields when using freshly isolated megagametophyte tissue, omitting the freezing step and proceeding immediately with cell lysis and DNA clean-up. Hence, using freshly isolated megagametophyte tissue from Douglas-fir should also be tested for possible improvements to DNA yields.

Ploidy of DNA extracts from megagametophyte tissue:
We have identified a number of SNP gene regions that have proved informative in our sample set with regards to assigning diploid or haploid status for a Douglas-fir DNA sample. This is important when using haploid DNA to identify non-specific DNA markers, as there is always a risk that haploid samples are contaminated with diploid DNA, particularly when diploid seed components are small and removed manually. Sequencing of two or three of these gene regions provides enough information on polymorphism to reliably assign haploid/diploid status of a DNA sample, based on the frequencies observed in the samples tested in this study. A wider sampling of Douglas-fir in New Zealand would further increase the confidence in this method.

Combination of three SNP gene regions tested on 16 samples achieved a potentially false negative threshold of 0%. Two further SNP gene regions may be used additionally or to replace one of the three SNP regions if required.

Overall, sufficient yields of DNA were extracted from Douglas-fir megagametophyte tissue and potential contamination by diploid tissue can be excluded successfully by sequencing two to three SNP gene regions.

During 2018-19, we intend to include a subset of DNA extractions (up to 96) obtained from haploid D.-fir megagametophyte tissue with known parentage to the genotyping panel to confirm confidence in SNP marker information. We recommend sequencing all DNA extracts obtained from Douglas-fir megagametophyte tissue using the SNP gene regions described in this report before samples are used for genotyping to ensure that DNA extracts are free from diploid DNA contamination.

ACKNOWLEDGEMENTS

We would like to thank you the Oregon State University for access to the SNP markers for genotyping and Ernslaw One providing seed for megagametophytes. Scion SSIF funding supported the work in this report. Thanks to the SWP FGR/MBIE funding for the continued support of alternative species research.

REFERENCES

Abdelkrim, J., B. C. Robertson, J. A. L. Stanton, and N. J. Gemmell. 2009. Fast, cost-effective development of species-specific microsatellite markers by genomic sequencing. BioTechniques 46: 185–191.

Doyle J. and Doyle J., 1990. Isolation of plant DNA from fresh tissue. Focus 12: 13–15.

FreezerPro® (7.2.99-r14002). Frederick, Maryland, USA: RuRo, Inc.

Klápšteˇ, J., Bradford, K-T., Graham, N.G., Telfer, E.J., Goeke, D., Suontama, M. and Dungey, H.S. 2017. Initiation of Genomic Selection Research, Collection of Cambium and Extraction of DNA from a Douglas-fir Breeding Programme. Scion Core SWP aligned Funding Technical Report.

Krutovsky K. V. and Neale D. B., 2005. Nucleotide diversity and linkage disequilibrium in cold hardiness and wood quality related candidate genes in Douglas-fir. Genetics 171:2029-2041.

Schuelke, M. 2000. An economic method for the fluorescent labeling of PCR fragments. Nature Biotechnology 18: 233–234.

Slavov, G. T., Howe, G. T., Yakovlev, I., Edwards, K. J., Krutovskii, K. V., Tuskan, G. A., Carlson, J. E., Strauss, S. H., and Adams, W. T., 2004. Highly variable SSR markers in Douglas-fir: Mendelian inheritance and map locations. Theor. Appl. Genet. 108: 873–880.

Telfer E., Graham N., Stanbra L., Manley S. T., and Wilcox P. 2013. Extraction of high purity genomic DNA from pine for use with a high-throughput genotyping platform. New Zealand Journal of Forestry Science 43:3.

ThermoSCIENTIFIC, T042-TECHNICAL BULLETIN
http://www.nhm.ac.uk/content/dam/nhmwww/our-science/dpts-facilities-staff/Coreresearchlabs/nanodrop.pdf

White, T. J., Bruns T., Lee S., and J. Taylor., 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p. 315–322. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (ed.), PCR protocols: a guide to methods and applications. Academic Press, Inc., New York.