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 multi-locus 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.
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 traits. 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.

**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 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:

<table>
<thead>
<tr>
<th>Type of Data Resource</th>
<th>Description</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>For all other data resources</td>
<td>SMarTForests web site (<a href="http://www.smartforests.ca/en-ca/publications.aspx">http://www.smartforests.ca/en-ca/publications.aspx</a>)</td>
<td></td>
</tr>
</tbody>
</table>
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)


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