

# Empirical evaluation of non-invasive capture–mark–recapture estimation of population size based on a single sampling session

SEBASTIEN J. PUECHMAILLE\*† and ERIC J. PETIT\*

\*Laboratoire Ethologie Evolution Ecologie, UMR CNRS 6552, Université Rennes I, Station Biologique, 35380 Paimpont, France; and †School of Biological and Environmental Sciences, University College Dublin, Dublin, Ireland

## Summary

1. Non-invasive genetic data analysed with capture–mark–recapture (CMR) models can be used to estimate population size, particularly for elusive and endangered species. Data generated from non-invasive genetic sampling are different, however, from conventional CMR data because individuals can be contacted several times within a single sampling session. Two methods have been proposed recently to accommodate this type of data, but no study has attempted to compare their estimates and evaluate their reliability compared with independent estimates of population size.

2. We investigated the reliability and accuracy of estimating the abundance of lesser horseshoe bats *Rhinolophus hipposideros* by genotyping DNA from droppings collected non-invasively at three colonies over 2 consecutive years. The number of times that each individual was ‘contacted’ (i.e. the number of droppings per individual) was used to estimate population size with two different published methods: a maximum likelihood and a Bayesian estimator.

3. Among the 586 samples extracted, 534 provided a complete genotype at six to eight microsatellite loci, which enabled a reliable discrimination of 165 individuals. Statistical estimates of colony sizes often included independent estimates obtained from visual counts, validating the method. Discrepancies appeared when capture heterogeneity was not taken into account while it occurred.

4. *Synthesis and applications.* We have taken a first step towards improving methods of estimating numbers of bats by demonstrating that genetic data produced from bat faecal DNA are of high quality and can provide accurate estimates of population size even when samples are taken during only one sampling session. Such protocols provide valuable management tools for elusive and rare species in general. The method is relatively easy and cost-efficient because only one sampling session is required.

*Key-words:* Bayesian, Chiroptera, CMR, elusive species, faecal DNA, genetic tag, likelihood, microsatellite, population size, *Rhinolophus hipposideros*

*Journal of Applied Ecology* (2007) **44**, 843–852

doi: 10.1111/j.1365-2664.2007.01321.x

## Introduction

Population size is important in wildlife management and conservation. However, obtaining population size estimates is not straightforward, particularly for endangered, elusive and cryptic species. Traditional capture–mark–recapture (CMR) methods, based on multiple sampling sessions, have proven to be an

efficient tool for population size estimation but are not without problems when applied to rare, elusive and capture-sensitive species. In such cases, non-invasive genetic methods can be valuable in estimating population size without the need for capture (Kohn *et al.* 1999; Banks *et al.* 2003; Frantz *et al.* 2003, 2004), for example when DNA is obtained from faeces or shed hair (Piggott & Taylor 2003). Each individual is recognized on the basis of its genetic fingerprint inferred from the non-invasively collected material (Taberlet & Luikart 1999). Because of the low quantity and/or quality of the non-invasively collected DNA, two main scoring errors, allelic dropout [ADO; one allele of a

Correspondence: Eric Petit, Laboratoire Ethologie Evolution Ecologie, UMR CNRS 6552, Université Rennes I, Station Biologique, 35380 Paimpont, France (fax + 33 2 99 61 81 88; e-mail eric.petit@Univ-rennes1.fr).

heterozygous individual is not amplified during a positive polymerase chain reaction (PCR)] and false allele (FA; a PCR-generated allele as a result of a slippage artefact during the first cycles of the reaction), may lead to incorrect genotypes and consequently to overestimated population sizes (Waits & Leberg 2000). Such errors must be corrected by repeating the genotyping process and comparing genotypes to each other (Frantz *et al.* 2003; Paetkau 2003) until a particular genetic fingerprint (i.e. multilocus genotype) is found. Two or more samples having the same fingerprint are then considered to originate from the same individual. At the same time,  $m$  samples from the same individual yield one 'capture' and  $m - 1$  'recaptures'. Recaptures can be obtained from a single non-invasive sampling session (Miller, Joyce & Waits 2005; Petit & Valière 2006). To our knowledge, no study has compared estimates obtained through a single sampling session treated with different models, or compared these estimates with independent estimates (e.g. direct counts) to test their reliability.

Bats make up to at least one-fifth of extant mammalian species (c. 1100 species; Simmons 2005); statistically defensible population size estimates are needed but they are often difficult to obtain (O'Shea, Bogan & Ellison 2003). Bats are difficult to census mainly because of their small size, high vagility and nocturnal lifestyle (O'Shea, Bogan & Ellison 2003), and thus represent a group of particular interest for testing the reliability of non-invasive sampling for population size estimates. Our study species was the lesser horseshoe bat *Rhinolophus hipposideros* Bechstein. Sensitive to human disturbance and classified as vulnerable by the World Conservation Union (Hutson, Mickleburgh & Racey 2001), the lesser horseshoe bat has undergone substantial declines in recent decades (Arbeitskreis Fledermause Sachsen-Anhalt 1997).

The goal of this study was to evaluate the feasibility and reliability of population size estimates based on a single non-invasive sampling session. For three colonies over 2 years (2003 and 2004), droppings of lesser horseshoe bats were sampled during a single sampling session and then typed at eight microsatellite loci. Each sample was assigned to an individual based on its multilocus genotype. The number of times that each individual was contacted was compiled and analysed to estimate population size and a 95% confidence interval (95% CI) of this estimate. Visual counts of the number of individuals present in the colonies were compared with the estimates obtained from the non-invasive genetic survey.

## Materials and methods

### COLLECTION AND STORAGE OF BAT DROPPINGS

Three colonies were investigated at Epiniac, Pluherlin and Saint-Thurial in Brittany, France, and were sam-

pled in 2003 and 2004. Each year, we collected about three times as many droppings as the number of bats visually counted in the colony (see below). This yielded enough material for extraction of 279 droppings in 2003 and 307 in 2004. Sampling involved cleaning up the ground under the main cluster around 20 May and then spreading newspaper on the ground beneath the roosting site. Dried droppings deposited by bats were collected by hand or using tweezers from the newspaper 10–15 days later. Sampling took place at this time of the year because all adult females of the colony were likely to be present but no young had been born.

After collection, droppings were stored individually in 2-mL microtubes, which were kept open except for transport. Samples were sent to the laboratory, where a silica gel fragment was added to each microtube in order to absorb humidity, thus avoiding DNA degradation (Wasser *et al.* 1997; Taberlet, Waits & Luikart 1999). Microtubes were stored open in a dry room until further analysis.

### COUNTING LESSER HORSESHOE BATS BY DIRECT OBSERVATION

The colonies studied were nurseries, where females congregate from March to May, then stay until August to complete their pregnancy, give birth and rear their young. We estimated the colony size by direct visual counting of the resting bats during the day. This was completed two to seven times from mid-May until mid-July in each colony and each year. The maximum number of adults observed during this period was considered as the most reliable visual estimate of colony size.

### GENETIC DATA

We extracted DNA from bat droppings using modifications to the QIAamp DNA Stool Mini Kit (QIAGEN) (Puechmaille, Mathy & Petit 2007). A 338-basepairs (bp) fragment of the Cytochrome *b* gene was amplified from each extract (Puechmaille, Mathy & Petit 2007). Samples that successfully amplified were then typed for microsatellites. Among the 14 microsatellites described by Puechmaille, Mathy & Petit (2005), the following eight tetra repeat loci were used: RHC108, RHD102, RHD103, RHD111, RHD113, RHD119, RHD2 and RHD9. All eight microsatellites were amplified together in a 7- $\mu$ L multiplex reaction and sized as reported in Puechmaille, Mathy & Petit (2005). We tested for Hardy–Weinberg and genotypic disequilibria with permutation tests using F-STAT 2.9.3 (Goudet 2001).

### PROBABILITY OF IDENTITY ( $P_{ID}$ )

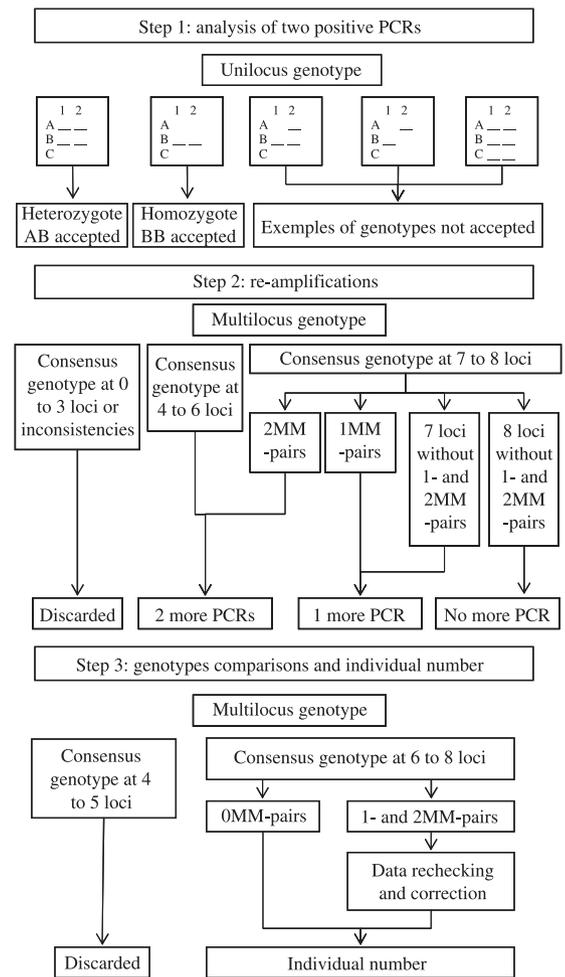
The probability that two individuals, drawn at random from a population, will have the same multilocus genotype ( $P_{ID}$ ; Waits, Luikart & Taberlet 2001) depends on

the number of loci used to construct the genotype, the heterozygosity of these loci, and the relatedness of individuals within the population (Waits, Luikart & Taberlet 2001). From multilocus genotypes of 91 individuals from two nurseries (Epiniac,  $n = 53$ ; Pluherlin,  $n = 38$ ), Puechmaille, Mathy & Petit (2005) showed that a set of eight loci is sufficient to discriminate all individuals at the nurseries studied. To confirm the power of the eight loci used, we calculated expected probabilities of identity for unrelated individuals ( $P_{ID-rand}$ ) and full siblings ( $P_{ID-sibs}$ ), the latter being more conservative (Waits, Luikart & Taberlet 2001), using GIMLET (Valière 2002).

#### REDUCING GENOTYPING ERRORS

When amplifying DNA from faeces, researchers must cope with genotyping errors, particularly ADO and FA (Taberlet *et al.* 1996). Although methods involving PCR replicates (multiple-tube approaches) have been developed to reduce genotyping error rates (Taberlet *et al.* 1996), these methods increase laboratory costs (Paetkau 2003). Other complementary methods, such as genetic profiles comparisons, have also been developed to detect genotyping errors (Frantz *et al.* 2003; Paetkau 2003). Comparisons of samples with each other identify samples with identical genotypes, referred to for convenience as zero-mismatch pairs (0-MM pairs), samples that match at all but one allele (1-MM pairs), samples that match at all but two alleles (2-MM pairs), and so on. As 1-MM pairs and 2-MM pairs are unlikely to happen when the probability of identity is very low (cf. the Results; Paetkau 2003), all samples involving such cases were suspected to contain one and two genotyping errors, respectively. In order to obtain reliable sample genotypes, we adapted a comparative multiple-tube approach in this study, combining PCR replicate and genetic profile comparisons (Fig. 1).

The first step of our approach followed Frantz *et al.*'s (2003) approach in performing two PCR per sample. A consensus genotype was defined for each locus, following the rule that an allele was accepted only if it had been recorded at least twice. If zero or more than two alleles met this criterion, no consensus genotype was accepted at this locus (Fig. 1). The second step shifted the decision level to the multilocus genotype. Samples having consensus genotypes accepted at less than four loci at step 1 were discarded because they were considered to be low-quality samples. Those with four to six loci accepted were re-amplified twice more in order to complete the genotype. Samples completed at seven or eight loci were compared using GeneCap (Wilberg & Dreher 2004) to confirm those identical at all alleles (0-MM pairs), and those differing at one (1-MM pairs) or two alleles (2-MM pairs). Pairs differing at one or two alleles were re-amplified once and twice more, respectively. Indeed, one more re-amplification was often sufficient to detect one error whereas two may



**Fig. 1.** Flow chart illustrating the comparative multiple-tube approach. By following this procedure it was possible to obtain reliable genetic profiles in this study (see text for further explanations). In step 1, numerals 1 and 2 identify positive PCR 1 and 2, while letters A, B, and C identify alleles that may be observed in a given PCR.

have been necessary to detect two errors. Samples identical at seven loci were re-amplified once more to complete the genotype. We discarded samples with inconsistencies between replicates (e.g. two PCR with genotype AB, then two PCR with genotype CD). In the third step, we removed samples having less than six loci accepted at the end of step 2. Samples with six, seven or eight loci were compared on a pairwise basis. Considering the very low probability of identity ( $P_{ID-sibs} < 10^{-3}$ ; cf. the Results), we identified two or more samples harbouring the same multilocus genotype as from the same individual. We considered samples having genotypes that differed by three or more alleles as distinct individuals. As in step 2, 1- or 2-MM pairs were likely to contain errors. We thus compared and rechecked their genotypes. If the mismatch persisted, an allele could be accepted if it was present only once, but only if it allowed reducing the number of mismatches between genotypes from 1 to 0 (reducing a 1-MM pair to a 0-MM pair) or from 2 to 1 (reducing a 2-MM pair to a 1-MM pair). After screening, if 1-MM pairs were

still present in the data set, they were declared as originating from the same individual, while 2-MM pairs were declared as originating from different individuals. The number of samples analysed being approximately twice the size of the population estimated visually, we expected on average one 'recapture' per individual. Thus multilocus genotypes that appeared only once in the data set were read again to check whether they contained errors.

GENOTYPING ERROR RATES

ADO and FA rates were computed by comparing PCR replicates to consensus genotypes, assuming that consensus genotypes are correct. The mean per-replicate probability of ADO at locus *j* was estimated as:

$$p_j = \frac{D_j}{A_{het_j}} \tag{eqn 1}$$

where *D<sub>j</sub>* is the number of amplifications involving the loss of one allele, and *A<sub>het<sub>j</sub></sub>*

 is the number of positive amplifications of individuals determined as heterozygous according to their consensus genotype at locus *j* (after Broquet & Petit 2004). The mean per-replicate probability of FA at locus *j* was estimated as:

$$f_j = \frac{F_j}{A_j} \tag{eqn 2}$$

where *F<sub>j</sub>* is the number of amplifications leading to the creation of one or more false allele at locus *j*, and *A<sub>j</sub>* is the number of positive amplifications (after Broquet & Petit 2004). After amplifications have been replicated *i* times for *N<sub>i</sub>* individuals at locus *j*, the probability of finding a consensus genotype that is wrong because of an ADO is, given the observed heterozygosity of locus *j* is *H<sub>O<sub>j</sub></sub>*

$$E_{ADO_j} = H_{O_j} \times \frac{\sum_i [(p_j)^i + i(p_j)^{i-1}(1 - p_j)] \times N_i}{\sum_i N_i} \tag{eqn 3}$$

because a consensus will be wrong when *i* or *i* - 1 replicates show an ADO in a heterozygote (remember that an allele was accepted when it was read at least twice over *i* replicates). Similarly, the probability that this consensus is wrong because of an FA is:

$$E_{FA_j} = (1 - H_{O_j}) \times \frac{\sum_i \sum_{n=2}^i C_i^n (f_j)^n (1 - f_j)^{i-n} \times N_i}{\sum_i N_i} \tag{eqn 4}$$

because such an error can affect only homozygous genotypes when an FA is read at least twice over *i* replicates (thus resulting in a false heterozygote).

Assuming that errors affect samples independently, we estimated the error probability (ADO or FA) for a consensus genotype at locus *j* as:

$$E_j = E_{ADO_j} + E_{FA_j} - (E_{ADO_j} \times E_{FA_j}) \tag{eqn 5}$$

As ADO and FA are not mutually exclusive, the probability of occurrence of an error (*E<sub>j</sub>*) is equal to the sum of probabilities of both events minus the probability of both events occurring simultaneously, the latter probability being equal to the term in parentheses (equation 5) as ADO and FA are considered as independent errors. Given a genotype consisting of *L* loci, the probability of this multilocus genotype being wrong (i.e. containing at least one error) is called the multilocus error rate (*E<sub>T</sub>*). Assuming that errors are independent between loci:

$$E_T = 1 - \prod_{l=1}^L (1 - E_j) \tag{eqn 6}$$

Equation 6 gives the probability of having at least one error in a consensus multilocus genotype, that is to say the sum of probabilities of having 1, 2, and so on up to 8 errors. Step 3 of our protocol (Fig. 1) enabled genotypes having exactly one or two errors to be detected. Thus, we calculated the probability of having a wrong genotype with exactly one or two errors (*E<sub>1or2</sub>*):

$$E_{1or2} = \sum_{j=1}^L \left[ E_j \prod_{k \neq j} (1 - E_k) \right] + \sum_{j=1}^L \left[ E_j E_{k \neq j} \prod_{l \neq j, l \neq k} (1 - E_l) \right] \tag{eqn 7}$$

ESTIMATION OF POPULATION SIZE

Two methods have been developed recently to accommodate the fact that individuals can be 'captured' more than once per sampling session in non-invasive genetic studies. Based on a model with *n* samples taken with replacement from *N* individuals, all individuals having the same probability 1/*N* of being sampled, Miller, Joyce & Waits (2005) developed a maximum likelihood estimator implemented in Capwire software ([www.cnr.uidaho.edu/lecg](http://www.cnr.uidaho.edu/lecg)). This corresponds to the null model of Capwire, also referred to as even capture model (ECM). Considering the same model, Petit & Valière (2006) described a sequential Bayesian method (BM) adapted from Gazey & Staley (1986). Computations of the Bayesian estimate were implemented in R (Ihaka & Gentleman 1996). Both methods yield a population size estimate and a 95% CI of this estimate.

The main assumptions of these models are: (i) a closed population (no birth, no death, no migration); (ii) a recapture probability equalling the capture probability; and (iii) an equal capture probability for all individuals. Assumptions (i) and (ii) were probably met in our study (see the Discussion). The assumption of capture homogeneity was more problematic and thus

tested in two different ways. First, when calculating the maximum likelihood estimator, we performed a likelihood ratio test using Capwire (Miller, Joyce & Waits 2005; threshold  $P$ -value set to 0.05). This test compares a model assuming a population comprising two groups of individuals with distinct capture probabilities (the two innate rate model; TIRM), with the null model assuming equal capture probability for all individuals. This likelihood ratio test (LRT) is based on a restrictive model of capture heterogeneity, and is thus likely to miss types of capture heterogeneity that do not conform to the two innate rates model implemented in Capwire. We therefore developed a second test that makes no assumption about the distribution of capture heterogeneity among individuals. We simply simulated the sampling process carried out during the study, under an assumption of homogeneous capture probability, and compared the expected with the observed number of captures per individual. Simulations were conducted as follows using the R package (Ihaka & Gentleman 1996). We randomly drew  $n$  samples from a pool of  $N$  individuals, each individual having the same probability of capture ( $1/N$ ). Sampling was done with replacement because the number of droppings per individual was great compared with the number of droppings taken (see the Discussion).  $n$  being greater than  $N$ , some individuals were sampled only once but others were sampled twice, three times, etc. We repeated each simulation 1000 times, from which an average and boundaries of the 95% CI of the number of captures per individual were computed. The null hypothesis of equiprobability of capture was rejected if, in the frequency distribution of the number of captures, the frequency of observed values lay outside the simulated 95% CI for at least one value of the number of captures. We ran the simulation test considering two different values for  $N$ : the maximum likelihood and the Bayesian estimates of  $N$ . The number of samples drawn ( $n$ ) was equal to the number of samples typed for each pair of colony and year.

We estimated  $N$  for each colony (Epiniac, Pluherlin and Saint-Thurial), each year (2003 and 2004) with both the maximum likelihood method under the null

model and the BM. Because tests carried out to detect capture heterogeneity might not be entirely reliable (Miller, Joyce & Waits 2005) or the threshold chosen might not be stringent enough ( $P = 0.05$ ),  $N$  was also estimated for each colony and year using the two innate rate model independently of the results given by the capture heterogeneity tests.

## Results

### GENETIC DATA

Amplification of mitochondrial DNA yielded positive controls for 567 of the 586 extractions (96.8%). Apart from four samples that were removed because they were mixed during extraction, all others (563 samples) were typed at eight microsatellite loci. Among the 563 samples, 25 were removed during step 2 (re-amplifications) and four during step 3 (genotype comparisons), resulting in 534 samples genotyped at six (five samples), seven (11 samples) and eight loci (518 samples).

The average expected heterozygosity of the eight loci was 0.72 at Epiniac (range 0.62–0.79), 0.70 at Pluherlin (range 0.56–0.83) and 0.72 at Saint-Thurial (range 0.53–0.82). When analysing each locus for each colony and each year, two loci showed heterozygosity deficiency and one heterozygosity excess, but no disequilibrium was detected at the colony level. As expected by chance from the number of tests carried out, the number of loci that showed a significant genotypic disequilibrium varied between one and three in each colony. No linkage between two particular loci was observed more than twice.

### PROBABILITY OF IDENTITY

The  $P_{ID}$  using all loci ranged between  $1.00 \times 10^{-8}$  and  $8.20 \times 10^{-8}$  for unrelated individuals ( $P_{ID-rand}$ ) and between  $7.37 \times 10^{-4}$  and  $1.23 \times 10^{-3}$  for full siblings ( $P_{ID-sibs}$ ) (Table 1). In each colony for both years, the minimal number of individuals that could be discriminated assuming full siblings was above 800 (Table 1).

**Table 1.** Mean observed ( $H_O$ ) and expected ( $H_E$ ) heterozygosities for three colonies of the lesser horseshoe bat over 2 years. Probabilities of identity and the number of individual distinguishable ( $1/P_{ID}$ ) are also presented

	Epiniac		Pluherlin		Saint-Thurial	
	2003	2004	2003	2004	2003	2004
$H_O$	0.74	0.73	0.78	0.76	0.72	0.69
$H_E$	0.72	0.72	0.7	0.7	0.74	0.7
$P_{ID-sibs}^*$	$9.84 \times 10^{-4}$	$1.01 \times 10^{-3}$	$1.20 \times 10^{-3}$	$1.23 \times 10^{-3}$	$7.37 \times 10^{-4}$	$1.20 \times 10^{-3}$
$P_{ID-rand}^\dagger$	$5.22 \times 10^{-8}$	$5.39 \times 10^{-8}$	$7.33 \times 10^{-8}$	$8.20 \times 10^{-8}$	$1.00 \times 10^{-8}$	$6.82 \times 10^{-8}$
$1/P_{ID-sibs}$	1016	995	833	813	1356	831
$1/P_{ID-rand}$	19 164 431	18 556 318	13 644 426	12 199 585	99 900 100	14 664 907

\*Probability of identity for full siblings (Waits, Luikart & Taberlet 2001).

†Probability of identity for unrelated individuals (Waits, Luikart & Taberlet 2001).

**Table 2.** Mean error rates (ADO and FA calculated after equations 1 and 2, respectively) and global error rates ( $E_T$ , equation 6;  $E_{1or2}$ , equation 7) are presented per colony and year. The number of samples typed is given in parentheses for each colony and year. The expected number of erroneous genotypes (error rate  $\times$  number of samples typed) is also presented in parentheses for each type of error rate

	Epiniac		Pluherlin		Saint-Thurial	
	2003 (114)	2004 (145)	2003 (97)	2004 (95)	2003 (46)	2004 (37)
ADO	0.049	0.044	0.054	0.031	0.128	0.153
FA	0.022	0.028	0.024	0.021	0.037	0.031
$E_T^*$	0.2426 (27.7)	0.2637 (38.2)	0.3113 (30.2)	0.1957 (18.6)	0.4578 (21.1)	0.6330 (23.4)
$E_{1or2}^\dagger$	0.2408 (27.5)	0.2613 (37.9)	0.3072 (29.8)	0.1948 (18.5)	0.4418 (20.3)	0.5801 (21.5)
$E_T - E_{1or2}$	0.0018 (0.2)	0.0024 (0.4)	0.0040 (0.4)	0.0009 (0.1)	0.0161 (0.7)	0.0529 (2)

\*Probability of a genotype being incorrect (containing at least one error, equation 6).

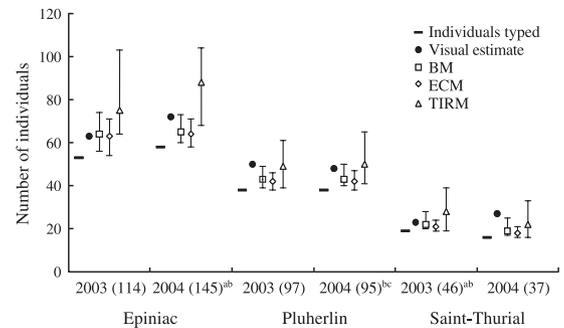
†Probability of a genotype containing exactly one or two errors (equation 7).

### GENOTYPING ERRORS

Mean ADO rates over loci were between 0.031 and 0.054 at Epiniac and Pluherlin and between 0.128 (2003) and 0.153 (2004) at Saint-Thurial. Mean FA rates over loci ranged between 0.021 and 0.037 for all three colonies over the 2 years. These averaged error rates were estimated per locus, replicate, colony and year (Table 2). Taking into account the number of replicates and following equation 6, we estimated the multilocus error rate ( $E_T$ ), which was between 0.1957 and 0.3113 at Epiniac and Pluherlin for both years and reached 0.4578 (2003) and 0.6330 (2004) at Saint-Thurial. Assuming that consensus genotypes are correct and according to these estimates of global error rates and the number of samples typed, the expected total number of samples containing one or more errors was as high as 159 over 534 (calculated after Table 2). But according to our equations 6 and 7, these 159 samples had a 97.7% chance of presenting exactly one or two errors (calculated after Table 2). By reading the samples again, screening all 1- and 2-MM pairs and also by checking genotypes that did not match any other sample (samples without recapture), all 1-MM pairs were removed and the number of 2-MM pairs was reduced to three. These 2-MM pairs were considered to be different individuals because in all three cases each genotype involved matched completely (0-MM pairs) another sample of the data set, which rendered unlikely the possibility that two identical errors occurred in two different samples from the same individual.

### ESTIMATION OF POPULATION SIZE

Among the 534 samples typed, 52 were unique (9.7%) and 482 had a genotype present at least twice in the data set. These 482 samples represented 113 different genotypes, which were attributed to as many individuals. In total, 165 different individuals were identified. For each colony and year, the number of samples typed and the number of unique individuals identified are reported in Fig. 2. The number of droppings sampled



**Fig. 2.** Colony size estimates in three nurseries of the lesser horseshoe bat. The Bayesian estimate (BM) is from Petit & Valière (2006). The maximum likelihood estimates (ECM and TIRM) are from Miller, Joyce & Waits (2005). For each year and nursery, the number of samples typed is presented in parentheses. (a) Denotes a significant likelihood ratio test of capture heterogeneity (see the Materials and Methods for further explanations) in the colony and year considered. (b) Denotes a significant simulation test of capture heterogeneity (see the Materials and Methods for further explanations) in the colony and year considered. (c) The  $P$ -value for the likelihood ratio test of capture heterogeneity is 0.076.

per individual and year varied between 1 and 10 (mean 2.41, SD 1.67).

Colony size estimates using BM, ECM and TIRM are represented in Fig. 2 for each colony and year. For all colonies, whatever the year, ECM and BM estimates were quite similar and agreed with visual counts, although they slightly underestimated the number of individuals counted visually. All visual counts lay within the 95% CI of the BM estimate, except two that were slightly higher (one more individual). With ECM, three visual counts were within the 95% CI of the estimate, and three were higher (plus one, four and six individuals).

Capture heterogeneity was detected by both tests (LRT and simulation) in two of the six pairs of colony and year tested (Epiniac 2004 and Saint-Thurial 2003; Fig. 2; see also Fig. S1 in the supplementary material). In these two cases, TIRM estimates were higher than

BM and ECM estimates, although their 95% CI overlapped, and comprised the visual estimate. At Pluherlin in 2004, the simulation test detected capture heterogeneity but not the LRT, although its probability was nearly significant ( $P = 0.076$ ). In this case,  $N$  estimated by the TIRM was very close to the visual estimate. At Epiniac 2003, no capture heterogeneity was detected and the TIRM 95% CI fell above the visual estimate. However, at Pluherlin 2003 and Saint-Thurial 2004, no capture heterogeneity was detected and the TIRM estimate was closer to the visual estimate than the ECM and BM estimates.

## Discussion

### QUALITY OF THE GENETIC DATA

In non-invasive genetics, DNA extract quality depends on the ability to reduce the presence of molecules that can inhibit the PCR (Taberlet, Waits & Luikart 1999). The amplification success of DNA extracted from faeces of the lesser horseshoe bat was quite high (96.8% of 586 samples), indicating that inhibitors were efficiently removed during extraction. This test was run with mitochondrial DNA, a genome present in high copy numbers in each cell. Among the samples that were successful for the mitochondrial DNA test, the amplification of nuclear loci, present in only two copies per cell, was also highly successful (98.6% of 12 592 PCR allowed one or more allele to be scored).

Because all loci were amplified together, our amplification method (Fig. 1) sometimes involved the re-amplification of all loci because one failed to amplify or contained one error. However, the average of 2.70 replicates conducted per sample (excluding failed reactions) was within the range of those published (3.4, Frantz *et al.* 2003; 2.4, Prugh *et al.* 2005). The success of the multiplex PCR was possible because all loci amplified equally (Puechmaille, Mathy & Petit 2007). Finally, a high percentage of samples extracted ( $534/586 \times 100 = 91.1\%$ ) provided a reliable genotype using our approach (Fig. 1). This efficiency is important in order to keep the costs reasonable for large-scale studies.

The genotyping error rates (mean ADO 0.0595, mean FA 0.0261) were among the lowest reported from faecal genotyping studies (Broquet & Petit 2004). These rates agree with error rates estimated during a pilot study (mean ADO 0.0492, mean FA 0.0328), where 22 samples originating from two colonies and 2 years were repeatedly typed eight times (data not shown). When performing steps 1 and 2 of the comparative multiple-tube approach (Fig. 1), multilocus error rates ( $E_T$ ) ranged from 0.1957 to 0.6330. The multilocus genotypes were compared to find these errors ( $E_{1,or2}$ ), and correcting them finally yielded final error rates ( $E_T - E_{1,or2}$ ) that were quite low, between 0.0009 and 0.0529 (Table 2). Hence our protocol (Fig. 1) is very efficient in reducing the expected number of wrong genotypes, which fell from 159 to less than four of 534

(Table 2) after 1-MM and 2-MM pairs had been checked for errors. These estimations of global error rates relied on assumptions of independence between ADO and FA errors and independence of errors between loci (cf. the Materials and Methods). The correlations between the number of ADO and the number of FA was significant but low for seven loci (Spearman rank correlation,  $r$  range 0.18–0.41, d.f. range 384–448,  $P < 0.001$ ) and not significant for locus D119 ( $r = 0.08$ , d.f. = 374,  $P = 0.14$ ). For the 518 complete multilocus genotypes, the number of errors per sample per locus (ADO + FA) was also significantly correlated between all possible pairs of loci (Spearman rank correlation,  $r$  range 0.18–0.49, d.f. = 516,  $P < 0.001$  for each correlation). Our estimates are thus biased, although not to a large extent because these correlations are weak. Furthermore, when correlations exist, errors concentrate on fewer genotypes than when errors occur independently (Bonin *et al.* 2004), leading to fewer wrong genotypes than expected. However, because these wrong genotypes contain more errors than expected, some of them might be undetected during step 3 of our comparative multiple-tube approach (Fig. 1).

### ESTIMATION OF POPULATION SIZE

Expected  $P_{ID-rand}$  and  $P_{ID-sibs}$ , considered, respectively, as the lower and upper bounds for the observed values of  $P_{ID}$  ( $P_{ID-obs}$ ), were quite low (Table 1). Those  $P_{ID}$  calculations could be slightly underestimated as we included individuals with exactly six (five individuals) or seven (six individuals) complete loci. However, calculations of weighed  $P_{ID}$  (Prugh *et al.* 2005), taking into account the number of individuals typed at six, seven or eight loci, showed that the minimal number of individuals that could be discriminated assuming full-siblings was still above 800 (data not shown). The mating system of the lesser horseshoe bat is polygynous, with each female having only one offspring per year (Gaisler 1965). Even if mate fidelity and intralocus polygyny occurs, as reported in the greater horseshoe bat (Rossiter *et al.* 2005), full-siblings will represent only a small proportion of all possible relationships between individuals within the colony. In such a case, the  $P_{ID-obs}$  may be closer to the  $P_{ID-rand}$  than the  $P_{ID-sibs}$ , as reported from the Australian northern hairy-nosed wombat *Lasiornis krefftii* (Waits, Luikart & Taberlet 2001). In our case,  $P_{ID}$  was low enough to avoid problems of underestimation linked to the 'shadow effect', when two individuals harbour the same genotype for the set of loci investigated (Mills *et al.* 2000; Waits & Leberg 2000).

Visual estimates of colony size can be considered reliable because they were consistent over the two to seven times each colony was counted each year (see Appendix S1 in the supplementary material). We can thus consider its value as the best estimate of the population size (true  $N$ ). BM and ECM provided very similar results and globally tended to underestimate the

true  $N$ , especially ECM, which furnished the narrowest 95% CI (Fig. 2). This suggests that genotyping errors, which could lead to overestimates of population size (Waits & Leberg 2000), were not a problem in this data set, even for Saint-Thurial 2004 which is the only colony and year where more than one erroneous genotype is expected (Table 2). Increasing the stringency of our protocol, and thus reducing the number of expected wrong genotypes, could be reached by accepting alleles observed in at least three (rather than two) independent PCR replicates. This could either mean running more PCR replicates, thus increasing the costs of the study, or eliminating more samples because of missing genotypes, which would reduce the accuracy of the population size estimates. At Epiniac 2003, the frequency distribution of the observed number of captures fitted the frequency expected under an even capture probability very well (see Fig. S1 in the supplementary material) and BM and ECM were accurate while TIRM overestimated the true  $N$ . When capture heterogeneity was detected, TIRM yielded higher estimates than BM and ECM methods, but in two out of three cases it tended to overestimate the true  $N$ , which was, however, included in its 95% CI (Fig. 2). Surprisingly, the TIRM estimate was the best at Pluherlin 2003 and Saint-Thurial 2004, whereas no capture heterogeneity was detected. The simulation test was, however, nearly significant, with two points of the observed number of captures reaching the boundaries of the expected 95% CI for Saint-Thurial 2004, and one point for Pluherlin 2003 (see Fig. S1 in the supplementary material).

#### MODEL ASSUMPTIONS

Our analyses were based on three assumptions [(TIRM relaxes assumption (iii)]: (i) a closed population during the sampling period (no births, deaths, immigration or emigration); (ii) a recapture probability equalling the capture probability; and (iii) an equal capture probability for all individuals, in other words an equal number of droppings deposited per individual throughout the sampling period. Because the respect of assumptions has an important impact on estimator performance (Miller, Joyce & Waits 2005), we assessed our results in the light of the lesser horseshoe bat biology and sampling design.

For lesser horseshoe bats, births occurred outside our sampling period (Gaisler 1965, 1966a; Schofield 1996). The bats are long-lived (Gaisler 1989; Gaisler *et al.* 2003) and our sampling period was short, so the probability of death was negligible. Finally, female lesser horseshoe bats are very faithful to their colony, even if they use satellite roosts occasionally (Schofield 1996; Kayikcioglu & Zahn 2004). To avoid missing those individuals using satellite roosts, the minimum sampling period lasted for 1 week. Thus, given what we know about the biology of the lesser horseshoe bat and what we observed, we can assume that assumption (i) was most probably met.

Using the number of pellets collected over one sampling period and the number of individuals counted visually, we estimated the number of droppings that an individual dropped each day to be five. So, for example, in a colony of 20 individuals and after a sampling period of 1 week, the capture probability of an individual (35/700) was very similar to subsequent recapture probabilities (first, 34/699; second, 33/698, etc.). We thus considered that drawing one dropping from an individual did not change the chance of drawing a second dropping from this same individual. We always selected entire droppings (about 5.3 mg) that were not moistened. We assumed that the fact that some droppings were broken or moistened was mainly the result of external factors (transport, size of the silica fragment, etc.) and not linked to the identity of the depositor. The sampling process was thus considered random and the second assumption was fulfilled.

The third assumption implied an equal capture probability for all individuals, in other words an equal number of droppings deposited per individual throughout the sampling period. Nothing is known about the defaecation rate of this species, particularly about interindividual differences. In other mammals, sex and age are the factors usually put forward to explain such differences (Kohn *et al.* 1999; Wilson *et al.* 2003). Adult females are believed to be in the majority in nursery colonies of lesser horseshoe bat (Gaisler 1963c, 1966b) but adult males or subadults can be present (Gaisler 1963a,b; F. Bontadina & G. Jones, personal communication). In mammals, and particularly bats, reproductive status greatly influences energy demands (Gittleman, Holroyd & Barclay 1988). Hence a colony comprising pregnant and non-pregnant females could generate capture heterogeneity. We cannot conclude anything about this question but assume that if all individuals are of the same sex (i.e. female) and same reproductive state (i.e. pregnant), the defaecation rate is likely to be the same for all individuals. It is possible that when at least one test detected capture heterogeneity between individuals, such as at Epiniac and Pluherlin in 2004 and Saint-Thurial in 2003, colonies comprised individuals harbouring different combinations of factors (sex, age and reproductive state). Considering that we sampled underneath the main cluster within the roost, it is also possible that for some reasons (e.g. thermal and social) individuals varied in their use of space and spent more or less time above the sampling area, leaving different numbers of droppings per individual.

When there is no capture heterogeneity (e.g. Epiniac 2003), BM and ECM methods are more precise and should be preferred. However, as mentioned by Miller, Joyce & Waits (2005), it is sometimes difficult to detect capture heterogeneity using tests. If a researcher has good biological reasons to believe capture heterogeneity is occurring, TIRM should be applied. In our study, where capture heterogeneity was moderate, ECM and BM generally underestimated the true  $N$  while TIRM mainly overestimated it.

## RECOMMENDATIONS FOR MANAGEMENT

We have demonstrated that genetic data produced from bat faecal DNA are of high quality and can provide accurate estimates of population size even when samples are collected during only one sampling session, given appropriate statistical tools. With such protocols, no more capture is required and, consequently, animals are not disturbed. Furthermore, the method is easy to execute and cost-efficient because only one sampling session is required.

However, we would like to emphasize the importance of the sampling design, particularly with regard to the assumptions of the models. Each sampling scheme has to be carefully designed by taking into account the biology of the species concerned and the assumptions of the model used for the estimation of population size, such as population closure. Additionally, some previous studies have shown that the sampling intensity influences the confidence of the estimate (see, for example, Miller, Joyce & Waits 2005). In this study, we collected and genotyped approximately twice as many samples as the number of bats counted visually and obtained relatively narrow confidence intervals. Thus, in agreement with other studies (Miller, Joyce & Waits 2005; Solberg *et al.* 2006), and when the information is available, the number of samples collected for future studies should be approximately three times the 'assumed' number of individuals, leaving one-third of samples for possible experimental failure (samples yielding no DNA after extraction or no reliable genotype). If the focus is not on a precise estimate of population size but on population trends, sampling effort can be reduced to the point that will give enough statistical power to detect the trend. Power analyses are then required to calculate which sample size is adequate. We also stress the importance of the reliability of individual identification based on genetic fingerprints, as genotyping errors can lead to an overestimation of population size (Waits & Leberg 2000). Thus, we present in this study seven equations dealing with error rates and we recommend that researchers meticulously estimate and report such error rates and test the assumptions behind their calculations.

Finally, the estimator performance has important implications for managers, who should be aware of the strengths and drawbacks of the methods employed. For example, applying BM and ECM models tends to be biased downwards when capture heterogeneity exists. However, this error is conservative and may be appropriate when dealing with rare and endangered species. On the other hand, applying TIRM when capture heterogeneity is low tends to be biased upwards, thus providing a maximum limit for population size.

Reporting on a workshop about bat population monitoring, O'Shea, Bogan & Ellison (2003) emphasized that 'major improvements are needed in methods of estimating numbers of bats'. We think that the protocols we present here are a step in that direction,

because we explore new techniques and apply a modern statistical design to improve the scientific basis for predictions about future bat population trends (O'Shea, Bogan & Ellison 2003). With the increasing attention paid to biodiversity and conservation, new methodologies are necessary to study species, particularly elusive and endangered ones. Non-invasive CMR protocols can readily be adapted to a broad range of species within chiropterans and beyond, and should become a useful tool in wildlife management and conservation.

## Acknowledgements

We thank all the people who helped collecting samples in the field and gave us the opportunity to participate in the project of Bretagne Vivante-SEPNB on the conservation of the lesser horseshoe bat. We are grateful to G. Jones, G. F. McCracken and N. Valière, who provided valuable comments on earlier versions of the manuscript, C. R. Miller and an anonymous referee, whose comments greatly enhanced the quality of this paper, and F. Bontadina, who shared with us some information on sex ratios in colonies of the lesser horseshoe bat. Thanks to J. Gaisler, who provided us with many interesting papers dealing with the lesser horseshoe bat biology. This project was funded by the Region Brittany, the departments of Côtes d'Armor, Ille-et-Vilaine and Morbihan (Contrat Nature 'Plan d'Action Régional en faveur du Petit Rhinolophe').

## References

- Arbeitskreis Fledermause Sachsen-Anhalt (1997) *Zur Situation der Hufeisennasen in Europa: Contributions À la Situation Des Rhinolophides Dans l'Europe*. IFA-Verlag, Stecklenberg-Berlin, Germany.
- Banks, S.C., Hoyle, S., Horsup, D., Sunnucks, P. & Taylor, A.C. (2003) Demographic monitoring of an entire species (the northern hairy-nosed wombat, *Lasiiorhinus krefftii*) by genetic analysis of non-invasively collected material. *Animal Conservation*, **6**, 1–10.
- Bonin, A., Bellemain, E., Eidesen, P.B., Pompanon, F., Brochmann, C. & Taberlet, P. (2004) How to track and assess genotyping errors in population genetic studies. *Molecular Ecology*, **13**, 3216–3273.
- Broquet, T. & Petit, E. (2004) Quantifying genotyping errors in noninvasive population genetics. *Molecular Ecology*, **13**, 3601–3608.
- Frantz, A.C., Pope, L.C., Carpenter, P.J., Roper, T.J., Wilson, G.J., Delahay, R.J. & Burke, T. (2003) Reliable microsatellite genotyping of the Eurasian badger (*Meles meles*) using faecal DNA. *Molecular Ecology*, **12**, 1649–1661.
- Frantz, A.C., Schaul, M., Pope, L.C., Fack, F., Schley, L., Muller, C.P. & Roper, T.J. (2004) Estimating population size by genotyping remotely plucked hair: the Eurasian badger. *Journal of Applied Ecology*, **41**, 985–995.
- Gaisler, J. (1963a) The ecology of lesser horseshoe bat (*Rhinolophus hipposideros* Bechstein, 1800) in Czechoslovakia. I. *Vestník Československe Spolenosti Zoologicke*, **27**, 211–233.
- Gaisler, J. (1963b) The ecology of lesser horseshoe bat (*Rhinolophus hipposideros hipposideros*) in Czechoslovakia. II. Ecological demands, problem of synanthropy. *Acta Societatis Zoologicae Bohemoslovenicae*, **27**, 322–327.

- Gaisler, J. (1963c) Nocturnal activity in the lesser horseshoe bat, *Rhinolophus hipposideros*. *Folia Zoologica*, **12**, 223–230.
- Gaisler, J. (1965) The female sexual cycle and reproduction in the lesser horseshoe bat (*Rhinolophus hipposideros hipposideros* Bechstein, 1800). *Vestník Československé Společnosti Zoologické*, **29**, 336–352.
- Gaisler, J. (1966a) Reproduction in the lesser horseshoe bat (*Rhinolophus hipposideros hipposideros* Bechstein, 1800). *Bijdragen tot de Dierkunde*, **36**, 45–64.
- Gaisler, J. (1966b) A tentative ecological classification of colonies of the European bats. *Lynx (Praha)*, **6**, 35–39.
- Gaisler, J. (1989) The r–K selection model and life-history strategies in bats. *European Bat Research 1987* (eds V. Hanak, I. Horacek & J. Gaisler), pp. 117–124. Charles University Press, Prague, Czech Republic.
- Gaisler, J., Hanák, V., Hanzal, V. & Jarsky, V. (2003) Results of bat banding in the Czech and Slovak Republics, 1948–2000. *Vespertilio*, **7**, 3–61.
- Gazey, W.J. & Staley, M.J. (1986) Population estimation from mark–recapture experiments using a sequential Bayes algorithm. *Ecology*, **67**, 941–951.
- Gittleman, J.L., Holroyd, S.L. & Barclay, R.M.R. (1988) Energy allocation in mammalian reproduction. *American Zoologist*, **28**, 863–875.
- Goudet, J. (2001) *FSTAT, A Program to Estimate and Test Gene Diversities and Fixation Indices*, Version 2.9.3. [http://www.unil.ch/dee/page6767\\_en.html](http://www.unil.ch/dee/page6767_en.html) (accessed 26 April 2007).
- Hutson, A.M., Mickleburgh, S.P. & Racey, P.A. (2001) *Microchiropteran Bats: Global Status Survey and Conservation Action Plan*. IUCN, Gland, Switzerland and Cambridge, UK.
- Ihaka, R. & Gentleman, R. (1996) r: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.
- Kayikcioglu, A. & Zahn, A. (2004) High temperature and the use of satellite roosts in *Rhinolophus hipposideros*. *Mammalian Biology*, **69**, 337–341.
- Kohn, M.H., York, E.C., Kamradt, D.A., Haught, G., Sauvajot, R.M. & Wayne, R.K. (1999) Estimating population size by genotyping faeces. *Proceedings of the Royal Society of London Series B, Biological Sciences*, **266**, 657–663.
- Miller, C.R., Joyce, P. & Waits, L. (2005) A new method for estimating the size of small populations from genetic mark–recapture data. *Molecular Ecology*, **14**, 1991–2005.
- Mills, L.S., Citta, J.J., Lair, K.P., Schwartz, M.K. & Tallmon, D.A. (2000) Estimating animal abundance using non-invasive DNA sampling: promise and pitfalls. *Ecological Applications*, **10**, 283–294.
- O’Shea, T.J., Bogan, M.A. & Ellison, L.E. (2003) Monitoring trends in bat populations of the United States and territories: status of the science and recommendations for the future. *Wildlife Society Bulletin*, **31**, 16–29.
- Paetkau, D. (2003) An empirical exploration of data quality in DNA-based population inventories. *Molecular Ecology*, **12**, 1375–1387.
- Petit, E. & Valière, N. (2006) Estimating population size with noninvasive capture–mark–recapture data. *Conservation Biology*, **20**, 1062–1073.
- Piggott, M.P. & Taylor, A.C. (2003) Remote collection of animal DNA and its applications in conservation management and understanding the population biology of rare and cryptic species. *Wildlife Research*, **30**, 1–13.
- Prugh, L.R., Ritland, C.E., Arthur, S.M. & Krebs, C.J. (2005) Monitoring coyote population dynamics by genotyping faeces. *Molecular Ecology*, **14**, 1585–1596.
- Puechmaillie, S., Mathy, G. & Petit, E. (2005) Characterization of 14 polymorphic microsatellite loci for the lesser horseshoe bat, *Rhinolophus hipposideros* (Rhinolophidae, Chiroptera). *Molecular Ecology Notes*, **5**, 941–944.
- Puechmaillie, S.J., Mathy, G. & Petit, E.J. (2007) Good DNA from bat droppings. *Acta Chiropterologica*, in press.
- Rossiter, S.J., Ransome, R.D., Faulkes, C.G., Le Comber, S. & Jones, G. (2005) Mate fidelity and intra-lineage polygyny in greater horseshoe bats. *Nature*, **437**, 408–411.
- Schofield, H.W. (1996) *The ecology and conservation of Rhinolophus hipposideros, the lesser horseshoe bat*. PhD Thesis. University of Aberdeen, Aberdeen, UK.
- Simmons, N.B. (2005) Order Chiroptera. *Mammal Species of the World: A Taxonomic and Geographic Reference* (eds D.E. Wilson & D.M. Reeder), pp. 312–529. Johns Hopkins University Press, Baltimore, Maryland.
- Solberg, K.H., Bellemain, E., Drageset, O.-M., Taberlet, P. & Swenson, J.E. (2006) An evaluation of field and non-invasive genetic methods to estimate brown bear (*Ursus arctos*) population size. *Biological Conservation*, **128**, 158–168.
- Taberlet, P. & Luikart, G. (1999) Non-invasive genetic sampling and individual identification. *Biological Journal of the Linnean Society*, **68**, 41–55.
- Taberlet, P., Griffin, S., Goossens, B., Questiau, S., Manceau, V., Escaravage, N., Waits, L.P. & Bouvet, J. (1996) Reliable genotyping of samples with very low DNA quantities using PCR. *Nucleic Acids Research*, **24**, 3189–3194.
- Taberlet, P., Waits, L.P. & Luikart, G. (1999) Noninvasive genetic sampling: look before you leap. *Trends in Ecology and Evolution*, **14**, 323–327.
- Valière, N. (2002) GIMLET: a computer program for analysing genetic individual identification data. *Molecular Ecology Notes*, **2**, 377–379.
- Waits, J.L. & Leberg, P.L. (2000) Biases associated with population estimation using molecular tagging. *Animal Conservation*, **3**, 191–199.
- Waits, L.P., Luikart, G. & Taberlet, P. (2001) Estimating the probability of identity among genotypes in natural populations: cautions and guidelines. *Molecular Ecology*, **10**, 249–256.
- Wasser, S.K., Houston, C.S., Koehler, G.M., Cadd, G.G. & Fain, S.R. (1997) Techniques for application of faecal DNA methods to field studies of Ursids. *Molecular Ecology*, **6**, 1091–1097.
- Wilberg, M.J. & Dreher, B.P. (2004) GENECAP: a program for analysis of multilocus genotype data for non-invasive sampling and capture–recapture population estimation. *Molecular Ecology Notes*, **4**, 783–785.
- Wilson, G.J., Frantz, A.C., Pope, L.C., Roper, T.J., Burke, T., Cheeseman, C.L. & Delahay, R.J. (2003) Estimation of badger abundance using faecal DNA typing. *Journal of Applied Ecology*, **40**, 658–666.

Received 7 April 2006; final copy received 8 March 2007  
Editor: Simon Thirgood

## Supplementary material

The following supplementary material is available for this article.

### Appendix S1. Visual estimates of colony size

**Fig. S1.** Results of the simulation test of capture heterogeneity

This material is available as part of the online article from: <http://www.blackwell-synergy.com/doi/full/10.1111/j.1365-2656.2007.01321.x>.

Please note: Blackwell Publishing is not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.