Publicly available

Initiation of genomic selection research, collection of cambium and extraction of DNA from a Douglas-fir breeding programme

Jaroslav Klápště, Ki-Taurangi Bradford, Natalie Graham, Emily Telfer, Dagmar Goeke, Mari Suontama, and Heidi Dungey

FGR Ref: SWP-T034

Report information sheet

Report title	**INITIATION OF GENOMIC SELECTION RESEARCH, COLLECTION OF CAMBIUM AND EXTRACTION OF DNA FROM A DOUGLAS-FIR BREEDING PROGRAMME**
Authors	**JAROSLAV KLÁPŠTĚ, KI-TAURANGI BRADFORD, NATALIE GRAHAM, EMILY TELFER, DAGMAR GOEKE, MARI SUONTAMA, AND HEIDI DUNGEY**
Client	MBIE/SWP
Client contract number	0000-00
MBIE contract number	If applicable
SIDNEY output number	
ISBN Number	
Signed off by	Heidi Dungey
Date	June 2017
Confidentiality requirement	Confidential (for client use only)
Intellectual property	
Disclaimer	The information and opinions provided in the Report have been prepared for the Client and its specified purposes. Accordingly, any person other than the Client uses the information and opinions in this report entirely at its own risk. The Report has been provided in good faith and on the basis that reasonable endeavours have been made to be accurate and not misleading and to exercise reasonable care, skill and judgment in providing such information and opinions. Neither Scion, nor any of its employees, officers, contractors, agents or other persons acting on its behalf or under its control accepts any responsibility or liability in respect of any information or opinions provided in this Report.

Published by: Scion, 49 Sala Street, Private Bag 3020, Rotorua 3046, New Zealand. www.scionresearch.com

Executive summary

The implementation of genomic selection (GS) in a Douglas-fir breeding programme has been initiated within the Specialty Wood 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 revision of the current breeding programme and selection of suitable field experiments to create a robust genomic selection training population.

The field experiments identified as suitable as a genomic selection training population, cover most of the genetic diversity of Douglas-fir deployed in New Zealand forestry. The progeny tests established in 1996 were defined as the most suitable, and phenotyping and collection of samples for DNA extraction was performed during the current financial year (16/17). The other material to be incorporated into the training population will be discussed further after consultation with industry partners.

DNA was extracted from cambium due to inaccessibility of needles with a DNA extraction protocol optimized to achieve DNA quality and quantity sufficient for genotyping through next generation sequencing platforms.

Initiation of genomic selection research, collection of cambium and extraction of DNA from a Douglas-fir breeding programme

Table of contents

Introduction

The implementation of genomic selection (GS) in a Douglas-fir breeding programme was initiated within the Specialty Wood 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 revision of the current breeding programme and the selection of suitable field experiments to create a robust genomic selection training population.

Genomic selection was initially proposed and successfully implemented in animal breeding (Meuwissen, Hayes, & Goddard, 2001;Goddard, Hayes, & Meuwissen, 2010; Solberg, Sonesson, Woolliams, Ødegard, & Meuwissen, 2009) and was also tested in forest trees (Grattapaglia & Resende, 2010; Resende et al., 2012; Gamal El-Dien et al., 2015; Ratcliffe et al., 2015; Bartholomé et al., 2016). The successful implementation of genomic selection depends on a trait's heritability, the size of training population and the effective number of genomic fragments (Hayes, Visscher, & Goddard, 2009). The generally large size of conifer genomes, high genetic diversity and fast decay in linkage disequilibrium requires dense and equally distributed genetic markers along the genome (Neale & Savolainen, 2004;Neale & Kremer, 2011). The development of next generation sequencing platforms has led to the rise of genotyping-by-sequencing (Elshire et al., 2011) or exome capture protocols (Neves, Davis, Barbazuk, & Kirst, 2013), which has allowed the implementation of genomics in forest trees. Implementation of genomics was also, in part, due to the considerable reduction in the need for sequencing of the whole genome in order to genotype an individual tree, which made it cost-effective.

The implementation of genetic markers allows researchers to construct the so called realized relationship matrix (VanRaden, 2008), which allows for tracking of not only temporary relatedness defined by pedigree, but also historical relatedness (Powell, Visscher, & Goddard, 2010). This benefit is extremely important and useful, especially in forest trees where the process of domestication is still in its initial stages. Therefore, we can precisely capture not only family means and Mendelian segregation effects available in family based populations (Visscher et al., 2006), but also provenance effects. These will be especially important in traits under natural selection that are regarded as adaptive traits.

Our main objective was to establish a training population that covered a broad range of the species' genetic diversity and allowing us to capture the best material for New Zealand's planting conditions. One challenge that needed to be overcome was the advanced age of the field experiments included in genomic selection training population. The height of the material precluded the collection of needles or buds for DNA extraction, and cambium via bark windows, was collected instead. Sampling the cambium required adaptation and modification of the standard DNA extraction protocols commonly used for DNA extraction from needles.

Materials and Results

Determining the genomic selection training population

A review of the current Douglas-fir breeding programme was performed to identify a set of suitable field experiments with solid phenotypic data available and the possibility to sample for DNA extraction. A major aim of the task was to find material that captured as much genetic diversity as possible, which is important for building a robust genomic prediction model. Our major attention was focused on a provenance/progeny test established as a collection of open-pollinated seed from original wild stands in Oregon and California that represented a broad resource of genetic diversity. The second important resource identified was the collection of 185 clones originating from provenance tests established at 1957, 1959, 1971 and 1974 capturing provenances from Washington, Oregon and California. Several progeny tests based on the collection of those 185 clones had been established in the past. However, the control-pollinated progeny trial established in 1999, including 70 seed lots, was compromised by unfortunate environmental factors (waterlogging) and survival of 66% was estimated (Low et al. 2015). Alternative training population options were progeny tests established in 2002 and 2009, however these have been transformed to pasture. Therefore, only a relatively small clonal test, containing 12 families and established in 2006 and 2007, was considered as part of a training population. Additionally, the seed stand established in 1981 (established within what was formerly called “superline B”) was selected as the third part of the genomic selection training population. This population was established from provenances in Fort Bragg and from California, and also NZ land races. Unfortunately, the phenotypic data were very limited and the pedigree was missing. Additional phenotyping and pedigree/sib-ship reconstruction would be required in this trial.

The GS training population was defined (Figure 1) and phenotyping of first two field experiments (Kaingaroa and Gowan Hill) initiated (Klápště et al. 2017). Due to financial limitations, only a subset of field tested individuals was considered for genotyping. The training population proposed included all open-pollinated families coming from the original wild stands that covered the Washington, Oregon and Californian provenances. The number of families included in the field experiments was 213 and each family was represented by 5 randomly selected individuals per site. This random sampling method allows us to capture the most genetic diversity, which is essential for construction of reliable GS prediction models. In total, 1125 and 1021 individuals were selection at Gowan Hill and Kaingaroa, respectively.

Due to the advanced developmental stage of the field tested trees, sampling of needles or buds was impossible or considered to be too dangerous, therefore, cambium was selected as the source of genomic DNA in our project. Cambium was collected by Scion’s field crew at Kaingaroa during February-March 2017 and at Gowan Hill in April 2017. The timing of cambium collection was critical and should be carried out at the time when cambium is actively developing to provide sufficient source material for DNA extraction. Two samples were collected for each individual to ensure the enough quantity tissue was available for DNA extraction. The samples were stored and refrigerated until the start of DNA extraction process. An optimization of the DNA extraction protocol was performed before DNA extraction from collected samples started to ensure the extracted DNA will be in sufficient quality and quantity for genotyping.

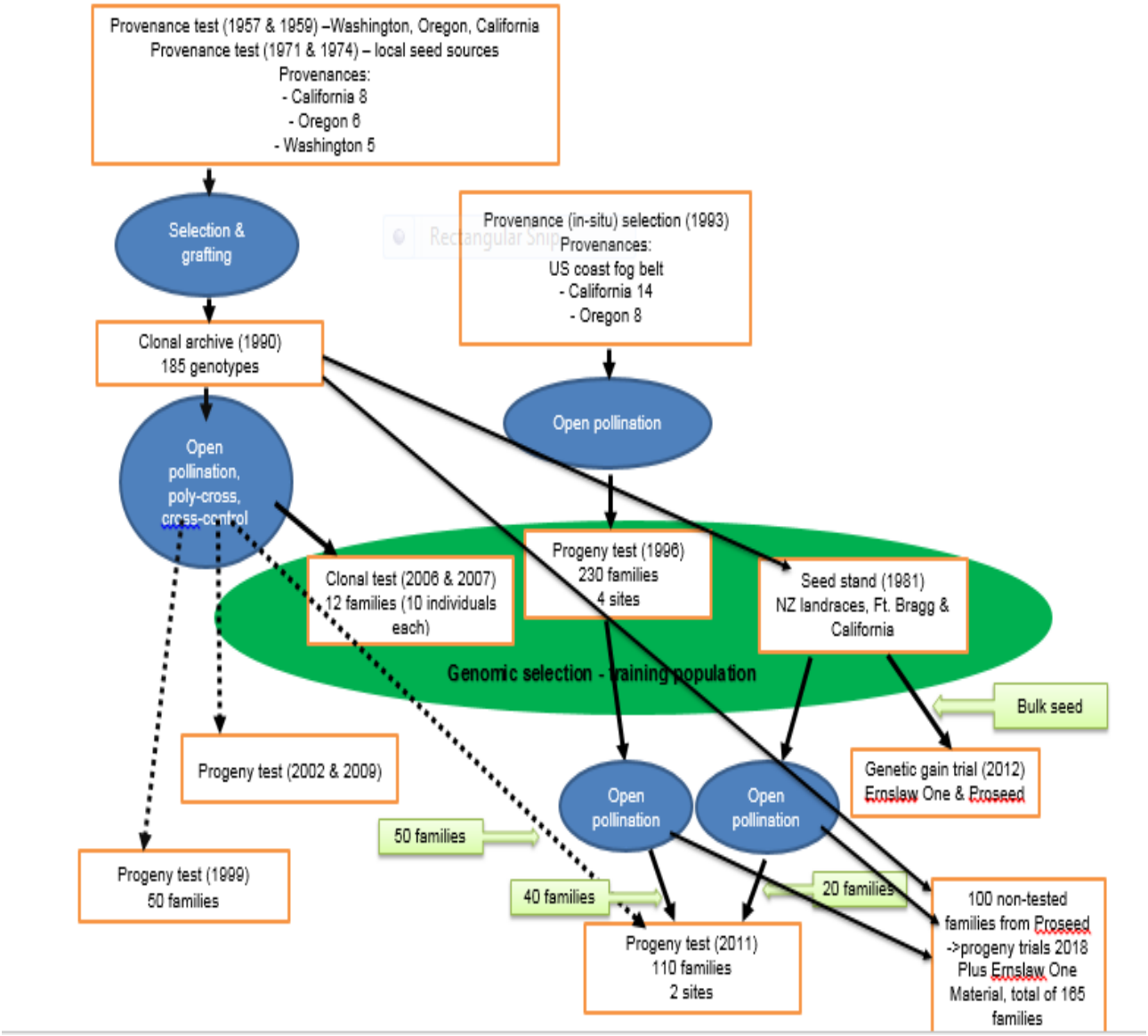

Figure 1. Definition of GS training population in the context of Douglas-fir breeding programme

DNA extraction protocol optimization

A pilot study to optimize the DNA extraction protocol was conducted on 91 Douglas-fir samples from the Long Mile archive on Scion grounds. The aim was to investigate which adjustments, if any, were required to ensure DNA quality and quantity were suitable for downstream genotyping applications.

General observations:

1. The DNA yields had a 78% success rate across the entire plate, which is similar to a standard (needle) extraction from this protocol. The average yield was 14.56 ng/ul per sample. We required a minimum of 5ng/ul per sample to be viable for downstream analysis.

Table 1: DNA Quantification sheet for pilot plate (ng/ul) (thawed sample concentrations in yellow)

		1	2	3	4	5	6	7	8	9	10	11	12
A	1000	0.756	8.816	0.268	4.333	20.263	16.976	16.397	8.482	13.983	29.384	20.563	29.911
B	750	14.135	10.744	12.952	0.527	2.246	9.962	24.109	16.437	17.570	14.499	12.655	34.737
C	500	53.852	14.003	8.408	14.799	2.445	7.528	18.813	2.961	12.460	17.138	2.314	33.936
D	200	5.637	5.020	-0.163	13.703	12.830	1.663	5.310	26.910	11.924	13.912	17.674	9.586
E	50	27.597	15.729	33.641	9.921	15.463	2.836	30.257	27.736	4.296	17.256	19.359	11.760
F	10	5.428	10.501	5.361	20.421	20.232	13.366	6.143	27.803	-0.203	8.576	18.483	7.960
G	0	17.950	3.756	10.589	21.655	19.413	7.050	3.918	4.521	-4.373	25.963	4.329	68.904
H		30.479	10.794	3.456	-0.089	23.563	12.840	7.012	19.369	14.462	6.524	7.559	57.093

Table 2: Plate layout for pilot plate (thawed samples highlighted in yellow**)**

	1	2	3	4	5	6	7	8	9	10	11	12
A	889.618	889.557	889.539	889.56	889.582	889.606	889.533	889.546	889.565	888.423	888.435	889.569
B	889.602	889.537	889.62	889.554	889.575	889.597	889.526	889.543	889.564	888.42	888.417	888.427
C	889.601	889.619	889.613	889.54	889.568	889.591	889.61	889.532	889.545	888.437	888.434	889.617
D	889.585	889.587	889.599	889.621	889.548	889.567	889.592	889.608	889.531	888.419	888.433	EN91
E	889.586	889.614	889.604	889.535	889.561	889.576	889.607	889.527	889.544	888.424	888.416	EN91
F	889.572	889.584	889.589	889.612	889.534	889.562	889.557	889.595	889.528	888.436	888.428	EN91
G	889.571	889.573	889.583	889.605	889.525	889.547	889.566	889.579	889.53	888.425	888.414	D.-fir
H	889.558	889.57	889.574	889.59	889.611	889.542	889.563	889.578	889.529	888.418	888.413	D.-fir

2. We observed that a layer of ice formed on the samples that proved hard to remove without also removing some of the cambium at the same time. It was essential to remove the ice layer before pre-weighing the wet-weight of each sample. We divided the samples between at 4°C and -20°C to ascertain the impact of storage temperature on yield and quality. The thawed (4°C) samples yielded an average concentration of 16.9ng/ul with 89% of the samples having over 5ng/ul, but proved hard to scrape the cambium when the material had thawed and hardened or dried out. The frozen (-20°C) material yielded an average concentration of 10.6ng/ul with 73% of the samples having over 5ng/ul and were easier to scrape. From these results, we proceeded to remove samples out of the freezer and let them thaw for approximately 30 minutes before scraping the material. This allows the ice to melt off without the sample hardening.

3. We had difficulty with differentiated wood being removed alongside the cambium. In the majority of samples with no DNA, this was the issue. We observed that it was possible to remove a layer of wood to reveals a layer of cambium underneath. This greatly minimised bad yields.

4. Adjustments were made to the protocol to suit the nature of the cambial tissue that was available.

5. As a result of these changes, we are now able to achieve a 98% success rate for sample extractions that yield in excess of 5ng/ul of DNA.

DNA extraction of the collected samples was completed. A list of selected sample can be found here:

Q:\Forest Genetics\aaaClients\FOA Diverse Forests\SWP 16-17\A02050 GS on D-fir breeding\phenotype files\GH_dna2017.csv

Q:\Forest Genetics\aaaClients\FOA Diverse Forests\SWP 16-17\A02050 GS on D-fir breeding\phenotype files\K_dna2017.csv

Recommendations and conclusions

The development of a Douglas-fir breeding programme suitable for testing a genomics-supported breeding approach, was initiated this year by searching for a suitable set of field experiments. The most suitable field experiments, which should capture the most species' genetic diversity, were identified and included into proposed genomic selection training population. A portion of the field experiments were measured to obtain the most up to date phenotypes (Klápště et al. 2017) and cambium was sampled from randomly selected individuals within every tested family. Each family is represented by 5 individuals at each site. In total 1125 individuals were selected at Gowan Hill site and 1021 individuals were selected at Kaingaroa site established within 1996 provenance/progeny trail.

The use cambium as a source of genomic DNA in Douglas-fir, required testing and modification of the DNA extraction protocol. The lab staff tested the current DNA protocols and performed modifications, which resulted in improvement of DNA quantity and quality required for genotyping through next generation sequencing platforms.

Acknowledgements

We would like to thank Anastasia Rickard for her great help and support to complete DNA extraction of the selected sample on time.

REFERENCES

Bartholomé, J., Van Heerwaarden, J., Isik, F., Boury, C., Vidal, M., Plomion, C., & Bouffier, L. (2016). Performance of genomic prediction within and across generations in maritime pine. *BMC Genomics*, *17*(1), 604.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, *6*(5), e19379.

Gamal El-Dien, O., Ratcliffe, B., Klápště, J., Chen, C., Porth, I., & El-Kassaby, Y. A. (2015). Prediction accuracies for growth and wood attributes of interior spruce in space using genotyping-by-sequencing. *BMC Genomics*, *16*(1).

Goddard, M. E., Hayes, B. J., & Meuwissen, T. H. E. (2010). Genomic selection in livestock populations. *Genetics Research*, *92*(5–6), 413–421.

Grattapaglia, D., & Resende, M. D. V. (2010). Genomic selection in forest tree breeding. *Tree Genetics & Genomes*, *7*(2), 241–255.

Hayes, B. J., Visscher, P. M., & Goddard, M. E. (2009). Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research*, *91*(1), 47–60.

Klápště, J., Suontama, M., Low, C.B., Stovold, G T, Miller, M. A., Fleet, K. and Dungey, H.S. (2017) Phenotypic assessment and quantitative genetic analysis of two Douglas-fir progeny tests. SWP Technical Report A02046 - 2.1.5

Low, C.B., Suontama, M., Miller, M.A., Fleet, K.R. (2015). Results of assessments in FR375, Beaumont controlled-pollinated 1999 progeny trial. SWP Technical Note.

Meuwissen, T. H. E., Hayes, B. J., & Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, *157*(4), 1819–1829.

Neale, D. B., & Kremer, A. (2011). Forest tree genomics: growing resources and applications. *Nature Reviews Genetics*, *12*(2), 111–122.

Neale, D. B., & Savolainen, O. (2004). Association genetics of complex traits in conifers. *Trends in Plant Science*, *9*(7), 325–30.

Neves, L. G., Davis, J. M., Barbazuk, W. B., & Kirst, M. (2013). Whole-exome targeted sequencing of the uncharacterized pine genome. *The Plant Journal : For Cell and Molecular Biology*, *75*(1), 146–56.

Powell, J. E., Visscher, P. M., & Goddard, M. E. (2010). Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics*, *11*(11), 800–805.

Ratcliffe, B., El-Dien, O. G., Klápště, J., Porth, I., Chen, C., Jaquish, B., & El-Kassaby, Y. A. (2015). A comparison of genomic selection models across time in interior spruce (Picea engelmannii$\times$ glauca) using unordered SNP imputation methods. *Heredity*, *115*(6), 547–555.

Resende, M. F. R., Munoz, P., Acosta, J. J., Peter, G. F., Davis, J. M., Grattapaglia, D., … Kirst, M. (2012). Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist*, *193*(3), 617–624.

Solberg, T. R., Sonesson, A. K., Woolliams, J. A., Ødegard, J., & Meuwissen, T. H. E. (2009). Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. *Genetics Selection Evolution*, *41*(1), 53.

VanRaden, P. M. (2008). Efficient methods to compute genomic predictions. *Journal of Dairy Science*, *91*(11), 4414–4423.

Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K. I., Zhu, G., Cornes, B. K., … Martin, N. G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet*, *2*(3), e41.