

ORIGINAL ARTICLE

Joint analysis of spatial genetic structure and inbreeding in a managed population of Scots pine

MR García-Gil¹, F Olivier², S Kamruzzahan¹ and P Waldmann¹¹Department of Forest Genetics and Plant Physiology, Umeå, Sweden and ²Faculty of Medicine, TIMC-IMAG, University Joseph Fourier Grenoble, La Tronche, France

We have investigated the fine-scale spatial genetic structure in a managed Scots pine forest. For this purpose, we perform a Bayesian genetic-cluster analysis of 96 geographically mapped individual seed trees of Swedish Scots pine based on 14 microsatellite loci. The analysis was carried out with the recently developed program GENECLUST (François *et al.*, 2006), which provides the facility to jointly incorporate both spatial information from a geographical neighborhood structure through a Potts–Dirichlet model and account for variable degrees of inbreeding within the clusters. To evaluate whether inbreeding and spatial interaction should be included in the

best-fitting statistical model for our data, we used the deviance information criterion (DIC), a weighted measure of model fit that accounts for an effective number of free parameters in a model. Analysis shows that a model with a single estimated cluster, with high levels of inbreeding (0.25) and with a moderate amount of spatial dependency within the unique cluster ($\Psi = 0.2–0.4$), best explains the data. We also carried out Bayesian parentage analysis, which enabled us to exclude the possibility that the sample constitutes one single full-sib family. *Heredity* (2009) **103**, 90–96; doi:10.1038/hdy.2009.33; published online 22 April 2009

Keywords: population genetic structure; natural regeneration; inbreeding; microsatellites

Introduction

Differences in genetic structure within and between populations in tree species are mainly because of the life form and breeding system. The availability of highly variable molecular markers has facilitated the analysis of fine-scale genetic structure in natural tree populations. The fine-scale structure has been found in some forest tree species (Sork *et al.*, 1993; Berg and Hamrick, 1995; Streiff *et al.*, 1998; Dutech *et al.*, 2002; Hardy *et al.*, 2006). These tree species are characterized by either limited seed dispersal or restricted pollen and seed dispersal.

Pines are wind-pollinated and the seeds generally have wings that facilitate wind dispersal (Ledig, 1998). Together with a predominant random mating system (Koski, 1970), these features contribute to little or no genetic structure being found in undisturbed pine forests: both in large (Gullberg *et al.*, 1985; Karhu *et al.*, 1996; Dvornyk *et al.*, 2002; García-Gil *et al.*, 2003) and in fine geographic scales (Knowles, 1991; Xie and Knowles, 1991; Parker *et al.*, 2001; Uchiyama *et al.*, 2006; Marquardt *et al.*, 2007). On the other hand, fragmentation and bottlenecks may cause a genetic structure because of self-fertilization and mating among genetically related individuals (Vogl *et al.*, 2002; Robledo-Arnuncio *et al.*, 2004; Boys *et al.*, 2005). When mating occurs between genetically related individuals, it increases inbreeding. Inbred individuals may have lower fitness because of the expression of recessive deleterious alleles (Charlesworth

and Charlesworth, 1987). Inbreeding also results in a decreased level of genetic diversity, which is of major concern in forest tree breeding and conservation programs.

Forest management practices have also been shown to increase genetic structure compared with natural forests, especially if the breeding practices imply drastic reduction of the effective population size (Young and Merriam, 1994; Finkeldey and Ziehe, 2004). Population size reduction could potentially increase the rate of self-fertilization because of the reduction in number of local compatible mates. Moreover, even under random mating, a smaller number of parent trees will increase the probability of seed cohorts with full-sib relationships (Surlles *et al.*, 1990; Muona and Harju, 1989; Robledo-Arnuncio *et al.*, 2004).

Recently, several Bayesian clustering methods for inference of population genetic structure have been developed. These methods are generally referred to as assignment methods and use allele frequency data of molecular markers to ascertain the population membership of individuals by assuming either fixed or variable numbers of population clusters (Manel *et al.*, 2005). In the original methods developed by Pritchard *et al.* (2000), Dawson and Belkhir (2001) and Corander *et al.* (2003), spatial information was not explicitly included in the modeling. However, some recent Bayesian assignment methods incorporate information from the geographical coordinates of individuals, by using prior distributions for the spatial distribution of individuals in a cluster (Wasser *et al.*, 2004; Guillot *et al.*, 2005; François *et al.*, 2006; Corander *et al.*, 2008). Simulation studies have shown that incorporation of geographical information into assignment methods can result in better statistical

Correspondence: Dr MR García-Gil, Department of Forest Genetics and Plant Physiology, SLU, Umeå SE-901 83, Sweden.

E-mail: M.Rosario.Garcia@genfys.slu.se

Received 27 August 2008; revised 23 December 2008; accepted 3 March 2009; published online 22 April 2009

performance (Chen *et al.*, 2007). It is well-recognized that inbreeding perturbs the Hardy–Weinberg equilibrium and can lead to spurious aggregates of population substructure (for example, Guinand *et al.*, 2006). With the exception of the methods developed by François *et al.* (2006) and Gao *et al.* (2007), the assignment methods are based on the assumption of Hardy–Weinberg equilibrium within clusters, and may therefore yield biased estimates of the number of clusters in the presence of inbreeding.

In this study, we jointly estimated the fine-scale genetic structure and inbreeding level in a managed tree population of Scots pine using a recently developed Bayesian hidden Markov model. We analyzed 96 geographically mapped individual seed trees of Swedish Scots pine using 14 microsatellite loci. The analysis was carried out using the program GENECLUST (François *et al.*, 2006), which provides the facility to jointly incorporate both spatial information from a geographical neighborhood structure through a Potts–Dirichlet model and account for variable degrees of inbreeding within the clusters. To evaluate whether inbreeding and spatial interaction should be included in the best-fitting statistical model for our data, we used the deviance information criterion (DIC), a weighted measure of fit that accounts for an effective number of free parameters in a model (Spiegelhalter *et al.*, 2002; Celeux *et al.*, 2006). We evaluated DIC statistics for several models with and without inbreeding, and with increasing levels of spatial connectivity.

Materials and methods

Scots pine material

Scots pine is a major conifer species across the northern boreal zone in Europe and Asia. Its distribution is the widest among the pine species, from southern Spain (38°N) to north Finland (68°N), and from western Scotland (6°W) to Okhotsk Sea in eastern Siberia (135°E) (Mirov, 1967). Within its distribution Scots pine grows at elevations from sea level to 2400 m and in many different environments in terms of temperature, soil quality and humidity. Scots pine is a keystone species, on which many other plants, insects, birds and animals species depend (Persson, 1980). Like the majority of the pine species, Scots pine has a diploid genome with a chromosome number of $2n = 24$ (Saylor, 1972). Scots pine is wind-pollinated and has wings on the seeds that facilitate wind dispersal, over distances that can be characterized by an exponential distribution with a tail that descends to a value close to zero within a few tens of meters. However, some long-distance animal-mediated seed dispersal cannot be ruled out (Lanner, 1998).

The trees for this experiment are situated in a population 25 km north-east of Umeå, Sweden and originate from wind-pollinated seed trees that were established in 1965. The number of seed trees was around 50 per hectare. Seedlings were allowed to establish until 1979, after which the seed trees were cut down. The population was thinned in 1989, resulting in a collection of trees with homogenous height and age.

We sampled needles, marked and estimated geographic positions with a satellite-based GPS system of 96 trees according to a square lattice. We sampled 25

hectares out of a total managed area of 65.9 hectares. The aim was to sample trees as close as possible to 50 m apart (that is, a lattice with 50 × 50 m cells). However, the lattice deviated slightly from this ideal because the seedlings had established naturally (see the Voronoi tessellation in Figure 1). Needles were sampled within 1 day in November 2004 and stored in a –80 °C freezer.

DNA extraction and microsatellite amplification

DNA was extracted from needles with the DNeasy Plant Mini Kit (Qiagen, Solna, Sweden, Cat number: 69104). Twelve nuclear microsatellite primers developed for *Pinus taeda* (Elsik *et al.*, 2000; Auckland *et al.*, 2002; Liewlaksaneeyanawin *et al.*, 2004; Chagn *et al.*, 2004) and *Pinus sylvestris* (Soranzo *et al.*, 1998) were selected to genotype all the individuals. The amplified microsatellite primers are PtTX2146, PtTX3107, PtTX3116, PtTX4001, PtTX4011, LOP1, LOP3, SPAC 12:5, SPAG 7:14, SPAC 11:8, SsrPt_ctg64 and Ssr_ctg4487b. Primer SsrPt_ctg64 amplified three different polymorphic microsatellite loci, namely ctg64a, ctg64b and ctg64c. The primers incorporated fluorescent dyes (D2, D3 and D4). The PCR volume was 25 µl and consisted of 50 ng of genomic DNA template, 0.2 mM of each primer, 0.2 mM of each dNTP, 2.5 µl of 10X *Taq* buffer (500 mM KCl, 100 mM Tris-HCl, 1% Triton X-100, Promega, Nacka, Sweden, Cat number: A3511), 2 mM of MgCl₂ (Promega) and two units of *Taq* polymerase (Fermentas, Helsingborg, Sweden, Cat number: EP0405). Amplifications were carried out using a Peltier Thermal Cycler PTC-225. The amplification protocol for SPAC 11:8, SPAC 12:5 and SsrPt_ctg64 primers was 5 min at 94 °C; followed by 35 cycles of 1 min at 94 °C, 1 min at 55 °C, 1 min at 72 °C; and finally one cycle for 10 min at 72 °C. The amplification conditions for PtTX4001, PtTX3107, LOP3 and SsrPt_ctg4487b primers were 5 min at 94 °C; followed by touch-down from 55 °C down to 45 °C and 25 cycles of 1 min at 94 °C, 1 min at 45 °C, 1 min at 72 °C; and finally one cycle for 10 min at 72 °C. Primers PtTX3116, PtTX4011, PtTX2146, SPAG 7:14 and LOP1 were amplified under the same touch-down protocol described before, except for the gradient temperature that started at 60 °C down to 50 °C. PCR amplifications were resolved in a Beckman Coulter CEQ-8000 using an internal size standard (400 bp size standard) and multiplexing the runs for a maximum of three different SSR loci, and allele scoring was done by using the CEQ system software.

Statistical analysis

GENECLUST (François *et al.*, 2006) is based on the concept of Hidden Markov Random Field (HMRF), which models the spatial dependencies in cluster membership. Hidden Markov models (HMMs) assume that the data are a noisy realization of an underlying process with Markovian dependence. In other words, a HMM is a one-dimensional Markov chain observed in noise (Cappé *et al.*, 2005). HMRFs are generalizations of HMMs to the two-dimensional plane and are therefore suitable for analysis of spatially structured observations. Markov random fields provide a statistically well-founded basis for modeling spatial autocorrelation, which is of major interest to many biological applications (Sokal and Oden, 1978). Markov random field models are motivated by the concept of conditional independence;

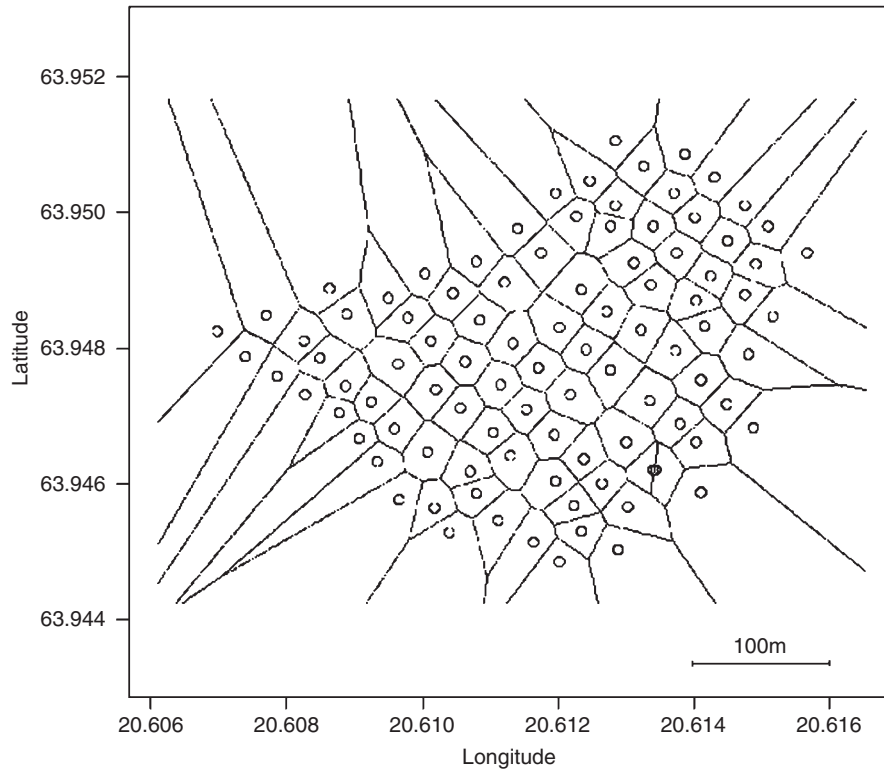


Figure 1 Neighborhood structure of the 96 sampled Scots pine trees obtained from a Voronoi tessellation. Trees sharing a border are considered as neighbors.

that is, the dependence of a random variable associated with a particular site on the random variables at all the other sites can be specified by the values of random variables in the neighboring sites only (Ripley, 1981; Cressie, 1993). In population genetics, HMRFs can account for the fact that individuals from spatially continuous populations are more likely to share cluster membership with their close neighbors than with distant individuals. GENECLUST can detect geographical discontinuities in allele frequencies and estimate individual population memberships as an unobservable parameter. To account for the dependencies among cluster labels, GENECLUST uses the Potts model, parameter Ψ of which specifies the importance of spatial interactions. The value of Ψ is generally non-negative. Zero values of Ψ indicate no special dependency; hence the statistical model used by STRUCTURE (Pritchard *et al.*, 2000) is recovered.

The first step in GENECLUST is to calculate a neighborhood structure from the geographical coordinates with Dirichlet tiling (also known as Voronoi tessellation). Two sampled individuals are neighbors if their Dirichlet cells share a border. The neighborhood structure for the sampled Scots pine trees is shown in Figure 1. Default priors were used for all parameters, that is, Dirichlet distributions $\text{Dir}(\alpha, \dots, \alpha)$ with $\alpha=1$ on allele frequencies f_k , $\beta(4, 40)$ prior on each f_k and fixed values of the spatial interaction parameter Ψ . GENECLUST also provides an estimate for the actual number of cluster in the data, K . For well-chosen values of Ψ , the hidden Markov model acts as a regularizer, and tends to empty spurious clusters when K_{\max} exceeds K . In order to avoid problems associated with specification of a

single K_{\max} (Evanno *et al.*, 2005), we used the two values, $K_{\max} = 2, 3$.

We fixed Ψ to different values (0, 0.2, 0.4, 0.6) and compared the DIC (Spiegelhalter *et al.*, 2002) of models with and without inbreeding, and for two values of K_{\max} . The basic principle of DIC is that models with smaller values are preferred to models with larger values. For a model with parameter θ and for some genetic data, y , the DIC can be computed by adding a penalty term, p_D , to the averaged deviance, $D(\theta) = -2 E \theta [\log p(\theta | y) | y]$. The penalty term is meant to represent an effective dimension for θ , which is estimated from the data. The penalty term is usually computed as $p_D = D(\theta) + 2 \log p(\theta_{\text{est}} | y)$, where θ_{est} represents an estimate of θ . Provided that the deviance, $-2 \log p(\theta | y)$, is available in closed form, $D(\theta)$ can easily be approximated from an MCMC run by taking the sample mean of the simulated values. With flat priors or when the likelihood overwhelms the priors, DIC behaves similarly as the Akaike Information Criterion (Akaike, 1974). In general, DIC contains a useful estimate of the effective number of parameters even when many of them are defined as latent variables, as is the case in many hierarchical models. We implemented the DIC for GENECLUST models, where the parameter θ comprised the set of allele frequencies, f_k , the set of individual cluster labels, z , and the set of inbreeding coefficients, ϕ , when inbreeding was included in the model. We used posterior average estimates for the allele frequencies and for the inbreeding coefficients as computed by GENECLUST, and estimated the cluster configuration from the cluster membership coefficients, after re-assigning each individual to their most likely cluster.

Paternal genotypes were reconstructed using the Bayesian program Parentage 1.0 (Emery *et al.*, 2001). We used two chains and the Metropolis-coupled MCMC option. Burn-in was set to 100 000 iterations, and sampling based on the next 100 000 (with a thinning of 10). Prior for the allele frequencies was the standard Dirichlet distribution and prior for the number of fathers and mothers was Unif(1,96) and Model 1. Each male is equally likely to be the father of any offspring.

Results

The number of alleles and their frequencies are available in the online supplement. The number of alleles per SSR locus ranges from 2 (ctg64a) to 47 (SPAC12:5). As expected, the SSR loci that originated from cDNA libraries (for example, ctg64) showed lower allelic richness than genomic SSR loci. As reported earlier in the literature, the SPAC12:5 locus turned out to be highly polymorphic (Soranzo *et al.*, 1998).

For each value of $K_{\max}=2-3$, for levels of the spatial interaction parameter Ψ ranging from 0 to 0.6, and considering models with and without inbreeding, we computed an average of the DIC over the best 20 values obtained after 100 replicates of GENECLUST runs. Using a total of 5000 sweeps for the MCMC program and discarding the 2500 first sweeps as a burn-in period, we carried out a total number of 1200 runs and compared 12 models.

Runs with Ψ greater than zero generally converged to a single cluster. Runs with $\Psi=0$ generally ended with a large majority of individuals assigned to a single cluster, and with a small minority, not exceeding five individuals, sometimes assigned to a second cluster. Table 1 reports DIC values corresponding to each model. Averaged DICs ranged from 9149 to 9270, with s.d.'s ranging from 7 to 30. The smallest values of the DIC were reached for models with inbreeding and for a spatial interaction parameter around 0.2–0.4. For these models, the effective number of parameters was estimated

around 505. The highest values were reached for models without inbreeding and with the spatial parameter Ψ set to 0. All models without inbreeding performed worse than those including inbreeding. Posterior estimates of the inbreeding coefficient were computed for models with $\Psi=0.4$ by pooling the 20 runs with the lowest DICs (DIC < 9140). The posterior mean of the inbreeding coefficient was equal to 0.248 (median = 0.249), with a 95% credibility interval ranging from 0.217 to 0.283. The posterior s.d. was estimated to be 0.018. Figure 2a shows a histogram for 10 000 simulated values from the posterior distribution. Using the same procedure, we computed posterior estimates from models with $\Psi=0$. The posterior mean of the inbreeding coefficient was equal to 0.249 (median = 0.249), with a 95% credibility interval ranging from 0.215 to 0.284. The posterior distribution was not different from the one obtained by using higher values of Ψ (Figure 2b). These results may indicate that, in the case of absence of population structure, the estimation of the inbreeding coefficient is robust to the presence of a small amount of spatial autocorrelation in the data. To conclude, the values of the DIC indicated that a model with a single estimated cluster, relatively high levels of inbreeding and a moderate amount of spatial dependencies within the unique population best explains the data.

Analyses of paternity were carried out in order to reconstruct the parental genotypes. The analyses supported 19 fathers and 15 mothers as the progenitors of the 96 sampled trees. The analyses identified a very small number of full- and half-sibs. By setting a probability limit at 0.9, we found the following full-sib pairs: 27–37 ($P=0.973$), 23–78 ($P=0.999$), 13–90 ($P=0.906$), 54–87 ($P=0.916$); and half-sib pairs: 27–37 ($P=0.984$), 47–49 ($P=0.933$), 13–90 ($P=0.956$), 24–78 ($P=1.000$) and 54–87 ($P=0.955$). Note that 27–37 can be both half-sibs and full-sibs with very high probability. Hence, we can conclude that seed trees do not belong to a single full-sib family, which is indicated by the inbreeding level. Instead, they must share some form of relationship before they are selected as seed trees.

Discussion

The fine-scale genetic structure of 96 geographically mapped Scots pine trees from a stand in northern Sweden was analyzed using 14 SSR loci. Our sample size and number of loci have earlier been shown to be sufficient for the study of spatial genetic structure in tree populations (Cavers *et al.*, 2005). Assignment analysis carried out with the program GENECLUST (François *et al.*, 2006) and model comparison based on the DIC show that a model with a single estimated cluster, with high levels of inbreeding and with a moderate amount of spatial dependencies within the unique cluster ($\Psi=0.2-0.4$), best explains the data. Although the DIC has been used earlier to decide which runs of a Bayesian clustering program should be kept after a multiple-run analysis (François *et al.*, 2008), its systematic use for deciding which model best fits the data is new in this context. The four versions of DIC implemented in this study were motivated by the fact that these measures could be directly and easily computed from the output of the program.

Different approaches have been used before to evaluate spatial clustering of genotypes within stands

Table 1 Average values of the deviance information criterion (DIC) for spatial genetic models with $K_{\max}=2-3$ clusters, with levels of the spatial interaction parameter ψ ranging from 0 to 0.6, and with absence (n) or presence (y) of inbreeding

Spatial parameter	Inbreeding coefficient	DIC	
$K_{\max}=2$	0	n	9270 (30)
		y	9163 (9)
	0.2	n	9238 (22)
		y	9151 (7)
	0.4	n	9233 (14)
		y	9149 (8)
0.6	n	9240 (23)	
	y	9154 (9)	
$K_{\max}=3$	0	y	9164 (18)
	0.2	y	9154 (11)
	0.4	y	9157 (16)
	0.6	y	9161 (15)

Values were averaged over the 20 best runs of the computer program GENECLUST (parenthesized values are s.d.). We used a total of 100 runs for each model.

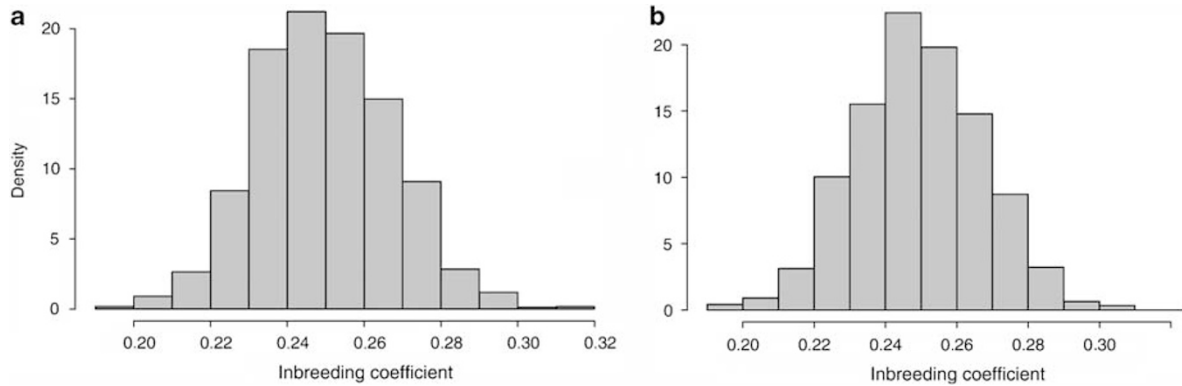


Figure 2 Posterior density of the inbreeding coefficient from the 20 models with lowest values of DIC, computed from GENECLUST using $K_{\max} = 2$ clusters (10000 simulations). (a) Model with spatial parameter $\Psi = 0.4$. (b) Model with spatial parameter $\Psi = 0$.

(or populations). The simplest method is to assess the degree of clustering by plotting the trees on a map (Knowles, 1991). Another method is based on dividing the stand into subplots and estimating the among-subplots differentiation by means of gene diversity (G_{st} or F_{st}) (Streiff *et al.*, 1998). The most commonly used procedure estimates the similarity between pairs of genotypes or subplots (based on allele frequencies) within a specified distance and evaluates whether the pairs are more similar than expected by chance under random spatial arrangement (Epperson, 1992; Parker *et al.*, 2001; Cavers *et al.*, 2005). However, these methods are not very useful when dealing with inbred populations.

Only few simulation studies have evaluated the effect of inbreeding on results from assignment and cluster analysis. Guinand *et al.* (2006) carried out a simulation study that investigated how different levels of inbreeding ($F = 0, 0.05$ and 0.15) influenced the accuracy of assignment analysis. They concluded that inbreeding had no effect on the accuracy of the assignments. However, it should be noted that Guinand *et al.* (2006) used a version of STRUCTURE that does not allow for proper modeling of inbreeding. Gao *et al.* (2007) presented a method (implemented in a program called InStruct) that extends the algorithm in STRUCTURE by eliminating the assumption of the Hardy-Weinberg equilibrium within clusters. Based on extensive simulations with various levels of selfing, they showed that their approach could avoid spurious signals of population substructure that could lead to biased assignments. However, the differences in assignment bias compared with STRUCTURE were mostly relatively small, and they did not evaluate how estimation of K was influenced by inbreeding. In addition, the program InStruct does not allow inference based on spatially explicit priors, as does GENECLUST, and is therefore less appropriate for analysis of our data.

Our results indicate a single estimated cluster and a relatively high overall inbreeding coefficient of 0.250, which correspond to co-ancestry of one full-sib family or a mixture of half-sibs and full-sibs established from already related seed trees. Based on the parentage analysis, the results of which support a total of 19 fathers and 15 mothers and only nine pairs of trees that were either full- or half-sibs, we can conclude that the trees do not form a single full-sib family. The high overall

inbreeding coefficient contrasts with the efficient mechanisms for purging inbreeds described in Scots pine (Muona *et al.*, 1987; Kärkkäinen and Savolainen, 1993). On the other hand, this apparent contradiction could be explained if some of the trees are the result of mating between already related parent trees, as supported by the parentage analysis.

In Scots pine, increased selfing is generally not a problem in natural stands because of heavy selection against inbreeds at the seed and seedling stages (Muona *et al.*, 1987; Kärkkäinen and Savolainen, 1993). However, in low-density stands (partially harvested forests), the remaining inbreed seedlings may be eliminated less efficiently due to lower competition. Natural regeneration from a few trees can potentially affect the population structure and mating system because of the reduced initial reproductive population size. Studies carried out in managed pine populations support changes in the level of inbreeding after natural regeneration from partially removed forests, but the degree and direction of the disturbance vary among reports. Although some studies indicate an increase in the inbreeding level (Rudin *et al.*, 1977; Farris and Mitton, 1984), others support absence (Yazdani *et al.*, 1989), or even decreased inbreeding (Marquardt *et al.*, 2007). Yazdani *et al.* (1989) concluded that seed trees had a low genetic contribution to regeneration compared with seeds from felled trees and surrounding trees. The discrepancies among studies may be because of factors such as percentage of tree removal (final forest density), level of gene inflow from the surrounding forest and level of genetic relatedness between the trees left after harvesting.

Spatial analyses have shown some level of clustering of genotypes within populations of other conifers (Knowles *et al.*, 1992; Cavers *et al.*, 2005). However, cluster sizes are quite small (5–50 m across), suggesting that they are primarily a result of limited seed dispersal, and that the clusters are made up of close relatives; Knowles *et al.* (1992) compared spatial genetic clustering in two stands of *Larix laricina*. The stand had naturally regenerated after clear-cutting, presumably by a few remnant individuals scattered within the stand, and showed significant spatial clustering of genotypes. No clustering, however, was observed in a nearby old-field stand.

It is possible to use molecular markers in combination with powerful Bayesian statistical methods for joint

estimation of spatial genetic structure and inbreeding in tree populations for estimation of genetic parameters that potentially could be used for monitoring forest-management practices, but results from comparative experiments with managed and non-managed stands are needed before we can draw a final conclusion. Results from this kind of study would be of special relevance in forestry, wherein the observation of long-term effects of forest management on genetic structures is retarded by the long rotation cycles.

Acknowledgements

Financial support was provided by the Research School in Forest Genetics and Breeding, SLU, Umeå and Kempefistelserna.

References

- Akaike H (1974). A new look at the statistical model identification. *IEEE Trans Automatic Control* **19**: 716–723.
- Auckland LD, Bui T, Zhou Y, Sheperd M, Williams CG (2002). *Conifer Microsatellite Handbook*. Corporate Press: Raleigh, NC.
- Berg EE, Hamrick JL (1995). Fine-scale genetic structure of a turkey oak forest. *Evolution* **49**: 110–120.
- Boys J, Cherry M, Dayanandan S (2005). Microsatellite analysis reveals genetically distinct populations of red pine (*Pinus resinosa*, Pinaceae). *Am J Bot* **92**: 833–841.
- Cappé O, Moulines E, Rydén T (2005). *Inference in Hidden Markov Models*. Springer-Verlag: New York.
- Cavers S, Degen B, Caron H, Lemes MR, Margis R, Slgueiro F *et al.* (2005). Optimal sampling strategy for estimation of spatial genetic structure in tree populations. *Heredity* **95**: 281–289.
- Celeux G, Forbes F, Robert CP, Titterton DM (2006). Deviance information criteria for missing data models (with Discussion). *Bayesian Anal* **1**: 651–706.
- Chagné D, Chaumeil P, Ramboer A, Collada C, Guevara A, Cervera MT *et al.* (2004). Cross-species transferability and mapping of genomic and cDNA SSRs in pines. *Theor Appl Genet* **109**: 1204–1214.
- Charlesworth D, Charlesworth B (1987). Inbreeding depression and its evolutionary consequences. *Annu Rev Ecol Syst* **118**: 237–268.
- Chen C, Durand E, Forbes F, François O (2007). Bayesian clustering algorithms ascertaining spatial population structure: a new computer program and a comparison study. *Mol Ecol Notes* **7**: 747–756.
- Corander J, Sirén J, Arjas E (2008). Bayesian spatial modeling of genetic population structure. *Comput Stat* **23**: 111–129.
- Corander J, Waldmann P, Sillanpää MJ (2003). Bayesian analysis of genetic differentiation between populations. *Genetics* **163**: 367–374.
- Cressie NAC (1993). *Statistics for Spatial Data*. Wiley: New York.
- Dawson KJ, Belkhir K (2001). A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genet Res* **78**: 59–77.
- Dutech C, Seiter J, Petronelli P, Joly HI, Jarne P (2002). Evidence of low gene flow in a neotropical clustered tree species in two rainforest stands of French Guiana. *Mol Ecol* **11**: 725–738.
- Dvornyk V, Sirviö A, Mikkonen M, Savolainen O (2002). Low nucleotide diversity at the *pall* locus in the widely distributed *Pinus sylvestris*. *Mol Biol Evol* **19**: 179–188.
- Elsik CG, Minihan VT, Scarpa AM, Hall SE, Willimans CG (2000). Low-copy microsatellite markers for *Pinus taeda* L. *Genome* **43**: 550–555.
- Emery AM, Wilson IJ, Craig S, Boyle PR, Noble LR (2001). Assignment of paternity groups without access to parental genotypes: multiple mating and developmental plasticity in squid. *Mol Ecol* **10**: 1265–1278.
- Epperson BK (1992). Spatial structure of genetic variation within populations of forest trees. *New Forest* **6**: 257–278.
- Evanno G, Regnault S, Goudet J (2005). Detecting the number of clusters of individuals using the software structure. A simulation study. *Mol Ecol* **14**: 2611–2620.
- Farris MA, Mitton JB (1984). Population density, outcrossing rate, and heterozygosity superiority in ponderosa pine. *Evolution* **38**: 1151–1154.
- Finkeldey R, Ziehe M (2004). Genetic implications of silvicultural regimes. *Forest Ecol Manage* **197**: 231–244.
- François O, Ancelet S, Guillot G (2006). Bayesian clustering using hidden Markov random fields in spatial population genetics. *Genetics* **174**: 805–816.
- François O, Blum MGB, Jakobsson M, Rosenberg NA (2008). Demographic History of European Populations of *Arabidopsis thaliana*. *PLoS Genet* **4**: e1000075.
- Gao H, Williamson S, Bustamante CD (2007). A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics* **176**: 1635–1651.
- García-Gil MR, Mikkonen M, Savolainen O (2003). Molecular Nucleotide diversity at phytochrome loci across a latitudinal gradient in *Pinus sylvestris* (Scots pine). *Mol Ecol* **12**: 1195–1207.
- Guillot G, Estoup A, Mortier F, Cosson JF (2005). A spatial statistical model for landscape genetics. *Genetics* **170**: 1261–1280.
- Guinand B, Scribner KT, Page KS (2006). Effects of coancestry on accuracy of individual assignments to population of origin: examples using Great Lakes lake trout (*Salvelinus namaycush*). *Genetica* **127**: 329–340.
- Gullberg U, Yazdani R, Rudin D, Ryman N (1985). Allozyme variation in Scots Pine (*Pinus sylvestris* L) in Sweden. *Silvae Genet* **34**: 193–201.
- Hardy OJ, Maggia L, Bandou E, Breyne P, Caron H, Chevallier MH *et al.* (2006). Fine-scale genetic structure and gene dispersal inferences in 10 Neotropical tree species. *Mol Ecol* **15**: 559–571.
- Karhu A, Hurme P, Karjalainen M, Karvonen P, Kärkkäinen K, Neale D, Savolainen O *et al.* (1996). Do molecular markers reflect patterns of differentiation in adaptive traits of conifers? *Theor Appl Genet* **93**: 215–221.
- Knowles P (1991). Spatial genetic structure within two natural stands of black spruce (*Picea mariana* (Mill.)B.S.P.). *Silvae Genetica* **40**: 13–19.
- Knowles P, Perry DJ, Foster HA (1992). Spatial genetic structure in two tamarack (*Larix laricina* (Du Roi) K Koch) populations with differing establishment histories. *Evolution* **46**: 572–576.
- Koski V (1970). A study of pollen dispersal a mechanism of gene flow in conifers. *Commun Inst Forestalis Fenniae* **70**: 1–78.
- Kärkkäinen K, Savolainen O (1993). The degree of early inbreeding depression determines the selfing rate at the seed stage model and results from *Pinus sylvestris* (Scots pine). *Heredity* **71**: 160–166.
- Lanner RM (1998). Seed dispersal in *Pinus*. In: Richardson DM (ed). *Ecology and Biogeography of Pinus*. Cambridge University Press: Cambridge, UK. pp 281–295.
- Ledig FT (1998). In: Ricahrdson DM (ed). *Ecology and Biogeography of Pinus*. Cambridge University Press: Cambridge.
- 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.
- Manel S, Gaggiotti O, Waples R (2005). Assignment methods: Which approaches best address which biological questions? *Trends Ecol Evol* **20**: 136–142.
- Marquardt PE, Echt CS, Epperson BK, Pubanz DM (2007). Genetic structure, diversity and inbreeding of eastern white pine under different management conditions. *Can J Forest Res* **37**: 2652–2662.

- Mirov NT (1967). *The Genus Pinus*. Ronald Press: New York.
- Muona O, Harju A (1989). Effective population sizes, genetic variability, and mating system in natural stands and seed orchards of *Pinus sylvestris*. *Silvae Genetica* **38**: 221–229.
- Muona O, Yazdani, Rudin D (1987). Genetic change between life stages in *Pinus sylvestris*: allozyme variation in seeds and planted seedlings. *Silvae Genet* **35**: 39–42.
- Parker KC, Hamrick JL, Parker AJ, Nason JD (2001). Fine-scale genetic structure in *Pinus clausa* (Pinaceae) populations: effects of disturbance history. *Heredity* **87**: 99–113.
- Persson T (1980). Structure and function of northern coniferous forests—an ecosystem study. *Ecol Bull* **32**: 604.
- Pritchard JK, Stephens M, Donnelly P (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.
- Robledo-Arnuncio JJ, Alía R, Gil L (2004). Increased selfing and correlated paternity in a small population of a predominately outcrossing conifer, *Pinus sylvestris*. *Mol Ecol* **13**: 2567–2577.
- Rudin D, Eriksson G, Rasmuson M (1977). *Inbreeding in a Seed Tree Stand of Pinus sylvestris L. in Northern Sweden. A Study by the Aid of Isozyme Technique*. Skogshögskolan nr 25. Royal College of Forestry.
- Ripley BD (1981). *Spatial Statistics*. Wiley: New York.
- Saylor LC (1972). Karyotype analysis of the genus *Pinus*-subgenus *Pinus*. *Silvae Genetica* **21**: 155–163.
- Sokal RR, Oden NL (1978). Spatial autocorrelation in biology I. Ethodology. *Biol J Linnean Soc* **10**: 199–228.
- Soranzo N, Provan J, Powell W (1998). Characterization of microsatellite loci in *Pinus sylvestris* L. *Mol Ecol* **7**: 1247–1263.
- Sork VL, Huang S, Wiener E (1993). Macrogeographic and fine-scale genetic structure in a North American oak species, *Quercus rubra* L. *Annales des Sciences Forestières* **50**(Suppl): 261s–270s.
- Spiegelhalter DJ, Best NG, Carlin BP, van der Linde A (2002). Bayesian measures of model complexity and fit (with discussion) *Journal of the Royal Statistical Society, Series B. Stat Methodol* **64**: 583–639.
- Streiff R, Labbe T, Bacilier R, Steinkellner H, Glössl J, Kremer A et al. (1998). Within-population genetic structure in *Quercus robur* L. & *Quercus petraea* (Matt.) Liebl. Assessed with isozymes and microsatellites. *Mol Ecol* **7**: 317–328.
- Surles SE, Arnold J, Schnabel A, Hamrick JL, Bongarten BC (1990). Genetic relatedness in open-pollinated families of two leguminous tree species, *Robinia pseudoacacia* L. and *Gleditsia triacanthos* L. *Theor Appl Genet* **80**: 49–56.
- Uchiyama K, Goto S, Tsuda Y, Takahashi Y, Ide Y (2006). Genetic diversity and genetic structure of adult and buried seed populations of *Betula maximowicziana* in mixed and post-fire stands. *Forest Ecol Manage* **237**: 119–126.
- Vogl C, Karhu A, Moran G, Savolainen O (2002). High resolution analysis of mating systems: inbreeding in natural populations of *Pinus radiata*. *J Evol Biol* **15**: 433–439.
- Wasser SK, Shedlock AM, Comstock K, Ostrander EA, Mutayoba B, Stephens M. et al (2004). Assigning African elephant DNA to geographic region of origin: applications to the ivory trade. *Proc Nat Acad Sci USA* **101**: 14847–14852.
- Xie CY, Knowles P (1991). Spatial genetic substructure within natural populations of jack pine (*Pinus banksiana*). *Can J Bot* **69**: 547–551.
- Yazdani R, Lindgren Y, Stewart S (1989). Gene dispersion within a population of *Pinus sylvestris*. *Scand J Forest Res* **4**: 295–306.
- Young AG, Merriam HG (1994). Effects of forest fragmentation on the spatial genetic structure of *Acer saccharum* Marsh (sugar maple) populations. *Heredity* **72**: 201–208.