Tree Genetics & Genomes
DOI 10.1007/s11295-013-0630-z

ORIGINAL PAPER

Comparison of genetic parameters from marker-based relationship, sibship, and combined models in Scots pine multi-site open-pollinated tests

Jiří Korecký · Jaroslav Klápště · Milan Lstibůrek · Jaroslav Kobliha · C. Dana Nelson · Yousry A. El-Kassaby

Received: 24 September 2012 / Revised: 1 March 2013 / Accepted: 8 April 2013

Abstract Nine microsatellite DNA markers (simple sequence repeats, SSRs) were used to estimate pairwise relationships among 597 Scots pine (*Pinus sylvestris*) trees as well as to generate a sibship structure for quantitative genetic parameters' estimation comparison. The studied trees were part of an open-pollinated progeny test of 102 first-generation parents. Three methods were used to estimate variance components and heritabilities, namely, structured pedigree (half- and full-sib), marker-based pairwise relationships (four pairwise estimators), and a combined pedigree and marker-based relationship. In each of the three methods, the same animal model was used to compute variances except when marker-based relationship was used wherein we substituted the average numerator relationship matrix (i.e., pedigree-based matrix) with that computed based on markers' pairwise relationships. Our results showed a high correlation in estimated breeding values between the pedigree (full-sib) and the combined marker-pedigree estimates. The marker-based relationship method produced high correlations when individual site data were analyzed. In contrast, the marker-based relationship method resulted in a significant decrease in both variance estimation and their standard errors which were in concordance with earlier published results; however, no estimates were produced when across-site analyses were attempted. We concluded that the combined pedigree method is the best approach as it represents the historical (pairwise) and contemporary (pedigree) relationships among the tested individuals, a situation that cannot be attained by any of the used methods individually. This method is dependent on the number and informativeness of the markers used.

Keywords Marker-based pairwise relationship · Combined pedigree–marker-based relationship · Quantitative genetic parameters · Scots pine

Communicated by Y. Tsumura

Electronic supplementary material The online version of this article (doi:10.1007/s11295-013-0630-z) contains supplementary material, which is available to authorized users.

J. Korecký · J. Klápště · M. Lstibůrek · J. Kobliha
Department of Dendrology and Forest Tree Breeding, Faculty of Forestry and Wood Sciences, Czech University of Life Sciences Prague, Kamycka 129, 165 21, Praha, Czech Republic

J. Klápště · Y. A. El-Kassaby (✉)
Department of Forest and Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, British Columbia V6T 1Z4, Canada
e-mail: y.el-kassaby@ubc.ca

C. D. Nelson
Southern Institute of Forest Genetics, USDA Forest Service, Southern Research Station, 23332 Success Road, Saucier, MS 39574-9344, USA

Introduction

Most tree improvement recurrent selection programs encompass a traditional sequence of activities, namely, selection, breeding, and testing leading to the identification of elite genotypes for further breeding as well as their inclusion in production populations (Namkoong et al. 1988; White et al. 2007). For example, in the Czech Republic, the Scots pine (*Pinus sylvestris* L.) breeding program was initiated by the establishment of several first-generation seed orchards, which was followed by progeny testing from orchards' produced open-pollinated offspring (Kaňák et al. 2009). The original plan was to evaluate all open-pollinated progeny following a half-sib model assumption and select a set of parents for incorporation in controlled mating scheme for further advanced generations' selection. This strategy was altered for two reasons. First, in the data analysis phase of

Published online: 12 May 2013

open-pollinated progeny trials, several important assumptions must be invoked, namely, parents are unrelated and noninbred, members of an open-pollinated family are equivalent to half-sib family with individuals sired by a large number of male parents, and absence of selfing. It is well known that these assumptions could not be fulfilled and thus the estimated additive genetic variance could be substantially inflated resulting in erroneous ranking of plus trees (see below and Squillace 1974; Namkoong et al. 1988; Askew and El-Kassaby 1994). The second reason was to take advantage of a novel approach in quantitative genetic data analyses, which uses molecular markers' pairwise relationship among individuals to estimate genetic parameters (Visscher et al. 2006).

Porth et al. (2013) utilized the marker-based relationship to convert trials with unknown (provenance testing) to full-fledged tests for estimating genetic parameters such as heritabilities and genetic correlations. Furthermore, Bömcke and Gengler (2009) introduced the concept of combining the molecular markers' relationship and pedigree in one analysis. The combined molecular marker–pedigree model is very effective as combining DNA fingerprinting data helps in providing historical information on co-ancestry with contemporary complete and/or incomplete parent/pedigree knowledge. This approach is unique as the combination of historical co-ancestry and contemporary pedigree could not be attained by the pedigree or marker-based relationship approaches individually. Thus, we expect the accuracy of genetic parameters to progressively increase from (1) open-pollinated to half- and full-sib families to (2) those based on markers' relationship estimates to (3) those combining genetic marker's relationship and pedigree information. This situation should change if a large number of loci are used, and the marker-based relationship should surpass all the listed approaches. It is noteworthy to mention that the often ignored Mendelian sampling term in structured pedigree is precisely accounted for with the use of molecular markers.

In this paper, we compared the genetic variance components, including heritability, and individuals' breeding value estimates generated from sibship composition (pedigree) generated using nine microsatellite markers (SSRs) to those based on molecular markers' pairwise relationship as well as the combined use of pedigree and marker information. Comparing these three methods is expected to provide insight on the best approach for assessing the Czech Republic's Scots pine progeny trials and to aid in the selection of elite genotypes for the second-generation breeding cycle as well as the establishment of advanced generations' production populations (seed orchards).

Materials and methods

Scots pine seed orchard's open pollinated (half-sib) offspring was used to establish two progeny trials (see details in Cappa et al. 2011). These progeny trials were established in 1994 (site 1) and 1991 (site 2). The offspring source for these progeny trials originated from a seed orchard established in 1982 and contained 102 parents (clones). In summary, the experiment consisted of half-sib progenies planted in a completely randomized design with 50-tree contiguous plots at a close spacing of 1.4×0.7 m. On average, 200 seedlings were planted per each half-sib family with four replicates per family (the number of maternal parents present in the two sites varied (site 1:85 and site 2: 38)). A total of 597 individuals (approximately 10 % of the progeny trials) were selected based on their height (ht), diameter at breast height (dbh), straightness, and branching (self-pruning and branch size) (i.e., truncation selection). These traits were considered as the selection criteria for establishing the next-generation seed orchard and facilitating future breeding (see Kaňák et al. 2009, for selection procedure details).

In the fall of 2011, dormant vegetative buds were collected from the selected 597 trees and immediately stored at 80 °C to avoid DNA degradation. DNA was extracted using Invisorb® Spin Plant Kit. Trees were genotyped using nine simple sequence repeat (SSR) markers: SPAC 11.4, SPAC 11.6, SPAC 12.5 (Soranzo et al. 1998: *P. sylvestris*-specific markers), LOP 1 (Liewlaksaneeyanawin et al. 2004: *Pinus taeda*-specific markers), PtTX 2146, PtTX 3025, PtTX 3107, PtTX 4001, and PtTX 4011 (Auckland et al. 2002: *P. taeda*-specific markers).

Pedigree method

In addition to the half-sib designation of each seedling, we used the COLONY program (Jones and Wang 2010) to construct the sampled trees' full pedigree according to the observed relationships in Supplement 1, and the COLONY program output that determined the posterior paternal likelihood for each tree and full-sib families was assembled for the subsequent sibship analysis. We assumed no selfing in the offspring generation as Scots pine is known for its high outcrossing rate of ≈0.987±0.005 (Burczyk 1998).

Marker-based relationship method

Four pairwise molecular relationship coefficient estimation methods were used to generate the all possible pairwise marker-based relationship matrices for their use in the subsequent quantitative genetic analyses (Li et al. 1993; Lynch and Ritland 1999; Queller and Goodnight 1989; Wang 2002). Pairwise relationship matrices were estimated using SPAGeDi v. 1.3 software (Hardy and Vekemans 2002). The details of the four pairwise relationship methods used are described in their respective references and in the work by Bink et al. (2008) for more details. The pairwise marker-

based relationship matrix which was used to generate all the possible estimators (Li et al. 1993; Lynch and Ritland 1999; Queller and Goodnight 1989; Wang 2002) was used to replace the average numerator relationship matrix in the quantitative genetic analyses (see below).

Combined pedigree and marker-based relationship method

Pedigree records from the sibship analysis (i.e., full-sib information) and molecular marker relationship were used to estimate revised relationships combining both sources of information (Bömcke and Gengler 2009). Briefly, the combined relationship estimates are computed as weighted pedigree-based and marker-based relatedness. The weights are based on the number of generation-equivalents (geq), which was computed for each individual as the sum of $(1/2)^n$, where n is the number of generations separating the individual from each known ancestor (Huby et al. 2003). The pedigree-based relationship was estimated according to Wright et al. (1925) relationship coefficient. The molecular relatedness was estimated as follows:

$$r_{mol,xy} = \frac{\mathrm{TA}_{xy}}{\sqrt{\mathrm{TA}_{xx}\mathrm{TA}_{yy}}} \quad (1)$$

where TA is the total allelic identity and x and y are indexes of individuals. The TA_{xy} is computed, following Nejati-Javaremi et al. (1997), as

$$\mathrm{TA}_{xy} = \frac{\sum_{l=1}^{L} \mathrm{TA}_{xy,l}}{L} \quad (2)$$

where $\mathrm{TA}_{xy,l}=2\ fM_{xy,l}$. The coefficient of 2 defines TA as twice the relationship coefficient (Malécot 1948) and is analogous to the average numerator relationship based on pedigree (Wright 1922). The $fM_{xy,l}$ is the molecular co-ancestry between two individuals (Caballero and Toro 2000) and is computed from similarity of alleles between individuals as

$$fm_{xy,l} = \frac{1}{4}[S_{ac} + S_{ad} + S_{bc} + S_{bd}] \quad (3)$$

where l indicates a specific locus, a and b are alleles of individual x, and c and d are alleles of individual y (Li et al. 1993). $S..=1$ if alleles at the allelic positions in the subscripts are the same and $S..=0$ if they are not. The weight based on the number of generation-equivalents (geq) follows

$$w_{x,y} = \frac{\sqrt{(1+\mathrm{geq}_x)(1+\mathrm{geq}_x)}}{\left(2+\left(1+\overline{\mathrm{geq}}\right)\right)} \quad (4)$$

where geq_i is defined as the number of generation-equivalents for individual i and $\overline{\mathrm{geq}}$ is the average number of generation-equivalents for the analyzed population. Finally, the combined relationship estimates are computed as

$$r_{\mathrm{comb},xy=} w_{xy} \times r_{\mathrm{ped},xy} + (1 - w_{x,y}) \times \mathrm{PIC}_{\mathrm{mean}} \times r_{\mathrm{mol},xy} \quad (5)$$

where $w_{x,y}$ is the above defined weight, $r_{\mathrm{ped},xy}$ is the pedigree-based relationship between individuals x and y, $\mathrm{PIC}_{\mathrm{mean}}$ (obtained from Cervus software; Kalinowski et al. 2007) is the average polymorphic information content coefficient (Botstein et al. 1980) (Supplement 1) for markers used for molecular relatedness calculations, and $r_{\mathrm{mol},xy}$ is the molecular relatedness between individuals x and y.

Molecular marker-based relationship estimates often produce not positive definite (PD) matrices due to internal inconsistencies; thus, we applied the "nearPD" function implemented in the R package "Matrix" to compute the nearest positive definite matrix and use it in the subsequent analyses (i.e., marker-based and combined methods) (Cheng and Higham 1998; Higham 2002; Knol and Tenberge 1989). The modified matrices were checked to be positive definite and the function nearPD in the R package Matrix was used to modify them as positive definite (Porth et al. 2013).

Data analyses

The phenotypic data were normalized and standardized to overcome issues related to variance heterogeneity across sites (Cappa et al. 2011). Variance component estimates, heritability, and individuals' breeding values and their level of precision (SE) were estimated with the restricted maximum likelihood animal model (Kruuk 2004) in ASReml v. 3.0 (Gilmour et al. 2002) as follows:

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{u} + \boldsymbol{e} \quad (6)$$

where $\mathbf{X}$ and $\mathbf{Z}$ are the incidence matrices relating the fixed effect in vector $\boldsymbol{\beta}$ (population mean and site) and random effects in $\boldsymbol{u}$ (individual breeding values) to measurements in vector, $\boldsymbol{y}$ and $\boldsymbol{e}$ is vector of residuals following $e \sim N(O,\sigma_e^2)$. Vector of breeding values follows $\mathrm{Var}(u) = \mathbf{A}\sigma_a^2$, where σ_a^2 is additive genetic variance and $\mathbf{A}$ is (1) the average numerator relationship matrix in the pedigree method based on both half-sib and sibship reconstruction which is considered as reference model, (2) the combined relationship matrix in the scenario where coefficients of relationship are estimated on both pedigree and marker information, and (3) the marker-based relationship matrix in scenarios based on pairwise relationships estimated with only molecular markers. This general model (6) is valid for single site, pedigree (half-sib)-, and/or marker-based relationship methods. It should be stated that the pedigree reconstruction has created a situation where dominance genetic variance can be estimated; thus, it required

additional modification (i.e., the inclusion of the dominance term). The dominance variance term was used when sibship and/or combined pedigree and marker-based methods were used. Furthermore, for across-site data analyses, an additional genotype x site interaction term is also added (see below).

The general model narrow-sense heritability was estimated as follows:

$$h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2} \quad (7)$$

where ${\sigma_e}^2$ is the additive genetic variance and ${\sigma_e}^2$ is the residual variance. The above-mentioned model modifications required adjustment of heritability estimates to account for the dominance genetic variance and genotype × site interaction terms for the sibship and across-site analyses, respectively.

The accuracy of the predicted breeding values across all the studied methods was calculated as follows:

$$r = \sqrt{1 - \frac{\mathrm{PEV}}{(1 + F_i) + \sigma_a^2}} \quad (8)$$

where PEV is the "prediction error variance" of predicted breeding values (Gilmour et al. 1995) and F_i is the inbreeding coefficient for the *i*th individual.

Results

Since the offspring used to establish the two open-pollinated progeny test trials originated from the same seed orchard's crop year, the selected individuals from both sites were used in the sibship analysis to construct their family membership (Supplement 2). Supplement 2 shows the formation of full- and half-sibship composition confirming offspring authenticity. The full description of familial composition at each and across sites is summarized in Table 1. Maternal (20 vs. 13) and paternal (62 vs. 21) distribution across the two sites, based on the selected offspring, showed increased representation in site 1 as compared to site 2, indicating that the phenotypic selection was concentrated on fewer maternal parents (28 out of 102 present in the seed orchard from which progeny trials originated) (Table 1). While site 1 has a greater parental representation, the sibship distribution in site 2 is better as indicated by greater mean full-sib family (3.5 vs. 2.6) and mean female (0.08 vs. 0.05) and male (0.05 vs. 0.02) contributions (Table 1). The number of common maternal and paternal parents common between sites is 9 and 17, respectively.

It is clearly shown that the utilization of the half-sib resulted in a substantial inflation to the additive genetic variance and subsequently height heritability compared to its full-sib counterpart (Table 2). Such inflation of the additive genetic variance was produced on the expense of the residual term (Table 2). Additional distinction between the two pedigree models (i.e., half and full-sib) is our ability to account for the dominance genetic variance and hence estimating the heritability with greater precision (site 1, 0.53 vs. 0.09 and site 2, 0.52 vs. 0.13). Generally, the heritability estimates of site 1 were lower than that of site 2; this is probably due to the differences in paternal and maternal representation between these two sites (Table 1).

The sibship analysis, which was considered as the reference model, produced similar height heritability estimates for each site individually (0.09–0.13) as well as across sites (0.11). The combined marker–pedigree estimator produced similar height heritability estimates for across sites (0.13); however, individual site's estimates varied between 0.10 and 0.17 (Table 2). It is interesting to note that even the differences in variance components apportionment and their level of precision, expressed by their standard errors, between the two sites were mirrored between the two methods (i.e., full-sib (${\sigma_e}^2$ 0.733–0.697) and combined marker–pedigree (${\sigma_e}^2$ 0.670–0.750)). While consistent estimates were produced for the residual variances between the two sites, differences in the dominance genetic variance were apparent (${\sigma_d}^2$ 0.358 vs.

Table 1 Description of familial structure at each site separately and across both sites (number of planted half-sib families/site)

Parameter	Site 1	Site 2	Across both sites
Number of individuals	284	313	597
Full-sib family size	1–30	1–20	1–30
Mean full-sib family size	2.6	3.5	3.0
Variance of FS size	15.376	14.071	15.122
Number of contributing females	20 (85)	13 (38)	28
Mean female contribution	0.05	0.08	0.04
Variance in female contribution	0.0013	0.0010	0.0004
Number of contributing males	62	21	66
Mean male contribution	0.02	0.05	0.02
Variance in male contribution	0.001	0.003	0.001

Table 2 Variance components (additive, dominant, genotype × site (G × S), residual, and total phenotypic and their standard errors) and comparison of heritability estimates among pedigree (half- and full-ship), combined marker and pedigree information, and all marker-based pairwise relationships as well as the accuracy of the predicted breeding values (r)

	Method	Additive	Dominance	G × S	Residual	Phenotypic	Heritability	r
Variance components – site 1								
Pedigree	Half-sib	0.546 (0.295)	N/A	–	0.483 (0.242)	1.028 (0.103)	0.531 (0.256)	0.506
	Full-sib	0.211 (0.166)	0.358 (0.368)	–	0.733 (0.122)	1.033 (0.099)	0.087 (0.088)	0.438
Combined		0.324 (0.172)	0.450 (0.358)	–	0.670 (0.104)	1.106 (0.122)	0.102 (0.080)	0.502
Marker-based relationship	Li et al.	0.053 (0.074)	N/A	–	0.946(0.105)	0.999 (0.086)	0.054 (0.073)	0.284
	Wang	0.117 (0.097)	N/A	–	0.886 (0.114)	1.003 (0.087)	0.116 (0.095)	0.386
	Lynch and Ritland	0.108 (0.087)	N/A	–	0.893 (0.107)	1.000 (0.087)	0.108 (0.085)	0.371
	Queller and Goodnight	0.148 (0.081)	N/A	–	0.844 (0.097)	0.992 (0.087)	0.150 (0.079)	0.478
Variance components – site 2								
Pedigree	Half-sib	0.530 (0.294)	N/A	–	0.489 (0.237)	1.019 (0.102)	0.520 (0.256)	0.485
	Full-sib	0.199 (0.150)	0.550 (0.324)	–	0.697 (0.106)	1.033 (0.099)	0.133 (0.075)	0.440
Combined		0.118 (0.121)	0.688 (0.345)	–	0.750 (0.088)	1.040 (0.100)	0.165 (0.079)	0.363
Marker-based relationship	Li et al.	0.175 (0.083)	N/A	–	0.818 (0.091)	0.993 (0.084)	0.176 (0.079)	0.507
	Wang	0.200 (0.095)	N/A	–	0.797 (0.099)	0.997 (0.084)	0.200 (0.090)	0.483
	Lynch and Ritland	0.214 (0.084)	N/A	–	0.773 (0.087)	0.987 (0.048)	0.217 (0.078)	0.500
	Queller and Goodnight	0.258 (0.099)	N/A	–	0.750 (0.093)	1.008 (0.088)	0.256 (0.089)	0.593
Variance components – both sites								
Pedigree	Half-sib	0.340 (0.240)	N/A	0.204 (0.187)	0.479 (0.175)	1.022 (0.074)	0.332 (0.224)	0.595
	Full-sib	0.226 (0.112)	0.443 (0.219)	0.000 (0.000)	0.698 (0.080)	1.034 (0.073)	0.107 (0.052)	0.595
Combined		0.213 (0.119)	0.567 (0.236)	0.024 (0.105)	0.702 (0.068)	1.081 (0.082)	0.131 (0.053)	0.560
Marker-based relationship	Li et al.	0.000 (0.000)	N/A	0.129 (0.057)	0.869 (0.067)	0.997 (0.060)	0.000 (0.000)	N/A
	Wang	0.000 (0.000)	N/A	0.167 (0.068)	0.833 (0.074)	1.000 (0.061)	0.000 (0.000)	N/A
	Lynch and Ritland	0.000 (0.000)	N/A	0.175 (0.061)	0.820 (0.067)	0.995 (0.060)	0.000 (0.000)	N/A
	Queller and Goodnight	0.000 (0.000)	N/A	0.195 (0.064)	0.803 (0.067)	0.998 (0.061)	0.000 (0.000)	N/A

N/A not available

0.550), and this was also mirrored in the combined analyses (σ_d^2 0.450 vs. 0.688), again reflecting the difference in family composition between the two sites. Marker-based relationship methods produced similar pattern to that of the full-sib where the reliability of breeding value estimates of site 2 was higher (0.48–0.59 vs. 0.28–0.48) (Table 2). With the exception of Wang's results (r=0.45), across-site analyses generally produced the highest reliable estimates of breeding values (0.51–0.58) (Table 2). This could be caused by either the better parental representation or greater sample size resulting from combining the two sites (Table 2). It should be stated that with the results obtained from these analyses, a clear method superiority cannot be determined based on the genetic parameters' estimates alone. However, it should be stated that the combined analysis offers greater information in tracking individuals' genetic background as it reflects both contemporary as well as historic pedigree and the resulting genealogical relationships among individuals could be better managed during selection.

The four pairwise relationship methods universally produced higher heritability estimates for site 2 (range 0.18–0.26) as compared to site 1 (range 0.05–0.15), mirroring site variation in parental contribution and additive and environmental variances (Table 2). It should be noted that all pairwise marker-based methods did not permit across-site additive and/or heritability estimation (Table 2). In these analyses (i.e., across sites), the additive variance component was reduced to zero as across-site equivalent genetic structure (i.e., clonal or family) was difficult to produce (Fig. 1). Thus, the observed additive variance observed from single site analyses was shifted to the G × S term (see Fig. 1).

Pearson's product moment and Spearman's rank correlation between individuals' breeding values were significant for all analyses with pedigree information (half- and full-sib and combined); however, across-site analyses produced higher correlations and were similar between the full-sib and combined pedigree–marker analysis (0.86 and 0.84) (Table 3, significance was determined using SAS program). It is interesting to note that the half- and full-sib correlations (Pearson and Spearman) were exceptionally high (Table 3). No correlations were estimated between the marker-based relationship and the other pedigree-type methods.

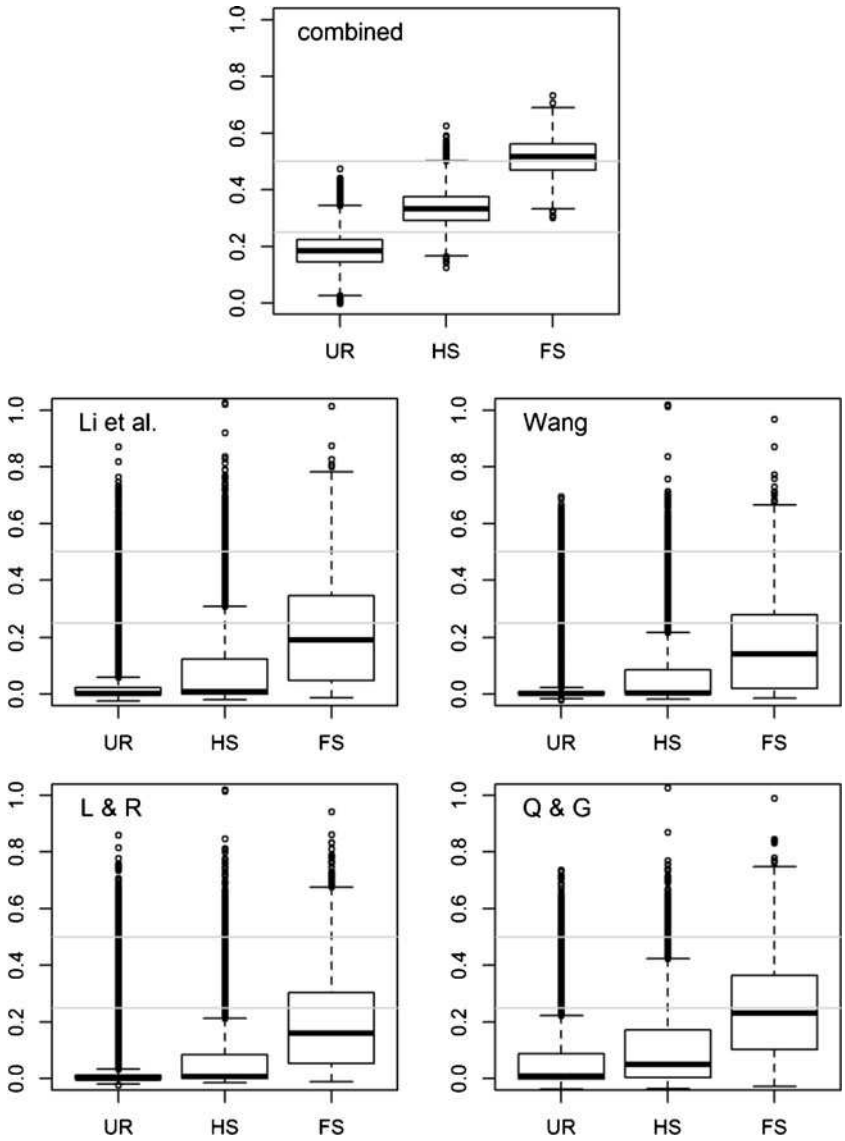

Fig. 1 Box plot comparison between combined (pedigree and marker-based) and marker-based relationship estimate distribution across relatedness classes (*UR* unrelated, *HS* half-sib, *FS* full-sib)

Discussion

Across the world, traditional forest tree breeding historically relied on open-pollinated (OP) testing scheme for its great simplicity. Simply, to evaluate the genetic worth (breeding value) of selected parents, one can proceed with collecting wind-pollinated seed from desirable parents (plus trees) and establish half-sib progeny trials, where a particular parent's breeding value is calculated as the mean deviation of the progeny from the population mean (Falconer and Mackay 1996). It should be understood that the actual merits of the OP testing scheme (low cost, logistical ease) are associated with significant limitations: (1) departures from "true" random mating of female parents with all male parents in the population (i.e., equal parental representation and close-to-zero correlation among mates in the selection criterion), (2) it is impossible to evaluate the impact of nonadditive gene actions (within and among traits of interest), and (3) sample of progenies from different families (different mother trees) does not necessarily guarantee the absence of half-sib relatedness or other types of relatedness, such as first cousin or even selfing (White et al. 2007). The over-inflated additive genetic variance and heritability estimates obtained from the present study further support this notion and highlight the caveats associated with this approach. However, it should be stated that if OP testing is considered as a filter for rouging inferior genotypes, then the true genetic merit of the selected individuals could be precisely determined using either full-sib (i.e., structured pedigree) or molecular marker pairwise relationship approach.

The development of molecular markers, coupled with statistical approaches to reconstruct pedigrees, made it possible to merge the advantages of traditional simple OP testing with those of more advanced full-sib family mating schemes. OP testing schemes with incorporating genetic markers therefore became a competitive platform to the

Table 3 Pearson's product–moment (below diagonal) and Spearman's rank-order (above diagonal) correlations between individuals' height breeding values among combined marker–pedigree (combined), sibship, and all marker-based pairwise relationships for site 1, site 2, and both sites (all correlations are significant at $P<0.0001$)

	Method	Half-sib	Full-sib	Combined	Li et al.	Wang	Lynch and Ritland	Queller and Goodnight
Site 1	Half-sib	1	0.904	0.680	0.414	0.485	0.441	0.360
	Full-sib	0.909	1	0.674	0.397	0.427	0.399	0.310
	Combined	0.737	0.737	1	0.608	0.624	0.625	0.757
	Li et al.	0.518	0.485	0.661	1	0.848	0.791	0.700
	Wang	0.585	0.512	0.681	0.882	1	0.740	0.703
	Lynch and Ritland	0.552	0.487	0.687	0.835	0.802	1	0.698
	Queller and Goodnight	0.480	0.408	0.787	0.748	0.749	0.749	1
Site 2	Half-sib	1	0.836	0.688	0.548	0.561	0.600	0.581
	Full-sib	0.872	1	0.821	0.550	0.552	0.567	0.547
	Combined	0.754	0.859	1	0.637	0.603	0.671	0.700
	Li et al.	0.646	0.638	0.725	1	0.875	0.806	0.763
	Wang	0.669	0.631	0.690	0.910	1	0.738	0.722
	Lynch and Ritland	0.680	0.652	0.752	0.862	0.806	1	0.787
	Queller and Goodnight	0.708	0.658	0.778	0.816	0.798	0.829	1
Both sites	Half-sib	1	0.902	0.770	N/A	N/A	N/A	N/A
	Full-sib	0.909	1	0.843	N/A	N/A	N/A	N/A
	Combined	0.792	0.861	1	N/A	N/A	N/A	N/A
	Li et al.	N/A	N/A	N/A	-	N/A	N/A	N/A
	Wang	N/A	N/A	N/A	N/A	-	N/A	N/A
	Lynch and Ritland	N/A	N/A	N/A	N/A	N/A	-	N/A
	Queller and Goodnight	N/A	N/A	N/A	N/A	N/A	N/A	-

N/A not available

traditional structured-pedigree-based (half- and full-sib crosses) breeding (see El-Kassaby and Lstibůrek 2009; El-Kassaby et al. 2011 for theoretical quantification).

The animal models were developed to facilitate classical breeding, but lately, the generalized mixed-model framework, combined with the use of molecular markers, has found great advantages in evolutionary genetics (Kruuk 2004). They can optimally accommodate variable family sizes (often found in open-family testing) and combine data from multiple locations and generations; furthermore, effects of nonrandom mating and selection are accounted for. For these reasons, the animal models provide a better platform to pedigree reconstruction approaches, as opposed to the simple half-sib evaluation which is based on simplified assumptions. Classical animal model incorporates the advantages of connectedness in the expected relationship (pedigree) matrix with relatedness matrix replacing the pedigree in the analysis. Frentiu et al. (2008) reported that such an approach, using reduced-rank relationship matrices, resulted in better estimates of genetic variances; however, estimating covariances remained a challenging task.

In our investigation, we combined the use of pedigree-based relatedness (**A** matrix) with marker-based relationship (**G** matrix) (Bömcke and Gengler 2009). This was done to minimize the impact of variable parental representation on marker-based relationships, which is based on only nine SSRs. It was suggested earlier that such a combined analysis leads to reliable ranking of individuals, when assessed by their respective additive genetic value (Blonk et al. 2010). Based on our results, high correlation coefficients between the full-sib and combined analysis confirm their reasoning as the combined analysis contains vital information from the full-sib. On the other hand, the four marker-based estimators used in the present study resulted in a significant decrease in both additive variance estimation and their respective standard errors. This goes in line with earlier findings by Frentiu et al. (2008), who discuss two major challenges when using the marker-based approach including (1) the intrinsic variability in within-population relatedness and (2) our ability to efficiently sample/detect these patterns. Moreover, the selection procedure implemented in this study might have reduced (i.e., downward estimates) the marker-based heritability estimates as previously observed by Rodriguez-Ramilo et al. (2007). In light of this, we investigated the distribution of marker-based pairwise relationships across all relatedness classes (i.e., unrelated and half- and full-sib) and compared them to that of the combined pedigree–marker relationship estimator (Fig. 1). The combined estimator

clearly produced three distinct unrelated and half- and full-sib classes, while the four marker-based estimators produced consistent and surprising results showing reduced mean values for both half- and full-sibs (Fig. 1). Additionally, this comparison also produced contrasting estimates' variances with reduced vs. wide variance for the combined and the marker-based methods, respectively (Fig. 1). This could be a reflection of the number and infomativeness of the markers used as well as parental contribution imbalance. We suspect that the observed lack of discrimination among relatedness classes is the cause for our inability to estimate the additive genetic variance when multiple sites were used (Table 2). It should be clearly stated that the implemented preselection of offspring prior to data analysis (i.e., restricting the analysis to a subset of individuals from the experiment) has restricted our ability to sample the "entire" variance and that might have affected the within-family variance estimation. However, this was done to maximize the genetic advance for the selection of parents for their inclusion in the second-generation seed orchard.

The main benefit of the marker-based approaches is their ability to sample the entire population history as it is not restricted by known pedigree and simplified probabilistic description of the relatedness based on the IBD likelihood. As there is no reference level to compare both pedigree-based and marker-based estimates, their pairwise comparison is not completely fair (Oliehock et al. 2006). Additionally, it should be highlighted that the marker-based methods, in this study, are not equivalent to their pedigree-based (full-sib) counterparts. The calculation of the dominance genetic variance can only be estimated through the determination of coefficient of fraternity which requires substantial number of markers (Lynch and Ritland 1999). So, we feel that the number of molecular markers used in the present study (which was 9) does not permit the precise construction of the coefficient of fraternity correctly; thus, we opted to avoid estimating the dominance genetic variance for the marker-based methods. Therefore, we combined the reconstructed pedigree with the pairwise marker-based relationship to overcome the drawback.

From the breeding perspective, marker-based methods, if done with precision, should be regarded advantageous to classical evaluation schemes as they free breeders from the task of building structured pedigree and directly permit genetic evaluation irrespective of genealogy. At the same time, more research is needed to encapsulate the effect of historical population structure (e.g., IBD vs. IBS: Powell et al. 2010), type, number, and informativeness of markers and sample size.

Finally, it must be clarified that the resulting offspring structure as revealed by pedigree reconstruction is merely a function of the number of parents participated in mating, their respective reproductive phenology and output, as well as the contamination level (gene flow). In case where high gene flow occurs, the pedigree created by the sibship analysis is anchored around maternal parents and it does not affect the quality of the maternal genetic parameters estimated.

Acknowledgments We would like to thank to T. Funda for his help with the selection of Scots pine primers. Funds from the NAZV (research grants no. QH81172 and QH81160) and the TAČR (grant no. TA01020512) to M. Lstibůrek and J. Kobliha and from the Natural Sciences and Engineering Research Council of Canada, Industrial Applied Forest Genetics and Biotechnology to Y. A. El-Kassaby, and partial support provided by USDA Forest Service, Southern Research Station through Cooperative Agreement SRS-05-IC-11330126-234 are hugely appreciated.

References

Askew GR, El-Kassaby YA (1994) Estimation of relationship coefficients among progeny derived from wind pollinated orchard seeds. Theor Appl Genet 88:267–272

Auckland LD, Bui T, Zhou Y, Shepherd M, Williams CG (2002) Conifer microsatellite handbook. Corporate, Raleigh

Bink MCAM, Anderson AD, van de Weg WE, Thompson EA (2008) Comparison of marker-based pairwise relatedness estimators on a pedigreed plant population. Theor Appl Genet 117:843–855

Blonk RJW, Komen H, Kamstra A, van Arendonk JAM (2010) Estimating breeding values with molecular relatedness and reconstructed pedigrees in natural mating populations of common sole, *Solea solea*. Genetics 184:213–219

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic-linkage map in man using restriction fragment length polymorphisms. Am J Hum Genet 32:314–331

Burczyk J (1998) Mating system variation in a Scots pine clonal seed orchard. Silvae Genetica 47:155–158

Bömcke E, Gengler N (2009) Combining microsatellite and pedigree data to estimate relationships among Skyros ponies. J Appl Genet 50:133–143

Caballero A, Toro MA (2000) Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet Res 75:331–343

Cappa EP, Lstiburek M, Yanchuk AD, El-Kassaby YA (2011) Two-dimensional penalized splines via Gibbs sampling to account for spatial variability in forest genetic trials with small amount of information available. Silvae Genetica 60:25–35

Cheng SH, Higham NJ (1998) A modified Cholesky algorithm based on a symmetric indefinite factorization. Siam J on Matrix Anal and Appl 19:1097–1110

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

El-Kassaby YA, Cappa EP, Liewlaksaneeyanawin C, Klápště J, Lstibůrek M (2011) Breeding without breeding: is a complete pedigree necessary for efficient breeding? PLoS One 6:e25737. doi:10.1371/journal.pone.0025737

Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics, 4th edn. Longman, London

Frentiu FD, Clegg SM, Chittock J, Burke T, Blows MW, Owens IPF (2008) Pedigree-free animal models: the relatedness matrix reloaded. Proc of the Royal Soc B-Biol Sci 275:639–647

Gilmour AR, Thompson R, Cullis BR (1995) Average information REML, an efficient algorithm for variance parameter estimation in linear mixed models. Biometrics 51:1440–1450

Gilmour AR, Cullis BR, Welham SI, Thompson R (2002) ASReml reference manual. NSW Agriculture, Australia

Hardy OJ, Vekemans X (2002) SPAGEDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Mol Ecol Notes 2:618–620

Higham NJ (2002) Computing the nearest correlation matrix—a problem from finance. IMA J Numer Anal 22:329–343

Huby M, Griffon L, Moureaux S, De Rochambeau H, Danchin-Burge C, Verrier E (2003) Genetic variability of six French meat sheep breeds in relation to their genetic management. Genet Sel Evol 35:637–655

Jones OR, Wang JL (2010) COLONY: a program for parentage and sibship inference from multilocus genotype data. Mol Ecol Resour 10:551–555

Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. Mol Ecol 16:1099–1106

Kaňák J, Klápště J, Lstibůrek M (2009) Úvodní genetické hodnocení semenných sadů borovice lesní v západních Čechách (Initial evaluation of seed orchards of Scots pine in the western Czech Republic). Zprávy lesnického výzkumu (Reports of Forestry Res) 3:189–204

Knol DL, Tenberge JMF (1989) Least-squares approximation of an improper correlation matrix by a proper one. Psychometrika 54:53–61

Kruuk LEB (2004) Estimating genetic parameters in natural populations using the 'animal model'. Philos Trans of the Royal Soc of London Series B-Biol Sci 359:873–890

Li CC, Weeks DE, Chakravarti A (1993) Similarity of DNA fingerprints due to chance and relatedness. Hum Hered 43:45–52

Liewlaksaneeyanawin C, Ritland CE, El-Kassaby YA, Ritland K (2004) Single-copy, species-transferable microsatellite markers developed from loblolly pine ESTs. Theor Appl Genet 109:361–369

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. Genetics 152:1753–1766

Malécot G (1948) Les mathematiques de I´heredite. Masson & Cie, Paris

Namkoong G, Kang HC, Brouard JS (1988) Tree breeding: principles and strategies. Springer, New York

Nejati-Javaremi A, Smith C, Gibson JP (1997) Effect of total allelic relationship on accuracy of evaluation and response to selection. J Anim Sci 75:1738–1745

Oliehock PA, Windig JJ, van Arendonk JAM, Bijma P (2006) Estimating relatedness between individuals in general populations with a focus on their use in conservation programs. Genetics 173:483–496

Porth I, Klápště J, Skyba O, Lai BSK, Geraldes A, Muchero W, Tuskan GA, Douglas CJ, El-Kassaby YA, Mansfield SD (2013) *Populus trichocarpa* cell wall chemistry and ultrastructure trait variation, genetic control, and genetic correlations. New Phytol 197:777–790

Powell JE, Visscher PM, Goddard ME (2010) Reconciling the analysis of IBD and IBS in complex trait studies. Nat Rev Genetics 11:800–805

Queller DC, Goodnight KF (1989) Estimating relatedness using genetic markers. Evolution 43:258–275

Rodriguez-Ramilo ST, Toro MA, Caballero A, Fernandez J (2007) The accuracy of heritability estimator using molecular information. Conserv Genet 8:1189–1198

Soranzo N, Provan J, Powell W (1998) Characterization of microsatellite loci in *Pinus sylvestris* L. Mol Ecol 7:1260–1261

Squillace AE (1974) Average genetic correlations among offspring from open-pollinated forest trees. Silvae Genetica 23:149–156

Visscher PM, Medland SE, Ferreira MAR, Morley KI, Zhu G, Cornes BK, Montgomery GW, Martin NG (2006) Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. PLoS Genetics 2:e41

Wang JL (2002) An estimator for pairwise relatedness using molecular markers. Genetics 160:1203–1215

White TL, Adams WT, Neale DB (2007) Forest genetics. CABI, Cambridge

Wright S (1922) Coefficients of inbreeding and relationship. Am Nat 56:330–338

Wright S, McPhee HC (1925) An approximate method of calculating coefficients of inbreeding and relationship from livestock pedigrees. J Agric Res 31:377–383