Supplemental Information for:

Brood-tending males in a biparental fish suffer high paternity losses but rarely cuckold

Aneesh P H Bose, Holger Zimmermann, Jonathan M Henshaw, Karoline Fritzsche, Kristina M Sefc

1. Outline of custom Perl script for multi-locus genotype comparisons

Here we provide an overview of the Perl script written to compare male genotypes for the purpose of identifying pair-bonded males that acted as cuckolders and any males that cuckolded in multiple territories. This script conducted pairwise comparisons between all male genotypes (i.e. observed pair-bonded male genotypes and reconstructed cuckolder male genotypes), while implementing the following rules:

1) Alternative genotypes reconstructed for the same cuckolder male were not compared to each other. Cuckolder male genotypes could not always be unambiguously reconstructed (see Methods) and in these cases cuckolder males sometimes had several alternative genotypes. For example, when the mother and extra-pair fry were both heterozygous for the same alleles, e.g. X and Y, we scored the paternal alleles (i.e. the cuckolder male's alleles) as X or Y, since the cuckolder could have contributed either of these two alleles. Therefore, this cuckolder male had two alternative genotypes, one containing the X allele and the other containing the Y allele. If a second ambiguity arose at a different locus, then there would be two ambiguous alleles (e.g. X or Y and A or B), and we would have generated four alternative genotypes to cover all combinations of X/Y and A/B. Out of all these alternative genotypes, one would be the true cuckolder male genotype.

2) So-called "partial genotypes" were also not compared to each other. When a cuckolder sired only a few offspring, genotype reconstructions were partial in that only one allele could be inferred at each locus. Any two such partial genotypes are always consistent with coming from a single male, rendering partial-partial genotype comparisons uninformative.

3) In order for a locus to be consistent with *identity* (i.e. the two genotypes belong to the same male), one of the following conditions had to be met:

   - In comparisons between two completely genotyped loci (both alleles scored for each genotype), both alleles had to match. If this was the case, we counted two matching alleles and one matching locus. If not, we counted one mismatching locus.

   - In comparisons between a completely genotyped locus (both alleles scored for this locus) and an incompletely genotyped locus (only one allele scored for this locus), the allele of the incomplete locus had to match one of the two alleles of the complete

locus. If this was the case, we counted 1 matching allele and one matching locus. If not, we counted one mismatching locus.

   - When both genotypes were incomplete at the locus under consideration, the locus was consistent with identity irrespective of whether the two alleles matched or not. We therefore always counted one matching locus. But if the two alleles were identical, we also counted 1 matching allele.

The number of matching alleles, matching loci, and *mis*matching loci were counted across all loci. The output reported all pairwise comparisons with the number of matching and mismatching loci, the number of matching alleles, the number of alleles scored for each genotype (the number of alleles scored sometimes varied due to partial reconstructions and missing data), and the number of alleles compared between each pair of genotypes (also variable due to missing data).

We considered two genotypes to be consistent with a single male when: 1) they were both identical, or 2) if they differed by up to two mismatching loci. This was to account for possible errors in genotyping or genotype reconstruction, though we also required at least eight matching loci. When we found a match between two such genotypes, we then considered this male to have sired offspring in both broods / territories.

2. Estimating average reproductive success of unpaired males

Here we derive maximum likelihood estimates for the number of unpaired males m_T in the entire quadrat and the probability p that any given male cuckolds any given territory (see Methods in main text). We assume that all males have the same propensity to cuckold, and that all territories have the same chance of being cuckolded. If we sample fry from n_S territories, then the number of sampled territories cuckolded by any given male follows a binomial distribution $\text{Bin}(n_S, p)$. The probability that he cuckolds exactly k territories is:

$$\binom{n_S}{k} p^k (1-p)^{n_S - k} \qquad \text{(S1)}$$

We write M_k for the random variable representing the number of males that cuckold exactly k of the sampled territories. The number of males cuckolding at least one of the sampled territories is then $M_S = M_1 + M_2 + \cdots + M_{n_S}$. The remaining males $M_0 = m_T - M_S$ do not cuckold any of the sampled territories. For fixed values of the parameters m_T and p, the random variables $M_1, \ldots, M_{n_S}$ jointly follow a multinomial distribution, with probability mass function given by:

$$f\left(m_1, \ldots, m_{n_S} \middle| m_T, p\right) = \frac{m_T!}{m_0! \, m_1! \, m_2! \cdots m_n!} \sum_{k=0}^{n_S} \left[\binom{n_S}{k} p^k (1-p)^{n_S-k}\right]^{m_k} \qquad \text{(S2)}$$

The log-likelihood function for (m_T, p) is then:

$$\mathcal{L}\left(m_1, \ldots, m_{n_S} \middle| m_T, p\right)$$
$$= \log(m_T!) - \sum_{k=0}^{n_S} \log(m_k!) \qquad \text{(S3)}$$
$$+ \sum_{k=0}^{n_S} m_k \left[\log\binom{n_S}{k} + k \log p + (n_S - k) \log(1-p)\right]$$

Using the fact that $m_0 = m_T - m_S$, the partial derivatives of the log-likelihood function with respect to m_T and p are:

$$\frac{\partial \mathcal{L}}{\partial m_T} = n_S \log(1-p) + \sum_{k=0}^{m_S-1} \frac{1}{m_T - k} \qquad \text{(S4)}$$

$$\frac{\partial \mathcal{L}}{\partial p} = \frac{1}{p(1-p)}\left(-p m_T n_S + \sum_{k=1}^{n_S} k m_k\right) \qquad \text{(S5)}$$

Setting both derivatives equal to zero then yields:

$$n_S \log(1-p) + \sum_{k=0}^{m_S-1} \frac{1}{m_T - k} = 0 \qquad \text{(S6)}$$

$$p m_T = \frac{1}{n_S} \sum_{k=1}^{n_S} k m_k \qquad \text{(S7)}$$

These equations can be solved numerically to obtain maximum likelihood estimates $\hat{m}_T$ and $\hat{p}$ given the sample values $m_1, \ldots, m_{n_S}$.

We now derive approximate standard errors for these maximum likelihood estimates based on the asymptotic variance-covariance matrix. The Hessian matrix for (m_T, p) is given by:

$$H = \begin{pmatrix} \dfrac{\partial^2 \mathcal{L}}{\partial m_T^2} & \dfrac{\partial^2 \mathcal{L}}{\partial m_T \partial p} \\[2ex] \dfrac{\partial^2 \mathcal{L}}{\partial m_T \partial p} & \dfrac{\partial^2 \mathcal{L}}{\partial p^2} \end{pmatrix}$$

$$= -\begin{pmatrix} \displaystyle\sum_{k=0}^{m_S-1} \dfrac{1}{(m_T-k)^2} & \dfrac{n_S}{1-p} \\[3ex] \dfrac{n_S}{1-p} & \dfrac{1}{(1-p)^2}\left(m_T n_S + \dfrac{1-2p}{p^2} \displaystyle\sum_{k=1}^{n_S} k m_k \right) \end{pmatrix} \tag{S8}$$

We wish to find the expected value of the Hessian matrix. For the first entry H_{11}, we note that the probability that a particular male cuckolds at least one of the sampled territories is $q = (1-(1-p)^{n_S})$. The total number of sampled cuckolders M_S thus follows a binomial distribution $\text{Bin}(m_T, q)$. Hence, we can write:

$$\mathbb{E}H_{11} = -\sum_{i=1}^{m_T} \binom{m_T}{i} q^i (1-q)^{m_T-i} \sum_{k=0}^{i-1} \frac{1}{(m_T-k)^2} \tag{S9}$$

For the last entry H_{22}, we note that $\frac{1}{n_S}\sum_{k=1}^{n_S} k m_k$ is just the average number of cuckolders per territory. Hence, we must have $\mathbb{E}\frac{1}{n_S}\sum_{k=1}^{n_S} k m_k = p m_T$ and so $\mathbb{E}\sum_{k=1}^{n_S} k m_k = p m_T n_S$. This means that:

$$\mathbb{E}H_{22} = -\frac{m_T n_S}{p(1-p)} \tag{S10}$$

The expectations of the non-diagonal elements are straightforwardly given by:

$$\mathbb{E}H_{12} = \mathbb{E}H_{21} = -\frac{n_S}{1-p} \tag{S11}$$

We now have everything needed to calculate the Fisher information matrix $\mathcal{I}(m_T, p) = -\mathbb{E}(H)$. The asymptotic variance-covariance matrix of the maximum likelihood estimates is then given by $A = \mathcal{I}(\hat{m}_T, \hat{p})^{-1}$. We approximated the variance in the estimate $\hat{p} n_T \bar{f}$ of the average number of fry per unpaired male as (see Results in main text):

$$\text{var}(\hat{p}n_T\bar{f}) = n_T^2\left(\hat{p}^2\text{var}(\bar{f}) + \bar{f}^2\text{var}(\hat{p}) + \text{var}(\bar{f})\text{var}(\hat{p})\right) \qquad \text{(S12)}$$

This assumes that the total number of territories n_T in the quadrat is measured with negligible error and that $\hat{p}$ is independent of $\bar{f}$.