University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

November 2004

Estimation of genetically effective breeding numbers using a rejection algorithm approach

Uma Ramakrishnan
University of California, San Diego, La Jolla, CA

Jay F. Storz
University of Nebraska - Lincoln, jstorz2@unl.edu

Barbara L. Taylor
National Marine Fisheries Service, Southwest Fisheries Science Center, 8604 La Jolla Shores Dr, La Jolla, California

Russell Lande
Department of Biology, University of California, San Diego, La Jolla, CA

Estimation of genetically effective breeding numbers using a rejection algorithm approach

Uma Ramakrishnan, Department of Biology, University of California, San Diego, La Jolla, CA 92093, USA. [Correspondence and current address: Uma Ramakrishnan, Anthropological Sciences, Bldg 360–361d, Stanford, California, 94305–2117, USA. Email: uramakri@stanford.edu]

Jay F. Storz, Department of Biology, San Francisco State University, 1600 Holloway Avenue, San Francisco, CA 94132, USA,

Barbara L. Taylor, National Marine Fisheries Service, Southwest Fisheries Science Center, 8604 La Jolla Shores Dr, La Jolla, CA 92037, USA

Russell Lande, Department of Biology, University of California, San Diego, La Jolla, CA 92093, USA

Abstract

Polygynous mating results in nonrandom sampling of the adult male gamete pool in each generation, thereby increasing the rate of genetic drift. In principle, genetic paternity analysis can be used to infer the effective number of breeding males (N_{ebm}). However, this requires genetic data from an exhaustive sample of candidate males. Here we describe a new approach to estimate N_{ebm} using a rejection algorithm in association with three statistics: Euclidean distance between the frequency distributions of maternally and paternally inherited alleles, average number of paternally inherited alleles and average gene diversity of paternally inherited alleles. We quantify the relationship between these statistics and N_{ebm} using an individual-based simulation model in which the male mating system varied continuously between random mating and extreme polygyny. We evaluate this method using genetic data from a natural population of highly polygynous fruit bats (*Cynopterous sphinx*). Using data in the form of mother–offspring genotypes, we demonstrate that estimates of N_{ebm} are very similar to independent estimates based on a direct paternity analysis that included data on candidate males. Our method also permits an evaluation of uncertainty in estimates of N_{ebm} and thus facilitates inferences about the mating system from genetic data. Finally, we investigate the sensitivity of our method to sample size, model assumptions, adult population size and the mating system. These analyses demonstrate that the rejection algorithm provides accurate estimates of N_{ebm} across a broad range of demographic scenarios, except when the true N_{ebm} is high.

Keywords: bats, effective number of breeding males, effective population size, individual-based model, mating system, polygyny, rejection algorithm and microsatellite DNA

Introduction

Effective population size is a parameter of central importance in evolutionary biology and conservation because it provides a predictive measure of the rate of loss of genetic diversity (Wright 1931; Lande & Barrowclough 1987). In populations of constant size, the rate of genetic drift is determined primarily by variance in reproductive success and the sex ratio of breeding adults. Comprehensive assessments of the genetic effective size of natural populations thus require estimation of the effective number of breeding individuals (N_{eb}) as well as the effective number of breeding males (N_{ebm}) and females (N_{ebf}).

Various methods have been developed to estimate the effective number of breeders. Moment-based *F*-statistics calculated from temporal genetic data (Pollock 1983; Waples 1989) are used to estimate N_{eb}. The temporal method can also be used to estimate effective size in a likelihood-based (MCLEEPS, Williamson & Slatkin 1999; Anderson *et al*. 2000), Bayesian (Berthier *et al*. 2002) and pseudo-likelihood context (MLNE, Wang 2001). N_{eb} can also be estimated using linkage disequilibrium (as in Bartley *et al*. 1992). The degree of heterozygote excess in same-age progeny arrays can be used to estimate N_{eb} using genetic data from a single time point (Pudovkin *et al*. 1996; Luikart & Cornuet 1999). If the number of males and females contributing to a progeny array is small,

allelic frequencies in male and female parents will differ due to binomial sampling error, which results in an excess of heterozygosity in the offspring (compared to Hardy–Weinberg expectations). This method was developed originally to model mating systems characterized by random union of gametes (Pudovkin *et al*. 1996), and has been modified to investigate the effects of monogamy and varying degrees of polygyny (Luikart & Cornuet 1999). Simulation-based evaluations of this method revealed that estimates of N_{eb} were characterized by an upward bias and very large confidence intervals. The temporal methods and the heterozygote excess method quantify the degree of uncertainty associated with point estimates of N_{eb} but do not allow separate estimation of N_{ebm} and N_{ebf}. The temporal method was used to estimate N_{ebf} using genetic data from mitochondrial DNA (as in Laikre *et al*. 1998) in fish populations. Similarly, a temporal approach based on Y-chromosome data could be used to estimate N_{ebm}, but we could not find any applications of the above approach.

In mammals and other animal taxa that are characterized by polygynous mating systems, the variance in male mating success can be expected to have an especially strong influence on effective size (Nunney 1993), making it particularly important to estimate N_{ebm}. In principal, genetic paternity analyses can be used to infer N_{ebm}, but this approach requires exhaustive sampling of candidate males. As illustrated by results of several recent marker-based studies of male mating success, even the most comprehensive sampling efforts typically leave the vast majority of offspring with unidentified sires (Coltman *et al*. 1998; Worthington Wilmer *et al*. 1999). Exhaustive sampling of candidate males is crucial to discerning paternity (Marshall *et al*. 1998). For example, when 50% of the adult males in a population are sampled, paternity is assigned successfully to only 10% of the offspring (Marshall *et al*. 1998). Such analysis might result in biased estimates of N_{ebm}. An alternative approach is to infer the number and proportional representation of breeding males from the frequency distribution of paternally inherited alleles in an exhaustive sample of offspring with known mothers. This approach avoids the problems of sampling bias with regard to candidate males and holds promise for empirical studies of animal species in which paternal care of young is minimal or absent.

Related approaches have been used to estimate the effective number of male breeders (or female breeders) contributing to a half-sib progeny array. Assuming that all breeding males (or females) contribute equally ($N_{ebm} = N_{bm}$), the "minimum fathers" (or "minimum mothers") method uses paternal alleles (or maternal alleles) in the offspring to calculate the minimum possible number of males (or females) that could have sired the sampled offspring. This method underestimates the true number of breeding males (or females) (DeWoody *et al*. 2000; Emery *et al*. 2001; Fiumera *et al*. 2001). DeWoody *et al*. (2000) developed a simulation-based method to es-

timate N_{bm} (or N_{bf}) by assessing the frequency distribution of paternally derived haplotypes in large broods of half-sib offspring. A similar method developed in a Bayesian framework (Emery *et al*. 2001) involves calculating the probability of each potential mother–father pair for each offspring, and then using Markov chain Monte Carlo methods to estimate a posterior probability distribution of paternity for each offspring. The applicability of these approaches is limited to half-sib progeny arrays and when the possible number of fathers (or mothers) is low (< 15).

Alternatively, levels of paternal relatedness between offspring can also be used to estimate the number of sibships present in a set of sampled offspring, which can then be used to calculate variance in male mating success (Storz *et al*. 2001b). However, this approach does not account for uncertainty in the assignment of offspring to paternal sibships.

All the methods described so far have one of the following shortcomings: (i) they do not estimate N_{ebm}; (ii) they have limited applicability to half-sib progeny arrays; and (iii) they do not quantify the degree of uncertainty associated with point estimates of N_{ebm}. Here we describe a new approach to estimate N_{ebm} that addresses all these issues. Estimation of N_{ebm} is based on a rejection algorithm, which compares observed values of three statistics (Euclidean distance between the distributions of maternally and paternally inherited alleles, average number of paternally inherited alleles and average gene diversity of paternally inherited alleles) to simulation-based expectations. These simulation-based expectations are generated using an individual-based model of male mating systems, which allows us to simulate a continuous distribution of variance in male reproductive success (between random mating and extreme polygyny). Best-fitting estimates of N_{ebm} are conditional on adult population size, adult sex ratio, number of sampled individuals, number of loci scored per individual and the allelic frequency distributions of the sampled loci. Multiple runs of the rejection algorithm provide a distribution of N_{ebm} estimates. The rejection algorithm has been used extensively in association with coalescent simulations (Pritchard *et al*. 1999; Estoup *et al*. 2001), but ours is the first application of this approach to individual-based models.

We evaluate this method by applying it to an empirical data set from a natural population of polygynous fruit bats, *Cynopterous sphinx*, where direct and indirect estimates of paternity were used to estimate N_{ebm} with reasonable certainty. We also assess the bias of N_{ebm} estimates associated with different sample sizes and model assumptions.

Materials and methods

We investigated changes in three test statistics (Euclidean distance, average gene diversity of paternally inherited alleles and average number of paternally inherited

alleles) as a function of N_{ebm} using an individual-based model and simulations of male mating. We then used a rejection algorithm to estimate N_{ebm} and evaluated this method for a test case of polygynously mating bats. Finally, we used sensitivity analyses to investigate how this approach was affected by various model assumptions.

Test statistics

Euclidean distance. Euclidean distance measures geometric distance between two points in multidimensional space. Here we calculated the Euclidean distance between the frequency distributions of maternally and paternally inherited alleles. Frequency distributions were inferred by pooling maternally and paternally inherited alleles for all offspring. For each locus, the squared Euclidean distance (D_l^2) was calculated as:

$$D_l^2 = \sum_{i=1}^{a} (p_i - q_i)^2 \tag{1}$$

where p_i and q_i are frequencies of the ith allele in the male and female gamete pools, respectively, and a is the maximum number of alleles at the given locus. The total Euclidean distance D is given by:

$$D = \left(\sum_{l=1}^{L} D_l^2 \right)^{1/2} \tag{2}$$

where D_l^2 is the squared Euclidean distance at the lth locus and L is the number of loci.

Number of alleles. The number of paternally inherited alleles was averaged across all loci for all the sampled offspring.

Gene diversity. Gene diversity was calculated using:

$$H = 1 - \sum_{i=1}^{a} p_i^2 \tag{3}$$

where p_i is the frequency of the ith allele in the male gamete pool and a is the maximum number of alleles at the given locus. Gene diversity was also averaged over loci.

Individual-based model

An adult population of size N was modelled, with N_m adult males and N_f adult females ($N = N_m + N_f$) and a same-age cohort of N_o offspring (with $N_o = N$). Male mating success was modelled using a geometric function. For the ith male, probability of reproduction, f_i, is given by:

$$f_i = \frac{c^{i-1}}{K} \tag{4}$$

where i varies from one to N_m and c varies from zero to one. K is a normalization constant given by:

$$K = \frac{1 - c}{1 - c^{N_m}} \tag{5}$$

where N_m is the number of adult males.

For a given value of c, the seasonal variance in male mating success (Var_{yr}) was given by:

$$Var_{yrm} = \left[(4N_m - Mean_{yrm}) \left(\frac{1 - c}{1 + c} \right) \left(\frac{1 + c^{N_m}}{1 - c^{N_m}} \right) \right] - 2 \tag{6}$$

where N_m is the number of adult males and $Mean_{yrm}$ in the mean mating success for males within a given year (or breeding season — the time unit depends on the reproductive biology of the species in question). Low values of c (close to 0) result in some males having a much higher probability of mating than other males in a given year. This resulted in a high variance in mating success and low N_{ebm}. Conversely, high values of c (close to 1) result in a more uniform distribution of male mating success and a relatively larger N_{ebm}.

The effective number of breeding males (N_{ebm}) was calculated using Var_{yr} and a given $Mean_{yr}$ using the following expression (Crow & Kimura 1970; Lande & Barrowclough 1987):

$$N_{ebm} = \left(\frac{(N_m Mean_{yrm}) - 1}{\left(\left(Mean_{yrm} + \left(\frac{Var_{yrm}}{Mean_{yrm}} \right) \right) - 1 \right)} \right) \tag{7}$$

where $Mean_{yr}$ is the mean seasonal mating success, Var_{yr} is the seasonal variance in mating success and N_m is the number of adult males. For females, we assumed no mating skew and that $c = 1$. Given $c = 1$, we evaluated $K = 1/N_f$ (using L'Hospital's rule for $c \geq 1$). Assuming that N_o offspring (where $N_o = N$) were born and the adult sex ratio was unity, $Mean_{yrf} = 2$. The seasonal variance in female mating success was given by $Var_{yrf} = 2(1 - 1/N_f)$ and the effective number of breeding females reduced to N_f.

Simulations

In order to assess the impact of mating system on the three statistics of interest (Euclidean distance, average gene diversity and number of alleles), we modelled a population of 250 adults (125 males and 125 females). Although Var_{yrm} (and consequently N_{ebm}) can be modelled continuously, trends in these statistics are illustrated better by modelling values of N_{ebm}/N_m and the associated mean of the three summary statistics. Specifically, we modelled N_{ebm}/N_m values of 0.02, 0.04, 0.06, 0.08, 0.1, 0.3, 0.5, 0.7 and 0.9 corresponding to c-values of 0.431, 0.665, 0.765, 0.819, 0.853, 0.948, 0.970, 0.981, 0.991 determined by solving equations 6 and 7 numerically. We assumed that $Mean_{yr} = 2.0$ for both male and females. Microsatellite loci modelled in the simulations were based

on empirically observed allele frequency distributions. Ten different loci were modelled using allele frequency distributions derived from a range of species [one locus each from rock wallabies (Eldridge *et al*. 2001), harbour seals (Coltman *et al*. 1998), pipefish (McCoy *et al*. 2001), gentle lemurs (Nievergelt *et al*. 2002), sunfish (Dewoody *et al*. 1998) and tree swallows (Conrad *et al*. 2001)] and two loci from sperm whales (S. Mesnick pers. com.) and chimpanzees (Gagneux *et al*. 1999). All these loci were characterized by $H > 0.85$. All loci were in Hardy–Weinberg equilibrium.

The simulations were conducted as follows:

1. Multilocus genotypes were assigned to all 250 adults assuming Hardy–Weinberg equilibrium (by random sampling from the allelic frequency distributions).
2. Two hundred and fifty offspring were born to the 125 adult females: all females had equal chances of reproducing. Offspring were recorded as part of each female's reproductive profile.
3. For a given value of N_{ebm}/N_m, all 250 offspring were assigned fathers based on a normalized probability distribution using the appropriate c-value and equations 4 and 5.
4. Offspring were assigned genotypes by sampling one allele each at random from the maternal and the paternal genotypes. This was repeated for all the loci and for all offspring.
5. Maternally and paternally inherited alleles were recorded for all offspring.
6. All three summary statistics (Euclidean distance, average gene diversity and averaged number of alleles) were calculated for the offspring cohort.
7. Simulations were repeated 2000 times (for a given N_{ebm}/N_m) to incorporate stochastic variation in the sampling of both maternal and paternal genotypes.

Application to empirical data

Rejection algorithm: estimating N_{ebm} for C. sphinx. Summary statistics provide insight into the population scenario. In our case, for example, increased Euclidean distance indicates polygyny. However, we are interested in parameter estimation, and the rejection algorithm allows us to estimate the population parameter of interest (in this case N_{ebm}) using a population modelling framework and the relevant summary statistics. The rejection algorithm has been applied extensively in the context of coalescent models to investigate historical effective population size, historical growth rates and mutational properties of loci. It is based on comparing a set of simulated (and relevant) summary statistics to observed values of the same statistics. If the simulated statistics are close enough to the observed values, the parameter values (in our case N_{ebm}) are saved as part of the parameter estimate.

To test the ability of our rejection algorithm to estimate N_{ebm} (using the above three statistics) we an-

alyzed an empirical data set from a natural population of fruit bats (Storz *et al*. 2000a,b, 2001a, 2001b). The data consisted of 10-locus genotypes for 67 mother–offspring pairs that were sampled during a single parturition period. Using genetic assessment of shared paternity among same-age offspring, Storz *et al*. (2001b) estimated that this cohort of 67 offspring was sired by a total of 15 males, and estimated the variance in male mating success to be 11.6. Using the formula, we calculated $N_{ebm} = 11$ for this population. We applied the rejection algorithm to these data using the following steps.

1. Genetic data: the 67 mother–offspring pairs were used to identify maternally and paternally inherited alleles. In cases where the paternal allele was uncertain (i.e. mother and offspring were both heterozygous for the same alleles), it was estimated by randomly assigning one of the offspring's alleles as the paternal allele. We then calculated Euclidean distance (E_O), paternally inherited allelic gene diversity (H_O) and number of paternally inherited alleles (Na_O).
2. Setting the initial bounds for N_{ebm}. In order to apply the rejection algorithm, we investigated the correspondence between simulated and observed summary statistics for a range of N_{ebm} values. The upper bound for N_{ebm} was set to be N_m and the lower bound was defined by the minimum number of males. An extremely conservative estimate of the minimum number of males was used. The number of paternally inherited alleles at all the loci were tabulated, and the minimum of these was determined. Further, this number was divided by two (assuming all fathers were heterozygous). This results in the absolute minimum number of fathers that could have contributed to the offspring. For example, if the number of paternally inherited alleles at each of three loci are 10, 6 and 12, at least three (6/2) males have contributed genetically to the offspring.
3. Other model parameters: adult population size (257), number of adult males (121), number of adult females (136), number of offspring born (257) and the allele frequency distributions for 10 loci (Storz *et al*. 2001a) were used as parameters for the individual–based model described above.
4. Following Pritchard *et al*. (1999), we implemented the rejection algorithm according to the following steps:
   (i) Pick a value of N_{ebm} from the initial distribution.
   (ii) Run the individual-based simulation model for this value of N_{ebm} and sample 67 mother–offspring pairs. Calculate the associated values of Euclidean distance (E_S), gene diversity (H_S) and number of alleles (Na_S) for the simulated data.
   (iii) Calculate $|E_O - E_S|/E_O$, $|H_O - H_S|/H_O$ and $|Na_O - Na_S|/Na_O$.
   (iv) If these values are less than a given value δ (we assumed $\delta = 0.01$ for Euclidean distance for gene

diversity and for number of alleles) for all three statistics, save the parameter value as part of the estimated distribution of N_{ebm}.

(v) Repeat steps (i) to (iv) many times. The saved values of N_{ebm} correspond to the estimated parameter distribution.

Sensitivity analyses

Sample size. The effect of reducing the number of sampled mother–offspring pairs was investigated by repeating the above rejection algorithm for a sample of 40 mother–offspring pairs. These 40 mother–offspring pairs were sampled at random (without replacement) from the available 67, and steps (i) to (v) were repeated to obtain parameter estimates for N_{ebm}.

Number of loci. The effect of decreasing the number of loci from 10 to five was investigated similarly. The five most polymorphic loci (of the 10) were identified for the bat data set, and observed summary statistics were calculated based on these five loci. The rejection algorithm was then used to obtain estimates for N_{ebm}.

Female mating system. The rejection algorithm approach presented above assumes a model where $N_{ebf} = N_f$. This may not be the case (although the researcher who collects data from these populations is unaware of this). Simulations were used to generate 'observed genetic data' assuming a population model with skewed female mating. Three levels of skew were modelled ($N_{ebf} = 0.8N_f$, $N_{ebf} = 0.6N_f$ and $N_{ebf} = 0.4N_f$). The rejection algorithm described above (assumes $N_{ebf} = N_f$) was used to estimate N_{ebm}. The simulation considered 67 mother–offspring pairs and 10 loci. All other parameters were kept the same as the earlier bat model.

Male mating function. We investigated the sensitivity of our estimates to the specific function used to model male mating success. The probability of reproduction for males is given by:

$$f_i = \frac{1}{i} \tag{8}$$

where i is the male number and varies between one and N_m. In the individual-based model for bats, this resulted in $N_{ebm} = 18$. A single run of this model was used to generate "observed" mother–offspring genetic data ($n = 67$, $L = 10$). The rejection algorithm (based on the original function describing male mating success; equations 4 and 5) was then used to estimate N_{ebm}.

Higher true $_{Nebm}$: For the fruit bat population, results suggest that true N_{ebm} is low. We investigated the performance of the rejection algorithm for three situations where the true N_{ebm} was higher ($N_{ebm} = 25$, 50 and 75 males). "Observed" mother–offspring genetic data were

generated using c-values of 0.923, 0.962 and 0.976 (corresponding to $N_{ebm} = 25$, 50 and 75 males, respectively) for 67 samples and 10 loci. All other parameters of the model were retained. The rejection algorithm was used to estimate N_{ebm}.

Higher population size: The number of adults in the bat population was small. We investigated performance of the rejection algorithm when the adult population size was much larger. We modelled three populations of sizes 500, 1000 and 2000 adults. We assumed $N_f = N_m$ in all cases. We assumed $N_{ebf} = N_f$ and $N_{ebm} = 0.1N_m$ for all three cases. "Observed" genetic data were generated for mother–offspring pairs (number of mother–offspring pairs = 100, 200 and 400, respectively, number of loci = 10) using a single simulation. The rejection algorithm was implemented to estimate N_{ebm} as described earlier.

Results

Changes in Euclidean distance, gene diversity and number of alleles as a function of N_{ebm}/N_m

Modelled values of N_{ebm}/N_m are presented in two separate graphs, corresponding to extreme polygyny ($N_{ebm}/N_m = 0$–0.1) and moderate polygyny ($N_{ebm}/N_m = 0.1$–1.0). Euclidean distance was highest for extreme polygyny, and decreased as the distribution of male mating success became less skewed (Figure 1). Conversely, average H and the average number of paternally inherited alleles both increased with N_{ebm} (Figs 2 and 3). The rate of change for all three statistics was higher for extreme polygyny compared with moderate polygyny.

Rejection algorithm estimates of N_{ebm}/N_m for C. sphinx

One million iterations of the rejection algorithm yielded a distribution of estimates of N_{ebm}/N_m with 1000 values (Figure 4). This corresponds to a rejection rate of 0.001 (Table 1). The estimate of N_{ebm} from the rejection algorithm (median value of the estimated distribution) was very close to the true value (estimated $N_{ebm} = 12$, true $N_{ebm} = 11.7$, bias = +5%, Table 1). Additionally, the 95th and 5th percentile of the estimated distribution allow quantification of parameter uncertainty (Table 1), indicating that seven to 25 males contributed genetically to this population.

Sensitivity analysis

Using fewer samples ($n = 40$ compared to $n = 67$) to estimate N_{ebm} resulted in a slightly higher bias (Table 1, bias =+16%) as well as a less confident estimate. Results were similar when mother–offspring genetic data from
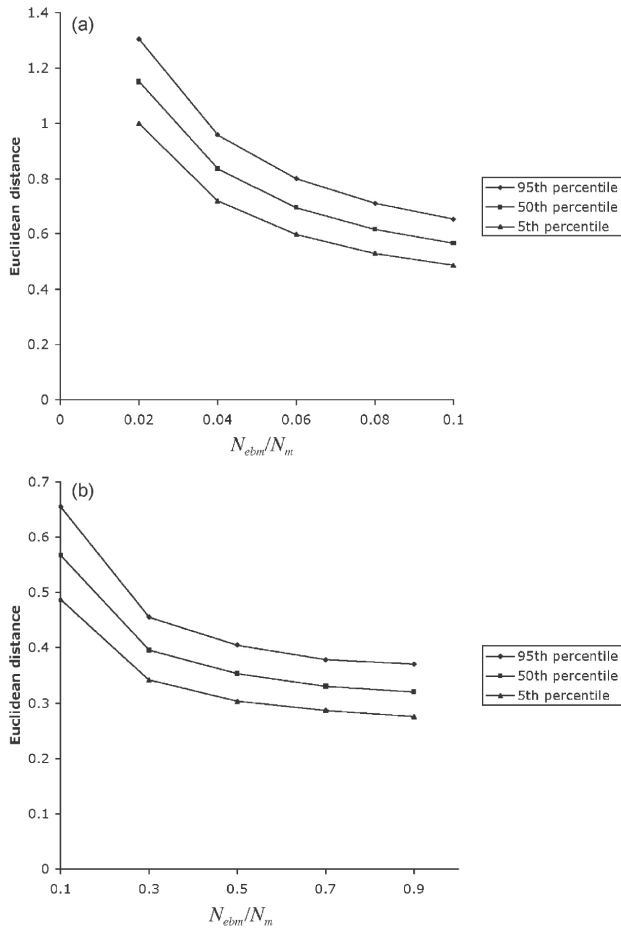
Figure 1. Change in Euclidean distance with N_{ebm}/N_m for (a) extreme polygyny (N_{ebm} = 0–0.1) and (b) moderate polygyny (N_{ebm} = 0.1–1.0) Figure shows the 95th, 50th and 5th percentile for 2000 simulation runs.

Figure 2. Change in average number of paternally inherited alleles with N_{ebm}/N_m for (a) extreme polygyny (N_{ebm} = 0–0.1) and (b) moderate polygyny (N_{ebm} = 0.1–1.0). Figure shows the 95th, 50th and 5th percentile for 2000 simulation runs.
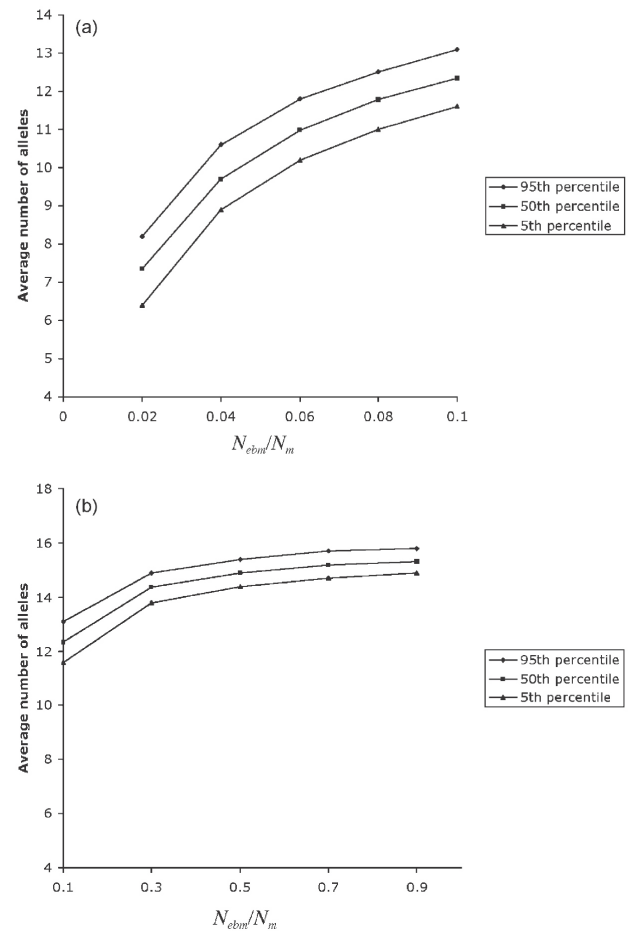
only five loci (compared with 10) were used, although the bias was slightly greater (Table 1, bias = +24%). Uncertainty in estimates of N_{ebm} were influenced more strongly by the number of sampled individuals than the number of loci (Table 1).

Assuming incorrectly that N_{ebf} = N_f (true N_{ebf} = 0.8N_f, 0.6N_f and 0.4N_f) resulted in slight underestimates of N_{ebm} (bias = +10%, −5% and −13%, respectively, Table 1) although the uncertainty of the estimates was low.

Using a different function for male mating resulted in slight underestimates of N_{ebm} (bias = −11%, Table 1), and the uncertainty of this estimate was higher. Both these analyses were characterized by lower acceptance rates for the rejection algorithm (Table 1).

When the rejection algorithm was applied to mother–offspring genetic data where true N_{ebm} was higher (N_{ebm} = 25, 50 and 100 males), N_{ebm} was underestimated (bias =−25%, −35% and −42%, respectively, Table 1). In all cases, the confidence intervals of the estimate included the true value.

Using the rejection algorithm to estimate N_{ebm} when the number of adults is much higher (N = 500, 1000 and 2000) resulted in slightly biased estimates of N_{ebm} (bias = +25%, +5% and +24%, Table 1). The uncertainty of these estimates was high.

Discussion

Our results reveal that the rejection algorithm used in association with three statistics (Euclidean distance calculated between frequency distributions of maternally and paternally inherited alleles, average number of paternally inherited alleles and average gene diversity of paternally inherited alleles) provides accurate estimates of the effective number of breeding males (N_{ebm}). Euclidean distance is related inversely to N_{ebm}, whereas gene diversity and number of paternally inherited alleles are directly proportional to N_{ebm}. As a result, these three statistics can be used in association with the rejection algorithm to estimate N_{ebm}.
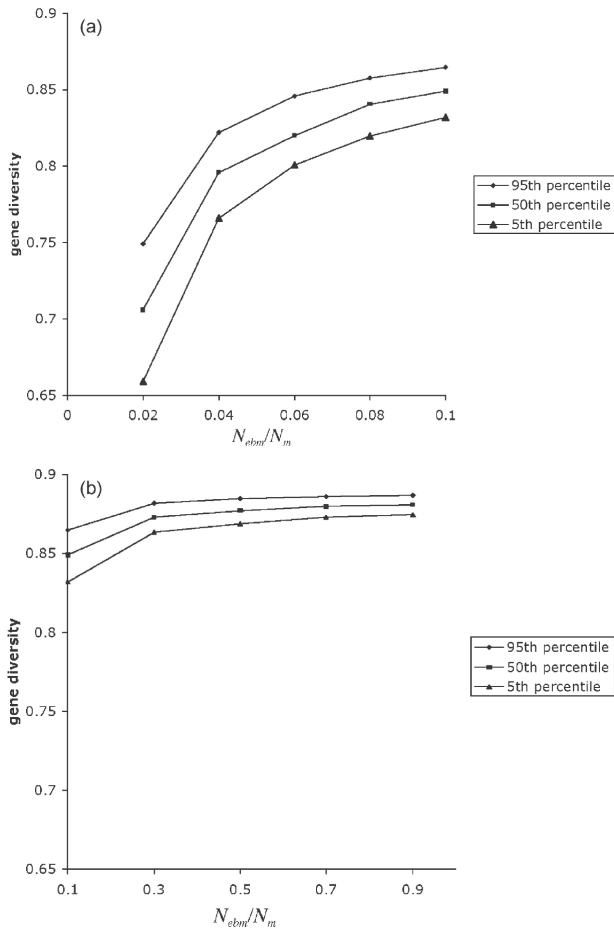
Figure 3. Change in average gene diversity of paternally inherited allelic distribution with N_{ebm}/N_m for (a) extreme polygyny ($N_{ebm} = 0$–0.1) and (b) moderate polygyny ($N_{ebm} = 0.1$–1.0). Figure shows the 95th, 50th and 5th percentile for 2000 simulation runs.

Figure 4. The estimated distribution of N_{ebm}/N_m from the C. *sphinx* data set. The median N_{ebm}/N_m estimate (0.096) corresponds closely to the true value of N_{ebm} (11 males).

Using data from mother–offspring pairs and potential fathers and a likelihood-based paternity test (which assigned paternity to 80% of the offspring) in conjunction with indirect inferences based on levels of paternal relatedness within offspring cohorts, Storz *et al*. (2001b) estimated that $N_{ebm} = 11$ for a population of polygynous fruit bats. Application of our rejection algorithm-based method to only mother–offspring genetic data from this population indicated that 12 males contributed genetically, although the range could be between seven and 25 (Table 1). These results confirm that using the rejection algorithm results in reasonably accurate estimates of N_{ebm} given this test case. Additionally, the rejection algorithm can be used to quantify the uncertainty around a given estimate.

Although we do not explore biases exhaustively in our approach due to variation in data characteristics (including number of samples and number of loci), adult population size and mating system and violati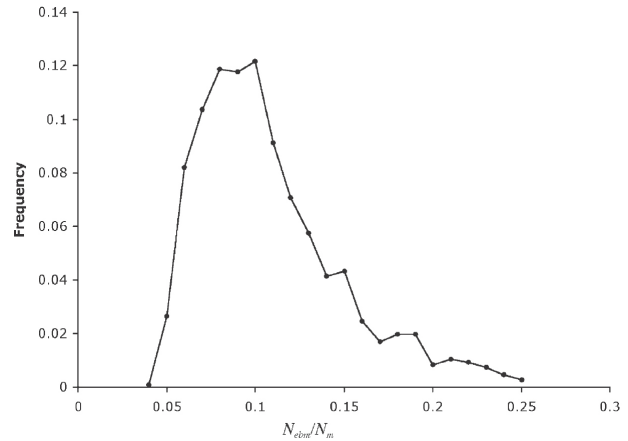on of model assumptions, the sensitivity analyses presented here provide an indication of possible impacts. These sensitivity analyses revealed that reducing the number of sampled individuals and the number of sampled loci resulted in slightly biased estimates of N_{ebm} (Table 1). Although the number of loci had the greatest impact on bias, the level of uncertainty of a given estimate of N_{ebm} was influenced more strongly by the number of sampled individuals. Such trade-offs between accuracy (bias) and precision (uncertainty) are relevant for practical application of this method.

We used sensitivity analyses to investigate parameters that neither the individual-based model nor the researchers collecting genetic data can control: the true N_{ebm}/N_m and the adult population size. Results revealed that our method is biased when the true N_{ebm} is high, with bias increasing with true N_{ebm}. This can be understood in the context of our model as follows. When c (the parameter in our model that controls mating skew) is low, large changes in c result in small changes in N_{ebm}/N_m. For example, changing N_{ebm}/N_m from 0.02 to 0.04 (at $N_m = 125$) requires a 55% increase in the value of c, whereas changing N_{ebm}/N_m from 0.7 to 0.9 requires only a 1% increase in c. Because c measures sampling skew, models with similar c-values correspond to similar sampling skews, resulting in similar statistic values. Based on the male mating function, we predict that estimates of N_{ebm} will be more biased as N_{ebm}/N_m approaches 1. The sensitivity analyses demonstrate this to be the case. However, an increase in bias as N_{ebm}/N_m approaches 1 will be true for any statistical method used to estimate N_{ebm}. As N_{ebm}/N_m increases, the correlation between the number of paternally inherited alleles and their distribution and the actual number of fathers decreases. In cases where N_{ebm}/N_m is high, we recommend that alternate methods (including heterozygosity excess (Luikart & Cornuet 1999) and the indirect sibship method (Storz

Table 1 Sensitivity analyses to investigate the impact of model assumptions*

Model	Lower N_{ebm}–upper N_{ebm}	True N_{ebm}	Median estimated N_{ebm}	Bias (%)	5th–95th percentile	Acceptance rate
Bat data						
$L = 10$, $n = 67$	3–121	11	11.6	+5	7–25	0.001
Bat data						
$L = 10$, $n = 40$	3–121	11	12.8	+16	6–40	0.001
Bat data						
$L = 5$, $n = 67$	3–121	11	13.7	+24	7–30	0.0009
Simulated data						
$N_{ebf} = 0.8 N_f$	3–121	11	12.1	+10	7–23	0.0005
$N_{ebf} = 0.6 N_f$	3–121	11	10.4	−5	6–18	0.0006
$N_{ebf} = 0.4 N_f$	3–121	11	9.6	−13	7–16	0.0005
Simulated data	4–121	18	16	−11	8–40	0.0006
Different function for male mating						
Simulated data						
$N_{ebm} = 25$	4–121	25	18.7	−25	10–46	0.001
$N_{ebm} = 50$	4–121	50	32.2	−35	11–62	0.0008
$N_{ebm} = 75$	4–121	75	43.1	−42	18–101	0.0015
Simulated data						
Number of adults = 500	5–250	25	30.2	+25	22–106	0.001
Number of adults = 1000	5–500	50	52.7	+5	26–192	0.0005
Number of adults = 2000	5–1000	100	75.3	+24	50–137	0.0004

*Sensitivity analyses investigating the impact of certain parameter values on bias of estimates for N_{ebm} using the rejection algorithm. The true model parameters, initial distribution of N_{ebm}, true N_{ebm}, median estimates of N_{ebm}, bias (based on median estimate) in N_{ebm}, the 5th and 95th percentiles of estimates and the acceptance rate are shown.

et al. 2001b) be used in association with the rejection algorithm approach presented here to arrive at multiple, independent estimates of N_{ebm}.

When the rejection algorithm is applied to data from larger populations (number of adults = 500, 1000 and 2000), the estimate of N_{ebm} is slightly biased. Bias in estimates of N_{ebm} did not reveal a discernible trend with adult population size. However, estimates of N_{ebm} for larger populations took significantly more computational time. Comparing these results to other sensitivity analyses revealed that bias in estimates of N_{ebm} depends more strongly on the value of N_{ebm}/N_m, not on the value of N_{ebm} alone. That is, even in a large population, the rejection algorithm will provide better estimates of N_{ebm} when the degree of polygyny is high.

Our results suggest that estimates of N_{ebm} are only slightly biased when distributions of male or female mating success differ from the assumed model.

Possible sources of bias that not addressed in this paper are uncertainty in estimates of adult population size and adult sex ratio. Our model assumes that these parameters can be estimated for the population. However, uncertainty in these parameters will affect estimates of N_{ebm}. If true adult population size is higher than the estimated size (used as input for the rejection algorithm), the estimates of N_{ebm} would be negatively biased (the model will underestimate the mean and overestimate variance incorrectly in male mating success, underestimating N_{ebm}). Similarly, if the true sex ratio were male biased, the rejection algorithm would underestimate N_{ebm}. Uncertainty in adult population size and adult sex ratio can be incorporated into our model if initial distributions of these two parameters were used as input for the rejection algorithm. Such approaches might increase uncertainty of estimates of N_{ebm} and further simulations are necessary to investigate this more completely.

Although the models explored in this study focus on male mating systems, results presented here are also valid for polyandrous species, where maternal care of offspring is rare. In such cases, the frequency distributions of maternally and paternally inherited alleles could be calculated from father–offspring pairs, and the rejection algorithm could be used to estimate effective number of breeding females.

This method does not attempt to reconstruct paternal genotypes. As a result, it does not provide any genetic information on the potentially successful males and does not allow for future genetic identification of these males. Thus, it will probably be most useful in cases where sampling adult males is difficult and sampling mother–offspring pairs is relatively easier. Additionally, like all previous methods, our method assumes that the population is closed.

We expect that the method presented here will be particularly relevant to populations of conservation concern. For such populations, estimating N_{ebm} could be a crucial part of the management process. For example, we might require information on N_{ebm} for input into a population viability analysis that would determine the probability of extinction of a species. In such cases, uncertainty in estimates of N_{ebm} would prove very useful to set upper and lower bounds on both N_{ebm} and the probability of extinction.

Acknowledgements

References

Anderson EC, Williamson EG, Thompson EA (2000) Monte Carlo evaluation of the likelihood for Ne from temporally spaced samples. *Genetics*, 156, 2109–2118.

Bartley D, Bagley M, Graham G, Bentley B (1992) Use of linkage disequilibrium to estimate effective size of hatchery and natural fish populations. *Conservation Biology*, 6, 365–375.

Berthier P, Beaumont MA, Cornuet JM, Luikart G (2002) Likelihood-based estimation of the effective population size using temporal changes in allele frequencies: a genealogical approach. *Genetics*, 160, 741–751.

Coltman DW, Bowen WD, Wright JM (1998) Male mating success in an aquatically mating pinniped, the harbour seal (*Phoca vitulina*), assessed by microsatellite DNA markers. *Molecular Ecology*, 7, 627–638.

Conrad KF, Johnston PV, Crossman C et al. (2001) High levels of extra-pair paternity in an isolated, low-density, island population of tree swallows (*Tachycineta bicolor*). *Molecular Ecology*, 10, 1301–1308.

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

DeWoody JA, Dewoody YD, Fiumera AC, Avise JC (2000) On the number of reproductives contributing to a half-sib progeny array. *Genetical Research*, 75, 95–105.

DeWoody JA, Fletcher DE, Wilkins SD, Nelson WS, Avise JC (1998) Molecular genetic dissection of spawning, parentage, and reproductive tactics in a population of red-breast sunfish, *Lepomis auritus*. *Evolution*, 52, 1802–1810.

Eldridge MDB, Kinnear JE, Onus ML (2001) Source population of dispersing rock-wallabies (*Petrogale lateralis*) identified by assignment tests on multilocus genotypic data. *Molecular Ecology*, 10, 2867–2876.

Emery AM, Wilson IJ, Craig S, Boyle PR, Noble LR (2001) Assignment of paternity groups without access to parental genotypes: multiple mating and developmental plasticity in squid. *Molecular Ecology*, 10, 1265–1278.

Estoup A, Wilson IJ, Sullivan C, Cornuet J, Moritz C (2001) Inferring population history from microsatellite and enzyme data in serially introduced cane toads, *Bufo marinus*. *Genetics*, 159, 1671–1687.

Fiumera AC, DeWoody YD, DeWoody JA, Asmussen MA, Avise JC (2001) Accuracy and precision of methods to estimate the number of parents contributing to a half-sib progeny array. *Journal of Heredity,* 92, 120–126.

Gagneux P, Boesch C, Woodruff DS (1999) Female reproduction strategies, paternity and community structure in wild West African chimpanzees. *Animal Behaviour*, 57, 19–32.

Laikre L, Jorde PE, Ryman N (1998) Temporal changes of mitochondrial DNA haplotype frequencies and female effective size in a brown trout (*Salmo trutta*) population. *Evolution*, 52, 910–915.

Lande R, Barrowclough GF (1987) Effective population size, genetic variation and their use in population management . In: *Viable Populations for Conservation* (ed. Soulé M), pp. 87–124. Cambridge University Press, New York.

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

Marshall TC, Slate J, Kruuk LEB, Pemberton JM (1998) Statistical confidence for likelihood-based paternity inference in natural populations. *Molecular Ecology*, 7, 639–655.

McCoy EE, Jones AG, Avise JC (2001) The genetic mating system and tests for cuckoldry in a pipefish species in which males fertilize eggs and brood offspring externally. *Molecular Ecology*, 10, 1793–1800.

Nievergelt CM, Mutschler T, Feistner ATC, Woodruff DS (2002) Social system of the Alaotran gentle lemur (*Hapalemur griseus alaotrensis*): genetic characterization of group composition and mating system. *American Journal of Primatology*, 57, 157–176.

Nunney L (1993) The influence of mating system and overlapping generations on effective population size. *Evolution*, 47, 1329–1241.

Pollock E (1983) A new method for estimating the effective population size from allele frequency changes. *Genetics*, 104, 531–548.

Pritchard JK, Seielstad MT, Perez-Lezaun A, Feldman MW (1999) Population growth of human Y chromosomes: a study of Y chromosome microsatellites. *Molecular Biology and Evolution*, 16, 1791–1798.

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

Storz JF, Balasingh J, Nathan PT (2000a) Dispersion and site-fidelity in a tent-roosting population of the short-

nosed fruit bat (*Cynopterus sphinx*) in southern India. *Journal of Tropical Ecology*, 16, 117– 131.

Storz JF, Bhat HR, Kunz TH (2000b) Social structure of a polygynous tent-making bat, *Cynopterus sphinx* (Megachiroptera). *Journal of Zoology* (London), 251, 151– 165.

Storz JF, Bhat HR, Kunz TH (2001a) Genetic consequences of polygyny and social structure in an Indian fruit bat, *Cynopterus sphinx*, I. Inbreeding, outbreeding, and population subdivision. *Evolution*, 55, 1215– 1223.

Storz JF, Bhat HR, Kunz TH (2001b) Genetic consequences of polygyny and social structure in an Indian fruit bat, *Cynopterus sphinx*. II. Variance in male mating success and effective population size. *Evolution*, 55, 1224– 1232.

Wang J (2001) A pseudo-likelihood method for estimating effective population size from temporally spaced samples. *Genetics Research*, 78, 243– 257.

Waples RS (1989) A generalized approach for estimating the effective population size from temporal changes in allele frequency. *Genetics*, 121, 379– 391.

Williamson E, Slatkin M (1999) Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics*, 152, 755– 761.

Worthington Wilmer J, Allen PJ, Pomeroy PP, Twiss SD, Amos W (1999) Where have all the fathers gone? An extensive analysis of paternity in the grey seal (*Halichoerus grypus*). *Molecular Ecology*, 8, 1417– 1429.

Wright S (1931) Evolution in Mendelian populations. *Genetics*, 16, 97– 159.

Uma Ramakrishnan and **Barbara Taylor** develop statistical methods in conservation genetics.

Jay Storz is interested in the molecular basis of adaptive evolutionary processes.

Russell Lande works on quantitative genetics and evolutionary theory.