SMarTForests Newsletter

Edition 4 July 2015

Contents:

Genomic selection and application to spruce breeding – some revealing tools

by Patrick Lenz, Jean Beaulieu and Jean Bousquet

Spruces have enormous ecological and economic importance across Canada. Since the early 1950s, breeding programs have been established to improve spruce planting stock for growth and, more recently, for wood traits (Mullin et al. 2011). Every year, more than 300 million spruce seedlings that originate from improved stock are planted in Canada. However, breeding efforts are often hampered by the costly and time-consuming evaluation of mature traits, such as wood quality, borne by the slow growth of boreal spruce species. Part of the SMarTForests project's mandate was to address the problems of the long duration in spruce breeding and the associated costs. The rise of genomics tools during the last 10 years and the discovery of prolific genetic markers at reasonable cost offer a set of new selection approaches to tree breeders.

Genetic association studies: We have conducted association studies (correlating genetic markers to traits) in prior work under the he Arborea II project and have extended the study to the entire genome as a genome-wide association study (GWAS) in our current SMarTForests project. Beaulieu et al (2011) reported on a number of significant associations between gene SNPs (single nucleotide polymorphisms) and wood quality traits. Statistical issues, namely correction for multiple testing and non-linearity of marker effects, led to small numbers of significant associations, even though quantitative traits, such as growth and wood quality, are controlled by large numbers of loci (Beaulieu et al. 2011, Pelgas et al. 2011). Consequently, associated SNP markers only explained a small percentage of trait variation. This is a major drawback for the application of association genetics in tree improvement, where breeders seek to predict traits on which selections are performed. Nevertheless, association studies are useful in discovering key genes and developing an understanding of the molecular control of complex traits.

Genomic selection – markers applied to tree breeding: In an applied tree breeding context, the low number of associated markers hinders the exact identification of superior individuals and led us to experiment with genomic selection (Meuwissen et al. 2001, Grattapaglia and Resende 2011). In order to predict the future genetic worth of trees at the seedling stage, we used genomic profiles of candidate trees obtained from several thousand markers and a multilocus model correlating phenotypic to genotypic information. We have conducted pilot studies of genomic selection in both white spruce and in black spruce (see Box 1) in the SMarTForests project.

Mission and Project Goals

The SMarTForests project builds on a decade of research discoveries in spruce genomics by previous projects: Arborea (Université Laval) and Treenomix (University of British Columbia). The new team combines the strengths of the two previous projects, building on their extensive background knowledge and experiences. Our mission is to break new ground in spruce genome sequencing and strongly represent Canada in international conifer genome initiatives, and to achieve efficient translation of results toward end-users from across Canada.

The SMarTForests project has three major goals:

1. Develop marker systems to aid in MAS.
2. Sequence the white spruce genome.
3. Analyze impacts of forest genome on economics and society.

Box 1

Development of genotyping resources in black spruce (*Picea mariana*): Scientists have developed considerable genomic resources for white spruce (*Picea glauca*) during the past 10 years. However, there has been a lack of genomic information for black spruce (*Picea mariana*), a sympatric species to white spruce and of high ecological and economic importance in Eastern Canada. Prior to using genomic selection in black spruce, we needed to develop a catalogue of SNPs (Single Nucleotide Polymorphisms) presentative of the black spruce transcriptome. To this end, we compiled SNPs from previous studies on white spruce (Pavy et al. 2008, Prunier et al. 2012, Prunier et al. 2013, Pavy et al. 2013a) and discovered new SNPs using the genomic tool called **exome capture**. Two capture approaches (solid phase capture with 454 GS-FLX-Titanium sequencing and liquid-phase capture with Illumina HiSeq sequencing) were used with *Picea glauca* designed probes (Rigault et al. 2011) and generated more than 680M sequences. Our bioinformatic analyses allowed us to identify more than 97K high-confidence SNPs from 21K gene sequence contigs, with a 96% success rate in genotyping that exceeded the 92% rate obtained for white spruce (Pavy et al. 2013b). Nearly 5000 SNPs were assembled on an Illumina Infinium iSelect array, which is now being used for genotyping progeny for genomic selection experiments in black spruce.

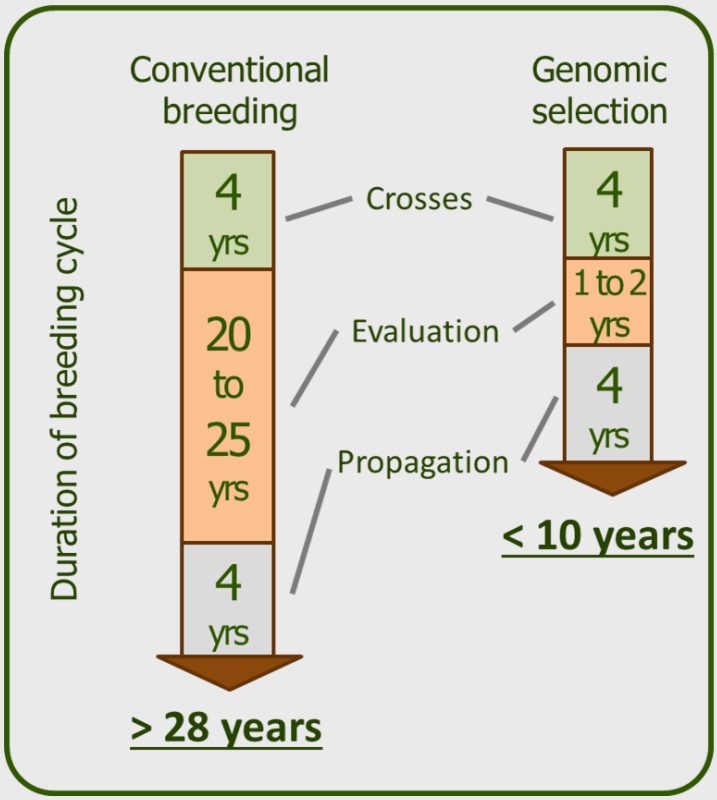
A first study in white spruce (Beaulieu et al. 2014a) was based on 1,694 trees from 215 half-sib families that came from different seed sources across Québec. Two other studies in white spruce (Beaulieu et al. 2014b) and black spruce (Lenz et al., in prep.) were performed with advanced-breeding populations using 1,748 and 734 trees from 59 and 34 controlled crosses established on various environmental sites. We genotyped trees in the three studies with 6,385, 6,932 and 4,993 SNP markers, respectively by using the Infinium iSelect genotyping platform at Genome Quebec Innovation Centre. The growth and wood traits determined in all three studies were chosen for their high interest for conifer breeders. Wood traits such as wood density and cellulose microfibril angle were determined from increment cores. These traits are good indicators of mechanical strength and stiffness of wood; both traits are highly sought after by forest industries.

Precision of genomic prediction models: The precision of prediction models is usually determined using cross-validation methods with independent trees and the accuracy is estimated through the comparison of the predicted genetic value with the known genetic value of an individual in the study populations. The studies based on controlled crosses in white and black spruce produced genomic selection models with high accuracies and correlations between the predicted and the known genetic value varied from 0.70 to 0.80. Accuracies of marker based models were thereby comparable (>90%) to accuracies achieved through conventional selection. This means that wood and growth traits could be effectively improved through genomic selection which can already be applied at the seedling stage. The accuracy of genomic selection models for the white spruce study based on half-sib families was moderate (as may be expected) and correlations ranged between 0.33 and 0.44. Marker-based models were somewhat less accurate than pedigree models and reached on average 90% of the predicted improvement achievable through conventional pedigree-based selection.

Our results underline how genomic selection can be more advantageous when using advanced-generation controlled crosses. At this stage, the genetic value of progeny can be predicted with greater precision, as indicated by the differences in accuracy between results obtained from half-sib families versus controlled crosses. There is an additional advantage: when building genomic selection models for different test sites, we found the overall accuracy was maintained, especially for wood quality traits. This indicates that little genotype-by-environment interaction for these species and consequently allows the application of genomic selection models across different environments.

The role of parentage for genomic prediction: The genetic structure for model training and the choice of the validation sets is crucial from both an experimental point of view and from a practical one. In the standard scenario, individuals used for model training and individuals used for model validation came from the same crosses or families. It is noteworthy that model accuracy will drop dramatically, if the model is validated with individuals are from a different lineage than individuals used for the model construction. This is expected, given that much of the predictive power of genomic selection at the current marker density is based on long-range linkage disequilibrium and relatedness. However, building genomic selection models encompassing different breeding groups could be possible with only a small loss in accuracy.

Figure 1: Time needed to complete a breeding cycle



Ongoing work: We are currently investigating the optimal number of markers and individuals necessary for the construction of genomic selection models in advanced-breeding populations. This will help balance the need for precision in our models against the cost for large breeding populations comprising of thousands of individuals. So far, our findings indicate that the number of markers and thus the genotyping cost may be reduced without a significant loss of model accuracy, further indicating that much of the accuracy is due to both long-range linkage disequilibrium and the strength of relatedness existing between trees of the training and testing populations.

Implications for conifer breeding: Our results highlight the power of genomic selection in tree breeding, especially for small advanced-breeding populations. The primary advantage is the possibility of early forward selection on thousands of candidate trees very early in the breeding cycle, since marker data can be obtained at the seedling stage. There is no longer a need to wait until mature traits can be evaluated, once genomic selection models are calibrated. Expensive phenotyping of all individuals will become obsolete. The length of the selection cycles can be reduced by three fold (see Figure 1) from ~ 28 years to ~ 10 years and the genetic gain per year that can be achieved through marker selection is multiplied by a similar ratio, when compared to conventional selection at the mature stages. A secondary advantage for operational application in breeding programs is that selection models can be easily calibrated for other phenotypes that may become important in the future, given that genotypes have already been obtained. Any amount of time reduction in tree breeding would result in cost savings translating to large benefits for tree breeders.

References:

- Beaulieu J, Doerksen T, Boyle B, Clement S, Deslauriers M, Beauseigle S, Blais S, Poulin PL, Lenz P, Caron S, Rigault P, Bicho P, Bousquet J, MacKay J (2011). Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. *Genetics*, 188: 197-214.
- Beaulieu J, Doerksen T, Clément S, MacKay J, Bousquet J (2014a). Accuracy of genomic selection models in a large population of open-pollinated families in white spruce. *Heredity*, 113: 342-352.
- Beaulieu J, Doerksen TK, MacKay J, Rainville A, Bousquet J (2014b). Genomic selection accuracies within and between environments and small breeding groups in white spruce. *BMC Genomics*, 15: 1048.
- Grattapaglia D, Resende MDV (2011). Genomic selection in forest tree breeding. *Tree Genetics & Genomes*, 7:241-255.
- Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157:1819–1829.
- Mullin TJ, Andersson B, Bastien J-C, Beaulieu J, Burdon RD, Dvorak WS, King JN, Kradowski J, Lee SJ, McKeand SE, Pâques L, Raffin A, Russel JH, Skroppa T, Stoehr M, Yanchuk A. (2011). Economic importance, breeding objectives and achievements, Chapter 2. In: Plomion C, Bousquet J, Kole C (eds), Genetics, Genomics and Breeding of Conifers. CRC Press and Science Publishers Inc., New York, pp 40-127.
- Pavy N, Pelgas B, Beauseigle S, Blais S, Gagnon F, Gosselin I, Lamothe M, Isabel N, Bousquet J (2008). Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC Genomics*, 9: 21-38.
- Pavy N, Gagnon F, Rigault P, Blais S, Deschênes A, Boyle B, Beaulieu J, Isabel N, Mackay J, Bousquet J (2013a). Development of high-density SNP genotyping arrays for white spruce (*Picea glauca*) and transferability to subtropical and nordic congeners. *Molecular Ecology Resources*, 13: 324-336.
- Pavy N, Deschênes A, Blais S, Lavigne P, Beaulieu J, Isabel N, Mackay J, Bousquet J (2013b). The landscape of nucleotide polymorphisms among 13,500 genes of the conifer *Picea glauca*, relationships with functions, and comparison with *Medicago truncatula*. *Genome Biology and Evolution*, 5: 1910-1925.
- Pelgas B, Bousquet J, Meirmans PG, Ritland K, Isabel N (2011). QTL mapping in white spruce: gene maps and genomic regions underlying adaptive traits across pedigrees, years and environments. *BMC Genomics*, 12: 145.
- Prunier J, Gérardi S, Laroche J, Beaulieu J, Bousquet J (2012). Parallel and lineage-specific molecular adaptation to climate in boreal black spruce. *Molecular Ecology*, 21: 4270-4286.
- Prunier J, Pelgas B, Gagnon F, Desponts M, Isabel N, Beaulieu J, Bousquet J (2013). Genomic architecture and association genetics of adaptive characters using a candidate SNP approach in boreal black spruce. *BMC Genomics*, 14: 368.
- Rigault P, Boyle B, Lepage P, Cooke JEK, Bousquet J, MacKay JJ (2011). A white spruce gene catalogue for conifer genome analyses. *Plant Physiology*, 157:14-28.

**Ressources Naturelles Canada, Département des sciences du bois et de la forêt, Université Laval*

Data Release:

Type of Data Resource	Description	Citation
White Spruce (*Picea glauca*) PG29 4th draft assembly	Fourth draft assembly of *Picea glauca* (PG29) is available (http://www.ncbi.nlm.nih.gov/bioproject/83435) The data was generated using Illumina Hiseq technology.	Warren, R., Keeling, C.I., Yuen, M.M.S., Raymond, A., Taylor G.A., Vandervalk, B.P., Mohammadi, H., Paulion D., Chiu R., Jackman S.D., Roberson, G., Yang, C., Hoffmann, M., Weigel, D., Nelson, D.R., Ritland, C., Isabel, N., Jaquish, G., Yanchuk, A., Bousquet, J., Jones, S.J., MacKay, J., Birol, I. and Bohlmann J. July 2015. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. The Plant Journal Vol 82 Issue 6 pp.xxx (currently online: http://onlinelibrary.wiley.com/journal/10.1111/%28ISSN%291365-313X/earlyview)
White Spruce (*Picea glauca*) WS77111 1st draft assembly	Revised first draft assembly of *Picea glauca* (WS77111) is available (http://www.ncbi.nlm.nih.gov/bioproject/242552). The data was generated using Illumina Hiseq technology.	Warren, R. Keeling, C.I., Yuen, M.M.S., Raymond, A., Taylor G.A., Vandervalk, B.P., Mohammadi, H., Paulion D., Chiu R., Jackman S.D., Roberson, G., Yang, C., Hoffmann, M., Weigel, D., Nelson, D.R., Ritland, C., Isabel, N., Jaquish, G., Yanchuk, A., Bousquet, J., Jones, S.J., MacKay, J., Birol, I. and Bohlmann J. July 2015. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer terpenoid and phenolic defense metabolism. The Plant Journal Vol 82 Issue 6 pp.xxx (currently online: http://onlinelibrary.wiley.com/journal/10.1111/%28ISSN%291365-313X/earlyview)
Transcript profiling of early and late wood in four conifer species	Expression profiling experiment, GEO accession GSE51884.	Beaulieu,J,. Doerksen, T., Boyle, B., Clément, S., Deslauriers, M., Beauseigle, S., Blais, S., Poulin, P-L., Lena, P., Caron, S., Rigault, P., Bicho, P., Bousquet, J. and MacKay J. May 2011. Association genetics of wood physical traits in the conifer white spruce and relationships with gene expression. Genetics Vol 188 Issue 1 pp. 197-214.
Transcript profiling of seven tissues in *Picea glauca*	Expression profiling experiment, GEO accession GSE60277.	Raherison, E., Rigault, P., Caron, S., Poulin, P-L., Boyle, B., Verta, J-P., Giguère, Is., Bomal, C., Bohlmann, J. and MacKay, J. 2012. Transcriptome profiling in conifers and the PiceaGenExpress database show patterns of diversification within gene families and interspecific conservation in vascular gene expression. BMC Genomics Vol 13 pp. 434-450/
For all other data resources	SMarTForests web site (http://www.smartforests.ca/en-ca/publications.aspx)	

VISIT US AT:
www.smartforests.ca

TO CONTACT US:

PROJECT MANAGERS

Sophie Laviolette
Université Laval
1030 av, de la Médecine
Bureau 2153
Québec (Qc), G1V 0A6
Canada
Phone: 418.656.2408
Fax: 418.656.7176
Email:
Sophie.laviolette@sbf.ulaval.ca

Carol Ritland
#3611-2424 Main Mall
Department of Forest
Sciences
University of British Columbia
Vancouver, B.C. Canada
V6T 1Z4
Phone: 604.822.3908
Fax: 604.822.2114
E-Mail:
critland@mail.ubc.ca

News

End of SMarTForests

The project SMarTForests officially ended June 30, 2015, after four years of research. We thank the members of our team, granting agencies, educational institutions involved as well as all our partners and stakeholders. We are very proud of the work performed and scientific advances of recent years. Our success is a team effort! Follow us in the coming months, new initiatives resulting from the project will be announced soon!

Notable Award

International Bioinformatics Resource Award 2015 awarded by the Swiss Institute of Bioinformatics (SIB) was awarded to Dr. Inanc Birol and his team for the ***ABySS*** (Assembly By Short Sequences) resource.

New articles (for a complete list, please go to www.smartforests.ca)

- Warren, R.L., Kelling, C.I., Yuen, M.M.S., Raymond, A., Taylor, Greg A., Vandervalk, B.P., Mohamadi, H., Paulino, D., Chiu, R., Jackman, S.D., Robertson, G., Yang, C., Boyle, B., Hoffmann, M., Weigel, D., Nelson, D.R., Ritland, C., Isabel, N., Jaquish, B., Yanchuk, A., Bousquet, J., Jones, S.J.M., MacKay, J., Birol, I., Bohlmann, J., 2015. Improved white spruce (*Picea glauca*) genome assemblies and annotation of large gene families of conifer defense metabolism. *The Plant Journal* (http://doi.org/10.1111/tpj.12886)
- Raherison, E.S.M., Giguère, I., Caron, S., Lamara, M., MacKay, J., 2015. Modular organization of the white spruce (Picea glauca) transcriptome reveals functional organization and evolutionary signatures. *New Phytologist* (http://doi.org/10.1111/nph.13343).
- Porth, I., Bull, G., Ahmed, S., El-Kassaby, A.Y., Boyland, M., 2015. Forest Genomics Research and Development in Canada: Priorities for Developing an Economic Framework. *The Forestry Chronicle* (http://doi.org/10.5558/tfc2015-011)
- Mageroy, M.H., Parent, G., Germanos, G., Giguère, I., Delvas, N., Maaroufi, H., Bauce, E., Bohlmann, J., MacKay, J., 2015. Expression of the beta-glucosidase gene PgβGlu-1 underpins natural resistance of white spruce against spruce budworm. *The Plant Journal* (http://doi.org/10.1111/tpj.12699).

Keep visiting our website, many new publications to be revealed over the next months!

Upcoming events

Workshop: 3rd Annual Conifer Genome Summit

September 28-30 2015 in Gysinge, Sweden
To register: http://login3.axaco.se/C12571F400166B19/registrationForm?openagent&unid=4F49A0412A5778A4C1257E1400560261

Special thanks to our sponsors

GenomeCanada
GenomeQuébec
GenomeBritishColumbia
GenomeAlberta