Genealogical Relationship among Members of Selection and Production Populations of Yellow Cedar (*Callitropsis nootkatensis* [D. Don] Oerst.) in the Absence of Parental Information

Nasim Massah, Jinliang Wang, John H. Russell, Annette Van Niejenhuis, and Yousry A. El-Kassaby

From the Department of Forest Sciences, Faculty of Forestry, The University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4, Canada (Massah and El-Kassaby); Institute of Zoology, Zoological Society of London, London, UK (Wang); British Columbia Ministry of Forests and Range, Research Branch, Cowichan Lake Research Station, Box 335, Mesachie Lake, BC, Canada (Russell); and Saanich Forestry Centre, Western Forest Products Limited, Saanichton, BC, Canada (Van Niejenhuis).

Address correspondence to Y. A. El-Kassaby at the address above, or e-mail: yel-kassaby@ubc.ca.

**Abstract**

We used DNA fingerprinting and pedigree reconstruction to determine the genetic relationship among members of 3 yellow-cedar (*Callitropsis nootkatensis* [D. Don] Oerst.) selection populations in the absence of their parental genotypes. Selection population members consisted of the tallest individuals within seedling crops originated from natural stand seed collected from multiple seed donors covering wide areas within 3 distinct locations (phenotypic mass selection). Pairwise relative kinship estimates indicated the presence of extensive coancestry among the selected seedlings, and pedigree reconstruction grouped each selection members into multiple full-sib families of different sizes (1–10) nested within several half-sib families (19–21). The “STRUCTURE” program (Pritchard JK, Stephens M, Donnelly P. 2000. “Inference of population structure using multilocus genotype data.” Genetics. 155:945–959.) provided a pictorial classification of the 3 selection populations and grouped their individuals into multiple cohorts (9–10). The STRUCTURE program’s results corresponded with that of the pedigree reconstruction, indicating that members of the selection populations originated from a subset of the seed donors forming the natural stand seed collections. The species’ silvics, reproductive biology, methods of natural stand seed collection and seedling production, and the high selection intensity applied to form the selection populations contributed to limiting the selection to a subset of the original donor trees. The associated buildup of coancestry in selection and production populations is expected to result in inaccurate estimation of genetic parameters and an unintentional reduction in genetic diversity in reforestation stocks.

**Key words:** *Callitropsis nootkatensis*, clonal testing, DNA fingerprinting, pairwise relative kinship, pedigree reconstruction, selection

The recurrent selection scheme of tree improvement programs follows 3 main steps: phenotypic selection of candidate trees from natural stands or plantations; breeding, wherein mating designs are commonly used to create structured pedigreed material for testing, evaluation, and ranking of selected parents based on their offspring’s performance; and the identification of elite genotypes for production population(s) establishment and/or starting a new round of selection (Namkoong et al. 1988). The robustness of this approach has been proven, and substantial genetic gain has been captured for many species for a multitude of attributes (e.g., Namkoong 1979; Xiang et al. 2003). This classical approach with its predictable mechanics (i.e., breeding, testing, and selection) remained unchanged; however, species biology such as predisposition to cloning through vegetative propagation or somatic embryogenesis provided opportunities to increasing testing fidelity (Russell and Cartwright 1991; Heilman 1999; Sutton et al. 2004) and
the application of clonal multiplication for deployment (Karlsson and Russell 1990; Russell et al. 1990; Heilman 1999; Cyr et al. 2001).

Yellow cedar (*Callitropsis nootkatensis* [D. Don] Oerst.), formerly known as *Chamaecyparis nootkatensis* [D. Don] Spach), is an important timber species in Western North America covering wide geographical and ecological range spanning up to 20° in latitude and elevation from sea level up to tree line (Harris 1990) (Figure 1). The species taxonomy has gained recent attention since the discovery of new species (*Xanthocyparis vietnamensis* [Farjon et Hiep]) in Northern Vietnam (Farjon et al. 2002; Little et al. 2004), and its contemporary taxonomy is the subject of intense debate (Gadek et al. 2000; Ritland et al. 2001; Wang et al. 2003). The species has a 3-year reproductive cycle as adaptation to its high elevation habitat (Owens and Molder 1974, 1975, 1977); however, this cycle is plastic and is reduced to 2 years in its southern and at low elevation in its northern range (El-Kassaby et al. 1991).

An innovative approach to reducing efforts as well as shortening the length of the breeding cycle was attempted for yellow cedar. This scheme involved substituting structured mating designs with wind-pollination matings among trees in natural stands and the application of 2 phases of selection. The first was a high intensity phenotypic mass selection (Allard 1960, p. 252) among natural crosses’ bulked offspring conducted at very early age (1-year-old seedlings), whereas the second took place after intense field testing over multiple sites and years using clonal material produced by vegetative propagation of the first selection phase members. The intent of the second selection phase is to identify elite genotypes for their inclusion into production populations. This approach was successful in capturing appreciable amount of genetic gain; however, the extent of this gain cannot be ascertained specifically when the genetic relationship among members of the first and second selection phases is unknown. Genetic gain calculation for clonal selection differs from that of half-sib families. The former is similar to that for monozygotic twins and assumes that the covariance among clones equals 100% of the additive genetic variance, whereas the latter assumes the covariance among half-sibs equals 25% of the additive genetic variance (Falconer and Mackay 1996). Thus, if genetic gain calculations are based on clonal material, as in the case when the genealogy among selections is unknown, the estimated gain will be higher than that based on structured half- and full-sib families.

The objectives of this study are to determine 1) the genealogical relationships among the first and second phase selection members and 2) effective population size (extent of genetic diversity) for the testing and deployment populations. If members of the first or second phase selections are related, then the original genetic parameters derived before considering the presence of coancestry require adjustment, and care must be directed toward the vegetative multiplication of the production population members to avoid potential reduction in genetic diversity in reforestation material through the amplification of the related material.

**Figure 1.** The range of yellow cedar in the Pacific North West (Harris 1990) and the location of the 3 studied populations.
Materials and Methods

Seedlings representing 3 first-phase selection populations originated from natural stand seed collections made in Vancouver Island, British Columbia (Table 1; Figure 1), were used in this study. At each location, seed was collected using a helicopter equipped with cone rakes to strip seed cones-bearing branches from tree tops. Collections were made from 29, 30, and 36 nonneighboring trees to form seedlots #1, 2, and 5, respectively; thus, it was assumed that seed donors acted as pollen recipients from their restricted neighborhood. For each location, the collected seed cones were bulked prior to seed extraction; therefore, the genetic identity of the seedlings produced from these seedlots was unknown. The seedlots were sown in a commercial nursery to produce seedlings (several thousands) for reforestation purposes. At the end of the nursery growing season, the healthiest and tallest 1-year-old seedlings within each seedlot were selected to form the testing populations (phase 1 selection). A total of 170, 271, and 133 individual seedlings were selected from seedlots #1, 2, and 5, respectively. The selection populations (seedlings) were planted in an arboretum and were repeatedly topped to form “stool beds” for the mass production of small shoots that were used to produce clonal material for field testing through rooting. The resulting seedlings (ramets) were planted in multiple replicated clonal trials over Vancouver Island. Nine years after the trial establishment, several growth and yield attributes were assessed for each individual within all test sites, and quantitative genetics analyses were conducted to identify the top performing individuals (clones) within each selection population. The top 5, 17, and 8 clones representing seedlots #1, 2, and 5, respectively, were selected as members of the production population.

Fresh foliage samples from each member of the selection populations were collected from 5 test sites near Port McNeill (3 sites) and Jordan River (2 sites), Vancouver Island, British Columbia. Collected samples were placed in marked plastic bags, stored on ice, and shipped to UBC for further processing. The tips of each branchlet sample were placed in a vial tube (2 ml) and stored at −20 °C until DNA extraction.

DNA extraction followed the Doyle and Doyle (1987) protocol after optimizations to remove organic compounds such as terpenes, proteins, and phenols (Barton 1976). Extracted DNA was qualified and quantified, and each sample was diluted to 20 ng/μl and stored at −20 °C for polymerase chain reaction (PCR) amplification. Four yellow cedar–specific microsatellite primers (Y1F10, Y1G09, Y2C12, and Y2H01) (Bérubé et al. 2003), one (Cos2590) of 19 primers tested on Japanese cypress (Chamaecyparis obtusa), a closely related species (Nakao et al. 2001; Matsumoto et al. 2006), and 1 of 6 tested primers obtained from the C. obtusa available expressed sequence tag (EST) data, accessed via genebank, were selected for genotyping. We used the SSR finder program (Robinson et al. 2004) to search through the C. obtusa available EST clones to detect microsatellite sequences with the following search criteria: 1) primer minimum, optimum, and maximum temperature were 50, 55, and 65 °C, respectively, and maximum temperature difference was set at default, 2) primers’ G-G content % (GC%) was set at minimum, optimum, and maximum of 40%, 50%, and 60%, respectively, and 3) the default settings were used for the rest. Because GC has more accurate genotyping and minimizing scoring errors, a search for allelic variation was done using 2 subsets of 30 individuals from seedlots #1 and 2, and then the selected 6 loci were amplified and genotyped with the SAGA GT (LiCor Biosciences, Lincoln, NE, United States) software. In addition to the standard marker/DNA ladder, samples that captured allelic variation within each locus were selected, mixed, and used as a DNA reference of known allele size. Each individual PCR plate contained samples from 2 seedlots at a time to aid in correct allelic detection, and all gels were scored twice.

Data analyses were conducted using 3 independent approaches: the first to explore if genetic relationship existed among the selected seedlings within each seedlot using Ritland’s (1996) “MARK” program, the second to construct the pedigree within each seedlot and classify the selected seedlings into multiple full-sib families nested within half-sib families using Wang’s (2004) “COLONY” program (it was considered as the primary analysis), and finally, Pritchard et al.’s (2000, 2007) STRUCTURE program to provide a pictorial representation of cohorts within each seedlot. The first 2 analyses were used to assist in hypothesis development (MARK) and testing (COLONY), whereas the third analysis (STRUCTURE) was used to verify the results obtained from the COLONY program.

The “MARK” program (Ritland 1996) was used to calculate the genetic relationship among each possible dyad within each seedlot by pairwise kinship analysis. The kinship

<table>
<thead>
<tr>
<th>Seedlot</th>
<th>Location</th>
<th>Latitude</th>
<th>Longitude</th>
<th>Elevation</th>
<th># Of seed donors</th>
<th># Of selected seedlings</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Port McNeill</td>
<td>50°19’</td>
<td>127°08’</td>
<td>550</td>
<td>30</td>
<td>170</td>
</tr>
<tr>
<td>2</td>
<td>Shawnigan</td>
<td>48°45’</td>
<td>124°00’</td>
<td>800</td>
<td>29</td>
<td>271</td>
</tr>
<tr>
<td>5</td>
<td>Adams River</td>
<td>50°25’</td>
<td>126°05’</td>
<td>570</td>
<td>36</td>
<td>133</td>
</tr>
</tbody>
</table>

Table 1. Location of the 3 yellow-cedar seed collections, the number of seed donors, number of seedlings selected during the first-phase selection, and their seedlot registration numbers.
coefficients were classified into 5 biologically meaningful distinct groups (0.0, 0.125, 0.250, 0.375, and 0.5, representing unrelated, half-sib, full-sib, full-sib with inbred parents, and inbred, respectively). The COLONY program ver. 1.3 (Wang 2004) was used to reconstruct all sample relationships as a cohort by applying their genotype data after setting genotyping error rate at 5%. The COLONY program uses group-likelihood methods, and individuals are studied as a group after integrating the genotypic information of all samples. However, it is restricted to distinguish first-degree relationships such as full-sibs and half-sibs in a cohort. The STRUCTURE program is a model-based method, which utilizes the Bayesian algorithm to cluster individuals within a population into genetically similar cohorts based on their multilocus genotypes (Pritchard et al. 2000). The program calculates every locus allele frequency separately for each population and clusters individuals to cohorts based on their posterior probabilities ($Q$) calculated from their genotype data (Pritchard et al. 2000). This analysis is commonly used in population structure-related studies to allocate each entity to its genetically related cohort even in cases where there is no prior parental or population structure information or in cases where entities are originated from more than one population. STRUCTURE program ver. 2.2 (Pritchard et al. 2000) was used with a length of burn-in period parameter set at 50,000 (burn-in length between 10,000 and 100,000 were recommended by the author), and the same number was taken for Markov chain Monte Carlo replications after burn-in. We set $K$ (cluster) to range between 1 and 20 with number of independent runs set at 40 (i.e., each $K$ was simulated 40 times). The number of most likely cohorts was determined using the ad hoc statistic known as $\Delta K$, which is based on the log probability change across successive $K$ values (Evanno et al. 2005).

Genetic diversity of a population (selection or production) can be expressed as the effective population size, which describes the proportion of individuals from various parents and their genetic relationships. Considering related and/or inbred individuals, the concept of status number was used (Lindgren et al. 1996) and denoted as $N_e$ in this study ($N_e = \frac{1}{2} \Theta$). Group coancestry, $\Theta$, represents the average coancestry of all pairs of population members including self-coancestry and reciprocals (Cockerham 1967).

**Results**

The kinship analysis (Ritland 1996) estimated all possible pairwise relative kinship coefficients among individuals within each selection population ($n(n - 1)/2$) (Figure 2). The mean kinships over all individual pairs within the studied 3 populations were 0.061 (standard error = 0.00067), 0.052 (0.00041), and 0.098 (0.00140), respectively. This analysis provided insight of the genetic relationship among individuals within each of the 3 selection populations. In cases where no genetic relationship exists among the selected seedlings (relative kinship ($r$) 0.0), it is expected that all of the pairwise combinations will produce a value of zero. However, the pattern obtained from the 3 selection populations indicated that the selected seedlings belonged to groups of half-sib families ($r = 0.125$) with several smaller full-sib families ($r = 0.250$) nested within these half-sib families (Figure 2). It should be pointed out that the pairwise kinship estimate representation of Figure 2 is done to portray biologically meaningful scenarios as the obtained kinship estimates of every selection population covered a continuum between 0 and 1 (for instance, 0.125 class represents a range of relative kinship estimates between 0.064 and 0.188). Additionally, the 3 seedlots produced 2 very small peaks of higher coancestry, indicating that some individuals resulted from mating between inbred parents and/or were a product of selfing (Figure 2). This pattern is consistent with the grouping of each population’s individuals into sets of...
full-sib families with different sizes nested within their respective half-sib families that were obtained from the pedigree reconstruction analyses (see below).

The pedigree reconstruction analysis (Wang 2004) was conducted based on the following assumptions: 1) seed donors received pollen from multiple male parents present in their vicinity (i.e., mother trees were polygamous) and 2) self-fertilization would fail to produce viable seed due to the high genetic load, or selfed seedlings would have a selective disadvantage and thus would not be included in the selection populations (see Discussion). The resulting analyses grouped the selected seedlings into multiple full-sib families nested within 19, 21, and 17 half-sib families, for seedlots #1, 2, and 5, respectively (Figure 3). The full-sib family size within these half-sib families varied and ranged between 2 and 5, 1 and 10, and 1 and 4, for seedlots #1, 2, and 5, respectively (Figure 3). The production population members’ genetic status was determined from the pedigree reconstruction, and half-sib membership was detected for selection populations #2 (6 HS families accounting for 14 of the 17 selections) and 5 (2 HS families accounting for 5 of the 8 selections) (Table 2).

Following the recommendations of Bradbury et al. (2008), the number of iterations for running the STRUCTURE program was set at 40 to allow reaching stable results. The pictorial classification of cohorts within each selection population was presented graphically, where every cohort is defined with a specific color, and individuals are represented by vertical bars demonstrating their cohort membership posterior probability. The STRUCTURE program grouped the selected population members into 9 (seedlot #2) and 10 (seedlots #1 and 5) cohorts (Figure 4).

For each selection population, some correspondence was observed between the COLONY and STRUCTURE results as presented by the size of the largest single half-sib family within a specific genetic cohort (Table 3). Although half-sib

Figure 3. Pictorial presentations of pedigree reconstruction results for 3 yellow-cedar selection populations (top: Port McNeill [seedlot #1]; middle: Shawnigan [seedlot #2]; and bottom: Adams River [seedlot #5]). Individuals with each half-sib family are grouped (nested) within their respective full-sib families.
family membership was not restricted to a single genetic cohort, some family representation was high and ranged between 18% and 61% of the cohort’s size, indicating that seed donors contributed unequally to the resultant seedlots and selection reflected their seedlings phenotypic superiority.

Discussion

The attempt to circumvent the arduous long-term classical breeding program by substituting parent-tree selection and structured mating designs with extensive early selection from naturally produced population of seedlings is novel. It has good potential to shorten the length and substantially reduce both costs and efforts of breeding programs. This approach is based on several assumptions that include 1) the nursery seedling population genetic structure is not affected by the species silvics and its peculiar seed germination and dormancy, 2) members within each selection population are genetically independent, 3) nursery truncation selection does not limit the selected seedlings to few seed donors (families), and 4) the age–age correlation between nursery and field performance is reasonably high.

Several factors must be considered during the interpretation of results obtained from pairwise, pedigree reconstruction, and STRUCTURE analyses. These are related to the genetic structure within the 3 natural stand seedlots as well as the expected level of coancestry present among their respective seedlings. Although the number of seed-cone donors forming each seedlot is known, their proportionate contribution to the bulk seedlot and subsequently their successful input to the seedling crops are unknown. The reproductive biology of yellow cedar is characterized by a 3-year cycle, an adaptation to its high elevation habitat (Owens and Molder 1974). When seed cones are collected, trees often carry a mixture of 1- and 2-year cones that are difficult to differentiate. The collected cone crops contain both developmentally mature and immature seed, and therefore, the volume of cones collected from a particular seed donor is not indicative of its seed contribution. Even when a donor contributes substantially

Table 2. Number of genetic classes within the production population (elite genotypes) and the production population effective population size ($N_e$) estimates under genetic independence (no genetic relationship) and the COLONY family classification assuming equal contribution to the reforestation planting stock

<table>
<thead>
<tr>
<th>Seedlot #</th>
<th>Elite genotypes</th>
<th>$N_e$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Unrelated (size = 2)</td>
<td>Half-sib (size = 3)</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>—</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>

$^a$ Wang 2005.

Figure 4. Results from the STRUCTURE program for 3 yellow-cedar selection populations (top: Port McNeill [seedlot #1]; middle: Shawnigan [seedlot #2]; and bottom: Adams River [seedlot #5]). Scale represents individuals’ group membership posterior probability.
Three half-sib families with the same sizes.

Two half-sib families with the same size.

developed seed (Pawuk 1993; El-Kassaby 1995).

in seed germination and the deep dormancy of fully

representation in the seedling crop, due to genetic variation

to the seed crop, this does not guarantee increased

representation in the seedling crop, due to genetic variation

in seed germination and the deep dormancy of fully
developed seed (Pawuk 1993; El-Kassaby 1995).

The species is commonly present in small patches
(clusters) with many scattered individuals (Harris 1990);
thus, nonneighboring seed-cone donors are expected to
receive pollen from their adjacent locations (i.e., they are
spatially and may be temporally isolated). Seed-cone
collections originating from multiple nonneighboring trees
are expected to represent multiple “local” gene pools, thus
increasing the chance for greater differentiation among the
wind-pollinated maternal half-sib families originated from
these seedlots. Owens and Molder (1977) estimated that
mature cones averaged 7.2 seeds per cone with only 29%
being filled, thus yielding only about 2 filled seeds per cone.
This low yield may result from sparse pollination, increased
abortion of selfed seed (Bishir and Pepper 1977), and the
rarity of 2 consecutive years of good environmental
conditions, needed for pollination, fertilization, and embryo
and seed development (El-Kassaby 1995).

The collective genetic structure of seedling populations
originating from natural stand seed is expected to harbor
high level of variability, reflecting the variation among the
maternal parents (seed donors) as well as their multiple local
pollen pools. British Columbia’s 2003–2008 average number
of yellow-cedar seedlings produced from natural stand seed
sources is 900 K (SPAR: BC Ministry of Forests and Range
2009). Natural stand seedlots are used to produce seedling
crops of size varying between 50 and 100 K seedlings.
Under mass selection for outcrossing species (Allard 1960,
p. 252), factoring in the total number of seedlings selected
from either the COLONY or STRUCTURE programs
these factors collectively, it is not surprising that the results
from either the COLONY or STRUCTURE programs
produced a smaller number of parents/cohorts as compared
with the original seed donors forming the seedlots.

The pattern obtained from the pairwise kinship analyses
(Ritland 1996) confirmed the presence of half- and full-sib
families (Figure 2) and allowed the formulation of the main
hypotheses that no genetic relationship exists among the
nursery stage selected seedlings (selection populations) or
among the selected elite genotypes forming the production
populations. Naturally, the presence of sib-ship relationship
among these individuals formed our alternate hypothesis.
We therefore employed the COLONY program (Wang
2004) to determine genetic membership among the selected
seedlings within each population. We assumed that the seed
donors receive pollen from multiple pollen donors
(polygamous) and that self-fertilization does not result in
viable seed, or self-seedlings are selected against at the
nursery stage during the removal (thinning) of excess
germinals, or seedling development during the growing
season resulted in inferior seedlings (El-Kassaby and
Thomson 1995; El-Kassaby 2000). Thus, self-seedlings
would not be included in the selection populations. This
assumption is supported by reports of inbreeding depression
causing embryo development failure or selective disadvan-
tage of selfed germinals, seedlings, and saplings has been
reported for multiple conifers including yellow cedar
(Sorensen 1969, 1982; Bramlett and Popham 1971; Koski
1971, 1973; Bishir and Pepper 1977; Bishir and Namkoong
1987; Namkoong and Bishir 1987; Savolainen et al. 1992;
Williams and Savolainen 1996; Williams et al. 1999).

Forest trees commonly practice a mixed-mating system
with reported high levels of outcrossing rate (O’Connell
2003) and often show low levels of correlated matings
(O’Connell et al. 2001; Liewlaksaneeyanawin 2006). How-
ever, low selfing might be contradictory to the application
Like other members of the Cuperssaceae family, yellow
cedar can self-fertilize, and high inbreeding coefficients were
reported by Ritland et al. (2001) (mean rangewide
FST), we estimated an
inbreeding coefficient of
0.18, also see Thompson et al. 2008). Using Wright’s (1965)
approach (\(t = (1 - F_{ST})/(1 + F_{ST})\)), we estimated a
inbreeding coefficient of \(F = 0.08\) for the 3 studied
populations. The difference between Ritland et al. (2001)
and the present study’s inbreeding coefficient estimates may
reflect the high selection intensity applied in the greenhouse
to form the 3 test populations. The observed pattern from
the pairwise analyses (Figure 2) as well as reported inbreeding
coefficient estimates for the species and that of the present
study could support the notion that some members of the
classified full-sib families may be the result of selfing.

The pedigree reconstruction analyses (Wang 2004)
grouped the test population members into multiple full-sib
families nested within half-sib families (Figure 3). A total
of 19, 21, and 17 half-sib families were observed within
seedlots #1, 2, and 5, respectively, and multiple full-sib
families with variable sizes (2–5, 1–10, and 1–4, for seedlots #
1, 2, and 5, respectively) were nested within them (Figure 3).
The observed number of half-sib families with their
variable sizes confirmed that the strict truncation selection
applied in the nursery during the test population formation
successfully eliminated some as well as restricted the
selection to a subset of seed donors. The observed family
structure reduced the testing populations’ effective popula-
tion size to 39.9, 37.6, and 28.0 for seedlots # 1, 2, and 5,

<table>
<thead>
<tr>
<th>Seedlot #</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genetic cohorts K (cluster)</td>
<td>20</td>
<td>23</td>
<td>31</td>
<td>42</td>
<td>55</td>
<td>23</td>
<td>23</td>
<td>47</td>
<td>33</td>
<td>29</td>
</tr>
<tr>
<td>2</td>
<td>22</td>
<td>61</td>
<td>44</td>
<td>44</td>
<td>38</td>
<td>50</td>
<td>35</td>
<td>29</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>38</td>
<td>33</td>
<td>33</td>
<td>22</td>
<td>43</td>
<td>18</td>
<td>19</td>
<td>22</td>
<td>21</td>
<td>36</td>
</tr>
</tbody>
</table>

* Two half-sib families with the same size.

* Three half-sib families with the same sizes.

Table 3. Correspondence between the COLONY and
STRUCTURE programs represented by the size of the largest
half-sib family within a single genetic cohort (%)

The observed number of half-sib families with their
variable sizes confirmed that the strict truncation selection
applied in the nursery during the test population formation
successfully eliminated some as well as restricted the
selection to a subset of seed donors. The observed family
structure reduced the testing populations’ effective popula-
tion size to 39.9, 37.6, and 28.0 for seedlots # 1, 2, and 5,
respectively, as compared with their census numbers (average 191; range: 133–271).

The STRUCTURE analyses provided pictorial classification of each seedlot presented by the program’s bar plots (Figure 4). Cohorts are defined by different colors where individuals are represented by vertical bars demonstrating the posterior probability \( q \) of each individual belonging to a specific cohort (Figure 4). It should be noted that the posterior probability of some individuals indicates that they belong to more than one cohort (i.e., individuals are represented by 2 or more colors), or it could be the result of incomplete information caused by missing data (genotype) (Bergel and Vigilant 2007). In general, the number of cohorts within a test population was smaller than the half-sib family number obtained from the pedigree reconstruction and ranged between 9 and 10 (Figure 4; Table 2). This reflects the difference between COLONY and STRUCTURE, where the former is based on the exact nature of jointly assigning parentage and sib-ships based on their multilocus genotypes and the latter on providing a group membership posterior probability of individuals.

Combining pedigree reconstruction and STRUCTURE program in analyzing natural populations has been attempted by Vähät et al. (2007) in studying the natal homing of Atlantic salmon populations. They repeatedly used the STRUCTURE program to stratify the populations in sequential manner, and finally, they used pedigree reconstruction (the COLONY program) to identify the genealogical relationship among the spawning individuals within each stratum. In the present study, the test populations were already stratified as they originated from seedlots collected from different locations (Figure 1). Our approach was similar and the use of the 2 analytical approaches (i.e., pedigree reconstruction and STRUCTURE) assisted in unraveling the genetic similarity/differences among each seedlot’s selected seedlings. However, it should be noted that the STRUCTURE and COLONY program results did not completely match in spite of using the same multilocus genotypic data. These discrepancies could be caused by violating the STRUCTURE program’s assumptions, specifically those pertaining to the presence of family structure within the seedling populations (Pritchard and Wen 2004; Camus-Kulandaivelu et al. 2007). Under increasing number of groups, the STRUCTURE program analysis starts by identifying the most genetically distinct groups followed by the identification of the less distinct groups. The presence of related individuals creates sufficient correlation among these individuals, forcing them to be identified as a genetic group. This situation was unintentionally encountered by Vähät et al. (2007) when they repeatedly applied the STRUCTURE analysis to stratifying their Atlantic salmon populations into subgroups and finally realized that the final “Structure” represented full- and half-sib families. In essence, we followed Yu et al.’s (2006) approach to identify possible family structure in the analyzed populations.

Results from the COLONY and STRUCTURE programs were compared to investigate the presence of any correspondence between the 2 approaches (Table 2). Although the COLONY and STRUCTURE programs produced different number of groups (half-sib families and cohorts, respectively), some correspondence between their results is apparent by the size of the largest single half-sib family within a specific genetic cohort (Table 2). Although half-sib family membership was not restricted to a single genetic cohort, some family representation was high and ranged between 18% and 61% of the cohort’s size, indicating that seed donors may have contributed differently to the seedlots and that some members of different cohorts shared common alleles. Additionally, some cohorts included members of more than one half-sib family (cohorts #1 and 4 encompassed 2 half-sib families, each representing 38% of the cohort size, whereas cohorts #6 and 7 included 3 half-sib families representing 18% and 19% of the cohort size, respectively; Table 2).

The ultimate goal of the implemented selection program was to identify elite genotypes for their inclusion in production populations. After field testing, the second selection phase resulted in identifying 5, 17, and 8 individuals, representing seedlots #1, 2, and 5, respectively, as determined by the COLONY analyses (Table 3). These selected genotypes will be used to produce large numbers of rooted cuttings for reforestation. If these genotypes are related, then it is expected that the genetic diversity of the deployed material will be lower than that assumed under no relationship; in fact, the effective population sizes of the selection populations were found to be lower than their perceived census numbers (see Table 3). With the exception of seedlot #1 where no genetic relationship was detected, seedlots #2 and 5 contained 5 and 2 half-sib families with different sizes, respectively, resulting in effective population sizes of 13.1 and 6.4 as opposed to census numbers of 17 and 8 (Table 3).

The utilization of matings among parents in their natural settings coupled with the implementation of 2 rounds of selection present a novel approach to breeding and selection of forest trees and provides an opportunity to reducing breeding time, efforts, and resources. The present study demonstrates the necessity of applying molecular biology techniques such as DNA fingerprinting and pedigree reconstruction as essential tools to advance these new breeding methods. The pioneering work of Lambeth et al. (2001) and Grattapaglia et al. (2004) opened new avenues to molecular tree breeding, and this work has been followed by the development of even bolder approaches such as “Breeding Without Breeding” (El-Kassaby and Lstiburek 2009). Tree breeding must capitalize on ongoing scientific developments in molecular biology, spatial statistics, parentage assignment, and association genetics to advance population improvement to new frontiers.

**Conclusion**

The implementation of an unconventional approach to tree improvement that deviates from classical breeding schemes...
was attempted. This approach forfits the phenotypic selection of “superior genotypes” phase as well as the use of conventional mating designs for creating structured pedigree for testing. The method involves high selection intensity at very early age (1 year) from offspring originated from natural crosses among parent trees in their native habitats, followed by exploitation of the species’ biology (easy vegetative propagation) for the production of clonal material for long-term extensive field testing. Using this approach, the pedigree of the offspring selected for testing is unknown, which could cause inappropriate estimation of quantitative genetics parameters as well as the deployment of genetically related material, thus reducing genetic diversity. We applied DNA fingerprinting and pedigree reconstruction to unravel the presence of extensive family structure among the selected offspring. In light of the findings, we recommend revision of 1) genetic parameter calculations based on assumed independence among tested individuals (i.e., clones) and 2) genetic diversity estimates of deployment material, as member propagules of half-sib families are expected to produce lower effective population size (a clone’s effective population size is equal to census number, whereas a half-sib’s effective population size is lower than the census number).

Funding

Johnson’s Family Forest Biotechnology Endowment, British Columbia Forest Investment Account Forest Genetic Conservation and Management program, Natural Sciences and Engineering Research Council of Canada—Discovery and Industrial Research Chairs Program grants to Y.A.K. and Natural Resources Canada—Science and Technology Internship Program to N.M.

References


Received April 2, 2009; Revised October 3, 2009; Accepted October 22, 2009

Corresponding Editor: James L. Hamrick