

Extra-pair paternity in birds: review of the genetic benefits

Erol Akçay* and Joan Roughgarden

*Department of Biological Sciences, Stanford University,
371 Serra Mall, Stanford, CA 94305, USA*

ABSTRACT

Question: How well are genetic benefits hypotheses for extra-pair paternity supported by empirical evidence?

Data incorporated: Almost all published studies testing for genetic benefits from 1980 onwards (121 papers, 55 species).

Analysis methods: Collected key features and findings of each study in a database. Determined overall level of support for both good genes and compatible genes hypotheses. Conducted a formal meta-analysis on a subset of studies asking the following questions: (1) Do extra-pair mates of females have different phenotypes than their within-pair mates? (2) Do extra-pair offspring differ in viability from within-pair offspring? (3) Is there a correlation between the genetic similarity of a social pair and the incidence of extra-pair paternity?

Results: Both the good genes and compatible genes hypotheses failed to be supported in more than half of the species studied. The meta-analysis shows that extra-pair males are on average larger and older than within-pair males, but not different in terms of secondary sexual traits, condition or relatedness to the female. No difference was found between extra-pair and within-pair young in survival to the next breeding season. We found no significant correlation between pair genetic similarity and extra-pair paternity.

Conclusions: Genetic benefits are not strongly supported by available empirical data. New hypotheses are needed.

Keywords: compatible genes, direct benefits, extra-pair paternity, good genes, meta-analysis.

BACKGROUND

Most birds are socially monogamous, yet most are not completely so genetically (Griffith *et al.*, 2002). This realization, brought about by molecular methods utilized in the last two decades, has spurred great interest in elucidating the evolution and maintenance of extra-pair reproductive activity in birds. Initially, interest was focused on whether and why males do not or cannot guard their mates to prevent extra-pair copulations, generating a variety of predictions on the relation of breeding demographics with extra-pair paternity (Griffith *et al.*,

* Author to whom all correspondence should be addressed. e-mail: erol@stanford.edu
Consult the copyright statement on the inside front cover for non-commercial copying policies.

2002). It has since been realized that females may be active players in creating the observed pattern by choosing whether to have extra-pair copulations with different males. This realization focused the research effort on what females might be gaining from extra-pair copulations (Petrie and Kempenaers, 1998; Westneat and Stewart, 2003). The most prominent hypotheses postulate an indirect, genetic benefit for the female's offspring from having an extra-pair sire. These hypotheses observe that interactions concerning ecological benefits mostly occur within the social pair bond and contend that extra-pair copulations are unlikely to result in substantial gains or losses in such. This leads to the conclusion that the gains, if any, from extra-pair paternity are likely to be genetic in nature. Two types of genetic benefits are commonly considered: one that is additive and one that is non-additive, referred to as the 'good genes hypothesis' and 'compatible genes hypothesis', respectively (Neff and Pitcher, 2005).

The genetic benefits perspective has dominated the research effort on extra-pair paternity for almost two decades, but recently it has been challenged as empirically unsupported and inadequate. In their review of extra-pair copulation behaviour, Westneat and Stewart (2003) argued that the behavioural evidence for female pursuit of extra-pair copulations in many species is weak, and that such copulations might be the result of mainly male pursuit. More recently, Arnqvist and Rowe (2005) collated data from six species and estimated that the magnitude of positive indirect selection acting through offspring is close to zero and non-significant. They also pooled data from 12 other species and found that extra-pair paternity is correlated with loss of paternal care in those species, which might cause a significant selection pressure against extra-pair copulation behaviour. They concluded therefore that females are unlikely to benefit from having extra-pair copulations, and that such copulations should instead be regarded as a case of sexual conflict, where male strategy is actually harming the females. However, Griffith (2007) objected to this conclusion, and argued that there might be other benefits of extra-pair copulations to females, mostly focusing on protection against infertility of the social mate. Meanwhile, papers on genetic benefits hypotheses keep appearing in increasing numbers.

We aim to contribute to the debate by compiling and presenting an exhaustive documentation of two decades of published research. Such a comprehensive compilation will allow researchers to assess objectively the published evidence when designing their studies and interpreting their results. We also conduct a meta-analysis to evaluate three important predictions of the genetic benefits hypotheses:

- extra-pair males should differ from within-pair males in a trait reflecting genetic quality or in relatedness to the females;
- extra-pair young should survive better than within-pair young; and
- the occurrence of extra-pair paternity should be positively correlated to the relatedness of a social pair.

To our knowledge, the first and third of these predictions have never been evaluated before using meta-analysis. The meta-analysis by Arnqvist and Rowe (2005) had addressed the second item in this list; however, our analysis differs from theirs in that we use direct measurements of survival and recruitment. In the final section, we discuss possible conclusions that can be drawn from our results. We end the paper with a call for alternative hypotheses, and suggest that more research should be focused on interactions between involved parties.

Summary of the genetic benefits hypotheses

According to the good genes hypothesis (Jennions and Petrie, 2000; Neff and Pitcher, 2005), there exists additive genetic variation in fitness among males. In other words, there is a genetic hierarchy with some males having 'better' genetic quality than others, which results in higher viability or attractiveness of those males. The good genes hypothesis states that males will be selected to advertise their genetic quality and females to prefer copulating with males carrying good genes. In most bird species, however, not all females can pair with the most preferred males because of the constraint of social monogamy. Consequently, females paired to non-preferred males might try to 'trade up' – that is, have copulations with a more attractive male – to obtain better genes for her offspring. In this way, females might secure both the direct benefits of parental care and the genetic benefits of the 'good genes' at the same time. In the conventional interpretation of the good genes hypothesis, males advertise their genetic 'quality' by some signalling mechanisms, which might be behavioural (such as song or courtship) or morphological (such as bright feathers or a larger body). Consequently, studies on the good genes hypothesis often look for relationships between within- and/or extra-pair paternity and male behavioural and/or morphological traits. A more precise test is to compare extra-pair sires with the social mate of the females they sired offspring with. Finally, the ultimate test for the good genes hypothesis is that extra-pair young should have higher fitness than their maternal half-sibs, the within-pair young.

An alternative to the good genes hypothesis is that not all females choose the same genes, but rather, each of them prefers different ones. Such differences in preference can be due to genetic incompatibility that might arise from sexually antagonistic co-evolution (Zeh and Zeh, 1996) or due to high relatedness between paired individuals (Blomqvist *et al.*, 2002). In this scenario, genetic benefits to females are due to the interaction between maternal and paternal genomic contributions. Females are predicted to pursue extra-pair copulations to increase the chances of securing a paternal contribution that is more compatible with their own contribution (Zeh and Zeh, 1997). In practice, most studies of the compatible genes hypothesis use relatedness or genetic similarity between partners as the test variable. Empirical tests of the compatible genes hypothesis are similar to those of the good genes hypothesis, with one important difference when comparing extra-pair young and within-pair young: under the compatible genes hypothesis, in addition to having higher fitness than their maternal half-sibs, extra-pair young can also have higher fitness than their paternal half-sibs (e.g. Johnsen *et al.*, 2000).

THE DATABASE

We have compiled a database of studies that tested either the good genes hypothesis or compatible genes hypothesis. We searched three databases (Biosis, ISI Science Citation Index, and ISI Social Science Citation Index) using multiple keywords related to extra-pair paternity and genetic quality in August 2006 and March 2007. We included years from 1980 to 2007 and selected studies that satisfied the following criteria: (1) the study had to report molecular exclusion and/or assignment of paternity; (2) the rate of extra-pair paternity needed to be non-zero; and (3) the study had to present at least one test of the genetic benefits hypotheses. From studies satisfying these conditions, we eliminated several that tested the good genes hypothesis using offspring sex-ratio (Trivers and Willard, 1973), since it is unlikely that this mechanism is functioning in the context of extra-pair paternity.

We supplemented the database with studies satisfying the above criteria that were cited in previous reviews (Møller and Ninni, 1998; Griffith *et al.*, 2002). In total, 121 studies were included in the resulting database (the full database can be found in the Online Appendix at <http://evolutionary-ecology.com/data/2203appendix.pdf>).

For each of the studies, the following data were extracted: name of the species, percentage of broods with extra-pair paternity, percentage of extra-pair paternity among all offspring, whether good genes effects were found, how good genes effects were inferred, whether compatible genes effects were found, and how compatible genes effects were inferred. We also noted whether any empirical evidence for direct benefits or a behavioural interaction between the extra-pair male and extra-pair young were reported. Reports of direct benefits were rare (e.g. Gray, 1997), while behavioural data on the interaction of the extra-pair mate with the extra-pair young is virtually non-existent.

The overall pattern

The 121 studies contained in the database were conducted on 55 different species. Some species and genera were more heavily represented in the database than others, reflecting the fact that researchers tend to focus on particular organisms. Altogether, 106 studies, conducted on 51 species, contained a test of the good genes hypothesis. To decide whether the good genes hypothesis was supported in a given species, we took into account all studies available in the database. For 32 species, the good genes hypothesis was tested with only one paper, whereas for 19 species, more than one study was available. The overall pattern is summarized in Table 1. The compatible genes hypothesis, on the other hand, was tested in 34 of the studies in the database, representing 24 species. The results for those studies are summarized in Table 2.

The most straightforward test of the good genes hypothesis is to compare the paternity rate of males within their own broods and look for correlations with a male trait that might

Table 1. Number of studies and species where the good genes hypothesis (GGH) was supported or not

	Supports good genes hypothesis (%)	Doesn't support good genes hypothesis (%)
Papers	45 (42%)	61 (58%)
Species	22 (43%)	29 (57%)

Table 2. Number of studies and species where the compatible genes hypothesis (CGH) was supported or not (for three species, the results were mixed)

	Supports compatible genes hypothesis (%)	Doesn't support compatible genes hypothesis (%)
Papers	15 (44%)	19 (56%)
Species	11 (44%)	11 (44%)

indicate genetic quality. Low-quality males are expected to have lower within-pair paternity than high-quality males, and males gaining extra-pair paternity should be of higher quality than males not gaining extra-pair paternity. Tests of the first prediction mirror the overall pattern in the database: of 75 studies testing this prediction, 30 (40%) support it, while 45 (60%) find either no significant relationship or a significant relationship in the reverse direction. For the second prediction, we found that of 23 studies, 12 (52%) found a significant relationship between male trait and extra-pair paternity gained, while 11 (48%) found no significant positive relationship. These results suggest that the prediction of the good genes hypothesis that, in general, males losing paternity should be of lower quality and males gaining paternity should be of higher quality, does not hold true in more than half of the cases.

A more precise prediction from the good genes hypothesis is that female choice for extra-pair mates will, on average, favour males of higher quality than their social (within-pair) partners. This would still be true if the female was paired to a higher than average male, but there were males of even higher quality nearby. Similarly, males of lower than average quality might gain extra-pair copulations from neighbouring females who are mated to even lower quality males. In both of these cases, the correlation between paternity and male traits might be masked by such sampling patterns, but extra-pair males would still be, on average, of higher quality than within-pair males. Of 38 studies testing this prediction, 19 (50%) found that extra-pair males were different than within-pair males in some trait that might reflect genetic quality, and in 19 (50%) no such difference was found.

The good genes hypothesis predicts that extra-pair young will have higher fitness than within-pair young either because of increased viability or increased mating success. This allows for direct testing of the good genes hypothesis by measuring offspring fitness or its components. In practice, almost all studies that we identified use viability, or some trait that may be correlated with it, due to the difficulties of measuring total fitness in the wild. Only one study (Schmoll *et al.*, 2005) in the database compared the number of grandchildren produced to a female via her extra-pair young and within-pair young, a much more complete measure of fitness. This study produced equivocal results, with within-pair young having higher fitness among eggs hatching earlier in the season, and extra-pair young having higher fitness among later hatching eggs. Overall, the results of 23 studies testing for differences between extra-pair young and within-pair young in fitness-related traits suggest that such differences are rarely found: only 6 studies (26%) found that extra-pair young were different than within-pair young, with 17 (74%) reporting either non-significant or inconsistent differences. Thus, direct evidence for indirect viability benefits to offspring is very scarce.

META-ANALYSIS

The summary presented above provides a useful overview to the field, indicating that empirical support for genetic benefits is equivocal overall. We have also conducted a formal meta-analysis on a subset of studies to get a more quantitative assessment of the published empirical results. First, we focused on the differences between extra-pair and within-pair males in various trait categories, because this prediction of the good genes hypothesis has yet to be subjected to meta-analysis (for meta-analyses dealing with the other two predictions, see Møller and Ninni, 1998; Arnqvist and Rowe, 2005). Moreover, this measure is expected to reflect female choice more closely, making it a more precise test of genetic benefits hypotheses than simply correlating paternity to male traits. Second, we asked whether there

is any relationship between the relatedness of paired individuals to extra-pair paternity, providing the first evaluation of the compatible genes hypothesis with meta-analytic methods. Lastly, in the most direct test of the genetic benefits hypotheses, we analysed whether extra-pair young and within-pair young were different from each other in terms of survival.

Meta-analytic methods

We have extracted effect sizes and sample sizes from 36 studies included in the database. Many studies reported comparisons between within-pair and extra-pair males in multiple traits, and we have included all effect sizes. We transformed all effect sizes into correlation coefficients r , and then into Fisher's z , using transformations by Rosenthal (1994) (tables with effect sizes can be found in the Online Appendix at <http://evolutionary-ecology.com/data/2203appendix.pdf>). Most tests for the extra-pair male/within-pair male and extra-pair young/within-pair young comparisons were paired t -tests, in which case the formulae given by Rosenthal (1994) can lead to overestimation of effect size (Dunlap *et al.*, 1996). This should not be a problem if the true effect size is around zero, but can have an effect with true effect sizes different than zero. Therefore, the magnitude of our effect sizes should be interpreted as upper bounds. We categorized traits into four: size, age, condition, and secondary sexual traits. We analysed each trait group separately. When more than one trait was measured for each trait category (e.g. both tarsus and wing sizes for size), we combined the effect sizes by two methods. If one of the comparisons had a significantly larger effect size than others, we found the P level for that comparison and multiplied it with the number of comparisons that were made within the trait group. If the effect sizes were of similar magnitude, we took the mean of z , weighted by the sample size for each test, in which case the sample size for the study was also averaged. We used studies rather than species as the analysis unit, since the number of studies for each trait was not very large and there were few cases where more than one study on the same trait group of the same species was included. Some studies divided data into two groups and reported results separately for these (e.g. Foerster *et al.*, 2003). In these cases, we treated these subsamples as two studies. To estimate the overall effect size, we employed both fixed and random effects models (Shadish and Haddock, 1994). We report the results from the random effects analysis whenever there is significant heterogeneity between studies. The overall effect size for fixed effects models is computed as

$$\bar{z} = \frac{\sum_j (n_j - 3) z_j}{\sum_j (n_j - 3)}. \quad (1)$$

For random effect models, the mean effect is calculated as

$$\bar{z} = \frac{\sum_j w_j z_j}{\sum_j w_j}, \quad (2)$$

where $w_j = 1/[v^* + 1/(n_j - 3)]$ is the weighting factor, and v^* is the estimator for between-study variance, calculated according to equation 18.23 in Shadish and Haddock (1994). For fixed and random effect sizes, we calculated the 95% confidence intervals using equations given by Shadish and Haddock (1994).

In meta-analyses that calculate mean effect sizes, publication bias against non-significant or significant negative results is a potential problem that can affect the overall conclusions.

To test and correct for publication bias, we employed the ‘trim and fill’ method (Duval and Tweedie, 2000a; Jennions and Møller, 2002). This method formalizes the use of the funnel graph (i.e. a plot of effect size versus within-study variance) by trying to infer how many studies are missing and then accounting for them by filling in mirror images of existing studies with positive results. Statistically, there are three estimators of the number of missing studies, of which two are recommended by Duval and Tweedie (2000a) (L_0^+ and R_0^+). Following Duval and Tweedie (2000b), we have used both R_0^+ and L_0^+ in each case to judge the sensitivity of our estimate of k_0 .

Results from the meta-analysis

For differences between extra-pair and within-pair males, we have compiled effect sizes from a total of 45 studies. The mean effect sizes in different trait groups are summarized in Table 3. Mean effect sizes were significantly different than zero for secondary sexual traits, size, and age (Table 3, Fig. 1; but see below for corrections for publication bias). For condition and relatedness to females, the mean effect size was not statistically significant. In all trait groups except relatedness to females, there was a significant heterogeneity across the sample, suggesting that true effect sizes vary between studies and, presumably, species.

To evaluate how robust these results are against publication bias, we employed the trim and fill method (Duval and Tweedie, 2000a). Results are summarized in Table 4 and Fig. 2. We found that an indication of significant publication bias for secondary sexual characteristics [number of missing studies, $k_0 = 6$, $P \approx 0.025$ (see table 4 in Duval and Tweedie, 2000b)]. Accounting

Table 3. Mean effect sizes and 95% confidence intervals for comparisons between extra-pair males and within-pair males, as correlation coefficients

	Mean effect size	CI	Q (significance level)	N
Secondary sexual	0.2947	(0.0952, 0.4715)	69.65 ($P = 10^{-7}$)	20
Size	0.1274	(0.0147, 0.2369)	35.40 ($P = 0.035$)	23
Condition	0.1104	(-0.0484, 0.2637)	36.48 ($P = 0.002$)	17
Age	0.3147	(0.1758, 0.4536)	27.60 ($P = 0.016$)	15
Relatedness	0.0324	(-0.0487, 0.1131)	11.86 ($P = 0.53$)	14

Note: The statistic Q and the associated significance P measure heterogeneity of effect sizes in the sample. N is the number of studies included in the sample. For relatedness to females, effect sizes are from the fixed effects model; all other effect sizes are from random effect models.

Table 4. Mean effect sizes and 95% confidence intervals for extra-pair male/within-pair male comparisons after correcting for publication bias using the trim and fill method

	Mean effect size	CI	Q (significance level)
Secondary sexual	0.0788	(-0.1457, 0.2955)	123.10 ($P = 0$)
Size	0.1328	(0.0157, 0.2463)	39.01 ($P = 0.018$)
Condition	0.1104	(-0.0484, 0.2637)	36.48 ($P = 0.002$)
Age	0.2559	(0.1105, 0.3907)	39.35 ($P = 0.001$)
Relatedness	0.0118	(-0.0673, 0.0908)	17.45 ($P = 0.292$)

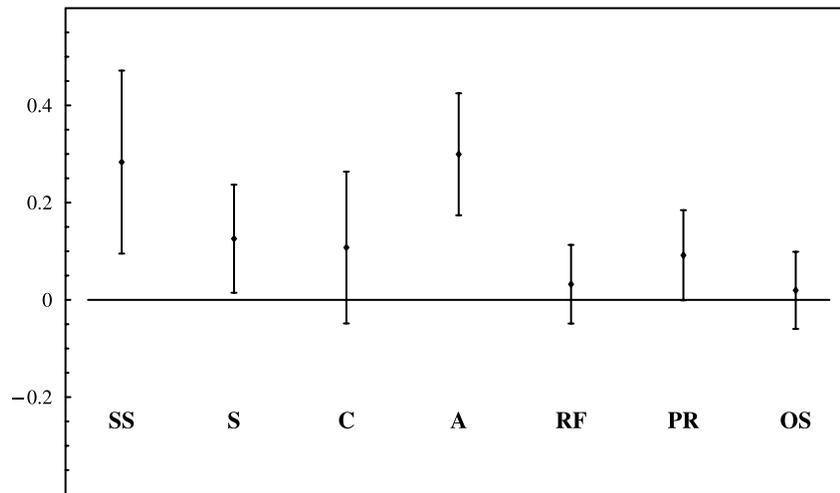


Fig. 1. Mean effect sizes and 95% confidence intervals before correction for publication bias. The first five bars compare extra-pair males to within-pair males (SS = secondary sexual traits, S = size, C = condition, A = age, RF = relatedness to the female). PR stands for the relation of pair relatedness to the occurrence or level of extra-pair paternity and OS (offspring survival) for differences in survival between extra-pair young and within-pair young.

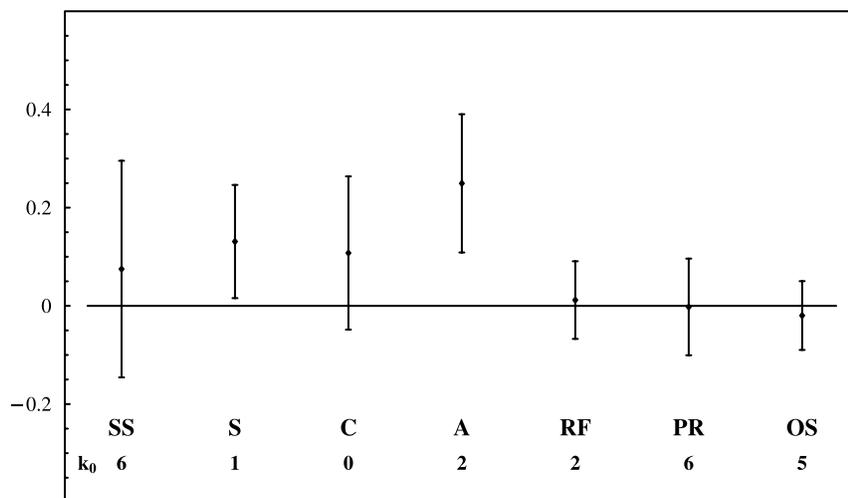


Fig. 2. Effect sizes and confidence intervals after correcting for publication bias with the trim and fill method (Duval and Tweedie, 2000a). Abbreviations as in Fig. 1; the estimated number of missing studies, k_0 , is given below each bar.

for this bias resulted in a qualitative difference in the results: the mean effect size was not different than zero anymore. Therefore, we cannot conclude that extra-pair males are different than within-pair males in secondary sexual traits. As a further test of publication bias, we checked the existence of a correlation between effect and sample sizes in the samples (Palmer, 1999). For secondary sexual traits, the correlation was not significant with the

complete sample, but did become so when two outliers with very large effect sizes (one negative, one positive) were removed ($r = -0.499$, one-tailed $P = 0.017$). Interestingly, both of these outliers (Hasselquist *et al.*, 1996; Marshall *et al.*, 2007) compared the same trait (repertoire size) in two species within the same genus (*Acrocephalus*). We conducted the meta-analysis again, with these outliers excluded, which resulted in a considerably smaller mean effect size before correction that also became non-significant after correcting for bias (not shown).

The most widely considered version of the compatible genes hypothesis predicts a correlation between incidence of extra-pair paternity and the relatedness of social mates. To evaluate this prediction, we compiled effect sizes from 23 studies. Effect sizes were highly heterogeneous in the sample ($Q = 86.99$, $P = 0$). The estimate for mean effect size using random effects models bordered on being significantly different than zero ($\bar{r}_{\text{relatedness}} = 0.0926$, 95% CI = -0.0009 , 0.1845 ; Fig. 1). There was a significant indication of publication bias [$k_0 = 6$, $P \approx 0.03$ (from table 4 in Duval and Tweedie, 2000b)]. When adjusted for bias using trim and fill, the effect size changed sign and became decidedly non-significant. The correlation between sample and effect sizes was nearly significant ($r = -0.3515$, one-tailed $P = 0.05$), and became highly significant when one isolated outlier with large negative effect size (Rätti *et al.*, 1995) was excluded ($r = -0.5408$, one-tailed $P = 0.005$). When the analysis was repeated with this outlier, the uncorrected effect size was positive and significantly different than zero, but the significance disappeared when publication bias was accounted for (not shown).

Lastly, we compare how well extra-pair young survive to the next breeding season relative to within-pair young. For that, we collate data from 10 studies, on six species. For this sample of studies, the hypothesis of homogeneity could not be rejected ($Q = 2.02$, $P = 0.996$), so we estimate the mean effect size using a fixed effects model. The average effect size was small, and not significantly different than zero ($\bar{r}_{\text{survival}} = 0.0197$, 95% CI = -0.0597 , 0.0989 ; Fig. 1). After accounting for publication bias, the mean effect size was negative, but was still not significantly different than zero (Fig. 2).

DISCUSSION

What conclusions can be drawn from the patterns documented? We can answer this question in a variety of ways. Below, we present some alternatives ranging from conservative interpretations to stronger inferences, first focusing on the good genes hypothesis and then the compatible genes hypothesis.

Do 'good genes' drive extra-pair paternity?

Perhaps the most conservative conclusion from our review would be that no conclusion can be made. One can argue that studies compiled here 'may not have been able to test the appropriate hypotheses adequately' (Anonymous). Specifically, genetic benefits need to be considered in a multi-player interaction scheme (Westneat and Stewart, 2003), where ecological constraints could restrict extra-pair copulation behaviour [e.g. the 'constrained female hypothesis' (Gowaty, 1996)]. This can mask the patterns of extra-pair paternity expected from the good genes hypothesis alone. However, if the good genes hypothesis still holds, its predictions about differences between extra-pair and within-pair males and extra-pair young and within-pair young should remain unaffected. This is because these comparisons are made among the females that did have extra-pair copulations despite constraints, which

the good genes hypothesis predicts should result in genetic benefits. Nonetheless, this argument is correct in pointing out that ecological constraints are important and remain largely understudied, regardless of the truth of the good genes hypothesis.

Another position is to ‘view the glass as half full’, and claim that the compiled data actually support the good genes hypothesis. If we had 100 studies that test an hypothesis at the 0.05 significance level, and the good genes hypothesis was not true, we would expect only 5 of them to show significant results. Therefore, the presence of 42% of papers showing significant effects might be interpreted as actually supporting the good genes hypothesis, or at least not rejecting it. A factor that would weaken this argument is that each study usually contains many tests of the good genes hypothesis without necessarily applying the relevant statistical corrections. Thus, acceptance of the hypothesis is usually more likely than the 0.05 significance level. Furthermore, it is common practice to ‘mine’ or partition data to find a significant correlation between paternity patterns and a variable that might be of relevance (e.g. Forstmeier *et al.*, 2002; Delhey *et al.*, 2003). While such practices do not necessarily affect the validity of individual studies or their conclusions, they would increase the frequency of studies that by chance alone support the good genes hypothesis, even under the null hypothesis that it is not true. The quantitative meta-analysis we present here should address such concerns. Its results lend only partial support to the good genes hypothesis (see below).

One can argue that the good genes hypothesis need not explain extra-pair paternity in all species, but may be valid only in a subset of species where it is supported. To be complete and testable, this argument needs a ‘meta-theory’ that specifies when to expect the good genes hypothesis to be important and when not. We are unaware of any such proposal, save one that predicts the good genes hypothesis should be more important when additive genetic variation is higher (Petrie *et al.*, 1998). Nonetheless, the idea of a meta-theory predicting different selection pressures in different taxa and contexts is an intuitively appealing one, again regardless of the truth of the good genes hypothesis.

Finally, one can conclude that genetic benefits are not supported with available evidence. This is the position we favour, for the following reasons. In more than half of the species, the good genes hypothesis has not been demonstrated. This is clearly not a result of smaller sample sizes or fewer studies. Moreover, in species where there is some evidence for the good genes hypothesis, different populations and closely related species frequently report evidence against it. Blue tits, *Parus caeruleus*, are a case in point. The relationship between paternity and male survival or strophe length that was found in one population (Kempnaers *et al.*, 1992, 1997) is not found in others (Krokene *et al.*, 1998; Charmantier *et al.*, 2004; Poesel *et al.*, 2006). Contradictory relations between paternity loss and gain and male ornamentation have been documented in the same population (Delhey *et al.*, 2003, 2007). Finally, no evidence in support of the good genes hypothesis was found in closely related species, the great tit [*Parus major* (Krokene *et al.*, 1998; Lubjuhn *et al.*, 1999)] and the coal tit [*Parus ater* (Dietrich *et al.*, 2004; Schmoll *et al.*, 2003, 2005)], despite comparable effort and similar or higher rates of extra-pair paternity. Such inconsistent patterns undermine the strength of support even in species which are listed as supporting the good genes hypothesis.

Further evidence against the good genes hypothesis comes from our meta-analysis on two of its important predictions. The first prediction, that extra-pair males should be of higher ‘quality’ than within-pair males, seems to be only partially true. Specifically, we find that extra-pair males on average tend to be larger and older than within-pair males, but not different in terms of expression of secondary sexual traits or body condition. The absence

of a difference in secondary sexual characteristics and phenotypic condition (which includes survival) is especially troubling for the good genes hypothesis: conventional wisdom regards secondary sexual traits as indicators of phenotypic condition, which is thought to stem from genetic quality (Hasselquist *et al.*, 1996; Forstmeier *et al.*, 2002). It is also interesting that we found significant publication bias in the test for secondary sexual traits. A possible explanation lies in the fact that the comparison between extra-pair and within-pair males is secondary to correlations between within-pair paternity and male traits, since additional analysis usually is required to assign extra-pair young to sires. It might be that such analysis is less likely to be conducted when correlations between paternity and secondary sexual traits is small or absent.

On the other hand, a difference in size and age can be taken to indicate that females might be gaining indirect benefits for their offspring by having extra-pair copulations. Specifically, if there is a correlation between size and fitness, and size is heritable, that might mean that extra-pair young have higher fitness than within-pair young. The magnitude of this difference can be estimated from our results as $h^2 \times r_{\text{fitness}} \times \bar{r}_{\text{size}}$, where h^2 is the heritability of size and r_{fitness} is the genetic correlation between size and fitness. Assuming $h^2 = 0.6$ and $r_{\text{fitness}} = 0.3$, both conservatively high figures (Qvarnström *et al.*, 2006), we arrive at 0.0235 as the estimated effect size for the difference in fitness between extra-pair young and within-pair young due to the difference in size of their sires. This figure is admittedly small, which means that indirect selection on extra-pair copulation behaviour through this effect will be weak at best. For age, the effect size is higher, but of course age itself is not heritable. There is an unresolved theoretical question whether age can indicate genetic quality (Kokko and Lindström, 1996; Beck and Powell 2000). Most empirical support for age as an indicator of quality comes from the extra-pair paternity literature (Brooks and Kemp, 2001), but that interpretation depends on acceptance of the good genes hypothesis as the main cause of extra-pair paternity, which would make the argument circular. For both size and age, alternative mechanisms might be creating these patterns, such as sexual conflict (Arnqvist and Rowe, 2005) or direct benefits (E. Akçay and J. Roughgarden, in preparation).

Finally, the absence of a difference between extra-pair young and within-pair young in survival provides the strongest evidence against the good genes hypothesis. Most studies of good genes deal with genetic viability benefits to the offspring. The studies we collated demonstrate that empirical evidence does not support this notion. This finding concurs with a previous meta-analysis by Arnqvist and Rowe (2005), but unlike Arnqvist and Rowe (2005), we use direct measurements of offspring viability. The good genes effect can also operate through increased attractiveness, rather than viability of offspring, but this possibility has not been tested adequately. The only study accounting for attractiveness as well as viability found no significant differences between number of grandchildren produced to the female through extra-pair young and within-pair young (Schmoll *et al.*, 2005).

Do females seek compatible genes?

The compatible genes hypothesis is a relative newcomer to the extra-pair paternity literature, but several studies have now accumulated. Our database shows that the overall pattern is not different than that for the good genes hypothesis, therefore the same arguments as the good genes hypothesis also apply to the compatible genes hypothesis. We also present here the first meta-analysis to our knowledge that evaluates the compatible genes hypothesis. We first compared extra-pair males with within-pair males in their relatedness to the female

they sired offspring with. Contrary to the compatible genes hypothesis, our combined estimate was not significantly different than zero. We also compiled correlation coefficients between genetic similarity and the occurrence of extra-pair paternity. This test also resulted in non-significant effect sizes, also contrary to the compatible genes hypothesis. Thus, despite some recent prominent reports (Blomqvist *et al.*, 2002), genetic similarity between socially paired individuals seems not to play any role in the occurrence of extra-pair paternity. Avoidance of inbreeding depression is not the only possible mechanism by which the compatible genes hypothesis might work. Other proposals include that females simply try to increase genetic variation with brood, as a 'bet-hedging' strategy. This predicts that extra-pair young should occur randomly with respect to broods, which is almost always rejected by the data. Additional alternatives such as avoiding genetic incompatibility that might arise from sexually antagonistic evolution (Zeh and Zeh, 1996) have been suggested, but to our knowledge they have not been tested in the context of extra-pair paternity so far.

Concluding remarks

We present the most comprehensive compilation to date of published evidence testing the genetic benefits hypotheses. Our aims were two-fold. First, we wanted to compile all available evidence to see the overall pattern. Too often, trends and accepted hypotheses in behavioural ecology are determined by the momentum of a few high-profile studies, even if their results are shown later not to be generalizable (e.g. Kempenaers *et al.*, 1992; Hasselquist *et al.*, 1996; Blomqvist *et al.*, 2002). Researchers working on the immediate question are often aware of such caveats, but workers asking different questions are not, and adopt such hypotheses as true for their reasoning. [An example is the discussion about whether age is an indicator of genetic quality (see above).] By documenting the overall pattern, we hope to help researchers to assess more accurately the state of the evidence when designing their studies and interpreting their results.

Our second aim was to contribute to the recently started discussion on the actual truth of genetic benefits hypotheses. Like previous reviews, we find that the evidence challenges genetic benefits status as the widely accepted cause of extra-pair paternity. Alternative hypotheses are needed. Like Westneat and Stewart (2003), we believe that more attention should be focused on the interactions between parties involved in the extra-pair paternity phenomenon. Unlike them, however, we believe that such interactions can also be of a cooperative nature, and involve exchange of direct benefits. We outline our reasoning for this in a companion paper that also presents a model of extra-pair paternity through agreements that stabilize a social offspring-rearing system with economic monogamy (Akçay and Roughgarden, in preparation). In any case, an approach focusing on interactions (involving either conflict or cooperation) seems to be the most promising direction to further our understanding of the phenomenon of extra-pair paternity.

REFERENCES

- Arnqvist, G. and Rowe, L. 2005. *Sexual Conflict*. Princeton, NJ: Princeton University Press.
- Beck, C.W. and Powell, L.A. 2000. Evolution of female mate choice based on male age: are older males better mates? *Evol. Ecol. Res.*, **2**: 107–118.
- Blomqvist, D., Andersson, M., Küpper, C., Cuthill, I.C., Kis, J., Lanctot, R.B. *et al.*, 2002. Genetic similarity between mates and extra-pair parentage in three species of shorebirds. *Nature*, **419**: 613–615.

- Brooks, R. and Kemp, D.J. 2001. Can older males deliver the good genes? *Trends Ecol. Evol.*, **16**: 308–313.
- Charmantier, A., Blondel, J., Perret, P. and Lambrechts, M.M. 2004. Do extra-pair paternities provide genetic benefits for female blue tits *Parus caeruleus*? *J. Avian Biol.*, **35**: 524–532.
- Delhey, K., Johnsen, A., Peters, A., Andersson, S. and Kempenaers, B. 2003. Paternity analysis reveals opposing selection pressures on crown coloration in the blue tit (*Parus caeruleus*). *Proc. R. Soc. Lond. B*, **270**: 2057–2063.
- Delhey, K., Peters, A., Johnsen, A. and Kempenaers, B. 2007. Fertilization success and UV ornamentation in blue tits *Cyanistes caeruleus*: correlational and experimental evidence. *Behav. Ecol.*, **18**: 399–409.
- Dietrich, V., Schmoll, T., Winkel, W., Epplen, J.T. and Lubjuhn, T. 2004. Cuckoldry and recapture probability of adult males are not related in the socially monogamous coal tit (parus ater). *J. Ornithol.*, **145**: 327–333.
- Dunlap, W.P., Cortina, J.M., Vaslow, J.B. and Burke, M.J. 1996. Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, **1**: 170–177.
- Duval, S. and Tweedie, R. 2000a. A nonparametric ‘trim and fill’ method of accounting for publication bias in meta-analysis. *J. Am. Stat. Assoc.*, **95**: 89–98.
- Duval, S. and Tweedie, R. 2000b. Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, **54**: 455–463.
- Foerster, K., Delhey, K., Johnsen, A., Lifjeld, J.T. and Kempenaers, B. 2003. Females increase offspring heterozygosity and fitness through extra-pair matings. *Nature*, **425**: 714–717.
- Forstmeier, W., Kempenaers, B., Meyer, A. and Leisler, B. 2002. A novel song parameter correlates with extra-pair paternity and reflects male longevity. *Proc. R. Soc. Lond. B*, **269**: 1479–1485.
- Gowaty, P.A. 1996. Battles of the sexes and origins of monogamy. In *Partnerships in Birds* (J.M. Black, ed.), pp. 21–52. Oxford: Oxford University Press.
- Gray, E.M. 1997. Female red-winged blackbirds accrue material benefits from copulating with extra-pair males. *Anim. Behav.*, **53**: 625–639.
- Griffith, S.C. 2007. The evolution of infidelity in socially monogamous passerines: neglected components of direct and indirect selection. *Am. Nat.*, **169**: 274–281.
- Griffith, S.C., Owens, I.P.F. and Thuman, K.A. 2002. Extra pair paternity in birds: a review of interspecific variation and adaptive function. *Molec. Ecol.*, **11**: 2195–2212.
- Hasselquist, D., Bensch, S. and von Schantz, T. 1996. Correlation between male song repertoire, extra-pair paternity and offspring survival in the great reed warbler. *Nature*, **381**: 229–232.
- Jennions, M.D. and Møller, A.P. 2002. Publication bias in ecology and evolution: an empirical assessment of the ‘trim and fill’ method. *Biol. Rev.*, **77**: 211–222.
- Jennions, M.D. and Petrie, M. 2000. Why do females mate multiply? A review of the genetic benefits. *Biol. Rev.*, **75**: 21–64.
- Johnsen, A., Andersen, V., Sunding, C. and Lifjeld, J.T. 2000. Female bluethroats enhance offspring immunocompetence through extra-pair copulations. *Nature*, **406**: 296–299.
- Kempenaers, B., Verheyen, G.R., Vandenbroeck, M., Burke, T., Vanbroeckhoven, C. and Dhondt, A.A. 1992. Extra-pair paternity results from female preference for high-quality males in the blue tit. *Nature (Lond.)*, **357**: 494–496.
- Kempenaers, B., Verheyen, G.R. and Dhondt, A.A. 1997. Extrapair paternity in the blue tit *Parus caeruleus*: female choice, male characteristics, and offspring quality. *Behav. Ecol.*, **8**: 481–492.
- Kokko, H. and Lindström, J. 1996. Evolution of female preferences for old males. *Proc. R. Soc. Lond. B*, **263**: 1533–1538.
- Krokene, C., Rigstad, K., Dale, M. and Lifjeld, J.T. 1998. The function of extrapair paternity in blue tits and great tits: good genes or fertility insurance? *Behav. Ecol.*, **9**: 649–656.
- Lubjuhn, T., Strohbach, S., Brun, J., Gerken, T. and Epplen, J.T. 1999. Extra-pair paternity in great tits *Parus major* – a long term study. *Behaviour*, **136**: 1157–1172.

- Marshall, R.C., Buchanan, K.L. and Catchpole, C.K. 2007. Song and female choice for extrapair copulations in the sedge warbler, *Acrocephalus schoenobaenus*. *Anim. Behav.*, **73**: 629–635.
- Møller, A.P. and Ninni, P. 1998. Sperm competition and sexual selection: a meta-analysis of paternity studies of birds. *Behav. Ecol. Sociobiol.*, **43**: 345–358.
- Neff, B.D. and Pitcher, T.E. 2005. Genetic quality and sexual selection: an integrated framework for good genes and compatible genes. *Molec. Ecol.*, **14**: 19–38.
- Palmer, A.R. 1999. Detecting publication bias in meta-analyses: a case study of fluctuating asymmetry and sexual selection. *Am. Nat.*, **154**: 220–233.
- Petrie, M. and Kempenaers, B. 1998. Extra-pair paternity in birds: explaining variation between species and populations. *Trends Ecol. Evol.*, **13**: 52–58.
- Petrie, M., Doums, C. and Møller, A.P. 1998. The degree of extra-pair paternity increases with genetic variability. *Proc. Natl. Acad. Sci. USA*, **95**: 9390–9395.
- Poesel, A., Kunc, H.P., Foerster, K., Johnsen, A. and Kempenaers, B. 2006. Early birds are sexy: male age, dawn song and extrapair paternity in blue tits, *Cyanista* (formerly *Parus*) *caeruleus*. *Anim. Behav.*, **72**: 531–538.
- Qvarnström, A., Brommer, J.E. and Gustafsson, L. 2006. Testing the genetics underlying the co-evolution of mate choice and ornament in the wild. *Nature*, **441**: 84–86.
- Rätti, O., Hovi, M., Lundberg, A., Tegelström, H. and Alatalo, R. 1995. Extra-pair paternity and male characteristics in the pied flycatcher. *Behav. Ecol. Sociobiol.*, **37**: 419–425.
- Rosenthal, R. 1994. Parametric measures of effect size. In *The Handbook of Research Synthesis* (H. Cooper and L.V. Hedges, eds.), pp. 231–244. New York: Russell Sage Foundation.
- Schmoll, T., Dietrich, V., Winkel, W., Epplen, J.T. and Lubjuhn, T. 2003. Long-term fitness consequences of female extra-pair matings in a socially monogamous passerine. *Proc. R. Soc. Lond. B*, **270**: 259–264.
- Schmoll, T., Dietrich, V., Winkel, W., Epplen, J.T., Schurr, F. and Lubjuhn, T. 2005. Paternal genetic effects on offspring fitness are context dependent within the extrapair mating system of a socially monogamous passerine. *Evolution*, **59**: 645–657.
- Shadish, W.R. and Haddock, C.K. 1994. Combining estimates of effect size. In *The Handbook of Research Synthesis* (H. Cooper and L. V. Hedges, eds.), pp. 261–281. New York: Russell Sage Foundation.
- Trivers, R.L. and Willard, D.E. 1973. Natural selection of parental ability to vary the sex ratio of offspring. *Science*, **179**: 90–92.
- Westneat, D.F. and Stewart, I.R.K. 2003. Extra-pair paternity in birds: causes, correlates, and conflict. *Annu. Rev. Ecol. Evol. Syst.*, **34**: 365–396.
- Zeh, J.A. and Zeh, D.W. 1996. The evolution of polyandry I: intragenomic conflict and genetic incompatibility. *Proc. R. Soc. Lond. B*, **263**: 1711–1717.
- Zeh, J.A. and Zeh, D.W. 1997. The evolution of polyandry II: post-copulatory defences against genetic incompatibility. *Proc. R. Soc. Lond. B*, **264**: 69–75.