

Increasing Breeding without Breeding (BwB) Efficiency: Full- vs. Partial-Pedigree Reconstruction in Lodgepole Pine

Yousry A. El-Kassaby^{1*}, Tomas Funda² and Cherdsak Liewlaksaneeyanawin¹

¹Department of Forest and Conservation Sciences, Faculty of Forestry, The University of British Columbia, Canada

²Department of Ecology and Environmental Science, Faculty of Science and Technology, Umeå University, Sweden

Received: 21 January, 2015; Accepted: 17 March, 2015; Published: 07 April, 2015

*Corresponding author: Yousry A. El-Kassaby, Department of Forest and Conservation Sciences, 2424 Main Mall, University of British Columbia, Vancouver, British Columbia, V6T 1Z4, Canada, Tel: 1-604-822-1821; Fax: 1-604-822-9102; E-mail: y.el-kassaby@ubc.ca

Abstract

The advantage of paternity assignment in assembling structured pedigree for breeding is investigated using two sampling methods; namely, family array (known maternal parent) and random offspring (unknown maternal and paternal parents) collected from an open-pollinated lodgepole pine experimental population with known parents (N = 74) using nuclear and chloroplast microsatellite markers. Offspring of equivalent sample sizes representing the family array (n = 619) and random offspring (n = 635) were genotyped and subjected to partial and full pedigree reconstruction, respectively. The full pedigree reconstruction assembled substantially larger number of full-sib families than the partial (446 vs. 268) and interestingly the two methods detected equivalent amount of external gene flow to the experimental population. The superiority of the random offspring over the family array sampling in producing more full-sib families was attributed to its better representation of the parental population, as random sampling included offspring from most parents as compared to the parent-limited family array. Owing to the observed advantages, the full pedigree reconstruction could be employed as an alternative to the breeding phase commonly required in conventional breeding programs for the development of structured pedigree needed for genetic parameters estimation.

Keywords: Molecular breeding; lodgepole pine; random sampling vs. family array; partial and full pedigree reconstruction

Introduction

Forest tree breeding is a long-term endeavor often adopting the recurrent selection scheme [1] where hundreds of parents are rigorously tested through the performance of several thousands of their offspring planted over vast geographic territories known as breeding zones [2]. Parental ranking, for forward selection, is often based on offspring's performance which is followed by the selection of elite genotypes for either new rounds of breeding (matings, testing, and selection) or the establishment of production populations (a.k.a., seed orchards) [3]. Breeding and testing are the most costly and time consuming aspects of tree breeding. Breeding is done following one of the established mating designs to generate "structured" pedigree (half- and full-sib families) needed for genetic parameters (e.g., traits' heritabilities and correlations, and parents and offspring's

breeding values) estimation [4]. The creation of structured pedigree is meticulous work requiring great care and often takes multiple years to complete owing to the large number of parents and the required numerous crosses. Completion of the breeding phase is often delayed by fertility and phenological differences among the breeding parents [5]. The authenticity of the resulting offspring affects the accuracy of the generated genetic parameters and ultimately the attained genetic gain; unfortunately, this process is never error-free [6,7].

Forest tree breeders attempted to simplify breeding through the use of "wind- /open-pollinated" families [8,9] and often treated them as half-sib families as maternal parents are known and assumed that offspring is sired by large number of male donors; however, the possibilities of having full-sibs or selfs within these "half-sib" families is high. Thus, treating wind- /open-pollinated families as half-sibs leads to an over inflated additive genetic variance estimation and subsequently breeding values and heritabilities, resulting to an inaccurate ranking of parents (seed donors) [10-12]. The availability of reliable, informative molecular markers coupled with paternity assignment methods [13] created an opportunity whereby the breeding phase of tree breeding could be effectively eliminated. Lambeth, et al. [14] were the first to capitalize on this development and used paternity assignment to unravel the paternal parents in a polymix breeding framework. This approach was further extended and the "Breeding without Breeding" concept was developed [15-18] and offered a viable option for breeding-phase avoidance in tree breeding programs.

Here we test two sampling methods for structured pedigree assembly; namely, partial- and full-pedigree reconstruction using equal sample sizes drawn from a 74-parent lodgepole pine parental population. Partial- and full-pedigree reconstruction were represented by family array (individuals generated from a subset of parental seed-donors) and random sampling (individuals drawn from a seedling population representing the reproductive output of the entire parental population), respectively. Pedigree reconstruction was based on using genomic and chloroplast DNA microsatellite markers.

Materials and Methods

Seed orchard population and offspring sampling

A 71-clone lodgepole pine seed orchard located near Armstrong, British Columbia (50° 23' N, 119° 17' E, 470 m a.s.l.) provided the material for this study. The orchard was established in 1994 following the permuted neighborhood design which maximizes the separation distances among ramets of the same clone, hence minimizing selfing [19]. At the time of sampling (2007), the orchard's population consisted of 1,047 ramets representing the 71 parents (13.9 ± 7.0 SD ramets per parent).

Dormant vegetative buds were sampled from the entire orchard's parental population (2 random ramets/parent) and two seed sampling methods; namely, 1) family array (known 11 seed-donors, each with 56.3 ± 7.3 SD seed/parent (N = 619)) and 2) bulk sample (random sample of 635 seeds from the entire orchard's seed crop with unknown maternal and paternal parentage). The dormant buds were stored at -80°C until DNA extraction while the seeds were stored at 4°C until germination.

DNA extraction and SSR genotyping

DNA was extracted from vegetative buds and germinating seed (2-3cm embryos) following Doyle and Doyle [20]. Parents and offspring were genotyped using 9 nuclear SSRs [21-23] and 6 cpSSRs chloroplast microsatellite loci [24].

Parentage analyses

For paternity assignment, we used a likelihood-based paternity inference method with a known level of statistical confidence and accounting for genotyping errors [25] (CERVUS 3.0.3). Two parentage analyses were carried out, one for the family array with known maternal parent and the other was a parent pair analysis with unknown sexes of the candidate parents for the bulk seed sample. The paternal population (N = 74) (the orchard's known 71 parents plus 3 additional alien genotypes detected during the orchard's parental genotyping). The parentage analysis for the known mother-offspring genotypes was based on 10,000 simulations with 74 sampled candidate parents, genotyping error rate of 0.01, and 95% (strict) confidence level using the 9 nuclear SSRs. We chose 6 cpSSRs to permit the identification of the paternal parentage from the most likely parent pair [24]. We conducted the identity analysis with cpSSRs (also in CERVUS 3.0.3), after creating dummy genotypes via converting the haploid profiles to a hypothetically complete homozygous offspring. For each offspring, the paternal parent determined by the identity analysis was compared with the two parents identified by the parent pair analysis. The maternity analysis with known fathers (although, strictly speaking, with fathers deduced from marker evidence) was then conducted for these offspring, using the same parameters described earlier.

Results

The paternity assignment analyses were successful in assigning the male parent for 528 out of 619 offspring (85.3%) and both male and female parents for 522 out of 635 offspring (82.2%) for the family array and random sample, respectively.

The inability to assign paternity or maternity to the remaining offspring was either due to insufficient informative genotypes to match the candidate parents with 95% confidence, or that seeds are sired by parents from outside the studied population (i.e., the product of gene flow/pollen contamination), or a combination of both. Since the 9 nuclear SSRs used are highly polymorphic and possess low null allele frequencies [23] and the fact that most of the unassigned offspring had mismatches on at least two loci, then it is conceivable to assume that the used loci provide the required statistical power.

The additional 6 uniparentally inherited cpSSRs (mean: 4.8 and SD: 1.3 alleles/locus, range: 4-7) produced unique 51 multi loci. These unique haplotypes were essential in providing the high discrimination power needed for the successful assignment of the male parents in the random sample and resulted in increasing the number of successfully assigned males to 545 offspring being successfully assigned to one of the candidate fathers (85.8%) (additional 23 offspring). The identity analysis fully corresponded with the parent pair analysis, as for each of the analyzed offspring the assigned candidate paternal parent was the same as one of the two most likely parents determined by the parent pair analysis (in total 545 offspring). The unassigned offspring on the male side are most likely a product of gene flow from non-sampled candidate paternal parents from outside the studied population, producing gene flow estimates of 14.7 and 14.2% for the family array and bulk seed sample, respectively. The close to identical estimates of gene flow sheds light on the accuracy of the pedigree reconstruction of assigning the male or female and male parents for family array and random sample, respectively. The utility of these unassigned individuals to quantitative genetic analyses is documented in the Discussion section (below). It should be noted based on these results that had we only used the nuclear markers and the standard parent pair analysis, we would have been able to identify which two parents produced a given offspring.

Pedigree reconstruction of the family array produced 268 full-sib families nested within the 11 sampled maternal half-sib families, ranging in number from 17 (maternal half-sib family #37) to 31 (#52) and in size from 1 to 15 individuals per full-sib family (Figure 1). Pedigree reconstruction of the random sample captured offspring of 65 out of the 74 candidate mothers present in the seed orchard (87.8%) and, consequently, revealing a considerably higher number of full-sib families than the family array analysis (446 full-sib families, ranging in size between 1 and 4 (Figure 2)). These results were anticipated as the random sample, unlike maternal family array, represented the entire population's reproductive output.

The paternal half-sib family sizes ranged from 1 (nine families) to 58 (family #52) and from 1 (three families) to 28 (family #61) for the family array and random sample, respectively, with a positive correlation ($r = 0.61$, $N = 74$, $p < 0.05$) (Figure 3). This represents Pearson's product-moment correlation between vectors of paternal HS family sizes (male reproductive success) estimated by the two approaches (i.e., family array and bulk sample) for all 74 paternal parents existing in the seed orchard.

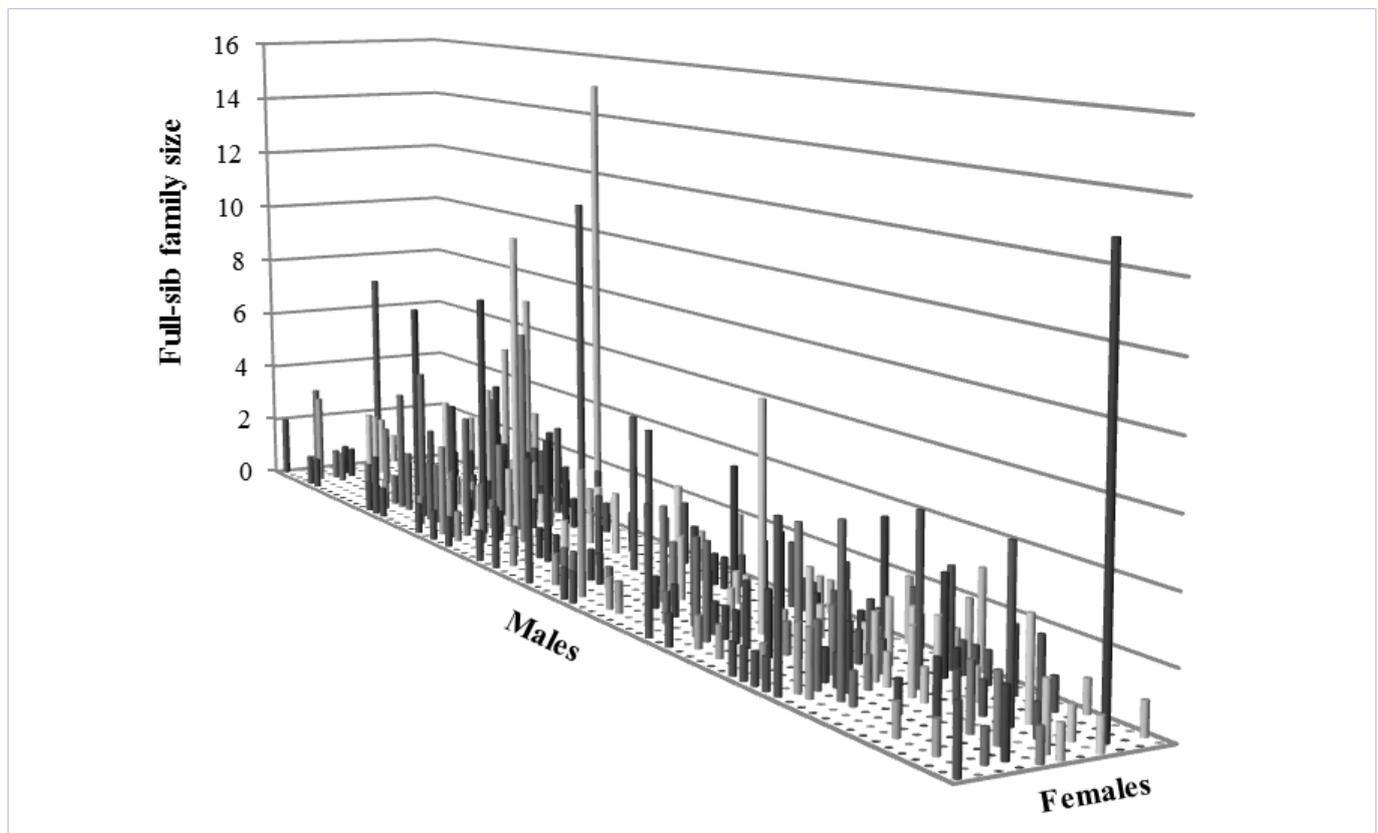


Figure 1: Distribution of 528 naturally occurred matings in a lodgepole pine seed orchard (74 parents) revealed by partial pedigree reconstruction of 11 wind-pollinated maternal half-sib families using nine nuclear microsatellite loci.

The variation in the paternal half-sib family sizes between these two approaches might have been due to the sampling methods of the individuals assayed, because seed representing each maternal half-sib family (i.e., family array) was only collected from one single ramet (i.e., one position) while the random sample was taken from a mixture of seed collected from the entire seed-producing population. Figure 2 illustrates the ability of the random sample to forming substantial number of full-sib families representing 87.8% of the population parents as well as demonstrates the restrictive ability of the family array sampling which is limited by the number of seed-donors sampled.

Discussion

Forest tree breeders utilize mating designs to create the “structured” pedigree needed for estimating the genetic parameters needed for elite genotypes identification and their selection for either breeding or seed production (seed orchards) [2]. Tree breeding programs often harbor large number of parents, thus, irrespective of which mating design is used; a substantial number of controlled crosses are needed. The physical task of controlled crosses itself is often hampered by parental fecundity and reproductive phenology variation, thus in most cases multiple years are needed for this phase completion and even when completed, cases of mistaken parental authenticity are common [6,7]. The partial or complete avoidance of using controlled crosses for structured pedigree formation would

be a favorable development to tree breeding programs. The combined use of DNA fingerprinting and pedigree reconstruction provided an opportunity for bypassing the breeding phase and “structured” pedigree can be assembled for quantitative genetics analyses. It should be stated that the resulting structured pedigree from pedigree reconstruction is often unbalanced favoring the more fecund parents and is greatly affected by the degree of gene flow from outside undesirable sources (i.e., wasted genotyping efforts) (Figures 1 and 2). However, the utilization of quantitative genetics’ algorithms such as ASReml [26] with their versatility to handle very large, multi-generational, and statically and genetically imbalance data sets made these analyses feasible and the restrictions of having balanced pedigree or statistical designs became unnecessary. This was clearly demonstrated by El-Kassaby, et al. [16] who presented an analysis for unbalanced structured pedigree that included a mixture of full- and half-sib families with various sample sizes. The inclusion of half-sib families in the analysis provide a situation where offspring from known mothers but unknown fathers (i.e., those sired by gene flow) could be effectively used to increase the precision of the estimated genetic parameters, thus the notion of “wasted” fingerprinting efforts is rectified.

The advantage of pedigree reconstruction, partial or full, is apparent from Figures 1 and 2. If the disconnected diallel mating design was used to create crosses for the 74 parents used in this study, then at least 12, 6-parent diallel unites is needed and a total

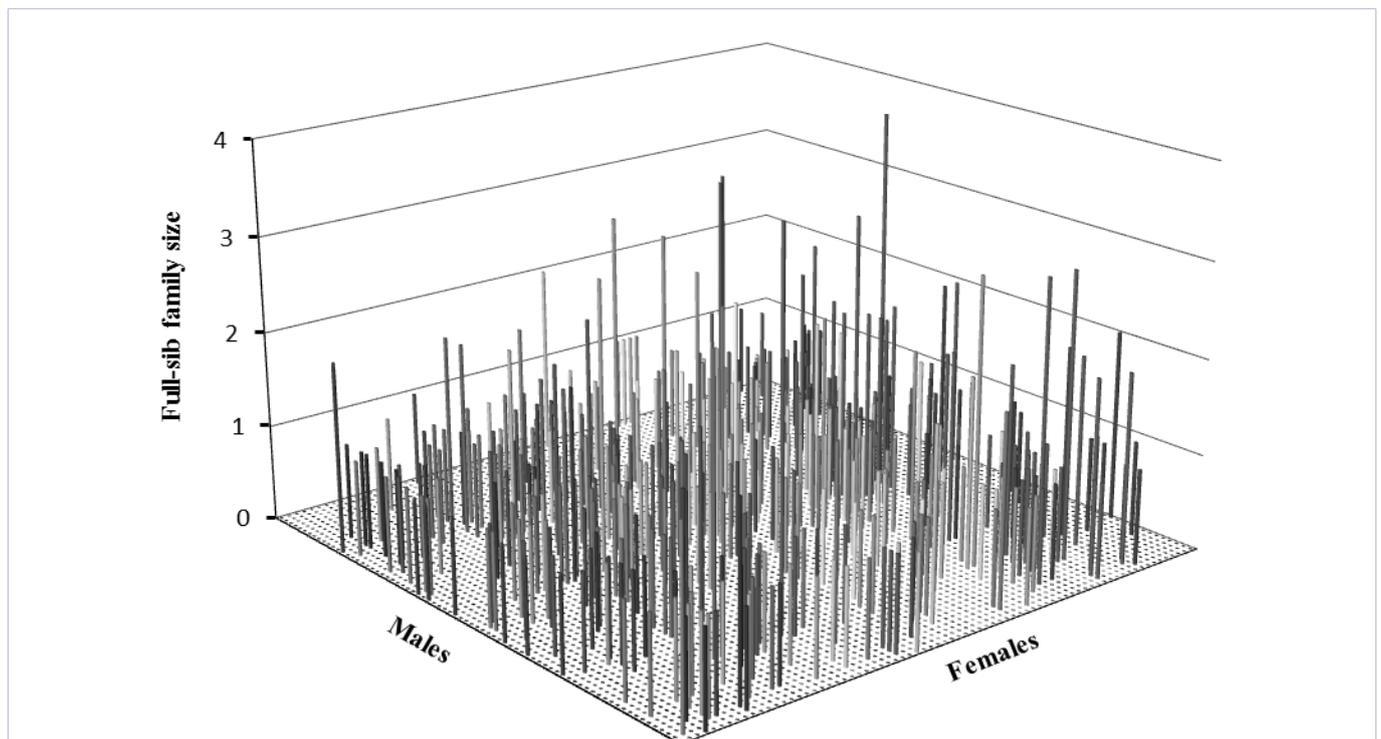


Figure 2: Distribution of 522 naturally occurred matings in a lodgepole pine seed orchard (74 parents) revealed by full pedigree reconstruction of random sample of offspring with unknown maternal and paternal parentage using a combination of nine nuclear and six chloroplast microsatellite loci.

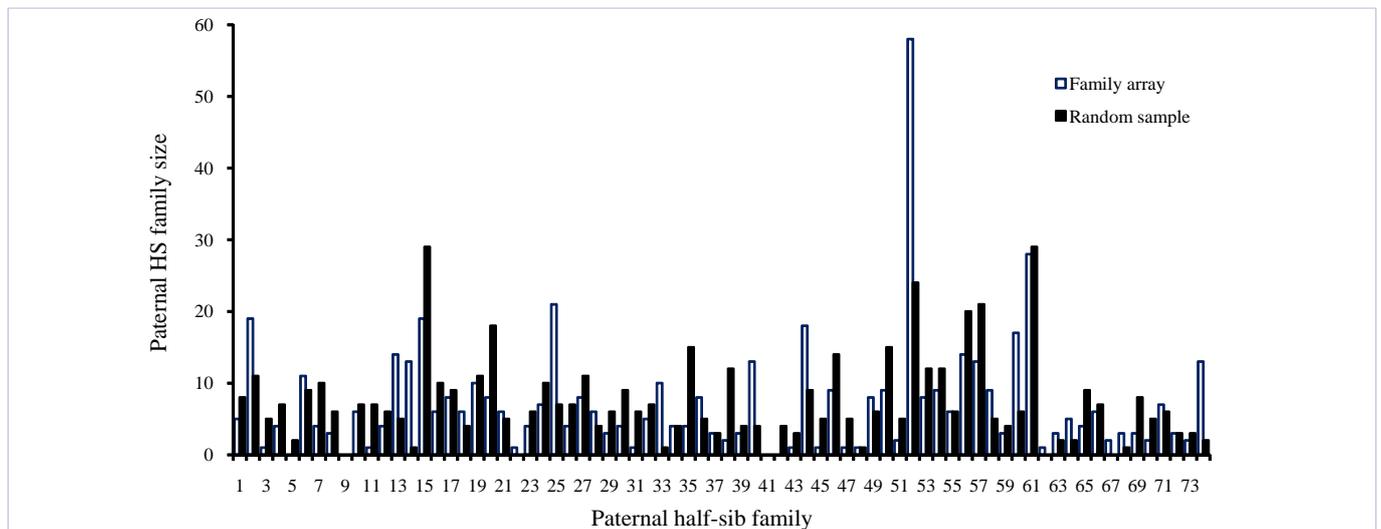


Figure 3a: Comparison of paternal half-sib family sizes obtained from partial (family array) and full (random sample) pedigree reconstruction of offspring from a lodgepole pine seed orchard ($r = 0.61, p < 0.05, N = 74$).

of 180 crosses would have been created. The family array and random sampling produced 268 and 446 crosses, respectively, exceeding that from the disconnected diallel mating design without making a single cross. The resulting crosses offered more mating combinations than those from the disconnected diallel mating design, thus eliminating the sampling caveat of

this design where crosses are restricted to within diallel unites and not among. It should be stated that the use of the nuclear SSR markers, alone, were sufficient in constructing the resulting crosses in the partial pedigree reconstruction as the offspring was collected from known maternal parents and thus the inference of parentage was restricted to the paternal component. On the

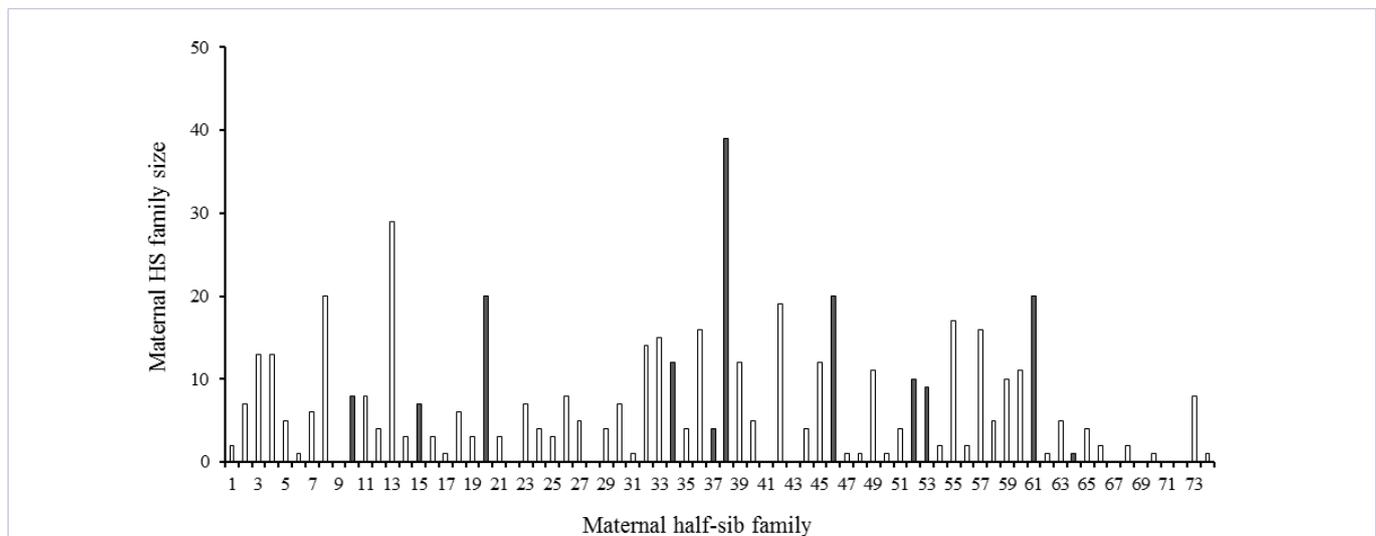


Figure 3b: Maternal half-sib family sizes obtained from full pedigree reconstruction of random sample offspring from a lodgepole pine seed orchard (black bars represent the 11 family arrays studied).

other hand, the identity of the paternal parentage in the bulk seed sample required supplement of an additional set of uniparentally inherited markers, thus cpDNA markers were used to separate males with similar nuclear genotypes [27-29].

Pedigree reconstruction has been extensively used to assess male and female fertility variation as well as selfing and gene flow rates in seed orchard populations [22,30-34]. The use of pedigree reconstruction as a platform for breeding was first proposed by El-Kassaby, et al. [15] and its theoretical foundation was illustrated by El-Kassaby and Lstibůrek [34] using a Douglas-fir retrospective study and was further demonstrated as an avenue for testing and selection of elite genotypes using a combination of assembled full-sib and wind-pollinated half-sib families from a western larch experimental population [16]. However, it should be stated that the work of Lambeth, et al. [14] was inspiring as it demonstrated the power of pedigree reconstruction in determining the male parents of crosses produced through polycross mating design (pollen consisted of a mixture from several male parents, thus paternity is unknown and the resulting families were considered half-sibs) and thus converted a set of half-sib to a full-sib families. Pedigree reconstruction as an aid to breeding has gained momentum and several retrospective studies on *Eucalyptus urophylla* [35], *Pinus pinster* [36], *Abies nordmanniana* [37] and *Picea rubens* [38] have been documented.

In conclusion, based on the present study results, we recommend the use of full pedigree reconstruction using individuals with unknown paternal and maternal parentage to enable the posterior assemblage of naturally occurring crosses among population's members, resulting in the creation of a mating design in the extent that would otherwise only be accomplishable by controlled pollination with extremely high costs and labor efforts.

Acknowledgments

This work was supported by the Johnson's Family Endowment

and the Natural Sciences and Engineering Research Council of Canada Industrial Research Chairs and Discovery Grants to YAE, and the University of British Columbia Graduate Fellowship to TF.

References

- Allard RW. Principles of Plant Breeding. New York: Wiley; 1960.
- White TL, Adams WT, Neale DB. Forest Genetics. Cambridge: CABI; 2007.
- Namkoong G, Kang HC, Brouard JS. Tree Breeding: Principles and Strategies. New York: Springer-Verlag; 1988.
- Falconer DS, Mackay TFC. Introduction to Quantitative Genetics. Harlow: Longman; 1996.
- El-Kassaby YA. Evaluation of the tree-improvement delivery system: factors affecting the genetic potential. *Tree Physiol.* 1995; 15(7-8): 454-50.
- Adams WT, Neale DB, Loopstra CA. Verifying controlled crosses in conifer tree-improvement programs. *Silvae Genet.* 1988; 37(3-4):147-52.
- Devey ME, Bell JC, Uren TL, Moran GF. A set of microsatellite markers for fingerprinting and breeding applications in *Pinus radiata*. *Genome.* 2002; 45(5):984-9.
- Burdon RD, Shelbourne CJA. Breeding populations for recurrent selection: conflicts and possible solutions. *N Z J For Sci.* 1971; 1:174-93.
- Jayawickrama KJS, Carson MJ. A breeding strategy for the New Zealand radiata pine breeding cooperative. *Silvae Genet.* 2000; 49:82-90.
- Namkoong G. Inbreeding effects on estimation of genetic additive variance. *For Sci.* 1966; 12(1):8-13
- Squillace AE. Average genetic correlations among offspring from open pollinated forest trees. *Silvae Genet.* 1974; 23(5):149-56.
- Askew GR, El-Kassaby YA. Estimation of relationship coefficients among progeny derived from wind-pollinated orchard seeds. *Theor Appl Genet.* 1994; 88(2):267-72. doi: 10.1007/BF00225908.

13. Jones AG, Ardren WR. Methods of parentage analysis in natural populations. *Mol Ecol.* 2003; 12(10):2511-23.
14. Lambeth C, Lee BC, O'Malley D, Wheeler N. Polymix breeding with parental analysis of progeny: an alternative to full-sib breeding and testing. *Theor Appl Genet.* 2001; 103(6-7):930-43.
15. El-Kassaby YA, Lstiburek M, Liewlaksaneeyanawin C, Slavov GT, Howe GT. Breeding Without Breeding: Approach, Example, and Proof of Concept. In: Fikret I, editor. Proceedings of the IUFRO Division 2 Joint Conference: Low Input Breeding and Conservation of Forest Genetic Resources; 2006 October 9-13; Antalya, Turkey. p. 43-54.
16. El-Kassaby YA, Cappa EP, Liewlaksaneeyanawin C, Klápště J, Lstibůrek M. Breeding without breeding: is a complete pedigree necessary for efficient breeding? *PLoS One.* 2011; 6(10):e25737. doi: 10.1371/journal.pone.0025737.
17. El-Kassaby YA, Lindgren D. Increasing the efficiency of breeding without breeding through phenotypic preselection in open pollinated progenies. In: Bryam TD, Rust ML, editors. *Tree Improvement in North America: Past, Present, and Future. 29th Southern Forest Tree Improvement Conference and the Western Forest Genetics Association*; 2007 June 19-22; Galveston, TX. p. 15-19.
18. El-Kassaby YA, Lstibůrek M. Breeding without breeding. *Genet Res.* 2009; 91(2):111-20.
19. Bell GD, Fletcher AM. Computer organized orchard layouts (COOL) based on the permuted neighborhood design concept. *Silvae Genet.* 1978; 27:223-5.
20. Doyle JJ, Doyle JL. Isolation of plant DNA from fresh tissue. *Focus.* 1990; 12:13-15.
21. Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K. Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. *Theor Appl Genet.* 2004; 109(2):361-9.
22. Funda T, Chen C, Liewlaksaneeyanawin C, Kenawy A, El-Kassaby YA. Pedigree and mating system analyses in a western larch (*Larix occidentalis* Nutt.) experimental population. 2008; *Ann For Sci.* 65(7):705.
23. Funda T, Liewlaksaneeyanawin C, El-Kassaby YA. Determination of paternal and maternal parentage in lodgepole pine seed: full versus partial pedigree reconstruction. *Can J For Res.* 2014; 44(9):1122-7. doi: 10.1139/cjfr-2014-0145.
24. Stoehr MU, Newton CH. Evaluation of mating dynamics in a lodgepole pine seed orchard using chloroplast DNA markers. *Can J For Res.* 2002; 32(3):469-76. doi: 10.1139/x01-222.
25. Kalinowski ST, Taper ML, Marshall TC. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Mol Ecol.* 2007; 16(5):1099-106.
26. Gilmour AR, Gogel BJ, Cullis BR, Thompson R. *ASReml User Guide, Release 2.0.* Hemel Hempstead: VSN International; 2006.
27. Neale DB, NC Wheeler, RW Allard. Paternal inheritance of chloroplast DNA in Douglas-fir. *Can J For Res.* 1986; 16(5):1152-4. doi: 10.1139/x86-205.
28. Wagner DB, Furnier GR, Saghai-Marooof MA, Williams SM, Dancik BP, Allard RW. Chloroplast DNA polymorphisms in lodgepole and jack pines and their hybrids. *Proc Natl Acad Sci USA.* 1987; 84(7):2097-100.
29. Sutton BC, Flanagan DJ, Gawley JR, Newton CH, Lester DT, El-Kassaby YA. Inheritance of chloroplast and mitochondrial DNA in *Picea* and composition of hybrids from introgression zones. *Theor Appl Genet.* 1991; 82(2):242-8. doi: 10.1007/BF00226220.
30. Moriguchi Y, Taira H, Tani N, Tsumura Y. Variation of paternal contribution in a seed orchard of *Cryptomeria japonica* determined using microsatellite markers. *Can J For Res.* 2004; 34(8):1683-90. doi: 10.1139/x04-029
31. Hansen OK, Kjær ED. Paternity analysis with microsatellites in a Danish *Abies nordmanniana* clonal seed orchard reveals dysfunctions. *Can J For Res.* 2006. 36:1054-8.
32. Doreksen TK, Herbinger CM. Male reproductive success and pedigree errors in red spruce open-pollinated and polycross mating systems. *Can J For Res.* 2008; 38(7):1742-9. doi: 10.1139/X08-025.
33. El-Kassaby YA, Funda T, Lai BS. Female reproductive success variation in a *Pseudotsuga menziesii* seed orchard as revealed by pedigree reconstruction from bulk seed collection. *J Hered.* 2010; 101(2):164-8. doi: 10.1093/jhered/esp126.
34. El-Kassaby YA, Lstibůrek M. Breeding without breeding. *Genet Res.* 2009; 91(2):111-20.
35. Grattapaglia D, Ribeiro VJ, Rezende GDSP. Retrospective selection of elite parent trees using paternity testing with microsatellite markers: an alternative short term breeding tactic for *Eucalyptus*. *Theor Appl Genet.* 2004; 109(1):192-9.
36. Gaspar MJ, de-Lucas AI, Alia R, Paiva JAP, Hidalgo E, Louzada J, et al. Use of molecular markers for estimating breeding parameters: a case study in a *Pinus pinaster* Ait. progeny trial. *Tree Genet Genomes.* 2009; 5(4):609-16.
37. Hansen OK, McKinney LV. Establishment of a quasi-field trial in *Abies nordmanniana* - test of a new approach to forest tree breeding. *Tree Genet Genomes.* 2010; 6(2):345-55.
38. Doreksen TK, Herbinger CM. Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. *Tree Genet Genomes.* 2010; 6(4):591-600.