Multi-varietal forestry integrating genomic selection and somatic embryogenesis

Yill-Sung Park[1*], Jean Beaulieu[2], Jean Bousquet[3]

[1] Natural Resources Canada, Canadian Forest Service, Canadian Wood Fibre Centre, PO Box 4000 Fredericton, New Brunswick, Canada
[2] Natural Resources Canada, Canadian Forest Service, Canadian Wood Fibre Centre, 1055 du P.E.P.S. Stn Sainte-Foy, POBox 10380 Quebec, QC, Canada (Emeritus)
[3] Canada Research chair in Forest and Environmental Genomics, Institute for Systems and Integrative Biology, Unversité Laval, Quebec, QC, Canada
**Corresponding Author: yillsung.park@canada.ca*

Abstract

Multi-Varietal Forestry (MVF) is defined as the deployment of a range of genetically tested tree varieties in commercial plantation forestry. Somatic embryogenesis (SE) and cryopreservation are the enabling technologies for the implementation of MVF. Recently, it has been shown that genomic selection (GS) has a great potential to be incorporated with MVF. MVF is well suited for intensively managed, high-productivity sites. MVF offers a much greater genetic gain than conventional tree breeding because it captures both additive and non-additive variations. Furthermore, MVF integrated with forward GS and SE eliminate the time required for producing seeds and, thus, gain per unit time is notably increased. In white spruce breeding in eastern Canada, for example, the gain is delivered 15 years sooner than by conventional seed orchard breeding. Moreover, GS will make the testing and selection efforts more efficient and streamlined through pre-screening. Sufficiently refined and efficient SE protocols for commercial MVF are available for a number of conifers, primarily the spruces, some pines, and a few larches, but more refinements are required for several economically important conifers. The main challenge for implementing industrial MVF, however, is the relatively high cost of SE seedling production due to manual handling of embryos, both pre- and post-germination. In order to be cost effective, it requires the development of a mechanized embryo handling system for transplanting into mini-plugs for greenhouse culture, which is under development. However, with the current lack of an automated transplanting system, complementary serial rooting of cuttings may be used as a mass propagation tool

Yill-Sung Park, Jan M Bonga, Heung-Kyu Moon (eds) (2016) Vegetative Propagation of Forest Trees. National Institute of Forest Science(NiFos). Seoul, Korea. pp 302-322

once the suitable varieties are developed from the SE-based system. In addition to obtaining a large genetic gain at a significantly reduced time, MVF offers flexibility to adapt to changing breeding goals, environment, disease and insect conditions, and this flexibility is provided by up-to-date information obtained from varietal field testing. Furthermore, in MVF, plantation diversity is dynamically managed over time by carefully balancing genetic gain and diversity based on the best available varietal field test data.

Keywords: Cryopreservation, deployment strategy, varietal field testing

1. Introduction

Somatic Embryogenesis (SE) is an important biotechnology in conifers for the development and production of tree varieties with desirable traits. The plant production by the SE process in conifers involves the initiation, proliferation, maturation, germination, and greenhouse culture steps with varying degrees of success at each step. Nonetheless, the SE system for several conifers has advanced to the stage where it can be implemented in a large-scale commercial production.

SE is not a tree breeding method, but it is a required biotechnology aiming at the development of tree varieties and their deployment in conjunction with tree breeding. The most important advantage of conifer production by SE is that the embryogenic cell lines can be cryogenically stored in a juvenile state indefinitely, which was not possible with other vegetative propagation techniques for trees. This allows for the long-term field testing and for the subsequent selection and retrieval of tested cell lines for mass propagation. This led to the operational implementation of Multi-Varietal Forestry (MVF) in eastern Canada by JD Irving Limited and the Quebec Ministry of Forests, Wildlife, and Parks.

The purpose of this paper is to review the successes and issues of implementing SE and how novel biotechnologies such as genomic selection can be integrated with SE into modern tree breeding systems.

2. Somatic embryogenesis

Since the first successful report of SE in conifers (Nagamani and Bonga 1985, Hakmann et al. 1985, Chalupa 1985), rapid progress has been made, particularly for most spruce species and some pine species. In these species, SE is initiated efficiently from immature zygotic embryos. In spruce species, SE can be obtained from mature zygotic embryos, but, in pines, SE from mature seed has met limited success. Several media formulations were successfully used including mLV (Litvay et al. 1985), DCR (Gupta and Durzan 1985), MSG (Becwar et al. 1990),

and these formula typically contain auxin and cytokinin, such as 2,4-D and BAP. The proliferation of embryogenic tissue is usually accomplished either on solid or in liquid medium of the same formulation. The initiation of SE in many conifers is influenced by additive genetic variation offering a possibility of breeding for increased SE initiation (Park 2002).

Maturation of somatic embryos is achieved by removing auxin and cytokinin and supplying ABA. In addition to the use of ABA, it was discovered that a critical factor for developing a large numbers of somatic embryos was the restriction of water availability either by physical, or osmotic, or both means. The most commonly used methods are the use of high molecular weight PEG and increased gel strength. The quality of mature somatic embryo is very important as it affects germination rates and somatic seedling quality. This is the most important but challenging step, because maturation success is widely variable from total recalcitrance to abundance.

Germination of somatic embryos is usually carried out on a semi-solid medium without growth regulators. Normal germination and zygotic-like development are common provided that mature somatic embryos are well formed and vigorous. This is the step linking the automated transplanting and greenhouse culture. With a lack of an automated system, the current transplanting process is a manual process, consequently time consuming and expensive. Thus, in order to be cost-effective, the development of a mechanized somatic seedling transplant system, or direct germination of somatic embryos into micro-plugs, or their incorporation into artificial seed is highly desirable. Despite these challenges, SE of many conifers, most of spruce and some pine species, is sufficiently refined to the point that it can be used in industrial production.

3. Cryopreservation

Cryopreservation is the key element of conifer SE programs that makes long-term storage of embryogenic tissue at an ultra-low temperature possible while lengthy field testing of cell line is being carried out. For most species, cryopreservation is a routine with an excellent recovery rate, using rather simple procedures. The current protocol entails incubating EM with sorbitol in liquid maintenance medium. Then, the cooled cell suspension, with added DMSO, is dispensed into cryo-vials, which are placed in an alcohol-insulated freezing container (Nalgene®). The freezing containers are pre-cooled and placed at -80 ºC for 1-2 hours, where slow cooling takes place. Subsequently the vials are immersed into liquid nitrogen (-140 ºC to -196 ºC). The recovery of EM involves a rapid thawing in water at 37 C for 1-2 minutes, then the EM suspension is poured over a filter-paper disk allowing the drainage of storage solutions, and placing of the disk with EM onto the semi-solid proliferation medium for regrowth.

The genetic stability of cryopreserved cell lines has been studied in various species (Cyr et al. 1994; DeVerno et al. 1999; Sutton and Plonenko 1999), showing no evidence of somaclonal variation. Harvengt et al. (2001) found no allelic difference, nor abnormal growth behavior, among *Picea abies* plants raised from somatic embryos obtained from up to 3-year-old plants and their ortets. However, a high mutation rate was detected during the *in vitro* phase. Nonetheless, owing to an effective selection for normally formed somatic embryos, the resulting plantlets were all normal. Cryopreservation of conifer cell lines is already used commercially (Cyr 1999).

Given the success of cryogenic storage for conifers, the production of identical genotypes consistently over time without somaclonal variation or loss of juvenility is now possible, which is analogous to the development of agronomic and horticultural varieties. Somatic embryogenesis in combination with cryopreservation offers the means for forward selection and mass producing tree varieties after the varietal field testing of an appropriate length has determined which cell lines have the desirable attributes. The development of tree varieties in conifers was not possible previously.

4. Conventional tree breeding

Conventional tree breeding typically employs a form of recurrent selection, and the production of genetically improved material is accomplished by wind pollinated seed orchards (White 1987). This procedure, for each generation, involves the formation of multiple breeding populations, controlled pollinations among parents within the breeding population, establishing, maintaining, and evaluating the progeny test at multiple sites, and the establishment of clonal seed orchards for the production improved seed, while the selected parents form a new breeding population for the next cycle of breeding. Therefore, tree breeding programs require extensive resources and an extended period of time. Also, the establishment of land-based seed orchards is expensive and remains fixed and inflexible until the establishment of the new next generation orchards. However, these orchards will deliver substantially increased productivity.

A typical breeding cycle using a subset of a breeding population for white spruce (*Picea glauca*) is illustrated in Figure 1 and takes about 15 years to complete. This is primarily due to time required to attain the flowering maturity needed to allow breeding. The time can be shortened by the use of stimulants but this has limitations. This seed orchard-based tree breeding scheme typically produces about 10% volume increase per generation (Fullarton 2015).

The most commonly used conventional genetic evaluation is based on the mixed linear model using pedigree information:

$$\mathbf{y} = \boldsymbol{Xb} + \boldsymbol{Zu} + \boldsymbol{e} \qquad (1)$$

where **y** is the vector of observed phenotype (trait); *X* and *Z* are known design matrices of fixed and random effects, respectively; ***b*** is an unknown parameter of

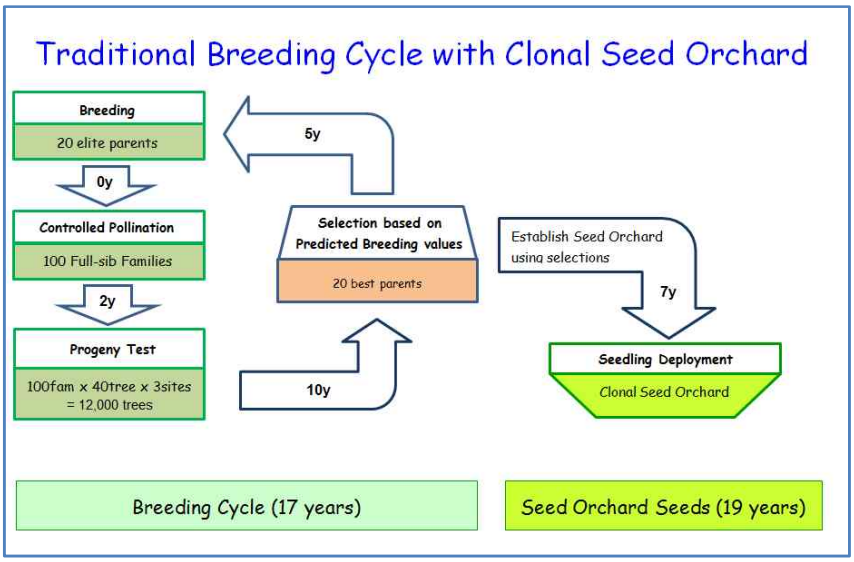

Figure 1. *Schematic presentation of white spruce breeding in New Brunswick, Canada: A white spruce breeding population is divided into 20-tree sub-populations and this example uses an elite sub-population thereof. Controlled-pollinations are performed within this elite sub-population to produce full-sib families (typically 100 families) that are planted in the field tests at multiple locations. When the test is about 10 years old, growth is assessed and the 20 best parents are selected to form the next generation breeding population. Since the flowering maturity for white spruce is about 15 years, it takes 17 years to begin the next generation breeding cycle. The genetic gain from this breeding cycle is delivered through the establishment of a clonal seed orchard using grafts of selected parents. Due to flowering constraints it takes 19 years after the beginning of the breeding cycle before appreciable amounts of seeds are produced.*

fixed values to be estimated; and ***u*** and ***e*** are vectors of breeding values (random) to be predicted and residuals, respectively, such that E(***u***)= E(***e***)=0, Var (***u***)=$\boldsymbol{A}\sigma^2_A$ and Var (***e***)= $\boldsymbol{I}\sigma^2_e$, where ***A*** is the numerator relationship matrix based on genes identical by descent, σ^2_A is additive genetic variance and σ^2_e error variance. The Best Linear Unbiased Prediction of ***u***, known as A-BLUP, is obtained by replacing the variance-covariance matrix of the individual trees in the mixed linear model by the ***A*** matrix (Henderson 1975). These breeding values are the genetic merits of the individuals, and thus the highest values should be used for selection. Alternatively,

the breeding values (BV) of an individual *i* in a population based on the narrow-sense heritability (h^2) is defined as:

$$\mathbf{BV_i = m_0 + h^2 (y_i - m_0) = m_0 + V_A/V_P (y_i - m_0)} \quad (2)$$

where y_i is the phenotypic value of individual *i*, m_0 is the population mean. The estimated narrow-sense heritability (h^2) is computed as the ratio of the estimates of additive variance (V_A) to total phenotypic variance (V_P) from the analysis model. The BV predicted in this manner is referred to as estimated BV (EBV), while the BV predicted by the genome-wide markers will be referred to genomic EBV (GEBV) (see below).

5. Molecular markers, marker-aided selection, and genomic selection

In the past 20 years, there has been a rapid development in marker technologies, and the availability of inexpensive molecular markers offers a possibility of using them to improve the efficiency of tree breeding. Various classes of DNA markers, such as Simple Sequence Repeats (microsatellites), Single Nucleotide Polymorphisms (SNP), Diversity Arrays Technology (DArT), Genotyping-By-Sequencing (GBS), and Restriction site associated DNA (RAD) have been developed for commercially important species such as spruces and pines (Pavy et al. 2013a; Liu et al. 2014; Neves et al. 2014), eucalypts (Sansaloni et al. 2010; Silva-Junior et al. 2015), and poplars (Schilling et al. 2014), among others. High-throughput genotyping technologies were also developed by companies such as Sequenom Inc. (San Diego, Ca, USA), Illumina Inc. (San Diego, Ca, USA) and Affymetrix (Santa Clara, Ca, USA). Thus, depending on availability of markers for a given species, a large number of individuals can be genotyped for a few dozen of DNA markers to many thousands of them. For species like eucalypts, a flexible multi-species genome-wide 60K SNP genotyping chip is available (Silva-Junior et al. 2015) for any genotyping purpose, while for other species, custom DNA chips must still be designed and built for specific needs using the DNA marker information that is available on public domain databases (Pavy et al. 2013b; Pavy et al. 2015).

Application of DNA marker technology in breeding covers two main areas: population management and selection. A wise use of molecular markers in the context of population management is in the pedigree reconstruction proposed by El Kassaby and Listiburek (2009), where they could reconstruct a full pedigree from the open-pollinated seed of a lodgepole pine seed orchard through genotyping using DNA microsatellite markers. When such pedigree reconstruction is implemented in breeding populations, it will circumvent the expensive controlled pollination step and the resulting pedigree can more inclusively of all available

cross combinations. The genetic evaluation of the progeny from the reconstructed pedigree can be carried out in the usual manner.

The use of molecular markers for selection in breeding was initially focused on marker-aided selection (MAS). The mapping of quantitative trait loci (QTL) and candidate gene association approaches have been explored to relate gene architecture and trait expressions, i.e., based on the presumption that causative mutations underlying genetic variation can be localized with DNA markers. The concept of MAS entails that if the QTL associated to a given trait is identified with corresponding molecular markers, they could be used to select superior genotypes in the breeding population. The general process of MAS consists of two phases, training and selection phases. In the training phase, phenotypes in the mapping population are investigated to identify significant associations of a phenotype with marker genotypes using statistical procedures and identify MAS markers for use in the selection phase. In the selection phase, genotyping is necessary for the targeted region of the quantitative trait of interest to screen for MAS markers and selection. However, QTL mapping and candidate gene association approaches in forest trees have not been used widely, primarily due to the fact that the most important traits are controlled by many QTLs, each with only a small effect, and because only a limited portion of the existing variation in a given trait can be explained by the several associations or QTLs detected (Beaulieu et al. 2011, Pelgas et al. 2011).

Genomic selection (GS) or genome-wide selection is a form of MAS; however, it is distinctly different from the traditional MAS based on QTLs. Indeed, GS aims to trace all the QTLs controlling an individual's phenotype and simultaneously estimate all marker effects across the entire genome to calculate its genomic estimated breeding value (GEBV). If the marker coverage is sufficiently dense, all the QTLs controlling the phenotype should theoretically be in linkage disequilibrium (LD) with at least one marker, and unlike the QTL-based MAS, prior information on the association between the phenotype and markers, and on the effects of QTLs is not necessary. However, GS also consists of two phases. A model to predict GEBV is first developed with a training population using both genotypic and phenotypic data. In the ensuing selection phase, only genome-wide genotypic data are needed to obtain GEBVs using the prediction model developed in the training phase. The selection is then based on the GEBVs. The stages of genomic selection are illustrated in Figure 2.

Various statistical methods have been developed for GS and they can be classified in two main groups (de los Campos et al. 2013). The first is based on the idea of Meuwissen et al. (2001) that it is possible to predict the genetic value of individuals by regressing phenotypes on all available markers using a regression model. However, because the number of available markers generally exceeds the number of individuals of the training population, variable selection or shrinkage

estimation procedures are required. Since then, several shrinkage estimation methods, using Bayesian estimation procedures, have been proposed to address this issue, such as ridge regression (RR) (Hoerl and Kennard 1970) and the least absolute angle and selection operator (LASSO) (Tibshirani 1996). The second group uses genomic relationships derived from markers in a mixed model framework to predict the genomic breeding values of individuals. Thus, contrary to the methods of the first group, the effects of individual markers are not estimated,

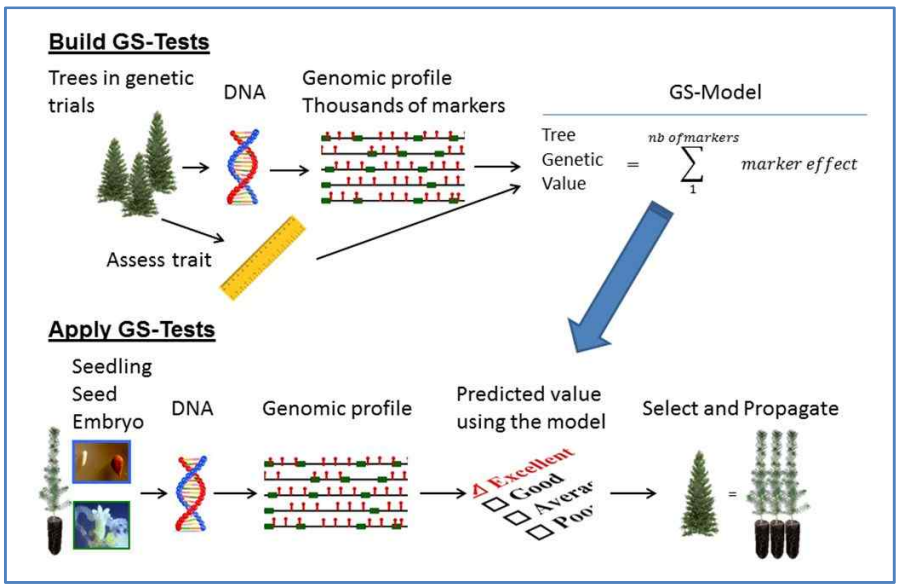

Figure 2. *The process of genomic selection and application in tree breeding: Genomic selection (GS) involves two phases, the development of a GS model in the training population and the application of a validated GS model to the breeding population. In forestry, training populations can be a well-established existing genetic test plantation or can be taken from a selection plantation. In the training population, both phenotyping (traits) and genotyping (e.g., SNP makers) are required to build a GS model. In the GS model, phenotype is typically considered as the sum of all marker effects and is validated by using a subset of the training population. The breeding population (BP) is an offspring population of the training population and the selection is to be made from this BP. In the BP, only genotype data are required to calculate the genomic estimated breeding value (GEBV) using the GS model developed from the training population. The best GEBV individuals are selected to form the next generation breeding population and are used to establish a seed orchard; however, in "Forward GS", the selections are mass propagated using SE or rooting of cuttings for immediate deployment. Thus, vegetative propagation techniques such as SE and/or rooting of cuttings are required to mass produce selections without sexual recombination.*

although they can be obtained with extra calculation. This method is usually referred to as Genomic Best Linear Unbiased Prediction or simply G-BLUP, and can be used in the context of an additive infinitesimal model in which the standard pedigree-based numerator relationship matrix is replaced with a marker-based estimate of additive relationships (Van Raden 2008, de los Campos et al. 2013).

In multiple-marker regression, many markers are simultaneously estimated as random effects in an individual tree model:

$$\mathbf{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e} \qquad (3)$$

y is the vector of observed phenotype (trait); *X* and *Z* are known design matrices of fixed and random effects, respectively; $\boldsymbol{b}$ is an unknown parameter of fixed values to be estimated (including an overall mean and population structure); and $\boldsymbol{a}$ is a vector of random marker effects with the *n* x *m* incidence matrix containing marker covariates coded as Z_{ik} = (0, 1, 2) so that the sum of marker effects approximates the individual (additive) genomic estimated breeding value (GEBV) $\hat{g}_i = \sum_{k=1}^{m} z_k \, a_k$, and e_i is the vector of residual error effects. It is assumed that E($\boldsymbol{u}$)=0, E($\boldsymbol{e}$), and that m follows a normal distribution (~ N (0, $I\sigma_a^2$)), and *I* is an identity matrix. Such a model with a normal distribution of marker effects is often called ridge regression best linear unbiased prediction or RR-BLUP (Meuwissen et al. 2001, Van Raden 2008). Under a Bayesian approach, all SNP effects are assumed to have a common variance by assigning a Gaussian prior as $a_{RR,k} \sim N(0, \sigma_a^2)$, and all markers are shrunk to the same extent toward the mean and the degree of shrinkage is controlled by the prior variance. This method appeared most appropriate when a quantitative trait is controlled by many QTLs, each with a small effect. Several GS studies following this approach have been recently published for conifers (Beaulieu et al. 2014a,b, Resende et al. 2012a,b, Zapata-Valenzuela et al. 2012). The LASSO is an alternate shrinkage method that minimizes the residual sum of squares constraining the sum of absolute values of the regression coefficients if the predictors are standardized. Some estimated regression coefficients can be zero, contrary to ridge regression.

The genomic-estimated breeding value (G-BLUP) can be obtained by using the same mixed linear model that is used to obtain A-BLUP, and by replacing the numerator relationship matrix ($\boldsymbol{A}$) with the realized genomic relationship matrix ($\boldsymbol{G}$) derived from the markers (Van Raden 2008, Legarra et al. 2009, Zapata-Valezuela et al. 2013). The $\boldsymbol{G}$ matrix can be computed as:

$$\boldsymbol{G} = \frac{\boldsymbol{ZZ'}}{2\sum_{j=1}^{m} \boldsymbol{p_i}(\boldsymbol{1}-\boldsymbol{p_i})} \qquad (4)$$

where $\boldsymbol{Z} = \boldsymbol{M} - \boldsymbol{P}$, $\boldsymbol{M}$ is a matrix which elements are set to 1, 0 and -1, i.e. the number of minor alleles minus 1, and $\boldsymbol{P}$ is a matrix that contains allele frequencies

as $\boldsymbol{P_i = 2(p_i - 0.5)}$ where $\boldsymbol{p_i}$ is the minor allele frequency of the marker i. The denominator of the formula scales the $\boldsymbol{G}$ matrix to be similar to the $\boldsymbol{A}$ matrix (Van Raden 2008). Computation could pose some challenges if GS involves tens of thousands SNPs, but the BLUP computations can be accomplished by using statistical software that fits linear mixed models using Residual Maximum Likelihood (REML) such as ASReml (Gilmour et al. 2009).

Several factors can affect prediction accuracy. Presence of linkage disequilibrium between markers and QTLs controlling the trait of interest (Habier et al. 2007) is of course crucial to maintain a high level of accuracy over the generations. The number of markers used for estimating the GEBVs can also have an important influence (Schaeffer 2006, Poland and Rife 2012). Grattapaglia and Resende (2011) for instance showed that a high accuracy level can be obtained even at a marker density of 2 markers per centiMorgan (cM) when the effective training population size (Ne) is as small as about 30. However, for larger effective population sizes, marker density must be considerably increased to obtain high prediction accuracy (Daetwyler et al. 2008, Jannink et al. 2010). Trait heritability also influences the prediction accuracy (Heffner et al. 2009), but its impact is less important than marker density and the effective population size (Grattapaglia and Resende 2011). The existence of relationships between training and testing sets has also been shown to be essential (Albrecht et al. 2011; Beaulieu et al. 2014, Zapata-Valenzuela et al. 2012) unless the marker density is very high.

6. Multi-varietal forestry (MVF)

Multi-Varietal Forestry (MVF) can be defined as the deployment of a range of genetically tested tree varieties in plantation forestry. It is also known as clonal forestry; however, with advances in conifer SE and cryopreservation, the term MVF is more descriptive when applied to commercial plantation forestry (Park 2004). In general, a clone refers to any genotype with its genetic copies or ramets, whereas a variety refers to a clone that is selected or bred for certain attributes (and has test data to show to what extent these attributes have been achieved). In the past, the MVF concept in conifers was not realistic because of our inability to produce the same genotype over time. With the use of SE and cryopreservation combined with varietal testing, it is now possible to produce the same test-proven genotypes consistently over time, similar to the production of agronomic and horticultural varieties.

For several conifers, particularly for spruce and several pine species, the SE process is sufficiently refined to the stage that it can be implemented in industrial production. In New Brunswick, Canada, MVF is being practiced with spruce species, e.g., *Picea glauca* and *P. abies*, by JD Irving Limited since the late 1990s. A schematic representation of MVF by JD Irving Limited is shown in

Figure 3. Briefly, the MVF process takes the following steps: it begins with controlled crossing of superior parents selected from the breeding population; the resulting seeds are subjected to somatic embryogenesis; once embryogenic cell lines are proliferated, they are cryopreserved; once a number of lines to test is determined, a portion of each line is thawed and propagated; using the plants from the thawed lines, varietal testing is conducted at multiple locations; based on the periodic evaluation of varietal test, superior varietal lines are identified and retrieved from cryogenic storage; a selected number of superior varietal lines are mass vegetatively propagated; and the varietal lines are deployed in commercial plantations using appropriate numbers of varietal line mixtures.

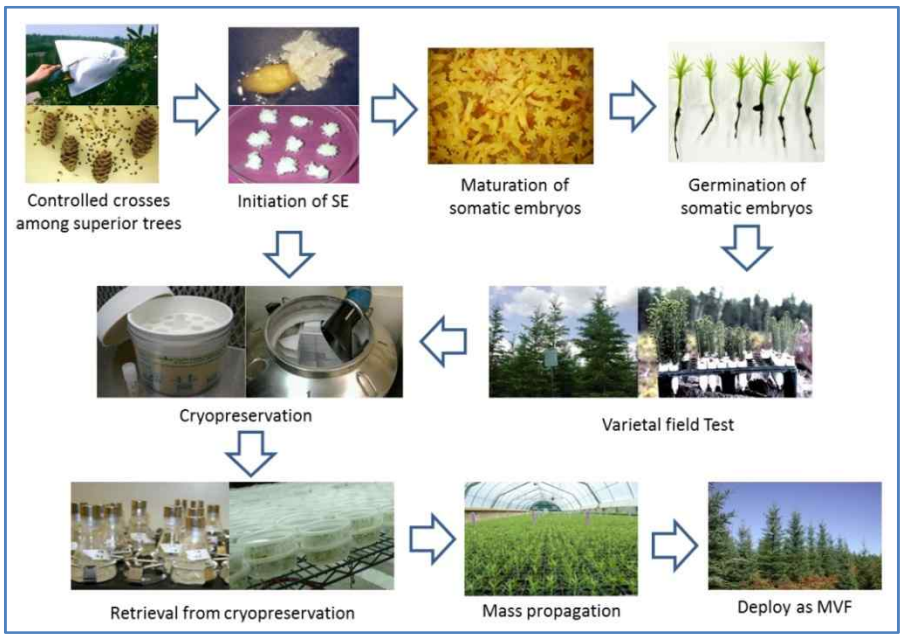

Figure 3. *The current MVF implementation using somatic embryogenesis at JD Irving Limited in eastern Canada. Selected parents from a long-term breeding program are controlled crossed and the resulting seeds are subjected to somatic embryogenesis for development of clonal lines. Embryogenic lines are cryopreserved and then a portion of each line is thawed and propagated to produce plants for varietal field testing. Once field testing shows which are the best lines, the corresponding embryogenic tissue are retrieved from cryopreservation, mass propagated, and deployed in the plantations.*

The field testing is an important phase of MVF because it is a critical part of selecting cell lines with desired attributes and of developing varietal lines. It is also the most time-consuming and expensive part of the process because trees are long-lived and, unlike agricultural crops, they grow slowly. Field tests are

evaluated at regular intervals, and the most current genetic information is used to amend the composition of multi-varietal mixtures thus offering the flexibility to adapt to changing conditions.

7. MVF incorporating forward genomic selection (GS)

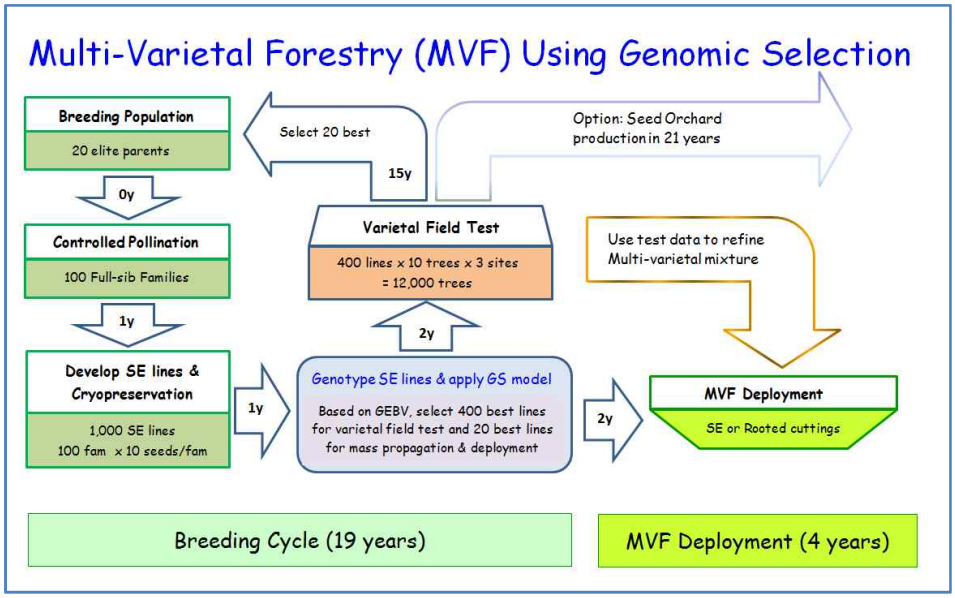

Figure 4. *Multi-varietal forestry integrating somatic embryogenesis and genomic selection: Similar to traditional breeding, controlled pollinations are performed to produce, e.g., 100 families from a subset of a breeding population. From each full-sib family, 10 embryogenic lines are developed (a total of 1,000 embryogenic lines) and cryopreserved. Genotyping of SE lines can be performed using embryogenic tissue, mature somatic embryos, or plantlets. The genomic data of SE lines are then applied to the GS model developed from the training population of the previous generation to obtain their GEBV. Based on best GEBV, about 400 SE lines, for example, may be selected to establish a varietal field test (VT), while the 20 lines that have the best GEBV may be selected for immediate deployment. In the VT, when trees begin to flower, the 20 best individuals are selected to form a new breeding sub-population. Thus, the breeding cycle is prolonged by 2 years due to the SE and cryopreservation steps that are required, thus resulting in a 19 year cycle. However, high GEBV SE lines are available for immediate mass propagation and deployment in MVF. This results in the deployment of genetically improved material 15 years sooner than via traditional seed orchard breeding by skipping one sexual reproduction cycle. Since the seed orchard is dependent on the time it takes to reach flowering maturity, GS has a limited advantage when production is based on sexual reproduction.*

With the availability of relatively inexpensive genotyping costs, GS is becoming attractive in tree breeding. In eastern Canada, integrated MVF using GS and SE in white spruce is based on forward GS and vegetative deployment. In this scheme, the GS model is developed in the mature genetic test plantations (training population) and the GS model is applied to the offspring population of the training population, hence forward GS as is illustrated in Figure 4: the elite individuals in a subset of breeding population are controlled-crossed to produce full-sib families; from these families, SE lines are developed and cryopreserved; after several months of cryo-storage, these lines are thawed, genotyped, and GEBVs are calculated; based on the GEBV, desirable individuals are mass propagated by SE or rooted cuttings from SE plants as proposed by Park et al. (1998) and deployed in the MVF. Thus, SE combined with GS can deliver genomically tested varietal mixtures for MVF in 4 years. This is a huge time saving when compared to delivery of genetic improvement by seed orchard, which may take 19 years even with GS.

Even though GS can identify superior genotypes at a very early stage without phenotyping, "varietal field testing (VT)" is necessary as it will verify the performance of the selections based on the GS model. Also, since GS can provide genetic information of the individuals at a very early stage, it can be used to pre-select genotypes to be included in VT. For example, instead of testing all available embryogenic lines obtained from seed produced by breeding, a breeder can select an upper 20-25% of high GEBV lines based on the genomic prediction, and propagate them to establish VT. This will reduce test establishment and maintenance cost drastically. VT is an important component of this MVF scheme because it provides continuously updated performance data that can be used to revise or modify varietal mixtures for deployment in the plantations, offering flexibility to adapt to changing conditions. Also, VT offers opportunities to capture non-additive variability as well as non-targeted traits (trait stacking) when observed during testing. Finally, the best selected trees in the VT will be used as the parents of the next cycle of breeding when they produce flowers and commence the next cycle of breeding.

8. Benefits of multi-varietal forestry

There are many benefits of MVF, but a few of the more important ones are:

1. Much greater genetic gain is possible than is obtained by using seed orchard seed. This is due to the capture of both additive and non-additive genetic variance.

2. MVF integrated with GS enables fast delivery of genetic gain and improves cost and efficiency of varietal testing. In turn, this will result in drastically higher genetic gain per unit time.
3. MVF can deliver trees with superior wood quality and uniformity
4. MVF offers flexibility to rapidly adapt to changing breeding goals, insect and disease conditions, and climate change through the use of continuously updated VT data.

9. Deployment strategies for MVF

The diversity of multi-varietal plantations is of concern, because there is a perception that narrow genetic variation may make MVF plantation more vulnerable to disease and insects than seedling-based plantations, and may results in plantation failure. However, for known diseases and insects, MVF has an advantage because more resistant varieties may be developed while simultaneously improving economic traits. But, for unknown or introduced pests, the protection is rather limited regardless of genetic variability existing within the species. It is difficult, if not impossible, to design protection against unknown diseases and insects. Nevertheless, it is generally assumed that, the more varieties in the MVF mixture, the lower the risk. However, the use of an increased number of varieties will reduce the genetic gain, Therefore, it is necessary to balance genetic gain and diversity, and this leads to a question of what is an appropriate number of varieties in a plantation.

Based on various assumptions, scientists generally agree that 10-20 varietal mixtures are sufficient for protection while providing benefits of MVF (Huhn 1987; Libby 1982; Zobel 1993; Roberds and Bishir 1997, Namroud et al. 2012). Such a threshold assures that alleles with population frequency of 10% or more are generally conserved, which are responsible for most of genetic variance in quantitative traits. Lindgren (1993) suggested some basic considerations: (1) if the species is used for short rotation, a lower number of varieties may be used because the exposures to the potential risk is short; (2) a lower number of varieties is acceptable if plantation management is intense and includes pest management; (3) the more well-known a variety, the more acceptable is its extensive use. Planting of varieties can be in varietal blocks or random mixtures, notwithstanding that they could also be used in mixed-species plantation schemes. In general, a random mixture is appropriate when varieties are not well-known or the future pest situation is uncertain (Lindgren 1993).

In eastern Canada, an approach called “Desired gain and diversity” is used to determine the number of varieties in a mixture. In this approach, the number is dynamically decided by selecting a desired or predetermined level of genetic gain and diversity based on the VT data (Figure 5). For example, a larger number of

varieties are included in the mixture at an early stage of VT in favor of diversity; however, at a later stage VT when the data are more reliable and varietal characterization is complete, a smaller number of varieties are used in the mixture in favor of larger genetic gain. This strategy is also combined with the previously proposed "Mixture of varieties and seedlings," which is mixture of selected varieties and seed orchard seeds (Park et al. 1998). This strategy will increase initial plantation diversity and reduce the stock cost as the seed orchard seeds are cheaper. Typically, in eastern Canada, about 40% of a plantation's basal area is commercially thinned at half-rotation age leaving superior quality trees for the final harvesting regardless of genetic origin. Thus, it is expected that the majority of trees are varietal trees with some exceptional trees of seedling origin. Therefore, the diversity of plantations is dynamically managed over time, where selection of varieties will be continuously revised based on the current VT data throughout the rotation age.

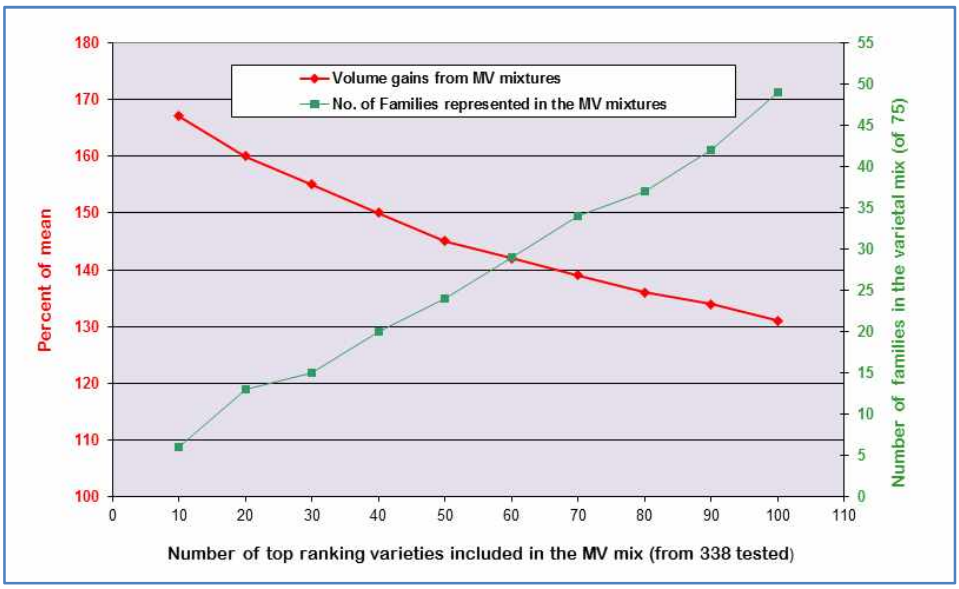

Figure 5. *Available genetic gains and diversity from MVF from a clonally replicated genetic test of white spruce assessed at age 14. The test contained 338 candidate varietal lines developed from 75 full-sib families. Since the parents of the crosses are the same parents that were used in the seed orchard, the overall mean (100%) represents the theoretical output of the orchard. If we take the 10 best clones in the varietal mix, the volume gain is 68% better than provided by the seed orchard but it contains only 5 families. If we take the 100 best clones, genetic gain is 30% better than the seed orchard output but it includes much greater genetic variation, i.e., 50 of 75 families. Thus, a breeder can set a desired level of genetic gain at given level of diversity.*

10. Hybrid varieties

Hybridization is a useful and widely used breeding approach in crop improvement, e.g., hydrid corn, through the crossing of usually different homozygous lines. In forestry breeding, hybridization usually refers to inter-specific or, sometimes, inter-provenance crosses. The main objectives of hybridization are to capture hybrid vigor and a combination of desirable characters. An example of hybrid vigor is demonstrated by the interspecific cross between Japanese (*Larix kaemferi*) and European (*L. decidua*) larch where certain lines are outperforming either parental species. In Korea, the pitch (*Pinus rigida*) and loblolly (*P. taeda*) pine hybrid was successfully used in reforestation to take advantages of the trait combination of the fast growth of loblolly pine and cold tolerance of pitch pine. However, despite the large potential, hybridization in conifers has rarely been used in modern tree breeding due to the labor intensiveness of hybrid seed production through mass controlled pollination and/or inefficiencies of bi-species seed orchards. SE appears as the ideal technology for developing hybrid varieties in conifers, because it can mass produce hybrid seedlings from a small number of seed obtained by interspecific controlled crosses. Moreover, with cryopreservation and VT, it offers further improvement through selection of the best individuals within the interspecific crosses. The development and deployment of hybrid varieties may be carried out similarly to the MVF as described here.

11. Commercial implementation of MVF

The industrial implementation of MVF is at an early stage. Many forestry companies and organizations are known to produce somatic seedlings from SE, including Arborgen (USA), Weyerhaeuser (USA), JD Irving Limited (Canada), FCBA (France), Arauco (Chile), Scion (New Zealand), Coillte (Ireland), Forestry Commission (UK), Government of Quebec (Canada) and others; however, their production rate is generally unknown but it seems relatively small in most cases. With the exception of JD Irving Limited and the Province of Quebec, the current SE production is mostly a laboratory-based system with *in vitro* germination, which is suitable for establishing varietal tests or small-scale commercial production but not for a large-scale production.

The primary challenge for MVF is the efficiency of the SE process from initiation to somatic seedling production. For many economically important species, SE initiation and maturation rates are too low; however, for most spruce and several pine species, the SE process is sufficiently refined to be used in the industrial MVF. For example, in white spruce, initiation of SE is at about 70%, proliferation in both liquid and semi-solid media generally works well. Usually, a

gram of embryonal mass produces on average about 500 mature embryos, and germination on appropriate culture media works well. However, there is a large variability in proliferation and maturation rates among embryogenic lines, and it is well-known that the SE process is affected by genetic background and culture conditions, offering a possibility of further refinements.

Cryopreservation of embryogenic lines using previously mentioned "Freezing Containers" is relatively simple. The recovery of cryopreserved lines is also satisfactory. For example, the recovery rate of 234 cell lines that were cryopreserved for 22 years was 95% (Park, unpublished data). The presence of contaminating microbes was also observed in the thawed cultures but the loss due to contamination was only about 1 percent of the total sample.

Perhaps, the most important challenge is the relatively higher cost of producing trees by somatic embryogenesis when compared to the seedling production using seed. In eastern Canada, it is estimated that SE trees cost more than 1.5 times the cost of seedlings, which is a net improvement compared to a generation ago, but still slightly too high even when the higher genetic gains are considered. Based on a series of crude assumptions, it was estimated that the SE production cost should not exceed 1.3 times the cost of seedlings in order to be profitable. Currently, the most expensive part of SE-derived trees is the manual transplanting of germinated embryos (*in vitro* state) into a commercial container system in the greenhouse. Therefore, it is critically important to develop either a semi-automated transplanting system or the means for direct germination into a growth substrate (micro plug) system; these options are currently being explored experimentally.

In the absence of a fully operational mechanized SE transplanting system, an alternative path to implementing MVF is the use of serial rooted cuttings from juvenile donor plants. Once superior embryogenic varieties are identified and thawed from cryopreservation, a small number of donor plants are propagated by SE, forming "stock" hedge plants (Park et al. 1998). Subsequently, mass propagation from the hedge stock can be accomplished by rooting of cuttings, which can be relatively inexpensive and automatable. These hedges can be used as stock hedges for about 5 years. The mass production of stecklings by rooting of cuttings from juvenile plants has been accomplished in several conifers (Park and Fowler 1987; Mullin et al. 1992; Kleinschimit et al. 1993; Russell 1993).

Finally, preliminary cost-benefit assessments of integrating SE with forward GS indicate that MVF will deliver unprecedented economic returns, much higher than achievable by any tree breeding effort (Beaulieu & Bousquet, unpublished data). This is the case because MVF can deliver much greater genetic gain than seed orchard breeding by capturing both additive and non-additive genetic variation without recombination through sexual reproduction. Furthermore, forward GS and SE eliminate the time required to produce seeds and, thus, gain per

unit time is notably increased. Therefore, SE becomes a key enabling technology for delivering the forward GS strategy.

12. Acknowledgements

This work was made possible through funding from the Natural Resources Canada – Canadian Wood Fibre Centre and from the Genomics Applications Partnership Program (GAPP)'s FastTRAC project of the Genome Canada and Genome Quebec awarded to the authors. The authors thank K. Klimaszweka, L. Caron and R. Wiart for reviewing the manuscript.

13. References

Albrecht T, Wimmer V, Auinger HJ, Erbe M, Knaak C, Ouzunova M, Simianer H, Schön CC (2011) Genome-based prediction of testcross values in maize. Theor Appl Genet 123:339–350

Beaulieu J, Doerksen T, Boyle B, Clement S, Deslauriers M, Beauseigle S, Blais S, Poulin P-L, 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**:**343-352

Beaulieu J, Doerksen T, 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

Becwar MR, Nagamani R, Wann SR (1990) Initiation of embryogenic cultures and somatic embryo development in loblolly pine (*Pinus taeda*). Can J For Sci 20:810-817

Chalupa V (1985) Somatic embryogenesis and plantlet regeneration from cultured immature and mature embryos of *Picea abies* (L.) Karst. Comm Inst For 14:57-63

Cyr DR, Lazaroff WR, Grimes SMA, Quan GQ, Bethune TD, Dunstan DI, Roberts DR (1994) Cryopreservation of interior spruce (*Picea glauca engelmanni* complex) embryogenic cultures. Plant Cell Rep 13:574-577

Cyr DR (1999) Cryopreservation of embryogenic cultures of conifers and its application to clonal forestry. *In* Jain SM, Gupta PK, Newton RJ (eds) Somatic Embryogenesis in Woody Plants. Kluwer Academic Publishers, the Netherlands Vol 4 239-261

Daetwyler HD, Villanueva B, Woolliams JA (2008) Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLos ONE 3: e3395

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL(2013) Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327–345

De Verno LL, Park YS, Bonga JM,Barrett JD (1999) Somaclonal variationin cryopreserved embryogenic clones of white spruce (*Picea galuca* (Moebch) Voss). Plant Cell Rep 18 :948-953

El-Kassaby YA, Lstiburek M (2009) Breeding without breeding. Genet Res 91:111- 120

Fullarton M (2015) Tree Improvement (NBTIC). http://www.nbforestry.com/ sustainability/tree-improvement/ Forest NB. Accessed Sept 12, 2015

Gilmour AR, Gogel B, Cullis B, Thompson R (2009) ASReml user guide, Release 3.0**.** VSN International Ltd, Hemel Hempstead, UK 2009, www.vsni.co.uk

Grattapaglia D, Resende MDV (2011) Genomic selection in forest tree breeding. Tree Genet Genomes 7: 241–255

Gupta PK, Durzan DJ (1987) Shoot multiplication from mature trees of Douglas-fir (*Pseudotsuga menziesii*) and sugar pine (*Pinus lamberiana*). Plant Cell Rep 4:177-179

Habier D, Fernando RL, Dekkers JCM (2007) The impact of genetic relationship information on genome- assisted breeding values. Genetics 177**:** 2389–2397

Hakman I, Fowke LC, von Arnold S, Eriksson T (1985) The development of somatic embryos in tissue cultures initiated from immature embryos of *Picea abies* (Norway spruce). Plant Sci 38:53-59

Harvengt L, Trontin J, Reymond I, Canlet F and Pâques M (2001) Molecular evidence of true-to type propagation of a 3-year-old Norway spruce through somatic embryogenesis. Planta 213:828-832

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Henderson CR (1975) Best linear unbiased prediction under a selection model. Biometrics 31:423-447

Hoerl AE, Kennard RW(1970) Ridge regression: biased estimation for non-orthogonal problems. Technometrics 12(1**):**55-67

Hühn M (1987) Clonal mextures juvenile-mature correlations and necessary number of clones. Silvae Genet 36:83-92

Jannink JL, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. Briefings in Functional Genomics 9: 166–177

Kleinschmit J, Khurana DK, Gerhold HD (1993) Past, Present, and anticipated applications of clonal forestry. *In*: Ahuja MR, Libby WJ (eds) Clonal Forestry II: Conservation and application, Springer-Verlag, Berlin Germany pp 9-41

Legarra A, Aguilar I, Misztal I **(**2009) A relationship matrix including full pedigree and genomic information. Journal of Dairy Science 92**:**4656–4663

Libby WJ (1982) What is a safe number of clones per plantation? In Haybrook HM, Stephan BR, von Weissenberg K (eds) Resistance to Disease and Pests in Forest Trees. Pudoc Wageningen The Netherlands pp 342-360

Lindgren D (1993) The population biology of clonal deployment. *In* Ahuja MR, Libby WJ (eds) Clonal Forestry I: Genetics and Biotechnology, Springer-Verlag, Berlin Germany. pp 34-49

Liu J-J, Sniezko RA, Sturrock RN, Chen H (2014) Western white pine SNP discovery and high-throughput genotyping for breeding and conservation applications**.** BMC Plant Biol 14:380.

Litvay JD, Verma DC, Johnson, MA (1985) Influence of loblolly pine (*Pinus taeda* L.) culture medium and its components on growth and somatic embryogenesis of wild carrot (*Daucus corota* L.) Plant Cell Rep 4:325-328

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157**:**1819–1829.

Mullin TJ, Morgenstern EK, Park YS (1992) Genetic parameters from a clonally replicated test of black spruce (italic cp). Can J For Res 22:24-34

Nagamani R, Bonga JM (1985) Embryogenesis in subcultured callus of *Larix decidua*. Can J For Res 15:1088-1091

Namroud M-C, Bousquet J, Doerksen T, Beaulieu J (2012) Scanning SNPs from a large set of expressed genes to assess the impact of artificial selection on the undomesticated genetic diversity of white spruce. Evol Appl 5:641-656

Neves LG, Davis JM, Barbazuk WB, Kirst M (2014) A high-density gene map of loblolly pine (*Pinus taeda* L.) based on exome sequence capture genotyping**.** Genes, Genomes, Genetics 4:29-37.

Park YS (2002) Implementation of conifer somatic embryogenesis in clonal forestry: technical requirements and deployment considerations. Ann For Sci 59:651-656

Park YS (2004) Commercial implementation of multi-varietal forestry using conifer somatic embryogenesis. *In* Proc IUFRO Joint Conf of Division 2: Forest genetics and tree breeding in the age of genomics: Progress and future Charleston, SC Nov 1-5 2004 p139

Park YS, Barrett JD, Bonga JM (1998) Application of somatic embryogenesis in high-value clonal forestry: deployment, genetic control, and stability of cryopreserved clones. In vitro Cell Dev Biol –Plant 34:231-239

Park YS, Fowler DP (1987) Genetic variances among clonally propagated populations of tamarack and its implication for clonal forestry. Can J For Res 17:1175-1180

Pavy N, Gagnon F, Rigault P, Blais S, Deschênes A, Boyle B, Pelgas B, Deslauriers M, Clément S, Lavigne P, Lamothe M, Cooke JEK, Jaramillo-Correa JP, 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**.** Molec Ecol 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 polymorphism among 13,500 genes of the conifer *Picea glauca*, relationships with functions, and comparison with *Medicago truncatula*. Genome Biol Evol 5(10):1910-1925.

Pavy N, Gagnon F, Deschênes A, Boyle B, Beaulieu J, Bousquet J (2015) Development of highly reliable *in silico* SNP resource and genotyping assay from exome capture and sequencing: an example from boreal black spruce (*Picea mariana*). Molec Ecol Resources (*in press*)

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.

Poland J, Rife TW (2012) Genotyping-by-sequencing for plant breeding and genetics. The Plant Genome 5:92–102

Resende MFR, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M (2012a) Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). Genetics 190**:**1503–1510.

Resende Jr MFR, Muñoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Resende MDV, Kirst M (2012b) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol 193**:**617–624

Roberds JH, Bishir JW (1997) Risk analysis in clonal forestry. Can J For Sci 27:425-432

Russel JH (1993) Clonal forestry with yelloe-cedar. *In* Ahuja MR, Libby WJ (eds) Clonal Forestry II: Conservation and application, Springer-Verlag, Berlin Germany. pp 188-201

Sansaloni CP, Petroli CD, Carling J, Hudson CJ, Steane DA, Myburg AA, Grattapaglia D, Vaillancourt RE, Kilian A (2010) A high-density Diversity Arrays Technology (DArT) microarray for genome-wide genotyping in *Eucalyptus***.** Plant Methods 6**:**16.

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123:218–223

Schilling MP, Wolf PG, Duffy AM, Rai HS, Rowe CA, Richardson BA, Mock KE (2014) Genotyping-by-Sequencing for *Populus* population genomics: An assessment of genome sampling patterns and filtering approaches. *PLoS ONE* 9(4):e95292.

Silva-Junior OB, Faria DA, Grattapaglia D (2015) A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 *Eucalyptus* tree genomes across 12 species. New Phytol 206**:**1527-1540

Sutton BCS, Polonenko DR (1999) Commercialization of plant somatic embryogeness. *In* Jain SM, Gupta PK, Newton RJ (eds) Somatic Embryogenesis in Woody Plants. Kluwer Academic Publishers, the Netherlands Vol 4 pp 263-291

Tibshirani R (1996) Regression shrinkage and selection via LASSO**.** J Royal Statistical Soc. Series B. 58**:**267-288

Van Raden PM (2008) Efficient methods to compute genomic predictions**.** J Dairy Sci 91**:**4414–4423. http://dx.doi.org/10.3168/jds.2007-0980

White T (1987) A conceptual framework for tree improvement programs. New For 1**:** 325-342

Zapata-Valenzuela J, Isik F, Maltecca C, Wegrzyn J, Neale D, McKeand S, Whetten R (2012) SNP markers trace familial linkages in a cloned population of *Pinus taeda* – prospects for genomic selection. Tree Genet Genomes 8:1307–1318

Zobel B (1993) Clonal forestry in eucalyptus. *In* Ahuja MR, Libby WJ (eds) Clonal Forestry I: Genetics and Biotechnology, Springer-Verlag, Berlin Germany pp139-148