Theor Appl Genet (2000) 100:858–865

C. Lexer · B. Heinze · S. Gerber · S. Macalka-Kampfer
H. Steinkellner · A. Kremer · J. Glössl

Microsatellite analysis of maternal half-sib families of *Quercus robur*, pedunculate oak: II. inferring the number of pollen donors from the offspring

Received: 5 June 1999 / Accepted: 30 July 1999

Abstract We present an approach to infer the number of pollen donors directly from genotype data of open-pollinated progeny of *Quercus robur* (pedunculate oak), a highly outcrossing tree species. The approach is based on closely linked, highly polymorphic codominant microsatellite markers. Initially the close linkages between three previously mapped microsatellite loci were confirmed by studies of linkage disequilibrium (LD). Then an approach to track the pollen donors contributing to maternal half-sib families (open-pollinated families) was developed by analysing haplotype arrays of closely linked microsatellite markers transmitted from the fathers to the progeny. Simulated data of five linked microsatellite loci segregating in eight open-pollinated families were used to study the relationship between the number of paternal chromosomes detected by this "haplotype approach" and the number of diploid fathers contributing to the families. The results showed that the number of diploid pollen donors can be expressed as an exponential function of the number of paternal chromosomes inferred from the progeny. The 95% confidence interval of this regression function is used to determine the minimum number of fathers contributing to a genotyped open-pollinated family of *Quercus robur*. Finally this open-pollinated family is used to demonstrate the resolution obtained with the "haplotype approach". Six independent microsatellite loci were used to study relatedness among all pairs of pollen gametes that share a haplotype of three linked markers. The results suggest that the majority of such gametes are identical by descent from the same father. The "haplotype approach" presented here can be used to monitor the number of contributing pollen donors in commercial seedlot samples from oak or any other outcrossing tree species for which closely linked, highly polymorphic, codominant genetic markers are available.

Key words Oak · *Quercus* · Relatedness · Linkage disequilibrium · Microsatellites

Communicated by P.M.A. Tigerstedt

C. Lexer · S. Macalka-Kampfer · H. Steinkellner · J. Glössl (✉)
Zentrum für Angewandte Genetik,
Universität für Bodenkultur Wien, Muthgasse 18,
A-1190 Vienna, Austria
Fax: +43-1-36006-6392
e-mail: gloessl@edv2.boku.ac.at

B. Heinze
Forstliche Bundesversuchsanstalt Wien, Hauptstraße 7,
A-1140 Vienna, Austria

S. Gerber · A. Kremer
INRA,
Laboratoire de Génétique et d'Amélioration des Arbres Forestiers,
B.P. 45, F-33611 Gazinet Cedex, France

Introduction

Seed material for the artificial regeneration of outcrossing tree species is usually composed of maternal half-sib families collected in natural stands or seed orchards (Fineschi et al. 1991; Hattemer 1993). Half-sib family structure within the seed material arises because seeds are harvested from a certain number of mother trees (seed parents) and because the seeds of each mother tree arise by mating events with a certain number of fathers (pollen donors).

In European forestry it is widely recognised that a broad genetic base is necessary for the long-term stability of forests (Anonymus 1990; Anonymus 1993). Genetic diversity at the within-stand-level is increasingly considered to be a major factor in determining the quality of forest seed material (Geburek and Heinze 1998; Anonymus 1996a, b). Given the half-sib family structures within the seedlots, measures to ensure a minimum level of genetic diversity within the seed material will have to focus on two key features: (1) the *number of mother trees* from which seeds are harvested and (2) the *number of father trees* contributing to the families. Concerning point (1), in a recent publication we demonstrated that the number of different maternal oak trees (*Quercus robur*) can be inferred directly from the microsatellite data of

extremely small anonymous acorn samples of five to eight acorns per family, given that the seeds from each putative mother tree have been supplied separately (Lexer et al. 1999). As no genotype information on the parent population is required, these approaches will allow the monitoring of the number of seed parents included in commercial acorn seedlots.

In the investigation described here we extended microsatellite analysis to the number of father trees (pollen donors) contributing to open-pollinated families of pedunculate oak (*Q. robur*). Our objective was to develop an approach that is suitable for application in practice, i.e. the monitoring of commercial seedlots. In order to be of practical value, such an approach should require no genotype information of the parent population, and it should be based on a moderate number of easily scorable polymerase chain reaction (PCR)-based genetic markers. We chose pedunculate oak (*Quercus robur*) as a model tree species due to its abundance and economic importance in Europe and because suitable genetic markers are available. Highly polymorphic, codominant microsatellite markers isolated from the genomes of *Quercus petraea* (Steinkellner et al. 1997) and *Quercus robur* (Kampfer et al. 1998) were used for genetic analysis.

The approach presented here focused on a group of closely linked microsatellite markers chosen from a genetic map of *Q. robur* (Barreneche et al. 1998). As these nuclear microsatellites are transmitted from the fathers to the progeny with very limited recombination, they allow the tracking of the fathers by analysis of the pollen haplotypes received by each of the progeny. Such pollen haplotypes can be identified easily because (1) it has been shown previously that the maternal alleles can be reconstructed from microsatellite data of open-pollinated progeny of oak even if no *a priori* information is available on the mother tree (Lexer et al. 1999); (2) once identified, the maternal alleles can be subtracted from the progeny genotypes, revealing the contributions that were inherited from the different fathers (Dow and Ashley 1998; Streiff et al. 1999); (3) the genetic variability of the closely linked markers can then be combined to define pollen haplotypes. Since recombination between these closely linked markers is very rare, this step is similar to an analysis of haplotypes in the chloroplast genome, where recombination is assumed to be completely absent (see Demesure et al. 1996; Dumolin-Lapègue et al. 1997 and Bucci et al. 1998 for recent applications on forest trees). Close linkage between nuclear genetic markers, often perceived as a factor complicating genetic analysis (Thompson 1991; Thompson and Meagher 1998), serves as a simplifying factor in this particular situation as it allows the tracking of the unknown pollen donors.

Since the approach presented here is based on close linkage between loci, our experiments were designed in the following manner. (1) First the close linkages between the microsatellite loci were confirmed by studies of linkage disequilibrium (LD) in a diploid population sample and a haploid sample of pollen gametes of *Q. robur*. (2) The simulated data of open-pollinated families were then used to develop and test an approach to analyse the pollen donors based on closely linked microsatellites. The simulation studies gave us an opportunity to study families with an identified number of pollen donors. (3) Finally, the approach was applied to data of a genotyped open-pollinated family of *Q. robur*.

We present an approach that may be used to infer the number of pollen donors directly from open-pollinated seeds of oak (*Q. robur*) using closely linked microsatellite markers.

Materials and methods

Plant materials and microsatellite analysis

The following samples were genotyped in the course of this study: (1) 40 seedlings from one natural population of *Q. robur* and (2) 43 open-pollinated progeny of *Q. robur* including their mother tree. The open-pollinated family originates from an arboretum with about 40 potential pollen donors on the site. For both samples, DNA was isolated from leaf material following techniques described in Lexer et al. (1999).

The two samples were genotyped for nine microsatellite markers. The microsatellite loci have been isolated from the genomes of *Q. petraea* (Steinkellner et al. 1997) and *Q. robur* (Kampfer et al. 1998), respectively. All of them have been located on a genetic linkage map of *Q. robur* (Barreneche et al. 1998). The same nine markers were selected for a genetic analysis of maternal half-sib families of oak in a previous study (Lexer et al. 1999). We selected the loci from the genetic map as follows: three loci from a cluster of closely linked microsatellites on linkage group 2, namely ssrQpZAG 46, 36 and 104, and six moderately linked or unlinked loci. PCR amplification and electrophoresis of the microsatellites were conducted as described in Barreneche et al. (1998) and Lexer et al. (1999).

Simulated datasets

Data of five linked microsatellites segregating in eight open-pollinated families were generated by computer simulations. Initially, a diploid population of 10000 individuals was simulated over 500 generations under the assumption of random mating. This was done in order to create polymorphism comparable to that found within a large natural population. The starting allele frequencies for the simulations were taken from the population sample of 40 seedlings (dataset i), using data of the markers ssrQpZAG 46, 36 and 104. These input data consisted of 10–18 alleles per locus occurring in roughly equal frequencies. The mutation rate was set to 10^{-3} for each locus using a single-step stepwise mutation model in which each allele x_i can give rise to x_i+1 or x_i-1, where x is the number of microsatellite repeat units (Valdes et al. 1993; Goldstein and Pollock 1997). The recombination rates between the markers were set according to the previous mapping experiments (Barreneche et al. 1998), where a cluster of five closely linked microsatellites was identified on linkage group 2. The linear order and the recombination rates of the first three markers on the simulated chromosome were Locus 1- (0.011) – Locus 2 – (0.044) – Locus 3, corresponding to ssrQpZAG 46, 36 and 104, respectively. Locus 4 and 5 were introduced simply to investigate the increase in resolution when more than three linked markers are used. These two loci were simulated using again input allele frequencies derived from ssrQpZAG 46 and 36, respectively. The simulated order and the recombination rates of the markers were Locus 3 – (0.020) – Locus 4 – (0.020) – Locus 5. These latter recombination frequencies are fictitious, but the whole region of five linked markers is very similar to the one observed in the mapping experiments.

After simulating for 500 generations, eight open-pollinated families of sample size 40 were created by mating a specified

Table 1 Characterisation of the two samples used for calculating linkage disequilibrium between pairs of microsatellite markers (datasets 1 and 2). The three closely linked markers that are essential to this study are indicated in boldface (*n.s.* not significant)

	Diploid population sample:		Haploid pollen sample:	
Locus	Fixation index	Significance (HW)[a]	Gene diversity	Observed no. of alleles
ssrQpZAG 46	**–0.03**	**n.s.**	**0.88**	**14**
ssrQpZAG 36	**–0.11**	**n.s.**	**0.85**	**10**
ssrQpZAG 104	**–0.06**	**n.s.**	**0.88**	**14**
ssrQpZAG 110	–0.11	n.s.	0.62	6
ssrQpZAG 15	–0.28	n.s.	0.78	7
ssrQpZAG 1/5	+0.04	n.s.	0.78	9
ssrQpZAG 9	+0.07	n.s.	0.83	9
ssrQpZAG 3/64	+0.13	$P<0.05$	0.90	12
ssrQrZAG 112	–0.08	n.s.	0.87	14

[a] HW: Hardy Weinberg equilibrium. Significance levels have been adjusted for multiple tests

number of pollen donors from that population randomly to one mother tree. The number of pollen donors per family was chosen to increase from 5 to 40 in steps of 5. The simulations were programmed using the C language.

Data analysis

Linkage disequilibrium between microsatellite loci

The population sample (dataset 1) and the open-pollinated family (dataset 2) were used to confirm the close physical linkages between three microsatellite loci by studies of genotypic and gametic linkage disequilibrium (LD), respectively.

Initially, both samples were characterised using standard population genetic procedures. This was considered to be necessary because preferential sampling of related individuals can result in significant LD values even in the absence of physical linkage. The population sample was characterised by exact tests for Hardy-Weinberg disequilibrium using GENEPOP 1.2 (Raymond and Rousset 1995). Furthermore, we calculated the fixation index F for each locus as $1\text{-}H_O/H_E$, according to Nei (1987), in order to detect departures from random mating. The open-pollinated family was characterised as follows. First, haploid pollen data were derived from the progeny by subtracting the known maternal alleles from the progeny genotypes (described below, section: analysing a real open-pollinated family). The haploid pollen sample was then characterised by counting the observed number of alleles A_o and by calculating genetic diversity according to Nei (1987).

Tests for genotypic and gametic LD were conducted as exact tests based on Markov chain simulations (Guo and Thompson 1992) in GENEPOP 1.2. The Markov chain was set in such a way that the standard error of each *P*-value was always below 0.006. Finally, the resulting *P*-values were adjusted for multiple tests using Bonferroni procedures as described in Weir (1996, pp 133–135).

Analysing the simulated data – The haplotype approach

Initially the simulated population was characterised by counting the observed number of alleles A_o and by calculating genetic diversity according to Nei (1987). After drawing different simulated families from that population, we obtained haploid pollen data for each family by subtracting the maternal alleles from the simulated progeny. This step was simplified in the simulation study because in the simulated progeny the maternal and paternal contributions were ordered. Therefore, it was possible to identify the paternal allele for each individual at each locus. In real datasets, a certain subset of missing data will arise because some of the individuals will bear, at a given locus, the same 2 alleles as the mother tree. For these individuals the paternal alleles will not be identified unequivocally. However, the simplified procedure chosen here may cause no severe deviation from reality, as suggested by our experiences with a real open-pollinated family (see below).

An approach to estimate the number of paternal chromosomes contributing to open-pollinated progeny was developed using the haploid pollen data of the simulated families in the following manner. First, the genetic variability of the closely linked microsatellite loci was combined in order to count paternal haplotypes. The haplotype sorting and counting was conducted with the software EXCEL 97 (Microsoft Corporation). Next, the haplotype count was corrected for rare recombination events between the linked markers. To obtain this correction, we assumed that every recombination event in one of the father trees mimicked a new paternal haplotype. This simplifying assumption was based on the high levels of diversity observed at these loci in previous studies (Streiff et al. 1998, 1999; Lexer et al. 1999) and in the simulations. According to this assumption, the number of paternal haplotypes that were expected to arise solely due to recombination was calculated as $\text{Theta}\times n_g$, where Theta is the recombination rate in the chromosomal interval under consideration, obtained in the previous mapping experiments, and n_g is the total number of pollen gametes studied. The number of paternal chromosomes detected (n_c), corrected for rare recombination events, is therefore given by $n_c = n_h - (\text{Theta}\times n_g)$, where n_h is the number of paternal haplotypes counted. The resulting value was adjusted to give discrete numbers of paternal chromosomes. A standard error for the estimate was derived from the standard error of Theta as calculated by the computer programme JOINMAP 1.3 (Stam 1993) during the previous mapping experiments. This approach (from now on: haplotype approach) was used to count the number of paternal chromosomes for the simulated datasets and for a genotyped open-pollinated family (dataset 2).

For the simulated data the haplotype approach was employed using different combinations of linked microsatellites in order to study the effect of the number of linked markers. Then the results were compared to the (known) number of fathers in each dataset. Regression analysis and curve estimation were conducted using the SPSS 8.0 software package (SPSS, Chicago).

Analysing a real open-pollinated family

The genotyped open-pollinated family (dataset 2) was used to apply the haplotype approach to real data. First haploid pollen data were obtained by subtracting the (known) maternal alleles from the progeny genotypes. In one case a seedling had the same 2 alleles as the mother tree, so the paternal alleles were not identified and the genotype for that locus was treated as missing data. Next the number of paternal chromosomes was calculated from the haploid pollen data following the haplotype approach. Finally, the number of diploid fathers was estimated using the regression function obtained with the simulated data.

Table 2 Results of exact tests for linkage disequilibrium (LD) in the diploid population sample (dataset 1) and the haploid pollen sample (dataset 2). Pairwise comparisons between the three closely linked markers that are essential to this study are indicated in boldface. The recombination rate Theta±standard error is given to show concordance between the LD results obtained here and the linkage results obtained previously

Locus-pair[a]	Genotypic disequilibrium[b] (population sample)	Gametic disequilibrium[b] (pollen sample)	Theta±standard-error
ssrQp36-Qp46	**0.00023****	**0.00000****	**0.011±0.008**
ssrQp36-Qp104	**0.00196**	**0.00061***	**0.044±0.021**
ssrQp46-Qp104	**0.03073**	**0.00086***	**0.056±0.024**
ssrQp1/5-Qp9	0.23553	0.00162	0.227±0.031
Unlinked comparisons (mean):	0.33686	0.11112	Theta>0.400

* $P<0.05$; ** $P<0.01$

[a] ssrQpZAG 46, 36 and 104 are closely linked on linkage group 2, ssr QpZAG 1/5 and 9 are moderately linked on linkage group 7 of the *Q. robur* genetic map. Comparisons between completely unlinked markers are represented by their mean values

[b] Asterisks indicate significant p-values after correction for multiple tests

In order to demonstrate the level of resolution obtained with three linked markers, we calculated genetic relatedness (Queller and Goodnight 1989) among pairs of pollen gametes. This relatedness measure was applied to all pairs of pollen gametes that shared a haplotype at a given combination of linked markers. The calculations were based on the six independent microsatellites listed in Table 1 using the computer programme KINSHIP 1.2 (Goodnight and Queller 1999). The rationale behind this procedure was as follows: if gametes share the same haplotype of linked markers due to identity by descent from the same father, then relatedness among them, based on a set of independent reference markers, will be about 0.5. This will be the case because at each of the independent loci, the pollen gametes will receive 1 of the 2 possible alleles of that father with a probability of about 0.5, given that the father is heterozygous. In case of the father being homozygous, the pollen gametes will receive this paternal allele with a probability of 1. On the other hand, if gametes share the same haplotype of linked markers just by chance, then relatedness based on a set of highly polymorphic independent reference markers is expected to be lower than 0.5. Therefore, such relatedness calculations can provide basic information on the potential of a set of linked microsatellites to assign a given gamete to a specific father.

The relatedness results were confirmed by likelihood calculations with KINSHIP 1.2, testing the hypothesis of full-sib ship (r=0.5) over that of being unrelated (r=0). To compare a given pair of gametes, we calculated reference allele frequencies using the haploid pollen data, excluding each time the pair of individuals under consideration.

Results and discussion

Linkage disequilibrium between microsatellite loci

Characterisation of the population sample (dataset 1) and the pollen sample (dataset 2) revealed no strong tendency of inbreeding or close relatedness among the sampled individuals. For the population sample, exact tests of Hardy-Weinberg proportions revealed only one significant departure (P=0.05) among the nine loci used in this study. The fixation indices for the nine loci ranged between –0.28 and +0.13. The F values were not concordant across loci. However, the low proportion of positive F values (3 out of 9 loci) suggests that there is no severe inbreeding among the sampled individuals (Table 1). For the pollen sample, genetic diversity ranged from 0.62 to 0.90 at each locus, the largest observed number of alleles per locus was 14 (Table 1). Based on these results, both samples were considered to be suitable for the LD studies.

In Table 2, the P values of pairwise tests for LD as well as their significance after correction for multiple tests are compared to the recombination rate Theta as obtained in the course of the previous mapping experiments. It can be seen that the P values of exact tests for LD between pairs of markers follow the same pattern as Theta. This trend can be observed for the population sample (genotypic LD) and for the open-pollinated progeny (gametic LD). In each of the two samples, the P values are smallest for the three closely linked loci ssrQpZAG 46, 36 and 104, reflecting the close physical linkage of these markers. The P values for the moderately linked marker pair ssrQpZAG 9 and 1/5 are intermediate in each of the two samples, whereas the probabilities for the unlinked marker pairs, represented by their mean values, are generally large. In each of the two samples, only 1 out of 32 comparisons between unlinked loci resulted in a ”false-positive” significant LD (not shown). Our results confirm the close physical linkage between the microsatellite markers ssrQpZAG 46, 36, and 104. Based on our results, these three markers were chosen for our analysis of the pollen donors contributing to open-pollinated families. The six unlinked or moderately linked microsatellites were used as a reference set of independent markers later in this study.

Analysis of the pollen donors in simulated families

Genetic diversity in the simulated family after 500 generations ranged between 0.77 and 0.93, and between 9 and 31 alleles per locus were observed. These levels of polymorphism are comparable to those observed for microsatellite loci in a mixed stand of *Q. robur* and *Q. petraea* (Streiff et al. 1998). As the input allele frequencies for the simulation study were derived from a real population sample, the differences in polymorphism between the simulated loci reflect the locus-specific polymorphism that was found in nature.

For each of the eight open-pollinated families that were drawn from this population the number of paternal chromosomes was first calculated by the haplotype approach (see Materials and methods), and then the number of paternal chromosomes was compared to the simu-

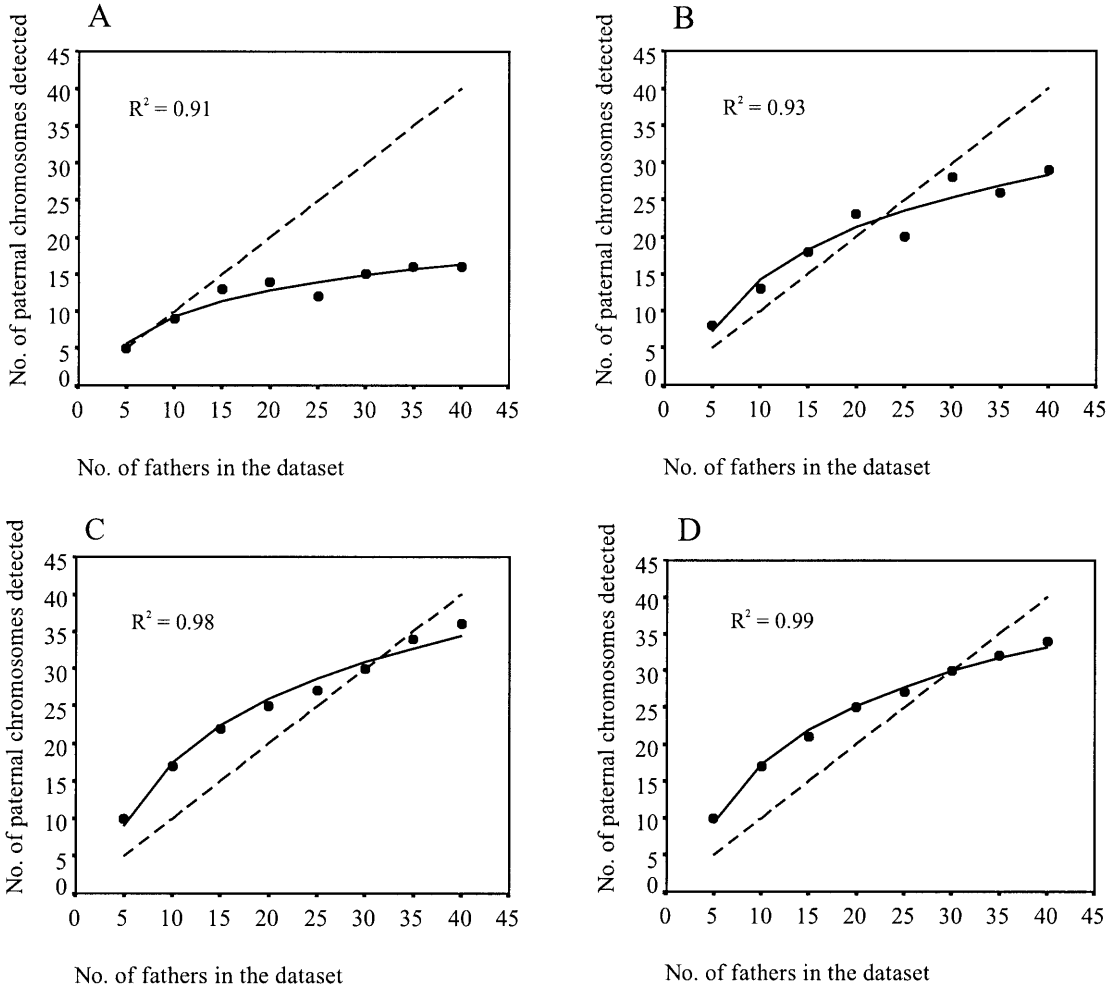

Fig. 1A–D The relationship between the number of paternal chromosomes detected and the number of fathers in different simulated datasets. The datapoints represent the results of eight different simulated families of 40 progeny each. *Solid lines* represent the logarithmic regression curve, *dashed lines* represent perfect collinearity. **A** One locus, **B** two linked loci, **C** three linked loci, **D** five linked loci

lated number of fathers in each family by regression statistics. Figure 1 shows the results of regression analysis for different combinations of linked loci. The number of paternal chromosomes detected is plotted as a function of the number of fathers in each dataset. For each of the locus combinations it is possible to fit a logarithmic function of the general formula $y=b_0+b_1\times\ln(x)$ to the data, where x is the number of fathers in the dataset and y is the number of paternal chromosomes detected. The regression coefficient r^2 for the relationship between the two variables ranges between 0.91 and 0.99, depending on the number of loci employed.

In order to validate the haplotype approach, we plotted the 1:1 line relative to the observed data for each locus combination. Any estimate that is completely collinear to the number of fathers in the dataset would be located on this line. As expected, using information from just one locus (i.e. counting paternal alleles at that locus) results in estimates that are far from the 1:1 line and are therefore far from the number of fathers in the dataset (Fig. 1A). When haplotypes are counted for two linked markers, the estimates are located closer to the 1:1 line (Fig. 1B) and after information from three linked loci is combined the observed data are scattered even more closely around that line (Fig. 1C). When four or five linked loci are employed, the increase in resolution is only small (only the results of 5 linked loci are shown; Fig. 1D). Note that for each locus combination the number of paternal chromosomes detected moves towards a plateau as the number of fathers approaches 40.

It would be beyond the scope of the present study to investigate in detail all of the causal relationships that are responsible for the shape of the observed function. Such investigations would require additional simulation studies, including different degrees of linkage between loci, in order to characterise the effect of linkage on the observed plateau effect. However, leaving linkage aside it seems obvious that the number of paternal chromosomes detected is influenced primarily by two variables: (1) the number of loci employed and (2) the number of fathers in the dataset. With respect to point (1), in this simulation study the number of paternal chromosomes detected increased only marginally when more than three linked loci were used. The importance of point (2) is best

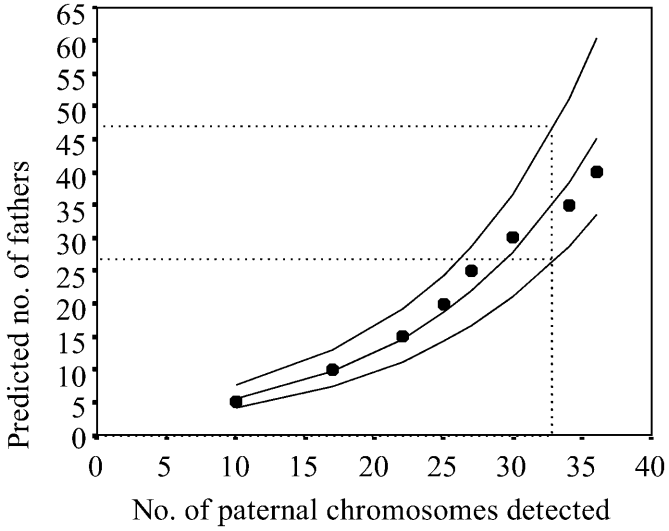

Fig. 2 Exponential regression curve and 95% confidence interval for the number of fathers as predicted from the number of paternal chromosomes detected with three linked loci. The datapoints are identical to the ones in Fig. 1C, only the axes have been swapped. *Dashed lines* are projections marking the 95% confidence interval for the number of fathers predicted for a genotyped open-pollinated family (dataset 2)

described by following the shape of the curve for three or five linked loci (Fig. 1C or 1D): when the number of fathers in the dataset is small, each father is likely to be represented by both of its complementary haplotypes. Therefore, for small numbers of fathers the haploid chromosome estimate tends to exceed the number of diploid fathers, and the estimate is located above the 1:1 line. As the number of fathers in the dataset increases, it becomes more likely that each father is represented by only one of its two complementary haplotypes. Then the chromosome estimate tends to reflect the real number of fathers in the dataset more closely, and the curve approaches the 1:1 line. On the other hand, as the number of fathers in the dataset increases, the potential of the linked markers to resolve all of the paternal chromosomes becomes smaller, leading to the above-mentioned plateau effect. Therefore, the shape of the curve may be interpreted as the combined effect of linkage, the number of linked markers and the number of fathers in the dataset.

In Fig. 1 the number of paternal chromosomes detected was placed on the y-axis for better visualisation. However, in a real dataset the number of fathers will be the unknown variable, and the number of paternal chromosomes detected will be the independent variable obtained from the genotype data. In such a situation it is desirable to express the number of fathers as a function of the chromosome count (Fig. 2). The relationship between the two variables can then be described by an exponential function of the general formula $y=b_0 \times e^{b1 \times x}$, where x is the number of paternal chromosomes detected and y is the number of fathers. This definition of the relationship is simply the inverse of the logarithmic function presented in Figure 1C. It allows the construction of confidence intervals for the number of fathers as predicted from the chromosome count (Fig. 2). The lower 95% confidence limit may then be used as a measure for the minimum number of fathers present in the dataset. Based on the simplifying assumptions of our simulations, we obtained a model to predict the minimum number of fathers contributing to open-pollinated progeny using haplotype counts of closely linked microsatellites.

Analysis of the pollen donors in a real family

In order to validate the haplotype approach using ”real” data, we studied an open-pollinated family whose paternal contributions were unknown (dataset 2). Combining genetic variability at the three closely linked microsatellites ssrQpZAG 46, 36 and 104 resulted in 35 haplotypes of linked markers among the 43 progeny. This number was subsequently corrected for recombination events between the linked markers as described above, resulting in an estimate of 33 (±1) paternal chromosomes; two haplotypes were assumed to arise solely due to recombination events.

The exponential model derived from the simulations was used to estimate the number of diploid fathers contributing to the progeny. If the exponential model is also valid for our real data, then the 33 chromosomes detected translate into a lower 95% confidence limit of 27 fathers (Fig. 2). Therefore, at least 27 fathers may have contributed to the progeny, a prediction that is compatible with the circumstances of our study site. About 40 potential pollen donors are located on the study site, and additional potential fathers are located in close vicinity.

Ideally such a result should be confirmed by comparison to the genotypes of the contributing fathers, but this was not possible for our study family. Given the high levels of pollen flow in temperate oaks (Dow and Ashley 1996, 1998; Streiff et al. 1999), we expected many of the pollen donors to be located outside our study site, making it difficult to obtain genotype information from all potential fathers. For this reason we did not attempt to identify the fathers by a ”classical” paternity analysis. Nevertheless we considered it important to demonstrate the level of resolution achieved by combining three linked microsatellites in a real open-pollinated family. Therefore we chose an approach based on genetic relatedness at the six unlinked microsatellite loci described in detail above.

Figure 3A shows the distribution of genetic relatedness values among all pairs of pollen gametes sharing an allele at locus ssrQpZAG 104. This locus was chosen as an example because it was the most polymorphic locus in this family. Figure 3B and 3C show the distribution of relatedness values among all pairs of pollen gametes sharing a haplotype at the linked locus combinations ssrQpZAG 46–36 (2 linked loci) and ssrQpZAG 46–36–104 (3 linked loci), respectively.

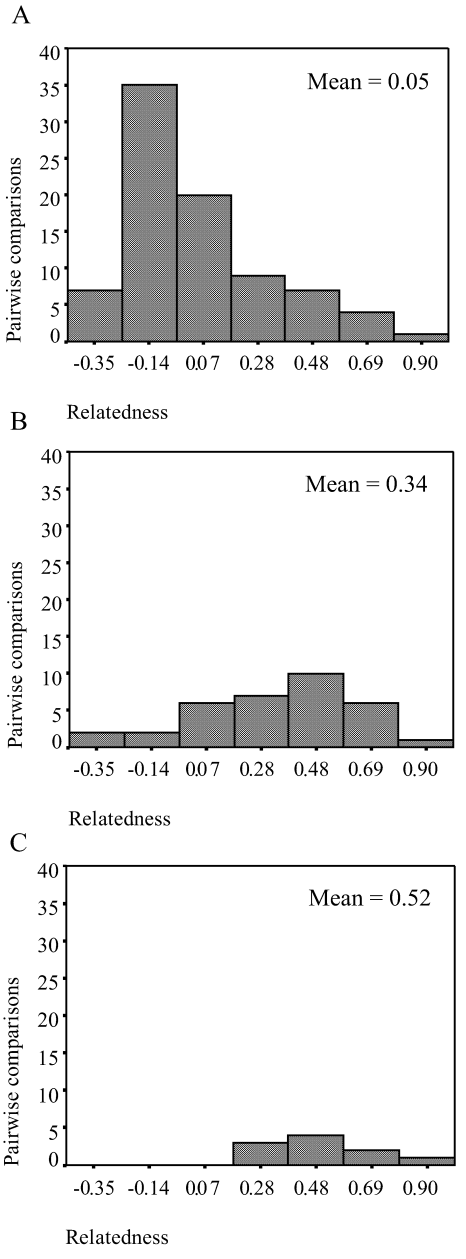

Fig. 3A–C Distribution of pairwise relatedness values among pairs of pollen gametes of a genotyped open-pollinated family (dataset 2). Relatedness was calculated among all pairs of pollen gametes that shared an allele at locus ssrQpZAG 104 (**A**) or a haplotype at the linked markers ssrQpZAG 46 and 36 (**B**) or a haplotype at the linked markers ssrQpZAG 46, 36 and 104 (**C**). Relatedness was calculated using six independent microsatellite loci

It can be seen that pollen gametes sharing alleles at locus ssrQpZAG 104 are only moderately related (Fig. 3A). Relatedness increases dramatically, when two closely linked loci are used in combination to define haplotypes (Fig 3B), and it increases even further when three linked loci are employed (Fig. 3C); mean genetic relatedness among the gametes is then about 0.5 (Fig. 3C). Note that the absolute number of pairwise comparisons decreases as more linked loci are added. This is the case because adding additional linked loci increases resolution and therefore leads to a decrease in the number of pollen gametes that share a haplotype of linked markers.

The results shown in Fig. 3 were supported by likelihood statistics. The proportion of positive log likelihood ratios, indicating that a relatedness of r=0.5 among the gametes is more likely than being unrelated (r=0.0), increased dramatically as more linked loci were added. Only 1 pairwise comparison had a negative log likelihood ratio when three linked loci were used in combination for definition of the haplotypes (not shown).

The results suggest that the majority of pollen gametes sharing a haplotype of three linked markers are identical by descent from the same father. We do *not* conclude that this is the case for *all* pairs of gametes sharing a haplotype at the loci ssrQpZAG 46, 36 and 104. The pairwise relatedness calculations based on six independent microsatellites would not even allow such conclusions. Furthermore, some of the pollen gametes may share the same allele due to (unknown) family structures among the pollen donors rather than by identity by descent from the same father. However, Fig. 3 certainly demonstrates the increase in resolution when the genetic variability of closely linked microsatellites is combined. How many linked loci are employed in practice to detect the paternal chromosomes will depend on the number of closely linked loci available and on the precision desired.

Practical considerations

We have presented an approach to estimate the number of paternal chromosomes contributing to open-pollinated families of *Q. robur*. Furthermore, we have used simulated data to construct a regression function that relates the haploid chromosome count to the number of diploid fathers in the dataset. Our data suggest that the approach may be useful for the sample sizes that were chosen in the present contribution. However, the results also show that the number of chromosomes detected with linked markers reaches a plateau very quickly when the number of fathers increases (Fig. 1). Therefore, the haplotype approach may be useful for moderate sample sizes like the ones chosen here.

The haplotype approach to detect paternal chromosomes may be highly useful to monitor the number of pollen donors in commercial acorn seedlots. The method is designed for a small number of PCR based markers (3 to 5 linked microsatellites). Furthermore, no genotype information is required of the parent population. Therefore, it should be applicable to practical situations with limited financial resources and a need for quick answers.

In particular, the approach presented here may help to decide whether a certain population should be included in the list of approved stands for commercial seed harvest or not. Such lists of approved stands for seed harvest are common in several European countries

(Geburek and Heinze 1998). If a stand repeatedly shows very low numbers of pollen donors per mother tree, then it may in some situations be wise to take it off the list and replace it by other stands. For instance, this may be the case for isolated stands with limited pollen flow from outside the stand.

Furthermore, if experimental samples are collected from the same mother trees over several years, it may be possible to relate fluctuations in the number of pollen donors to variations in the masting behaviour or to other factors. As no genotype information of the parent population is required, such studies are no longer restricted to model stands but can be conducted for a wide range of stands under more diverse conditions.

Finally, the approach presented here may be useful for genetic analyses of seedlots from other outcrossing tree species as well. Here, we have used *Quercus robur* as a model because numerous microsatellites are available for this species and because they are being mapped genetically. However, our approach should be applicable to any outcrossing tree species once closely linked highly polymorphic microsatellite markers have been identified.

Acknowledgements We are thankful to Elmar Kickingereder for skillful technical assistance and to Birgit Ziegenhagen, Institut für Forstgenetik, BFH Großhansdorf, D, for providing us with DNA from an open-pollinated family of *Quercus robur*. Furthermore we thank Manfred J. Lexer, Instiut für Waldbau from this University, for helpful discussions on the regression statistics. This work was supported by the Biotechnology Research Programme of the European Commission, DG XII, ERB-BIO4-CT960706. The experiments comply with the current laws of Austria.

References

Anonymus (1990) Ministerial Conference on the Protection of Forests in Europe (Strasbourg resolution). Ministère de l'Agriculture et des Forêts, Paris

Anonymus (1993) Sound forestry – sustainable development. Report on the follow-up ministerial conference on the protection of forests in Europe. Ministry of Agriculture and Forestry, Helsinki

Anonymus (1996a) Bundesgesetzblatt für die Republik Österreich. 1996/419. Bundesgesetz über forstliches Vermehrungsgut (Forstliches Vermehrungsgutgesetz)

Anonymus (1996b) Bundesgesetzblatt für die Republik Österreich. 1996/512. Verordnung, Forstliches Vermehrungsgut

Barreneche T, Bodenes C, Lexer C, Trontin JF, Fluch S, Streiff R, Plomion C, Roussel G, Steinkellner H, Burg K, Favre JM, Glössl J, Kremer A (1998) A genetic linkage map of *Quercus robur* L. (pedunculate oak) with RAPD, SCAR, microsatellite, isozyme and rDNA markers. Theor Appl Genet 97:1090–1103

Bucci G, Anzidei M, Madaghiele A, Vendramin GG (1998) Detection of haplotypic variation and natural hybridization in *halepensis*-complex pine species using chloroplast simple sequence repeat (SSR) markers. Mol Ecol 7:1633–1643

Demesure B, Comps B, Petit RJ (1996) Chloroplast DNA phylogeography of the common beech (*Fagus sylvatica* L.) in Europe. Evolution 50:2515–2520

Dow BD, Ashley MV (1996) Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, *Quercus macrocarpa*. Mol Ecol 5:615–627

Dow BD, Ashley MV (1998) High levels of gene flow in bur oak revealed by paternity analysis using microsatellites. J Hered 89:62–70

Dumolin-Lapègue S, Demesure B, Fineschi S, Le Corre V, Petit RJ (1997) Phylogeographic structure of white oaks throughout the European continent. Genetics 146:1475–1487

Fineschi S, Malvolti ME, Cannata F, Hattemer HH (eds) (1991) Biochemical markers in the population genetics of forest trees. [Proc Joint Meet Work Parties S2.04–01 Pop Genet Ecol Genet S2.04–05 Biochem Genet Int Union For Res Organizations (IUFRO)]. SPB Academic Publ, The Hague, The Netherlands

Geburek TH, Heinze B (eds) (1998) Erhaltung genetischer Ressourcen im Wald-Normen, Programme, Maßnahmen. Ecomed-Verlagsgesellschaft, Landsberg, Germany

Goldstein DB, Pollock DD (1997) Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. J Hered 88:335–342

Goodnight KF, Queller DC, Poznansky T KINSHIP 1.2 Rice University, Department Of Ecology and Evolutionary Biology, Houston, Texas. Available for downloading at http://www.bioc.rice.edu/~kfg/GSoft.html

Guo SW, Thompson EA (1992) Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics 48:361–372

Hattemer HH, Bergmann F, Ziehe M (1993) Einführung in die Genetik für Studierende der Forstwissenschaft. 2. Auflage. Sauerländer's Verlag, Frankfurt am Main, Germany

Kampfer S, Lexer C, Glössl J, Steinkellner H (1998) Characterization of $(GA)_n$ microsatellite loci from *Q. robur*. Hereditas 129: 183–186

Lexer C, Heinze B, Steinkellner H, Kampfer S, Ziegenhagen B, Glössl J (1999) Microsatellite analysis of maternal half-sib families of *Quercus robur*, pedunculate oak: Detection of seed contaminations and inference of the seed parents from the offspring. Theor Appl Genet 99:185–191

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

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

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. J Hered 86: 248–249

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: JOINMAP. Plant J 3: 739–744

Steinkellner H, Fluch S, Turetschek E, Lexer C, Streiff R, Kremer A, Burg K, Glössl J (1997) Identification and characterization of $(GA/CT)_n$ -microsatellite loci from *Quercus petraea*. Plant Mol Biol 33:1093–1096

Streiff R, Labbe T, Bacilieri R, Steinkellner H, Glössl J, Kremer A (1998) Within population genetic structure in *Quercus robur* L. and *Quercus petraea* (Matt.) Liebl. assessed with isoenzymes and microsatellites. Mol Ecol 7:317–328

Streiff R, Ducousso A, Lexer C, Steinkellner H, Glössl J, Kremer A (1999) Pollen dispersal inferred from paternity analysis in a mixed oak stand of *Quercus robur L.* and *Quercus petraea* (Matt.) Liebl. Mol Ecol 8:831–841

Thompson EA (1991) Estimation of relationships from genetic data. In: Rao CR, Chakraborty R (eds) Handbook of statistics, vol 8. Elsevier Science Publ, Amsterdam, pp 255–269

Thompson EA, Meagher TR (1998) Genetic linkage in the estimation of pairwise relationship. Theor Appl Genet 97:857–864

Valdes AM, Slatkin M, Freimer NB (1993) Allele frequencies at microsatellite loci: the stepwise mutation model revisited. Genetics 133:737–749

Weir, BS (1996) Genetic Data Analysis II. Sinauer Associates, Sunderland, Massachusetts, pp 133–135