Tree Genetics & Genomes (2012) 8:873–877
DOI 10.1007/s11295-012-0472-0

ORIGINAL PAPER

Breeding without Breeding

Effect of gene flow on fingerprinting effort

Milan Lstibůrek · Jaroslav Klápště · Jaroslav Kobliha · Yousry A. El-Kassaby

Received: 2 August 2011 / Revised: 3 January 2012 / Accepted: 30 January 2012 / Published online: 28 February 2012

Abstract We developed a deterministic model to optimize DNA fingerprinting effort in the presence of gene flow during the application of Breeding without Breeding. The method considers trait's heritability, level of gene flow, selection differential, and the proportion of progeny test subjected to fingerprinting (truncation). All the model's variables individually or in concert proved important; however, the trait's heritability magnitude played the most important role in minimizing the selection of contaminated individuals and it could be used as a surrogate to traits with low heritability.

Keywords Breeding without Breeding · Pedigree reconstruction · DNA fingerprinting · Contamination

Introduction

The concept "Breeding without Breeding" (BwB) (El-Kassaby and Lstibůrek 2009) is presented as an alternative tree improvement strategy, where conventional breeding is bypassed, and testing is simplified through using either wind-pollinated half-sib families or plantations. The pedigree of half-sib families (partial pedigree) or plantations (full pedigree) is determined using a combination of DNA fingerprinting and paternity assignment or full pedigree reconstruction, respectively. The seedling sources of these half-sib families and/or plantations are expected to be drawn from populations with known parentage such as seed orchards or breeding arboreta (El-Kassaby and Lstibůrek 2009; El-Kassaby et al. 2011).

Forest trees are characterized by their large populations, wind-pollinated mating system, and extensive gene flow (Hamrick et al. 1995). Seed orchard populations are known to be the subject of extensive gene flow (Slavov et al. 2005; Lai et al. 2010) substantially affecting their genetic efficiency (El-Kassaby 1989). The extent of gene flow is determined by within-orchard pollen production intensity (Chung 1981) and the degree of reproductive asynchrony among the orchard's parents as well as outside pollination sources (El-Kassaby et al. 1989; Nikkanen et al. 2002; Torimaru et al. 2009).

DNA fingerprinting of offspring produced from a seed orchard or breeding arboretum, followed by pedigree reconstruction, is expected to yield uninformative genotyping events that are proportional to the level of contamination (i.e., the higher the gene flow, the greater the number of uninformative genotyping events), thus resulting to wasted resources for the effective implementation of the BwB procedure. Since the genetic advance between the seed orchard's parental population and the contaminating pollen source varies between modest to high, it is expected that gene flow will diminish the genetic quality of orchard's crops. This prompted El-Kassaby and Lindgren (2007) to propose phenotypic preselection (truncation

Communicated by A. Kremer

M. Lstibůrek (✉) · J. Klápště · J. Kobliha
Faculty of Forestry and Wood Sciences,
Czech University of Life Sciences Prague,
Kamýcká 129, 165 21 Praha 6, Czech Republic
e-mail: lstiburek@gmail.com, lstiburek@fld.czu.cz

Y. A. El-Kassaby · J. Klápště
Department of Forest Sciences, Faculty of Forestry,
University of British Columbia, 2424 Main Mall,
V6T 1Z4 Vancouver, British Columbia, Canada

selection) prior to fingerprinting in order to minimize the inclusion of contaminant offspring. Thus, truncation selection is expected to focus on the offspring produced from within orchard's matings.

In the present study, we developed a deterministic model to investigate the efficiency of phenotypic preselection in an attempt to minimize the fraction of offspring sired by alien pollen. This was demonstrated for various values of gene flow, selection differential between the orchard's parental population and contamination source, traits' heritabilities, and intensity of preselection prior to genotyping. The present study represents a quantitative genetic approach in dealing with gene flow to reduce fingerprinting efforts, a method that is different from that of Lstibůrek et al. (2011) on effective population size-driven probabilistic model used to predict the minimum genotyping effort needed to satisfy requirement in target population.

Deterministic model

Parental population

Let us assume a base population (natural stands) consisting of an infinite number of unrelated and noninbred selection candidates. The phenotypic variance of a specific polygenic trait is as follows:

$$\sigma_P^2 = \sigma_A^2 + \sigma_E^2 \tag{1}$$

where

$$P \sim N(\mu, \sigma_P^2) \tag{2}$$

$$A \sim N(\mu, \sigma_A^2) \tag{3}$$

$$E \sim N(0, \sigma_E^2) \tag{4}$$

where μ is the average of the base unselected population, and σ_A^2 and σ_E^2 depict additive and environmental variances, respectively. If a portion of the base population is phenotypically selected to form a first-generation seed orchard with intensity of selection i (calculated based on the selected proportion p; Falconer and Mackay 1996), then the selection differential will be as follows:

$$S_{SO} = i\sigma_P \tag{5}$$

and the trait's distribution in the orchard's population could be represented by the following:

$$P \sim N(S_{SO}, (1-k)\sigma_P^2) \tag{6}$$

$$A \sim N(S_{SO}, (1-h^2k)\sigma_A^2) \tag{7}$$

$$E \sim N(0, \sigma_E^2) \tag{8}$$

where $k = i(i - x)$ and x is the deviation of the truncation point from the mean in standard deviation units, and h^2 represents narrow-sense heritability of the studied trait (Falconer and Mackay 1996).

If the seed orchard's pollen pool is subjected to external gene flow, then it can be assumed that a proportion r $(0 \le r \le 1)$ of the siring pollen will be the product of contamination with the attributes presented in Eqs. 2–4, while the remaining pollen contribution is from the orchard's male parents (Eqs. 6–8).

Offspring composition

Assuming the orchard is in panmictic equilibrium and contamination is random, then the offspring could be classified in two distinctive distributions (see Fig. 1 for conceptual representation) with the following properties:

1. When the offspring is produced from within-orchard's pollen pool, it can be described as follows:

$$P \sim N(h^2 S_{SO}, (1 - 0.5h^4k)\sigma_P^2) \tag{9}$$

$$A \sim N(h^2 S_{SO}, (1 - 0.5h^2k)\sigma_A^2) \tag{10}$$

$$E \sim N(0, \sigma_E^2) \tag{11}$$

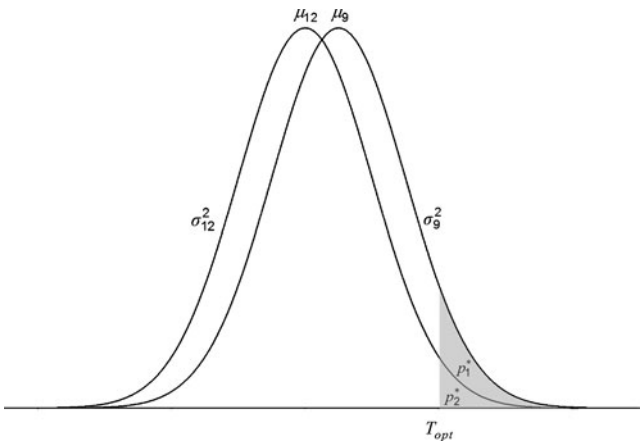

Fig. 1 Conceptual representation of two offspring populations produced from within the orchard's pollen pool (*right*, μ_9 and σ_9^2 corresponding to Eq. 9) and those sired by outside pollen (*left*, μ_{12} and σ_{12}^2 corresponding to Eq. 12). The area beyond the truncation point denotes the probabilities p_1^* (Eq. 9) and p_2^* (Eq. 12), representing within-orchard and outside pollination, respectively. The actual parameters (i.e., means and variances) depend on the input. T_{opt} is the truncation point produced by Solver optimization

2. When the offspring is sired by outside pollen, it can be described as follows:

$$P \sim N(h^2(S_{SO} + \mu)/2, (1 - 0.5h^4k')\sigma_P^2) \quad (12)$$

$$A \sim N(h^2(S_{SO} + \mu)/2, (1 - 0.5h^2k')\sigma_A^2) \quad (13)$$

$$E \sim N(0, \sigma_E^2) \quad (14)$$

where k' is calculated as the unweighted average (different selection intensity on male and female component).

Gene flow (contamination)

The orchard's offspring is used for progeny testing and evaluated (i.e., ranked) solely based on phenotypic performance. Then proportion (p^*) is selected following truncation (El-Kassaby and Lindgren 2007) and fingerprinted for paternity assignment (i.e., partial pedigree reconstruction). This proportion is expected to harbor an offspring from the two distinctive distributions (i.e., sired from pollen inside (p_1^*) and outside (p_2^*) the orchard's pollen pool), where

$$p^* = (1 - r)p_1^* + rp_2^* . \quad (15)$$

The value of r is set as an input in Eq. 15, and p_1^* and p_2^* are calculated using the normal distribution function in Excel (NORMDIST). This function returns the probability that a number falls at or below a given value of a normal distribution (for the specified number, mean, and standard deviation). Each distribution's mean and variance are known; thus, the only unknown variable is the truncation point (common to both distributions, Fig. 1). Next, with the use of the "Solver" optimization tool in Excel, the truncation point is determined, so that p^* is meeting the requested proportion of progenies that are fingerprinted (set as input). The Microsoft Excel Solver is an optimization tool, which is included in the Microsoft Office.

Finally, the expected gene flow rate (r^*) in the fingerprinted sample is calculated as follows:

$$r^* = \frac{rp_2^*}{(1 - r)p_1^*} \quad (16)$$

Input parameters

The interplay among h^2, r, p, and p^* on the expected contamination rate detected in the fingerprinted sample r^* is evaluated using the following values: $h^2 = 0.2$, and 0.6; $r = 0$, 0.1, ..., 1; $p = 0.01\%$, 1%; and $p^* = 1\%$, 5%, 10%, and 20%. A summary of the model's parameters is provided in Table 1.

Table 1 Model's parameters

σ_A^2	Additive genetic variance
σ_E^2	Environmental variance
σ_P^2	Phenotypic variance
h^2	Narrow-sense heritability
μ	Average of the base population
p	Proportion selected (seed orchard parents)
p^*	Proportion selected to fingerprinting (following truncation)
p_1^*	Proportion of fingerprinted offspring sired by orchard parents
p_2^*	Proportion of fingerprinted offspring sired by alien parents
r	Contamination rate in the orchard
r^*	Expected contamination rate in the fingerprinted sample

The deterministic model was developed in a Microsoft Excel worksheet and is available upon request (ML).

Results and discussion

The results from the various scenarios presented in Fig. 2 show several general trends meeting our expectations. Decrease in either trait's heritability or the selection differential (difference between the orchard and the base populations) was associated with higher proportion of offspring resulting from contaminant pollen. Within each scenario, the proportion of offspring resulting from contaminant pollen increased with lower intensity of preselection (i.e., higher truncation proportion) (Fig. 2). Additionally, as the gene flow increases, it is expected that the proportion of contaminated offspring also increases (see individual lines with all scenarios, Fig. 2).

After the selection was completed from the base population and seed orchards were established, it became apparent that several factors are out of the breeder's control such as (a) the selection differential between the base population and the selected seed orchard population, (b) the level of background pollen contamination, and (c) trait's heritability. Traits' heritabilities can be enhanced through the application of advanced genetic evaluation methodology (e.g., BLUP procedure) assisted by spatial statistics to remove undetected environmental noise (El-Kassaby and Lstibůrek 2009). Then, in addition to heritability improvement, the only major factor affecting the amount of uninformative genotyping events in the selected offspring is the truncation. So, if the truncation step is conducted using moderate to high heritable characters, the likelihood of wasted efforts is lowered. This could be visualized

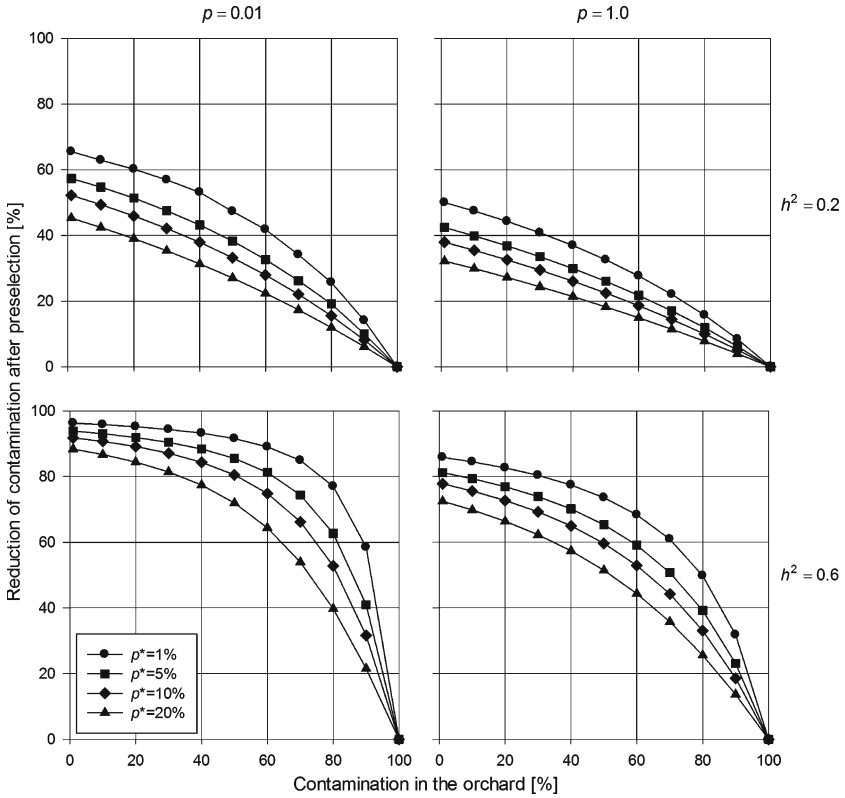

Fig. 2 Relationship between the original contamination rate in the orchard (r, X-axis) and the resultant reduction of contamination among the fingerprinting progenies after preselection ($(r^* - r)/r$, Y-axis). *Graphs* represent combinations of two h^2 values (*rows*), two selection proportion values p (*columns*), and four fingerprinted proportion values p^* (*individual lines*: 1, 5, 10, and 20% within each plot)

as a surrogate trait to reduce the gene flow-resulting individuals in the truncation sample even if this trait is not the target of selection. Thus, the genotyping effort results will be maximized. However, caution should be given to the possible negatively correlated response (Falconer and Mackay 1996). Therefore, under multi-trait selection, the narrow-sense heritabilities and genetic covariance among traits will determine whether there is a suitable surrogate trait in line with maximizing genetic response to selection.

This approach will eventually increase the number of individuals produced from inside the orchard in the genotyping step and thus will reduce type II error during the paternity assignment process, as compared to the anticipated, if high percentage of contaminated individuals were included (Pompanon et al. 2005). It should be stated that the paternity assignment phase needed for pedigree reconstruction could be completed under any contamination level; however, our object is to reduce the wasted fingerprinting effort (Lai et al. 2010).

Gene flow to seed orchards can be effectively reduced with the application of supplemental mass pollination (Wakeley et al. 1966). This pollen augmentation technique has proven effective (El-Kassaby and Davidson 1990) with a single application success rate of 8% (El-Kassaby et al. 1993). Therefore, in situations where gene flow is high and/or the seed orchard population requires expansion (i.e., addition of new parents), this technique could be considered as an option (El-Kassaby and Lstibůrek 2009). Additionally, roguing of first-generation seed orchards is expected to greatly increase the effectiveness of BwB, as the removal of inferior parents will widen the selection differential gap between the orchard's improved parental population and outside pollen pool.

Conclusion

It is plausible, even under high contamination levels in seed orchards, to successfully reconstruct the pedigree; however, high magnitude of uninformative genotyping events results in increased genotyping effort (sample size) as a compensation. In this study, we proved (using a standard quantitative trait and a full range of important variables) that high contamination rate is not fully imprinted into the elite part of the seed orchard progenies (subjected to screening and selection during the BwB implementation). Consequently, the number of uninformative genotyping events is reduced, and the genotyping effort is lower than the one expected based on the initial contamination rate in the orchard.

Acknowledgements Funds from the Natural Sciences and Engineering Research Council of Canada, Industrial Applied Forest Genetics and Biotechnology to senior research chair Y. A. El-Kassaby, and from the National Agency for Agricultural Research (research grants no. QH81172 and QH81160), and the Technology Agency of the CR (grant no. TA01020512) to M. Lstibůrek and J. Kobliha are hugely appreciated.

References

Chung MS (1981) Flowering characteristics of *Pinus sylvestris* L. with special emphasis on the reproductive adaptation to local temperature factor. Acta For Fenn 169:5–68

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

El-Kassaby YA, Lindgren D (2007) Increasing the efficiency of breeding without breeding through phenotypic pre-selection in open pollinated progenies. In: Joint meeting of the South. For. tree improve. conf. and the Western for. genetics association, Galveston, Texas, pp. 15–19

El-Kassaby YA, Barnes S, Cook C and MacLeod DA (1993) Supplemental-mass-pollination success rate in a mature Douglas-fir seed orchard. Can J For Res 23:1069–1099

El-Kassaby YA, Cappa EP, Liewlaksaneeyanawin C, Klápště J, Lstibůrek M (2011) Breeding without breeding: is a complete pedigree necessarily for efficient breeding? PLoS ONE 6(10):e25737

El-Kassaby YA, Davidson R (1990) Impact of crop management practices on the seed crop genetic quality in a Douglas-fir seed orchard. Silvae Genet 39:230–237

El-Kassaby YA, Rudin D, Yazdani R (1989) Levels of outcrossing and contamination in two *Pinus sylvestris* L. seed orchards in northern Sweden. Scan J For Res 4(1): 41–50

El-Kassaby YA (1989) Genetics of seed orchards: expectations and realities. In: 20th South for tree improv conf, Charleston, SC, USA, pp 87–109

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Longmans Green, Harlow, Essex, UK, 464 p

Hamrick JL, Godt MJW, Sherman-Broyles SL (1995) Gene flow among plant populations: evidence from genetic markers. In: Hoch PC, Stephenson AG (eds) Experimental and molecular approaches to plant biosystematics. Missouri Botanical Garden Press, St. Louis, pp 215–232

Lai BS, Funda T, Liewlaksaneeyanawin C, Klápště J, Van Niejenhuis A, Cook C, Stoher MU, Woods J and El-Kassaby YA (2010) Pollination dynamics in a Douglas-fir seed orchard as revealed by pedigree reconstruction. Ann For Sci 67(8):808

Lstibůrek M, Ivanková K, Kadlec J, Kobliha J, Klápště J, El-Kassaby YA (2011) Breeding without breeding: minimum fingerprinting effort with respect to the effective population size. Tree Genet Genomes 7:1069–1078

Nikkanen T, Pakkanen A, Heinonen J (2002) Temporal and spatial variation in airborne pollen and quality of the seed crop in a Norway spruce seed orchard. Forest Gen 9(3):243–255

Pompanon F, Bonin A, Bellemain E and Taberlet P (2005) Genotyping errors: causes, consequences and solutions. Nat Rev Genet 6:847–859

Slavov GT, Howe GT, Adams WT (2005) Pollen contamination and mating patterns in a Douglas-fir seed orchard as measured by simple sequence repeat markers. Can J For Res 35: 1592–1603

Torimaru T, Wang X-R, Fries A, Andersson B, Lindgren D (2009) Evaluation of pollen contamination in an advanced Scots pine seed orchard in Sweden. Silvae Genet 58:262–269

Wakeley PC, Wells OO, Campbell TE (1966) Mass production of shortleaf *x* slash pine hybrids by pollinating unbagged female flowers. In: Proc 2nd genetics workshop of Soc Am foresters and 7th Lake States for tree improve conf res pap NC, pp 78–79