npg

Heterozygote deficiencies in small lacustrine populations of brook charr *Salvelinus Fontinalis* Mitchill (Pisces, Salmonidae): a test of alternative hypotheses

V Castric[1], L Bernatchez[1], K Belkhir[2] and F Bonhomme[2]

[1]*GIROQ, Département de biologie, Université Laval, Ste-Foy, Québec, Canada, G1K 7P4;* [2]*Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université Montpellier II, Case Courrier 63, Place E. Bataillon, 34095 Montpellier, Cedex 5, France*

Empirical studies of natural populations have commonly reported departures from Hardy-Weinberg expected proportions of heterozygote individuals. Recent advances in statistical population genetics now offer the potential to exploit individual multilocus genotypic information to test more rigorously for possible sources of heterozygote deficiencies. In a previous study in lacustrine brook charr (*Salvelinus fontinalis*), we reported stronger deficits in small than in large lakes. In the present paper, we propose a methodology for empirically testing alternative hypotheses to identify the cause of the deficits observed in three of the smallest lakes (85, 109 and 182 ha) analysed. First, as in several salmonid species, brook charr may exhibit a trophic polymorphism in north temperate lakes. If morphs are genetically divergent, indiscriminate sampling of both forms would result in less heterozygote individuals than expected in a randomly mating population (Wahlund effect). Using an individual-based method aiming at detecting cryptic population structure, we can reject this explanation as the sole source of deficits for all three lakes. Secondly, mating among relatives could also be frequent in small lakes and lead to heterozygote deficiencies. Significantly more fish than expected at random had low individual multilocus heterozygosity in two of the lakes, suggesting that inbred fish may have been present. Thirdly, sampling of genetically related fish would also lead to departures from Hardy-Weinberg proportions. In the same two lakes, the distribution of pairwise relatedness coefficients departed from its random expectation, suggesting that non-random sampling of kin may have occurred.
Heredity (2002) **89,** 27–35. doi:10.1038/sj.hdy.6800089

Introduction

In large and randomly mating populations of diploid species, the pairing of alleles within individuals can be predicted using binomial sampling properties according to Hardy-Weinberg equilibrium (HWE). Interpretations of departures from these expected genotypic proportions have been a recurrent theme of both theoretical and empirical population genetics studies since their first formal stating by Hardy (1908) and Weinberg (1908). Once artefactual sources of deficits have been discarded (eg, entirely or sporadically non-amplifying alleles: Callen *et al* 1993; Pemberton *et al*, 1995; Brookfield, 1996; Van Treuren, 1998; Wattier *et al*, 1998), heterozygote deficiencies may be used to infer various aspects of breeding systems *sensu lato*, such as quantifying the rate of selfing (Jarne and Charlesworth, 1993; Viard *et al*, 1997; Enjalbert and David, 2000) or the levels of consanguineous matings in natural and captive populations of non-selfing organisms (Wright, 1921; Thornhill, 1993). In largely outcrossing species, heterozygote deficits have commonly been interpreted as indirect evidence of the mixing of differentiated gene pools (Wahlund effect; Wahlund, 1928), while, on the other hand, heterozygote excesses have been used to quantify the effective number of breeders in small populations (Pudovkin *et al*, 1996; Luikart and Cornuet, 1999).

Overall, heterozygote deficiency in non-selfing, diploid populations is relatively common (see for example, Waldman and McKinnon 1993 for a review in fishes, amphibians and reptiles; Zouros and Foltz, 1984 in marine bivalves). Yet very few studies have been specifically aimed at systematically testing the possible causes of deficiencies (but see Christein *et al*, 1979; Zouros and Folz, 1984; David *et al*, 1997). Typically, a biologically plausible explanation is suggested as the most likely source of heterozygote deficits, but there is often a lack of rigorous testing of the various alternative hypotheses (but see Overall and Nichols, 2001). The recent advances in statistical population genetics prompted by the use of highly polymorphic markers such as microsatellites now offer the potential of quantifying departures from HWE with high precision (Chakraborty and Zhong, 1993; Raymond and Rousset, 1995) and exploiting individual multilocus genotypic information to test more rigorously for possible sources of heterozygote deficiencies.

Heterozygote deficiencies have been recently documented in lake populations of the brook charr (Castric *et al*, 2001), a freshwater salmonid whose movements in riverine habitats are restricted (Bélanger and Rodriguez, 2001). Deficiencies were negatively correlated with lake size (ie, they were stronger in smaller lakes), so we focused here on three small lakes (size <200 ha) whose census size was sufficiently large to allow a temporal comparison by resampling a second year. Since technical artefacts, such as non-amplifying (null) alleles or selective amplification of shorter alleles would be systematic for all the analysed samples, they could not account for the relationship with lake size and could be ruled out (Castric *et al*, 2001). We thus focus here on three competing hypotheses related to the biology of brook charr inhabiting small lakes.

First, heterozygote deficiency could result from a subdivision of the local population into isolated and differentiated reproductive units (Wahlund effect). The coexistence of genetically isolated populations of the same species in single lakes is a common feature of many northern temperate freshwater fishes (reviewed in Taylor, 1999). Typically, genetic differentiation among such populations is accompanied by differences in trophic related traits. Dynes *et al* (1999) recently documented that such a differentiation of sympatric lacustrine populations may also occur in brook charr. However, the extent of morphological differentiation between the two forms (benthic and pelagic) appears much subtler than that reported in other species. Consequently, the occurrence of sympatric but differentiated populations of brook charr in our samples may have remained undetected. Although the underlying mechanisms are still debated (Diekman and Doebeli, 1999; Kondrashov and Kondrashov, 1999), intraspecific competition is believed to be one of the major causes of the divergence of sympatric populations of northern temperate lacustrine fishes (reviewed in Schluter, 2000). If intraspecific competition is increased in smaller lakes because of more limited resources, the probability and intensity of population divergence could be higher in smaller lakes, thereby explaining the observed correlation between lake size and heterozygote deficiency due to Wahlund effect.

A second possible cause for the observed heterozygote deficiency is the mating of close relatives (Wright, 1921). Although little is known about brook charr reproductive behaviour under natural conditions (but see Blanchfield and Ridgway, 1996), kin recognition has been shown to occur in controlled environments (Hiscock and Brown, 2000) and could potentially bias the reproductive system towards inbreeding.

Finally, a third hypothesis could be the non-random sampling of members from a limited number of families. When few parents contribute to a given sample, the discreteness of the number of possible different genotypes leads to a slight heterozygote excess in the progeny relative to HW proportions in an infinite population with the same allele frequencies (Robertson, 1965; Crow and Kimura, 1970, pp 55–56; Pudovkin *et al*, 1996; Luikart and Cornuet, 1999). However, when several familial groups are sampled, each involving only a limited number of progenitors, allele frequencies are expected to vary among such groups, even if progenitors are drawn from the same source population. Hence, when the progeny of several reproductive groups are pooled, heterozygote

excess may be counterbalanced by a 'family' Wahlund effect, whose relative magnitude is not known (Pudovkin *et al*, 1996). More specifically, it is unclear how the number of pooled reproductive groups may affect heterozygote proportions in the next generation. Because larger lakes potentially provide more suitable spawning grounds, the number of reproductive groups may be an important factor to consider in explaining the relationship between lake size and heterozygote deficiencies. Specifically, fish from small lakes could be derived from a few families only, and consequently the probability of sampling two related individuals is higher than expected in a randomly mating population .

In this study, we used an individual-based approach to analyse multilocus genotypic data from six microsatellite loci in order to test the above three competing explanations for the observed heterozygote deficiencies. We first tested the possibility of detecting a Wahlund effect using a maximum likelihood method and quantified the detection power of that method using permutation techniques. We then assessed the distribution of pairwise relatedness coefficients to test for the existence of a limited number of effective breeders producing a given cohort. Finally, we performed computer simulations to explore how the number of spawning grounds could influence departures from HWE.

Materials and methods

Lake size and interlocus variance of F_{IS}
We first extended the data analysis from the 22 lakes sampled in 1997 and included in Castric *et al* (2001), which provided evidence of stronger heterozygote deficiency in smaller lakes. A systematic investigation of variation in F_{IS} estimates f (in what follows, F_{IS} and F_{ST} will stand for their estimates by f and θ respectively, Weir and Cockerham, 1984) across loci was not performed in that study. We therefore plotted interlocus variance in f-values against lake size and tested the correlation using Spearman's R.

Sampling and microsatellite data
If deficiencies proceed from a single cause, then it should be easier to identify in small lakes in which the signal is expected to come out stronger. Thus, with the aim of identifying the cause of heterozygote deficiencies and their variance across loci in small lakes, all further analyses focused on three selected lakes in which additional sampling could be performed in 1999. These lakes (Clish Pond, Robbins Brook Pond and Third Wallagrass Lake) were small (85, 109 and 182 ha) and differed in the distribution of heterozygote deficiencies among loci (Figure 1). Clish Pond was representative of most small lakes in Castric *et al* (2001) with strong and highly variable heterozygote deficiencies across loci. Robbins Brook Pond exhibited a slightly different pattern, with constantly strong deficits for all six loci. The third lake, named Third Wallagrass Lake, exhibited slightly positive but non-significant F_{IS} estimates. In addition to those fish sampled in 1997 in Clish Pond ($n = 28$), Robbins Brook Pond ($n = 28$), and the Third Wallagrass Lake ($n = 18$), 34, 50 and 51 additional fish were respectively collected by successive trap net attempts at different locations of the same lakes in 1999. Total DNA was extracted from white muscle
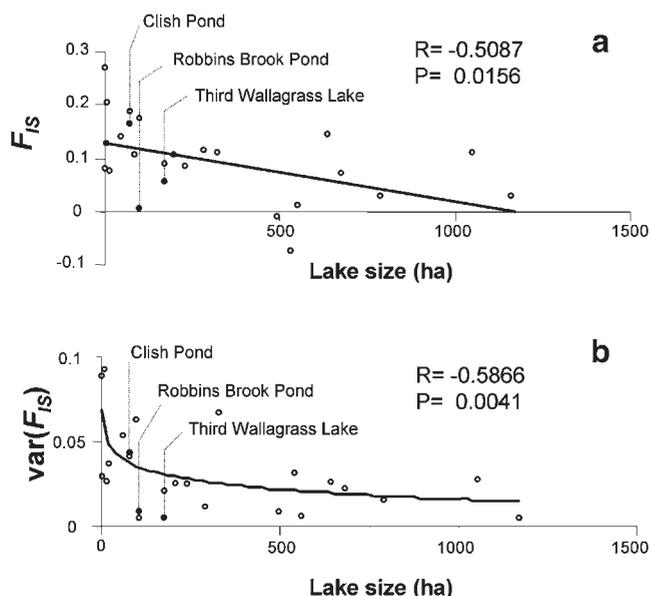
Figure 1 Correlations between lake size and (**a**) heterozygote deficiencies (*f*) (**b**) variance in *f* across six microsatellite loci. R is Spearman's non-parametric correlation coefficient. A log-linear curve was fitted on (b). Open circles are from Castric *et al* (2001) from which only populations with $n > 20$ were included, and black circles are additional data from this study.

tissue using phenol-chloroform protocol and fish were genotyped at six microsatellite loci as detailed in Castric *et al* (2001).

Wahlund effect?

Fewer heterozygotes are expected than under random mating when differentiated gene pools are sampled together (cryptic population structure). This hypothesis was tested using a maximum likelihood (ML) method aiming at unravelling cryptic population structure. The null hypothesis was that fish from each lake are a representative sample of a single Hardy-Weinberg population, and the alternative hypothesis is that the sample is a mixture of two differentiated populations (benthic and limnetic in the present case, Dynes *et al*, 1999). If fish collected in a lake actually belong to two differentiated subpopulations, the likelihood of a partition in two panmictic clusters should depart from its expected distribution under the hypothesis of no structure (a single panmictic population). We thus searched for the best (ie, most likely) possible partition in two independent panmictic clusters for each lake sample using a 'Simulated annealing' procedure (Kirkpatrick *et al*, 1983). The present ML-based algorithm (PartitionML, available at http://www.univ-montp2.fr/~genetix/partitionml.htm) provides the opportunity to compare conveniently the likelihood of different configurations and is a fast and complementary alternative to other approaches to the same problem that use more complex Bayesian statistics (Pritchard *et al*, 2000; Dawson and Belkhir, 2001). Following Paetkau *et al* (1995) and Rannala and Mountain (1997), the likelihood criterion for a monolocus genotype being drawn from a cluster is calculated as $p = x_i^2$ for homozygotes and $p = 2 x_i x_j$ for heterozygotes (where x_i is the frequency of the ith allele at this locus in the target cluster, the unknown x_i values being estimated by their

observed value in the cluster). These values are calculated for each locus and are then multiplied to give the likelihood of the individual's multilocus genotype. For a given partition of a sample into two clusters, the likelihood of the partition is given by the product of the likelihoods in the two clusters. This latter is the product of individual multilocus likelihoods calculated for each individual of the cluster. If fish sampled in single lakes actually belonged to two differentiated entities, this algorithm should sort individuals into these subpopulations. But because a model with two clusters (and hence more parameters) will tend to have a higher likelihood (ie, the algorithm will favour partitions with two clusters rather than one in any data set regardless of its relevance), the likelihood of partitions with one and two clusters cannot be directly compared. We thus tested whether the resulting clusters were significantly more likely than those found in the absence of subdivision. To this end, we generated the null distribution of the log-likelihood of the partitions proposed by the ML algorithm in each of 100 panmictic pseudo-samples generated by random permutation of monolocus genotypes of the real samples using GENETIX 4.02 (Belkhir *et al*, 2001).

Because an eventual lack of significance of the partitions could be due to insufficient statistical power of the algorithm, we empirically tested its ability to disentangle accurately individuals from composite populations at a given threshold of divergence (θ). To that end, simulations were performed using Easypop 1.7. (Balloux, 1999). Genetic drift was simulated for 60 generations between two populations of varying size (50–100–200–400–800 and 1000) to generate various levels of divergence starting from randomly generated populations at six loci with nine alleles each. Individuals were then pooled into a single composite population and the percentage of correctly classified individuals by the ML partitioning algorithm was plotted against θ. To ensure concordance of the present ML-approach with Bayesian methods (Pritchard *et al*, 2000), we ran the same analysis using both methods on the same simulated datasets.

Inbreeding?

When individuals differ in their inbreeding history, inbred individuals are more homozygous over all loci than expected based on their single-locus genotype frequency, and we should expect identity disequilibrium (Haldane, 1949; Bennet and Binet, 1956). The hypothesis that heterozygote deficiencies are due to the presence of inbred fish in samples was therefore tested by comparing the distribution of MLH (multilocus heterozygosity), ie the number of individuals heterozygous at 0, 1, 2…6 loci (Sved, 1968; Brown *et al*, 1980) with its expectation under random mating using a permutation test. The number of fish in each MLH interval in the sampled population was compared with its random distribution obtained from 1000 randomised populations in which alleles were randomly associated within individuals. Linear interpolation was used to include fish not analysed at all six loci, thus intervals rather than classes were used. A fish heterozygous at five loci over the six possible was included in the]4,5] interval, while a fish heterozygous at five loci over five genotyped was considered fully heterozygous and thus included in the]5,6] interval. Probabilities of observing fewer individuals in a given MLH interval than expected at random were then combined across the

six samples using Fisher's method (Sokal and Rohlf, 1995). Biparental inbreeding should result in a significant over-representation of low MLH classes compared to random expectations, that is a higher proportion of fish homozygous at several loci than expected by random mating.

Sampling from a limited number of families?

Genetic relatedness of every pair of fish in each lake was estimated according to the pairwise identity coefficient of Mathieu *et al* (1990), which computes the probability of identity by state at a locus of a pair's offspring weighted across loci.

$$I_{xy} = \frac{i_{xy}}{\sqrt{i_{xx}i_{yy}}} \text{ , where } i_{xy} = \frac{\sum_j n_{jx}n_{jy}}{2} \text{ ,}$$

where n_{jx} is the number of copies of allele j in individual x. The multilocus estimate is obtained by weighting loci by $1/\Sigma_j(p_j)^2$, p_j being the sample frequency of allele j.

Identity was chosen over other available measures of genetic relatedness (eg, that of Queller and Goodnight, 1989; Lynch and Ritland, 1999) because of its substantially reduced variance (Belkhir *et al*, submitted). Following a testing procedure first proposed by Blouin *et al* (1996) to classify individuals as full, half sibs or unrelated, we tested whether the relatedness pattern of fish collected in a given lake significantly differed from its null expectation in a panmictic population. The mean identity among all pairs of fish in a sample was compared with its expected distribution under the hypothesis of no relatedness obtained by random permutation of genotypes in 1000 randomised populations. Our approach thus differed from that of Blouin *et al*, (1996) in that relatedness was here considered at the population rather than at the individual level.

We also simulated the behaviour of heterozygote deficiency when only a few progenitors reproduce in each generation using Maple version 5.4 (Waterloo Inc, 1999). One to 25 discrete spawning grounds (S) were generated, each composed of five males and five females (Quinn, 1995; Blanchfield and Ridgway, 1997, 1999) whose genotypes were randomly drawn at a biallelic locus with frequency *p* and *q*. Fifty progeny were produced from each spawning ground using random drawings from male and female genotypes. All progenies were then pooled, and global heterozygosity of the system was computed and compared with Hardy-Weinberg expectations using observed allele frequencies in the progeny. F_{IS} was estimated by $(H_E-H_O)/H_E$. Because we were mainly interested in the change of F_{IS} and var(F_{IS}) across loci with S, 250 independent biallelic loci ($p = q = 0.5$) were replicated for each value of S.

Results

Lake size and interlocus variance of F_{IS}

When including all 22 lakes previously reported in Castric *et al* (2001), a significant correlation ($P = 0.0041$) was found between lake area and var(F_{IS}) indicating that F_{IS} were not only higher but also more variable among loci in smaller lakes (Figure 1).

Heterozygote deficiencies in the three lakes

Contrasted levels of heterozygote deficiencies were observed among lakes as well as among sampling years (Table 1). Clish Pond had the strongest and most variable levels of heterozygote deficiencies both in 1997 ($f = 0.178$, $P = 0.001$) and 1999 ($f = 0.157$, $P = 0.007$). Robbins Brook Pond also revealed strong positive f-values in 1997 ($f = 0.167$, $P<0.001$), but no significant departure ($f = 0.002$, $P = 0.469$) was observed in this lake in 1999. The Third Wallagrass Lake exhibited positive global heterozygote deficiencies, but they were not significantly different from zero following sequential Bonferroni correction either in 1997 ($f = 0.083$, $P = 0.068$) or in 1999 ($f = 0.049$, $P = 0.05$).

No evidence for allele frequency variation between sampling years was detected in Clish Pond and Robbins Brook Pond ($F_{ST} = 0.0051$, $P = 0.132$ and $F_{ST} = 0.0051$, $P = 0.201$), whereas both samples differed significantly, though slightly, in the Third Wallagrass Lake (interannual $F_{ST} = 0.026$, $P = 0.004$).

Wahlund effect

Using simulated pools of distinct populations with known levels of divergence, the partitioning algorithm could classify individuals to their population of origin without error when θ approximately reached 0.10 (Figure 2), in accordance with previous results (see Figure 3 in Cornuet *et al*, 1999). At θ = 0.05, over 90% of individuals were still correctly classified (Figure 2). Results provided by the software Structure (Pritchard *et al*, 2000) were very similar (Figure 2), suggesting that both methods have similar efficiency in such simple situations. Because (1-F_{IT}) = (1-F_{IS})(1-F_{ST}), a given F_{ST} between two pooled and panmictic ($F_{IS} = 0$) differentiated units translates into a F_{IT} of equal magnitude (regardless of the estimation bias of the parameter F_{ST} when only two populations are considered, Weir and Cockerham, 1984). Because the observed multilocus heterozygote deficiencies were superior to 0.10 in Robbins Brook Pond in 1997 and Clish Pond in 1997 and 1999, this suggests that the algorithm used should have been able to unambiguously detect a partition if the observed heterozygote deficiency resulted from a Wahlund effect between two differentiated entities.

Permutation tests indicated that this was not the case. In none of the lakes did the log-likelihood value depart clearly from that created by the partitioning algorithm in homogenised populations (Figure 3, $P > 0.05$ in all cases). These results therefore argued against a Wahlund effect as a systematic explanation for the observed deficiencies.

Inbreeding

Evidence was found for a global shift of MLH towards lower individual heterozygosity than expected in randomly mating populations (Table 2). More fish heterozygous at three or less loci were observed than would have been expected if mating had been random (Fisher's method, d.f. = 12; $P = 0.0022$, $P = 0.0001$ and $P = 0.0025$ respectively for the three lowest MLH intervals), while no such departure was observed in the three highest MLH intervals ($P = 0.8259$, $P = 0.9467$ and $P = 0.3834$). This global trend was mostly due to fish sampled in 1997 in Robbins Brook Pond and in 1999 in Clish Pond (Table 2), indicating that a higher proportion of inbred fish than

Table 1 Number of alleles (A), expected (H_{nb}) and observed (H_o) heterozygosity, F_{IS} estimate and exact test probability of HW-equilibrium (P(HW))

		Third Wallagrass Lake		Robbins Brook Pond		Clish Pond	
		1997	1999	1997	1999	1997	1999
SFO-12	n	18	51	28	50	28	34
	A	4	5	6	7	2	2
	H_{nb}	0.671	0.586	0.709	0.666	0.195	0.317
	H_o	0.500	0.600	0.571	0.740	0.214	0.323
	F_{IS}	0.261	−0.044	0.197	−0.082	−0.102	−0.017
	P(HW)	0.119	0.802	0.637	0.025	1.000	1.000
SFO-18	A	5	6	7	10	3	5
	H_{nb}	0.641	0.626	0.712	0.685	0.389	0.414
	H_o	0.750	0.588	0.571	0.604	0.429	0.226
	F_{IS}	−0.176	0.068	0.201	0.138	−0.049	0.341
	P(HW)	0.742	0.395	0.053	0.010	0.299	0.013
SFO-23	A	11	14	9	10	9	8
	H_{nb}	0.865	0.845	0.874	0.850	0.810	0.831
	H_o	0.750	0.776	0.778	0.800	0.643	0.735
	F_{IS}	0.137	0.068	0.111	0.059	0.207	0.117
	P(HW)	0.007	0.178	0.244	0.243	0.015	0.283
SFO-8	A	10	15	12	12	10	7
	H_{nb}	0.890	0.863	0.778	0.832	0.645	0.332
	H_o	0.800	0.750	0.583	0.809	0.375	0.273
	F_{IS}	0.106	0.135	0.254	0.005	0.423	0.579
	P(HW)	0.252	<0.0005	0.006	0.249	0.009	0.000
SSA-197	A	4	5	6	4	2	2
	H_{nb}	0.600	0.734	0.712	0.706	0.166	0.268
	H_o	0.611	0.700	0.577	0.771	0.179	0.250
	F_{IS}	−0.019	0.058	0.193	−0.093	−0.080	0.072
	P(HW)	0.373	0.247	0.037	0.484	1.000	0.551
Mu-85	A	6	9	7	10	5	5
	H_{nb}	0.765	0.812	0.758	0.681	0.607	0.678
	H_o	0.667	0.833	0.714	0.688	0.482	0.594
	F_{IS}	0.132	−0.019	0.058	−0.001	0.194	0.125
	P(HW)	0.558	0.475	0.425	0.709	0.606	0.349
Global	A	40	54	47	53	31	29
	H_{nb}	0.739	0.744	0.757	0.736	0.469	0.473
	H_o	0.680	0.708	0.633	0.735	0.387	0.400
	F_{IS}	0.083	0.049	0.167	0.002	0.178	0.157
	P(HW)	0.068	0.050	<0.0005	0.469	0.001	0.007
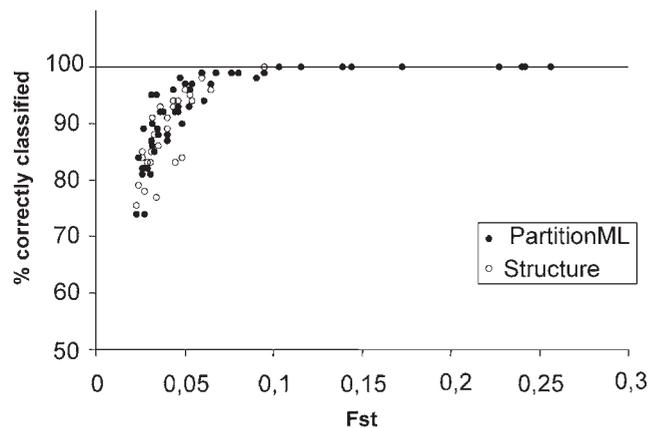
Figure 2 Proportion of correctly classified individuals by the ML-algorithm (PartitionML) and Structure (Pritchard *et al*, 2000) as a function of population divergence F_{ST}. Further details on simulation parameters are provided in the text.

randomly expected may have been collected in those two samples.

Relatedness
Evidence for relatedness was found in Robbins Brook Pond in 1997 and Clish Pond in 1999. The mean pairwise identity coefficient departed significantly from its expected distribution under the hypothesis of a random association of monolocus genotypes within individuals, as would be the case in a panmictic population (Figure 4; two-tailed test, $P = 0.006$ and $P = 0.024$, respectively).

Sampling from a limited number of families?
Computer simulations showed that F_{IS} was slightly negative (heterozygote excess) for low numbers of spawning grounds (S). There was however a rapid increase in F_{IS} with S, closely approaching 0 when S reached approximately 3 (Figure 5). The main effect of an increase in the number of spawning grounds was to asymptotically decrease the variance in F_{IS} across independent loci, while the effect on mean F_{IS} itself was much weaker

1997 **1999**
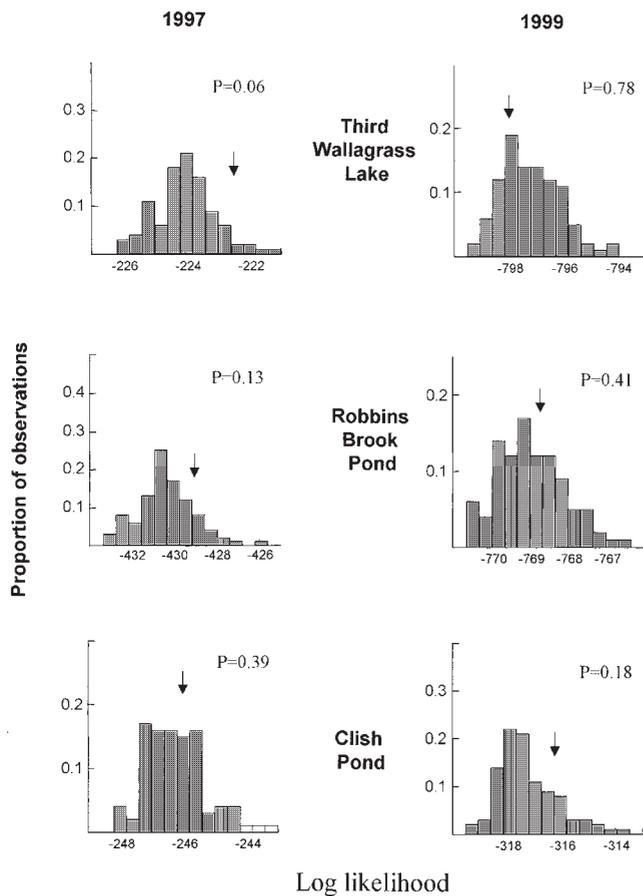
Figure 3 Distribution of loglikelihood of ML-partitions found in randomised populations. Arrows indicate observed values.
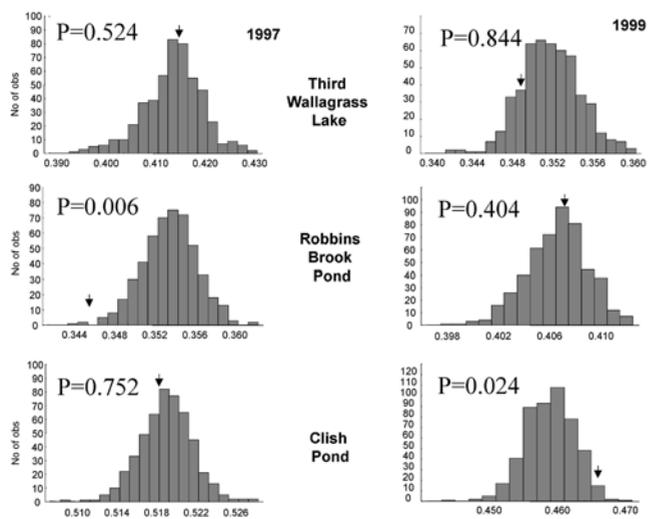
Figure 4 Distribution of pairwise relatedness coefficients estimated by Mathieu *et al* (1990) identity coefficient assuming random association of monolocus genotypes. Arrows indicate observed values.

unless S was very small. For any S, mean F_{IS} was at least one order of magnitude smaller than its standard deviation. An increase in the number of spawning grounds with lake size could thus account for the correlation between lake size and variance of F_{IS} estimates across loci (Figure 1b).

Discussion

The objective of this study was to find a biological explanation for the relationship between lake size and heterozygote deficiencies found in lacustrine brook charr populations in Maine. To that end, we focused on three relatively small lakes and tested three alternative hypotheses using multilocus microsatellite data. On the one hand, computer simulations provided support to the hypothesis that the observed decrease in variance of F_{IS} estimates across loci with lake size could be explained by an increase in the number of reproductive groups involved in producing each cohort. On the other hand, none of the proposed explanations for increased F_{IS} in smaller lakes could be given full support in all three lakes we used for hypothesis testing. A Wahlund effect was rejected in both years in all three lakes, while the presence of inbred fish (MLH distribution) in the samples, together with departures from the expected pairwise relatedness distribution, were detected in Robbins Brook Pond in 1997 and in Clish Pond in 1999, but not in any other sample. Although based on a restricted number of loci, our data thus suggested that no single explanation could account for both above-mentioned correlations.

Fish sampled in Robbins Brook Pond in 1997 and in Clish Pond in 1999 exhibited strong heterozygote deficiencies (Figure 1). Significant departures from the expected relatedness were also detected, indicating that the progeny of a limited number of families only may have been sampled at these locations. A significant shift toward lowered individual heterozygosity further suggested that the relatedness detected may be associated with mating among relatives. Small population size alone cannot account for both observations, as random mating even in a small isolated population would not lead to lowered individual multilocus heterozygosity compared to random expectations (just as ambient inbreeding is different from consanguinity). Shoaling behaviour of kin, whereby genetically related individuals swim in cohesive groups, could possibly account for the observed relatedness and has already been reported in other fish species (Brown and Brown, 1996; Pouyaud *et al*, 1999; Gerlach *et al*, 2001; Krause *et al*, 2001). As fish were collected in successive trap net attempts, kin swimming together in a shoal may have been sampled together. Kin recognition has recently been documented in the brook charr (Hiscock and Brown, 2000) and may provide further support to this explanation. Additional sampling and field behavioural observations in lakes of different size together with relatedness estimates would however be required to address this question, more specifically to compare brook charr shoaling and mating behaviour in lakes of different size.

A restricted number of spawning grounds in those two lakes may also potentially explain increased relatedness. If fish remain in the vicinity of their hatching location until maturity (see for example, Allendorf and Phelps, 1981; Hansen *et al*, 1997), most individuals collected in those two small lakes could likely be derived from a small number of progenitors only. Mating could then be biased toward relatives if most dispersal opportunities are restricted to the close surroundings of the lakes. Such consanguineous matings would thus be more frequent in small than in larger lakes with numerous tributaries, because fish have access to a lower number of unrelated

Table 2 Permutation test of the distribution of multilocus heterozygosity. Proportion of 1000 randomized populations in which a MLH class is underrepresented when compared with the real population

		\]0,1\]	\]1,2\]	\]2,3\]	\]3,4\]	\]4,5\]	\]5,6\]
				Multilocus heterozygosity			
Clish Pond	1997	0.105	0.036	0.304	0.963	0.796	0.120
	1999	0.019	0.006	0.974	0.644	0.736	0.083
Robbins Brook Pond	1997	0.081	0.004	0.001	0.732	0.969	0.719
	1999	0.197	0.086	0.224	0.583	0.649	0.301
Third Wallagrass Lake	1997	0.044	0.125	0.029	0.605	0.303	0.867
	1999	0.152	0.176	0.132	0.150	0.629	0.887
Combined test (df = 12)		0.0022	0.0001	0.0025	0.8259	0.9467	0.3834
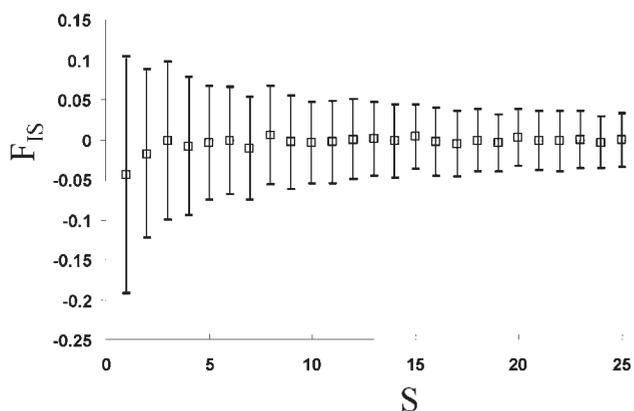
Figure 5 Deviation from Hardy-Weinberg proportions (mean $F_{IS} \pm$ s.d.) as a function of S, the number of spawning grounds (see text for details).

mates. This explanation is more parsimonious in that it only implies a limitation in dispersal opportunities in small lakes rather than a change in the pattern of mate choice. Because Hardy-Weinberg is an underlying assumption of the method aiming at detecting population substructure, such groups of related individuals (within which this assumption is not met, Figure 5) may have remained undetected.

Although samples analysed in the present study were all collected in small lakes, relatedness and consanguinity were detected in Robbins Brook Pond in 1997 and in Clish Pond in 1999 only. In all other samples, heterozygote deficiencies, when present, were not consistently associated with significant departures in individual multilocus heterozygosity or relatedness. A lack of power of the analytic tools we used could maybe account for this. Population geneticists' attention was drawn by Chakraborty (1984) to the loss of information due to the use of as summary a statistic as individual multilocus heterozygosity. We used a non-parametric approach (permutation test) to circumvent the low power of chi-square tests in this context. Although several outcomes are non-significant, the combined tests were significant and the three highest MLH classes were underrepresented in most instances when compared to random expectations ($P > 0.5$ in 13/18 cases), while the three lowest classes were overrepresented ($P < 0.5$) in 17/18 instances. Altogether, this suggested that there may still

exist a trend for lowered individual heterozygosity in our data. A second possible weakness of our analysis lies in the lack of precise knowledge of the detection power of methods designed to detect relatedness in natural populations. Identity, however, has been shown to exhibit considerably less variance than other published (Queller and Goodnight, 1989; Ritland, 1996; Lynch and Ritland, 1999) estimators of pairwise relatedness (Belkhir *et al*, submitted). A systematic evaluation of the power of methods based on pairwise estimates would however be required to address this issue more conclusively.

Beside the potential weaknesses of analytic tools, the contrasting results we obtained between lakes and between years for the same lake may reflect spatial and temporal variations in ecological conditions. Namely, the number of suitable spawning grounds in a lake was found to be non-significantly correlated with lake size in a study of nine lakes (10 to 986 ha) from Ontario, Canada (Rigdway and Blanchfield, 1998). Lake size alone may hence be an inaccurate predictor of the number of available spawning habitats, and some small lakes (such as for example, Third Wallagrass Lake) may potentially behave as lakes of a much larger area. Finally, inter-year variations in ecological conditions may also explain why the patterns in Robbins Brook Pond in 1997 and in Clish Pond in 1999 were not found again in 1999 and 1997, respectively. Ridgway and Blanchfield (1998) documented important inter-year variations in site use (less than 25% of all sites were used more than 2 out of the 4 years of the study despite strong competition for suitable spawning grounds) which may reflect extensive fluctuations in ecological conditions among years. If effects vary both among lakes and among years, a wider sampling design would be required to characterize finely the variation patterns in the specific mechanisms involved.

Altogether, the present test of alternative hypotheses allowed us to conclusively reject a Wahlund effect as the prime source of heterozygote deficiency, while providing indications that unexpected levels of relatedness and consanguinity occurred among fish in two of these small lakes. We therefore suggest that the correlation we observed between lake size and both heterozygote deficiencies and variance of heterozygote deficiencies across loci proceeds from two mechanisms. First, a direct increase in interlocus variance in F_{IS} estimates, because lakes of limited size generally, but not necessarily, provide a small number of suitable spawning sites. Second,

an indirect increase in F_{IS} estimates, because fish that emerged in smaller lakes have less opportunities for dispersal and have therefore a higher probability of mating with relatives when compared with fish emerged in larger lakes. Because this signal was only detected in two of the six samples analysed in the present study and because only six loci were used, care should be taken before extending the phenomenon to all small lakes. Following this analysis, however, it remains as the most coherent explanation for the increase in both F_{IS} estimates and interlocus variance in F_{IS} estimates observed in small lakes.

Acknowledgements

References

Allendorf FW, Phelps SR (1981). Use of allelic frequencies to describe population structure. *Can J Fish Aquat Sci* **38**: 1507–1514.

Balloux F (1999). *EASYPOP, a software for population genetics simulation*, Institut de Zoologie et d'Ecologie de l'Université de Lausanne, Lausanne (Switzerland).

Belanger G, Rodriguez M (2001). Homing behaviour of stream-dwelling brook charr following experimental displacement. *J Fish Biol* **59**: 987–1001.

Belkhir KP, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2001). *GENETIX 4.02, logiciel sous Windows TM pour la génétique des populations*, Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France).

Bennett JH, Binet FE (1956). Association between Mendelian factors with mixed selfing and random mating. *Heredity* **10**: 51–56.

Blanchfield PJ, Ridgway MS (1996). Reproductive timing and use of redd sites by lake-spawning brook trout (*Salvelinus fontinalis*). *Can J Fish Aquat Sci* **54**: 747–756.

Blanchfield PJ, Ridgway MS (1999). The cost of peripheral males in a brook trout mating system. *Animal Behav* **57**: 537–544.

Blouin MS, Parsons M, Lacaille V, Lotz S (1996). Use of microsatellite loci to classify individuals by relatedness. *Mol Ecol* **5**: 393–401.

Brookfield JFY (1996). A simple new method to estimate null allele frequency from heterozygote deficiency. *Mol Ecol* **5**: 453–455.

Brown AHD, Feldman MW, Nevo E (1980). Multilocus structure of natural populations of *Hordeum spontaneum*. *Genetics* **96**: 523–536.

Brown GE, Brown JA (1996). Does kin-biased territorial behavior increase kin-biased foraging in juvenile salmonids? *Behav Ecol* **7**: 24–29.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richard RI, Mulley JC *et al* (1993). Incidence and origin of "null" alleles in the (AC)n microsatellite markers. *Am J Hum Genet* **52**: 922–927.

Castric V, Bonney F, Bernatchez L (2001). Landscape structure and hierarchical genetic diversity in the brook charr *Salvelinus fontinalis*. *Evolution* **55**: 1016–1028.

Chakraborty R (1984). Detection of non-random association of alleles from the distribution of the number of heterozygous loci in a sample. *Genetics* **108**: 719–731.

Chakraborty R, Zhong Y (1993). Statistical power of an exact test of Hardy-Weiberg proportions of genotypic data at a multi-allelic locus. *Hum Hered* **44**: 1–9.

Christein DR, Guttman SI, Taylor DH (1979). Heterozygote deficiencies in a breeding population of *Bufo americanus* (Bufonidae: Anura): the test of a hypothesis. *Copeia* **1979**: 498–502.

Cornuet JM, Piry S, Luikart G, Estoup A, Solignac M (1999). New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* **153**: 1989–2000.

Crow JF, Kimura M (1970). *An Introduction to Population Genetics Theory*, Harper and Row: New York.

David P, Perdieu M-A, Pernot A-F, Jarne P (1997). Fine grained spatial and temporal population genetic structure in the marine bivalve *Spisula ovalis*. *Evolution* **51**: 1318–1322.

Dawson KJ, Belkhir K (2001). A Bayesian approach to the identification of panmictic populations and the assignment of individuals. *Genet Res* **78**: 59–77.

Dieckman U, Doebeli M (1999). On the origin of species by sympatric speciation. *Nature* **400**: 354–357.

Dynes J, Magnan P, Bernatchez L, Rodriguez MA (1999). Genetic and morphological variations between two forms of lacustrine brook charr. *J Fish Biol* **54**: 955–972.

Enjalbert J, David JL (2000). Inferring recent outcrossing rates using multilocus individual heterozygosity: application to evolving wheat populations. *Genetics* **156**: 1973–1982.

Gerlach G, Schardt U, Eckmann R, Meyer A (2001). Kin-structured subpopulations in Eurasian perch (*Perca fluviatilis* L.). *Heredity* **86**: 213–221.

Haldane JBS (1949). The association of characters as a result of inbreeding and linkage. *Annals Eugenics* **15**: 15–23.

Hansen MM, Nielsen EE, Mensberg K-L (1997). The problem of sampling families rather than populations: relatedness among individuals in samples of juvenile brown trout *Salmo trutta* L *Molec Ecol* **6**: 469–474.

Hardy GH (1908). Mendelian proportions in a mixed population. *Science* **28**: 49–50.

Hiscock MJ, Brown JA (2000). Kin discrimination in juvenile brook trout (*Salvelinus fontinalis*) and the effect of odor concentration on kin preferences. *Can J Zool* **78**: 278–282.

Jarne P, Charlesworth D (1993). The evolution of the selfing rate in functionally hermaphrodite plants and animals. *Ann Rev Ecol Systemat* **24**: 441–466.

Kirkpatrick S, Gelatt CD, Vecchi MP (1983). Optimization by simulated annealing. *Science* 220: 671–680.

Kondrashov AS, Kondrashov FA (1999). Interactions among quantitative traits in the course of sympatric speciation. *Nature* **400**: 351–354

Krause J, Butlin RK, Peuhkuri N, Pritchard VL (2001). The social organization of fish shoals: a test of the predictive power of laboratory experiments for the field. *Biol Rev Camb Phil Soc* **75**: 477–501.

Luikart G, Cornuet J-M (1999). Estimating the effective number of breeders from heterozygote excess in progeny. *Genetics* **151**: 1211–1216.

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

Mathieu E, Autem M, Roux M, Bonhomme F (1990). Epreuve de validation dans l'analyse de structures génétiques mulivariées: comment tester l'équilibre panmictique? *Revue de statistiques appliquées* **38**: 47–66.

Overall ADJ, Nichols RA (2001). A method for distinguishing consanguinity and population substructure using multilocus genotype data. *Mol Biol Evol* **18**: 2048–2056.

Paetkau D, Calvert W, Stirling I, Strobeck C (1995). Microsatellite analysis of population structure in Canadian polar bears. *Mol Ecol* **4**: 347–354.

Pemberton JM, Bancroft SJ, DR, Barrett JA (1995). Non-amplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molec Ecol* **4**: 249–252.

Pouyaud L, Desmarais E, Chenuil A, Agnèse J-F, Bonhomme F (1999). Kin cohesiveness and possible inbreeding in the mouth-brooding tilapia *Sarotherodon melanotheron* (Pisces Cichlidae). *Mol Ecol* **8**: 803–813.

Pritchard JK, Stephens M, Donnelly PJ (2000). Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Pudovkin AI, Zaykin DV, Hedgecock D (1996). On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics* **144**: 383–387.

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

Quinn NWS (1995). General features of brook trout, *Salvelinus fontinalis*, spawning sites in Algonquin Provincial Park, Ontario. *Canad Field Naturalist* **109**: 205–209.

Rannala B, Mountain JL (1997). Detecting immigration by using multilocus genotypes. *Proc Natl Acad Sci USA* **94**: 9197–9201.

Raymond M, Rousset F (1995). Testing heterozygote excess and deficiency. *Genetics* **140**: 1413–1419.

Ridgway MS, Blanchfield PJ (1998). Brook trout spawning areas in lakes. *Ecol Freshwater Fish* **7**: 140–145.

Ritland K (1996). Estimators for pairwise relatedness and individual inbreeding coefficients. *Genet Res Cambridge* **67**: 175–185.

Robertson A (1965). The interpretation of genotypic ratios in domestic animal populations. *Animal Production* **7**: 319–324.

Schluter D (2000). *The Ecology of Adaptive Radiation*, Oxford University Press: Oxford, UK.

Sokal RR, Rohlf FJ (1995). *Biometry*, 3rd edn. W. H. Freeman: San Francisco, CA.

Sved JA (1968). The stability of linked systems of loci with a small population size. *Genetics* **59**: 543–563.

Taylor EB (1999). Species pairs of north temperate freshwater fishes: evolution, taxonomy, and conservation. *Rev Fish Biol Fisheries* **9**: 299–334.

Thornhill NW (1993). *The Natural History of Inbreeding and Outbreeding*, University of Chicago Press: Chicago.

Van Treuren R (1998). Estimating null allele frequencies at a microsatellite locus in the oystercatcher (*Haematopus ostrealegus*). *Mol Ecol* **7**: 1413–1417.

Viard F, Justy F, Jarne P (1997). Population dynamics inferred from temporal variation at microsatellite loci in the selfing snail *Bulinus truncatus*. *Genetics* **146**: 973–982.

Wahlund S (1928). Zusammensetzung von Populationen und Korrelationserscheinungen vom Standpunkt der Vererbungslehre ausbetrachtet. *Hereditas* **11**: 65–106.

Waldman B, McKinnon JS (1993). Inbreeding and outbreeding in fishes, amphibians and reptiles. In Thornhill NW (ed) *The Natural History of Inbreeding and Outbreeding*. University of Chicago Press: Chicago.

Waterloo Inc (1999). *Maple release 5.4*. Waterloo Inc. Waterloo, ON, Canada.

Wattier R, Engel CR, Saumitou-Laprade P, Valero M (1998). Short allele dominance as a source of heterozygote deficiency at microsatellite loci. Experimental evidence at the dinucleotide locus Gv1CT in *Gracilaria gracilis* (Rhodophyta). *Mol Ecol* **7**: 1569–1573.

Weinberg W (1908). Über den Nachweis der Vererbung beim Menschen. *Jh Ver Vaterl Naturk Wurtemb* **64**: 369–382.

Weir BS, Cockerham CC (1984). Estimating *F*-statistics for the analysis of population structure. *Evolution* **38**: 1358–1370.

Wright S (1921). Systems of mating II. The effects of inbreeding on the genetic composition of a population. *Genetics* **6**: 124–143.

Zouros E, Foltz DW (1984). Possible explanations of heterozygote deficiency in bivalve molluscs. *Malacologia* **25**: 583–591.