Pedigree-free animal models: the relatedness matrix reloaded

Francesca D. Frentiu^{1,2,3}, Sonya M. Clegg⁴, John Chittock³, Terry Burke³,
Mark W. Blows¹ and Ian P. F. Owens^{4,5,*}

¹*School of Integrative Biology, University of Queensland, St Lucia, Queensland 4072, Australia*

²*Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92697, USA*

³*Department of Animal and Plant Sciences, University of Sheffield, Sheffield S10 2TN, UK*

⁴*Division of Biology and* ⁵*NERC Centre for Population Biology, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, UK*

Animal models typically require a known genetic pedigree to estimate quantitative genetic parameters. Here we test whether animal models can alternatively be based on estimates of relatedness derived entirely from molecular marker data. Our case study is the morphology of a wild bird population, for which we report estimates of the genetic variance–covariance matrices (**G**) of six morphological traits using three methods: the traditional animal model; a molecular marker-based approach to estimate heritability based on Ritland's pairwise regression method; and a new approach using a molecular genealogy arranged in a relatedness matrix (**R**) to replace the pedigree in an animal model. Using the traditional animal model, we found significant genetic variance for all six traits and positive genetic covariance among traits. The pairwise regression method did not return reliable estimates of quantitative genetic parameters in this population, with estimates of genetic variance and covariance typically being very small or negative. In contrast, we found mixed evidence for the use of the pedigree-free animal model. Similar to the pairwise regression method, the pedigree-free approach performed poorly when the full-rank **R** matrix based on the molecular genealogy was employed. However, performance improved substantially when we reduced the dimensionality of the **R** matrix in order to maximize the signal to noise ratio. Using reduced-rank **R** matrices generated estimates of genetic variance that were much closer to those from the traditional model. Nevertheless, this method was less reliable at estimating covariances, which were often estimated to be negative. Taken together, these results suggest that pedigree-free animal models can recover quantitative genetic information, although the signal remains relatively weak. It remains to be determined whether this problem can be overcome by the use of a more powerful battery of molecular markers and improved methods for reconstructing genealogies.

Keywords: quantitative genetics; animal model; relatedness; molecular markers; morphology; birds

1. INTRODUCTION

The application of animal models to wild populations promises to revolutionize our understanding of evolutionary genetics in natural environments (Kruuk 2004). This is because animal models, in their broadest sense, are simply individual-based mixed models that use a known pedigree to estimate relatedness among individuals and thereby estimate a range of quantitative genetic parameters (Lynch & Walsh 1998). The key reason that animal models offer such promise for the study of wild populations is that this approach can use a natural pedigree to extract quantitative genetic information under natural conditions. In contrast, most quantitative genetic techniques require breeding experiments and are consequently largely restricted to laboratory or

agricultural studies (Falconer & Mackay 1996). Animal models have now been applied to a number of populations to tackle questions as diverse as the heritability of fitness (Kruuk *et al.* 2000), evolutionary stasis (Merilä *et al.* 2001; Kruuk *et al.* 2002), sexual selection and coloration (Hadfield & Owens 2006; Hadfield *et al.* 2006, 2007), condition dependence (Gleeson *et al.* 2005), parental care (MacColl & Hatchwell 2003), the genetic consequences of harvesting (Coltman *et al.* 2003) and the evolutionary response to climate change (Brommer *et al.* 2005).

Such widespread interest in the animal model approach has, however, led to the realization that the need for a known pedigree is itself a limitation. It is no coincidence that most studies to date using the animal model concern populations that have been the subject of long-term projects (Kruuk *et al.* 2000, 2001, 2002; Merilä & Sheldon 2000; Merilä *et al.* 2001; Coltman *et al.* 2003; Garant *et al.* 2004, 2005; McCleery *et al.* 2004; Charmantier *et al.* 2006a,b). The need for long-term information on individual patterns of mating and reproduction limits the range and type of populations where an animal model can be used. One way potentially to overcome this limitation is to use molecular marker data

* Author and address for correspondence: Division of Biology, Imperial College London, Silwood Park, Ascot, Berkshire SL5 7PY, UK (i.owens@imperial.ac.uk).

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rspb.2007.1032> or via <http://journals.royalsociety.org>.

One contribution of 18 to a Special Issue 'Evolutionary dynamics of wild populations'.

to estimate the genetic relationships among individuals in a population and then use the resulting relatedness matrix, instead of a known pedigree, to construct the animal model (Lynch & Walsh 1998; Garant & Kruuk 2005; Rodríguez-Ramilo *et al.* 2007). This approach could allow the animal model framework to be extended to any population for which it was possible to obtain reliable estimates of relatedness based on molecular marker data (Moore & Kruuk 2002), which would greatly expand the range of potential applications if the approach proved to be robust. Such an approach has yet to be fully implemented in any population, however.

The overall aim of this study was therefore to test whether animal models can indeed be based on estimates of relatedness derived entirely from molecular marker data. The idea of estimating quantitative genetic parameters using relatedness estimates derived from molecular marker data has been explored by a number of workers (Mousseau *et al.* 1998; Thomas & Hill 2000; Thomas *et al.* 2000; Thomas 2005) and, in particular, has been developed by Ritland (1996, 2000a,b; Ritland & Ritland 1996). Although Ritland's method is conceptually similar to the pedigree-free animal models that we discuss here, there are key differences between the two. The most important of these is that Ritland's method is based on regressing pairwise estimates of phenotypic similarity on pairwise estimates of genetic relatedness (Ritland 1996). Limitations of this approach include difficulties in estimating significance due to non-independence of relatedness estimates and that method of moments relatedness measures do not provide estimates that are internally consistent across the entire population. In contrast, the pedigree-free animal model approach we explore here is based on a relatedness matrix that is positive definite (i.e. relatedness among multiple individuals are congruous), and allows the full power (and convenience) of the animal model to be applied when estimating quantitative genetic parameters.

We apply the pedigree-free animal model to the example of a free-living population of Capricorn silveryeyes (*Zosterops lateralis chlorocephalus*) on Heron Island, a small coral cay on the Australian Great Barrier Reef. This study population is well suited to our needs because we can construct a known pedigree from behavioural information (Kikkawa & Wilson 1983; Robertson *et al.* 2001), which we have previously used to estimate heritabilities and genetic correlations for a series of morphological traits using a pedigree-based animal model (Frentiu *et al.* 2007). In addition, we have a dataset that consists of individuals used in a cross-fostering experiment (Frentiu *et al.* 2007), minimizing the extent to which the effect of shared genes is inflated by the effect of shared environments. Finally, the quantitative genetic basis of morphology in this population of birds is of intrinsic interest because it is an example of an unusually large island race that shows the characteristic pattern for insular passerines (Clegg & Owens 2002), having evolved to be approximately 40% larger than its mainland counterpart in just 4000 years (Clegg *et al.* 2002a,b; Robinson-Wolrath & Owens 2003; Scott *et al.* 2003; Frentiu *et al.* 2007).

The specific aims of the study were to: (i) develop a series of polymorphic molecular markers and determine whether they were able to differentiate between close relatives (full siblings) and unrelated individuals,

(ii) estimate quantitative genetic parameters for six morphological traits using Ritland's pairwise regression method and compare these with the estimates from a pedigree-based traditional animal model, (iii) develop a pedigree-free animal model using a molecular genealogy, (iv) determine the effectiveness of the pedigree-free versus the traditional animal model in estimating quantitative genetic parameters, and (v) explore methods to increase the power of the molecular genealogy approach.

2. MATERIAL AND METHODS

(a) Study population and morphological measurements

The individuals used in this study ($N=479$) were the same as those described in a previous study of this population (Frentiu *et al.* 2007). Briefly, measurements and samples were taken from nestlings in the Heron Island population of Capricorn silveryeyes over three consecutive breeding seasons: 2000–2001, 2001–2002 and 2003–2004. The modal clutch size in this population is three, so our expectation was that the sample would contain a substantial number of full siblings. The presence of such full siblings was anticipated to be helpful because it would allow us to check the power of our molecular markers in identifying close relatives. As part of a long-term study of evolutionary genetics in this population, we also conducted reciprocal cross-foster manipulations in which one or two chicks were swapped between pairs of nests and this design was used to implement the traditional animal model (see Frentiu *et al.* 2007).

Morphological measurements were taken as described by Frentiu *et al.* (2007). Briefly, six morphological measurements were obtained: wing length; tail length; tarsus length; culmen (bill) length; culmen width; and culmen depth. All measurements were standardized to unit variance and a mean of zero before analysis.

(b) Molecular markers and genotyping

Genomic DNA was extracted from blood samples using a modified salting-out protocol (Nicholls *et al.* 2000). Eleven microsatellite markers were available from previous studies (Degan *et al.* 1999; Frentiu *et al.* 2003) or obtained by screening other passerine primers for amplification in the silveryeye (for details see Frentiu 2004). Loci were tested for deviations from Hardy–Weinberg equilibrium, linkage disequilibrium, selective neutrality and sex linkage using the software MSA (Dieringer & Schlötterer 2003). There was no evidence of null alleles, which can skew relatedness estimates, as indicated by patterns of inheritance across eight silveryeye families and by checking the entire dataset using the software MICROCHECKER (Van Oosterhout *et al.* 2004).

Individuals were genotyped at the selected 11 loci using multiplex PCR in 10 µl volumes containing 15 ng of DNA, 1.0 µM of each primer, 0.2 mM of each dNTP and 0.25 units DNA polymerase (Thermoprime^{Plus}, ABgene, UK) in the manufacturer's buffer, including 1.0 or 1.5 mM MgCl₂. PCR cycling consisted of 94°C for 3 min, followed by 33 cycles of 45 s at 94°C, 45 s at the appropriate annealing temperature (Frentiu *et al.* 2003) and 20 s at 72°C. One primer in each pair was labelled with one of three fluorescent dyes (Applied Biosystems, Warrington, UK). PCR products were analysed using an ABI 377 DNA sequencer and the GENOTYPER Software (Applied Biosystems, Warrington, UK).

Finally, as a measure of the informativeness of the markers, we calculated exclusion probabilities for parentage assignment, defined as the ability to exclude a 'random' individual from parentage when the other parent's genotype is known (Jamieson & Taylor 1997), for each individual marker and then all markers combined.

(c) Marker-based estimates of relatedness

In order to test the performance of both our markers and various methods of estimating relatedness, we estimated relatedness for pairs of individuals expected to be 'unrelated' and those expected to be full siblings. Pairs of full siblings were initially inferred on the basis of behavioural observations and were subsequently confirmed using the program COLONY (Wang 2004a,b). Relatedness was inferred from molecular markers using three widely used pairwise methods: Queller & Goodnight (QR; 1989); Lynch & Ritland (LR; 1999); and Wang (W; 2002) using allele frequencies based on all typed individuals using the program SPAGEDi v. 1.2 (Hardy & Vekemans 2002). We also estimated relatedness for both categories of relationship using an alternative algorithm based on the simulated annealing approach of Fernández & Toro (FT; 2006). We used full siblings and unrelated individuals as categories because these comprise the two main types of relationship categories present in this dataset.

In addition, because it has been noted that variance in relatedness in a population sample will have a substantial effect on the power to calculate quantitative genetic parameters (Ritland & Ritland 1996; Ritland 2000b; Garant & Kruuk 2005; Csilléry *et al.* 2006), we calculated the actual variance of relatedness, $\text{var}(r_{ij})$ (Ritland & Ritland 1996), for each estimator in the program SPAGEDi v. 1.2 (Hardy & Vekemans 2002). Standard errors of $\text{var}(r_{ij})$ were calculated by bootstrapping across loci.

(d) Estimating quantitative genetic parameters

(i) Traditional animal model

For the traditional animal model based on known pedigree information, we followed the methods described in Frentiu *et al.* (2007). Briefly, additive genetic variances and covariances were estimated from nestling data for all six morphological traits using a restricted maximum-likelihood (REML) animal model (Kruuk 2004), using ASREML v. 1.0 (Gilmour *et al.* 2002). This model included year of sampling, hatching date and their interaction as fixed effects, while nest of rearing was included as an additional random effect. The known pedigree was based on social interactions alone because a previous study has shown that extra-pair paternity is absent in this population (Robertson *et al.* 2001).

(ii) Ritland's pairwise regression method

For the pairwise regression method, we used the program MARK v. 3.0 (Ritland 2004) to calculate heritabilities and genetic correlations for the same six morphological traits based on nestling data. Thus, no pedigree information was used in this model. The Queller & Goodnight (1989) estimator of relatedness was used because, according to the method of Wang (2006), this estimator was the most appropriate to use in estimating relatedness (KINFOR software: Wang 2006). We used spectral decomposition to test whether the pairwise relatedness estimates would produce a positive-definite matrix, as indicated by all eigenvalues being positive. It is important to note that, owing to the constraints of the pairwise regression technique,

no additional effects were fitted to this model, whereas for the other two types of model we were able to include year of measurement, hatching date and nest of rearing. Nevertheless, this lack of additional factors fits with the rationale of the Ritland method, which aims to estimate quantitative genetic parameters from marker-based relatedness estimates, even in the absence of other types of information (Ritland 2000a,b).

(iii) Molecular genealogy method

There are at least two problems in attempting to apply a pedigree-free animal model based on a relatedness matrix inferred from molecular data alone. First, relatedness estimates need to be represented in a positive-definite symmetrical relatedness matrix, **R**, to take the place of the pedigree-based relatedness matrix used in the traditional animal model. Although pairwise estimates of relatedness could be arranged in such a fashion, the resulting relatedness matrix will not be positive definite due to internal inconsistencies in relatedness among multiple individuals generated by pairwise estimators. Second, pedigrees in animal models are assumed to be known without error, and although there has been some recent work on the effect of misassigned paternities on heritability estimates (Charmantier & Réale 2005), it is unclear in the context of an animal model how error will influence the performance of the mixed model (Lynch & Walsh 1998).

To overcome these problems, we implemented the simulated annealing approach to the estimation of relatedness developed by Fernández & Toro (2006) to generate a genealogy based on the molecular marker data, rather than a relatedness matrix *per se*. The key properties of a genealogy in this context are that, unlike relatedness matrices based on pairwise estimates, all relationships between pairs of individuals are symmetrical and all relationships across individuals are consistent with one another. The software MOL_COAN (Fernández & Toro 2006) was used to obtain this genealogy, with the following control parameters for the simulated annealing procedure: maximum number of steps allowed = 300; number of solutions tested at each step = 5000; initial temperature = 0.01; and rate of increase of temperature = 0.09. In addition, we specified the numbers of generations present as three, since the data contained three cohorts, and the maximum number of sires and dams as 221 in each category, based on our knowledge of the size of the breeding population. However, no pedigree information was included in this model.

Since the relatedness matrix generated by simulated annealing is guaranteed to be positive definite, this presented the opportunity to explore how error in the estimation of relatedness affected the performance of the animal model. One approach to reducing the error in **R** is to reduce the dimensionality (or rank) of this symmetrical matrix. In such a high-dimensional space, the vast majority of independent dimensions will represent very minor axes of variance that are likely to reflect measurement error in individual relatedness estimates. We therefore implemented a procedure that enabled an investigation of which subspaces of the relatedness space represented by **R** were more closely associated with phenotypic similarity. Our approach is similar to that implemented in investigations of the dimensionality of identity-by-descent matrices in the context of variance-component quantitative trait locus analysis (Rönnegård & Carlborg 2007; Rönnegård *et al.* 2007).

A spectral decomposition of $\mathbf{R}$ is given by

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T,$$

where $\mathbf{U}$ is an upper triangular matrix of coefficients representing the eigenvectors of $\mathbf{R}$; $\mathbf{\Lambda}$ is a diagonal matrix containing the eigenvalues of $\mathbf{R}$; and superscript T represents the transpose of a matrix. A reduced-rank $\mathbf{R}$ ($\mathbf{R}_{n-d}$) where d is the number of dimensions may then be obtained by defining a subspace of $\mathbf{R}$ by selecting a subset of columns of $\mathbf{U}$ ($\mathbf{U}_{n-d}$) and applying

$$\mathbf{R}_{n-d} = \mathbf{U}_{n-d}\mathbf{\Lambda}_{n-d}\mathbf{U}_{n-d}^T.$$

$\mathbf{R}_{n-d}$ can now be substituted for $\mathbf{R}$ in the animal model. All matrix manipulations were conducted in SAS IML (SAS 2001).

We present the results of analyses that used four reduced-rank $\mathbf{R}$ matrices, in addition to the full-rank $\mathbf{R}$: $\mathbf{R}_{90-d}$ and $\mathbf{R}_{25-d}$, and $\mathbf{R}_{13-d}$ and $\mathbf{R}_{6-d}$ representing 50, 32, 27 and 23% of the estimated variance in relatedness, respectively (appendix 2 in the electronic supplementary material). The first two of these reduced-rank matrices were chosen for the level of variance in relatedness they explained, while the last two were chosen after an examination of the scree plot of eigenvalues of $\mathbf{R}$ (see §3).

We implemented the animal model in this case using the MIXED procedure in SAS (2001). As in the traditional animal model, analyses were based on nestling data and included year of sampling, hatching date and their interaction as fixed effects, and nest of rearing as an additional random effect. Univariate animal models were run for all traits to estimate genetic variances, and genetic covariances were estimated by first estimating the genetic variance in the sum of a particular bivariate trait combination, then taking away the genetic variance of each trait and dividing by two (Mezey & Houle 2005). Substantial problems were encountered in the convergence of REML-based models when the $\mathbf{R}$ matrix was substituted into the animal model, with many parameters unable to be estimated in initial analyses. We subsequently determined that three individuals needed to be excluded from the dataset to allow convergence: two birds based on outlying phenotypic values and one based on the convergence diagnostics provided by the MIXED procedure in SAS (2001).

(iv) Comparison of $\mathbf{G}$ matrices

Finally, we compared the genetic variance–covariance ($\mathbf{G}$) matrices derived from each of the three methodologies. Since the three methods were not all implemented in the same model, log-likelihood ratio tests for testing differences among genetic parameters in the matrices were not available to us. We therefore adopted a geometric approach that compared the similarity of three-dimensional subspaces of the $\mathbf{G}$ matrices using

$$\mathbf{S} = \mathbf{A}^T\mathbf{B}\mathbf{B}^T\mathbf{A},$$

where the matrices $\mathbf{A}$ and $\mathbf{B}$ contain the first three eigenvectors of the two $\mathbf{G}$ matrices to be compared as columns (Blows *et al.* 2004). The sum of the eigenvalues of $\mathbf{S}$ is then a bounded measure of the similarity of the two subspaces (between 0 and 3, with 3 indicating coincident subspaces; Blows *et al.* 2004).

3. RESULTS

(a) Molecular markers and genotyping

In total, 44 alleles were found at 11 polymorphic loci in the silveryeye population (appendix 1 in the electronic

Figure 1. Performance of four different algorithms in estimating genetic relatedness between individuals: QR (Queller & Goodnight 1989), LR (Lynch & Ritland 1999), W (Wang 2002) and FT (Fernández & Toro 2006). Mean relatedness denoted by circles (full-sibling pairs) or triangles (putatively unrelated pairs). Bars indicate twice the standard error. Full-sibling pairs were inferred from behavioural observations and COLONY (Wang 2004a,b). Samples consist of 30 independent pairs of either full siblings or putatively unrelated individuals and are identical for each estimator.

supplementary material). Although allelic diversity was moderate, with a maximum of seven alleles at the most polymorphic locus, observed heterozygosities were high and did not deviate from those expected (appendix 1 in the electronic supplementary material). When we calculated exclusion probabilities for parentage assignment we found that, although each locus individually had a low exclusionary power, the exclusion power of the full set was 0.979.

(b) Marker-based estimates of relatedness

In general, all three of the widely used algorithms for estimating pairwise relatedness performed well in differentiating between known relatives and unrelated individuals (figure 1), with the mean estimated relatedness for full siblings being consistently significantly higher than that for ‘unrelated’ pairs of individuals (QR: $t=15.17$, d.f.=58, $p<0.001$; LR: $t=13.28$, d.f.=58, $p<0.001$; W: $t=11.95$, d.f.=58, $p<0.001$). The simulated annealing approach to estimating relatedness also successfully differentiated between known relatives and unrelated individuals (figure 1), with mean estimated relatedness for full siblings being significantly higher than for putatively unrelated pairs of individuals (FT: $t=8.86$, d.f.=58, $p<0.001$). However, this approach tended to underestimate the absolute level of relatedness of full siblings, for which it estimated an average relatedness of 0.11.

We were unable to detect significant variance in relatedness using any of the traditional estimators, with $\text{var}(r_{ij})$ ranging from -0.001 to 0.008 and all standard errors being large (QR: $\text{var}(r_{ij})=0.005$ (± 0.012), LR: $\text{var}(r_{ij})=0.001$ (± 0.005) and W: $\text{var}(r_{ij})=0.008$ (± 0.011)). Despite the dataset consisting of full-sib families, the proportion of all pairwise comparisons that were among full-sib pairs was approximately 1%.

Table 1. Additive genetic variance–covariance matrices derived from three different methodologies: (a) traditional pedigree-based animal model, (b) Ritland's pairwise regression, and (c) pedigree-free animal model based on $\mathbf{R}_{13-d}$, the 13-d reduced-rank molecular genealogy. (Genetic variances are shown in bold on the diagonal with genetic covariances below the diagonal.)

	wing length	tail length	tarsus length	culmen length	culmen depth	culmen width
<i>(a) traditional pedigree-based animal model (from Frentiu et al. 2007)</i>						
wing length	0.17					
tail length	0.21	0.24				
tarsus length	0.14	0.16	0.23			
culmen length	0.17	0.22	0.22	0.22		
culmen depth	0.12	0.11	0.12	0.13	0.16	
culmen width	0.07	0.06	0.12	0.14	0.14	0.13
<i>(b) Ritland pairwise regression method</i>						
wing length	−0.001					
tail length	−0.001	−0.001				
tarsus length	0.000	0.000	0.000			
culmen length	−0.001	0.000	−0.001	0.000		
culmen depth	−0.001	−0.001	0.000	0.000	0.001	
culmen width	0.000	−0.001	0.000	0.001	0.002	0.002
<i>(c) pedigree-free animal model based on $\mathbf{R}_{13-d}$, the 13-d reduced-rank molecular genealogy</i>						
wing length	0.06					
tail length	−0.01	0.00				
tarsus length	−0.16	−0.14	0.47			
culmen length	−0.03	−0.01	−0.19	0.08		
culmen depth	−0.02	−0.05	−0.08	−0.10	0.26	
culmen width	−0.07	−0.06	−0.10	−0.05	−0.10	0.26

(c) Estimating quantitative genetic parameters

(i) Traditional animal model

The results of the traditional animal model analysis based on known pedigree information were reported in detail in Frentiu et al. (2007). Briefly, this method revealed substantial additive genetic variance for all six morphological traits, with genetic variances ranging from 0.13 to 0.24 (table 1a). There was also positive genetic covariance between all six morphological traits (range 0.06–0.22; table 1a). Finally, this method also revealed significant common environmental effects (nest effects) for four of the six traits, the two non-significant traits being culmen width and culmen depth (range 0.03–0.24; table 2).

(ii) Ritland pairwise regression method

Spectral decomposition showed that pairwise relatedness estimated using the QR estimator results in a non-positive-definite relatedness matrix, with approximately half of the eigenvalues being negative (figure 2a).

Ritland's pairwise regression method provided estimates of genetic variance (range −0.001 to 0.002) and covariance (range −0.001 to 0.002) that were very much smaller than those derived from the traditional animal model, with many of the estimates being effectively zero or negative (table 1b). Note that these results supersede those previously reported for this population (Frentiu 2004), and reviewed elsewhere (Garant & Kruuk 2005), which were based on a subset of the data used here.

(iii) Molecular genealogy method

Spectral decomposition confirmed that the simulated annealing algorithm successfully returned an $\mathbf{R}$ matrix that was positive definite (i.e. the eigenvalues were positive for all eigenvectors; figure 2b). The scree plot in figure 2b suggested that there may have been a qualitative (although small) difference in the amount of variance explained by

Table 2. Estimates of common environmental effects (nest effects) from the traditional animal model and the pedigree-free animal based on $\mathbf{R}_{13-d}$. (Significance is indicated by ** $p < 0.01$ and *** $p < 0.001$.)

trait	type of model	
	traditional animal model	reduced-rank molecular genealogy
wing length	0.244***	0.196***
tail length	0.178***	0.120***
tarsus length	0.231***	0.292***
culmen length	0.213***	0.304***
culmen depth	0.075	0.136**
culmen width	0.030	0.016

the eigenvectors after eigenvectors 6 and 13. We therefore included the $\mathbf{R}_{6-d}$ and $\mathbf{R}_{13-d}$ reduced-rank matrices in animal models, in addition to two further matrices, $\mathbf{R}_{25-d}$ and $\mathbf{R}_{0-d}$ that explored the effect of many more dimensions in $\mathbf{R}$ on the estimation of genetic variance, but still removed 68 and 50% of the estimated variance in relatedness from the analysis, respectively.

When we used the full $\mathbf{R}$ matrix to estimate genetic parameters, we found that only three of the six traits had estimates of genetic variance that were non-zero (appendix 2 in the electronic supplementary material). Subsequent tests indicated that the $\mathbf{R}_{25-d}$ and $\mathbf{R}_{13-d}$ were the best performing of the models. The $\mathbf{R}_{13-d}$ analysis returned all non-zero genetic variances, while the $\mathbf{R}_{25-d}$ analysis returned only three non-zero genetic variances and the $\mathbf{R}_{0-d}$ analysis returned smaller genetic variances than the $\mathbf{R}_{13-d}$ analysis for each trait.

We analysed the data using the $\mathbf{R}_{13-d}$ in two further respects. First, we tested for a common environment

Figure 2. Scree diagrams of eigenvectors for (a) the relatedness matrix used in the Ritland pairwise regression analyses and (b) the relatedness matrix used for the pedigree-free animal model based on the molecular genealogy. The arrow shows the point along the eigenvector axis after which eigenvalues become consistently negative.

variance component (nest effect) for each trait in turn (table 2). The reduced-rank $\mathbf{R}$ matrix indicated significant common environment effects for five of the six traits (range 0.016–0.304), with the non-significant trait being culmen width, which is also one of the two traits that were non-significant in the traditional animal model (table 2). Second, we estimated the genetic covariance among traits that were found to be very different from those obtained using the traditional animal model, all being negative (table 1c). It is worth noting, therefore, that the overall $\mathbf{G}$ matrix estimated by the pedigree-free method, shown in table 1c, breaks the rule that variance–covariance matrices must be positive definite because it contains estimates of covariance that are negative. This is presumably because we estimated each component separately, rather than in a single multivariate analysis, and the covariance estimates themselves are not very robust.

(iv) Comparison of $\mathbf{G}$ matrices

The comparison of the three-dimensional subspace between $\mathbf{G}$ from the traditional animal model and Ritland's regression method gave a value of only 1.20 for the sum of the eigenvalues of $\mathbf{S}$, indicating a low level of similarity between the two $\mathbf{G}$ matrices. The comparison between the traditional animal model and the pedigree-free animal model based on $\mathbf{R}_{13-d}$ gave a value of 1.62, indicating a marginally better performance in recovering the genetic covariance structure found by the traditional animal model.

4. DISCUSSION

Our results offer a mixed assessment of the potential use of pedigree-free animal models in estimating quantitative genetic parameters in natural populations. On one hand, we have shown that it is possible to implement animal models using a genealogical matrix based exclusively on molecular marker data. In addition, the estimates of genetic variance obtained using this method were of a similar overall magnitude to those obtained using a traditional animal model. Reducing the dimensionality of the molecular $\mathbf{R}$ matrix increased the ability of the pedigree-free animal approach in recovering genetic variance, presumably by improving the signal to noise ratio. On the other hand, the estimates of covariance obtained using the pedigree-free animal approach tended to be negative, whereas the traditional animal model showed positive covariance among most traits, and the overall similarity in $\mathbf{G}$ matrices between the two techniques was only moderate. Taken together, these results indicate that, while pedigree-free animal models can indeed recover quantitative genetic information, the full pattern of covariance might be difficult to detect.

Our results were less encouraging regarding the potential use of Ritland's pairwise regression method, at least for this population. The pairwise regression method also commonly yielded negative estimates of heritability, additive genetic variance and additive genetic covariance. In addition, the overall structure of the genetic variance–covariance matrices obtained using the pairwise regression technique was more different from that obtained using the traditional animal model method than that recovered from the molecular genealogy approach. The relatively poor performance of the pairwise method should, of course, be assessed cautiously because both the other types of model included additional explanatory factors (year of measurement, hatching date and the interaction between these two effects), which may have improved the performance of those models. On the other hand, we feel that the difficulty of including such factors in pairwise regression models is an inherent limitation to the development of this approach, along with the fact that the pairwise relatedness estimates cannot be rearranged in matrix form and used in an animal model, as such matrices are negative definite. Thus, in combination with the difficulty of estimating statistical significance using the pairwise regression method, and the fact that several other tests based on this approach have also reported problems in obtaining reliable estimates (Klaper *et al.* 2001; Thomas *et al.* 2002; Wilson *et al.* 2003; Andrew *et al.* 2005; Coltman 2005; Shikano 2005; van Kleunen & Ritland 2005), we agree with previous reviews that the prospect of applying this method to a wide range of populations may be limited (Garant & Kruuk 2005; Owens 2006).

Why did the pedigree-free approach fail to recover the full pattern of quantitative-genetic covariance? The core challenge in studies of this type is lack of significant variance in relatedness among individuals in the populations according to estimates based on molecular marker data (Ritland 1996; Csilléry *et al.* 2006). There are, however, two different types of explanation for low variance in relatedness, which have very different implications for the prospects of the pedigree-free approach. The first explanation for low relatedness variance is that

this is an intrinsic property of the population. Under this scenario, even an improved set of molecular markers would not increase relatedness variance because such variation is simply not present (Csilléry *et al.* 2006). In our particular case, for instance, it might be suspected that low relatedness variance may be due to the island-dwelling nature of the population. Indeed, it has recently been suggested that low relatedness variance may be a general property of wild populations (Csilléry *et al.* 2006), and therefore that the pedigree-free animal model may always struggle to obtain sufficient power to estimate complex genetic patterns. However, our finding that reducing the dimensionality of the molecular **G** matrix increased our ability to detect genetic variance argues against this possibility.

The second explanation for the apparent inability to recover the level of heritability found by the traditional animal model is that sufficient variation in relatedness does exist in the population, but we were unable to detect that variation (Csilléry *et al.* 2006). This could be because either the battery of molecular markers employed here was not sufficiently powerful to reveal that variation or the variation is obscured through limitation in the algorithm used to estimate the genealogy. There is some evidence to support the influence of both these sources of error variance. For instance, the molecular markers used have relatively low polymorphism so, although we have shown they are sufficiently powerful to potentially identify first-degree relatives, they may fail to differentiate among more distant relatives. Similarly, it is clear from the absolute size of the relatedness estimates in figure 1 that, although the algorithm used to construct the molecular genealogy (Fernández & Toro 2006) can, on average, distinguish successfully between close genetic relatives and unrelated pairs of individuals, the same algorithm tends to underestimate the genetic relatedness of full sibs. This underestimation is likely to be caused by a combination of a weak signal due to the low polymorphism of the molecular markers and the search procedure of the algorithm itself. Improving the performance of such methods is clearly a priority for future work if molecular genealogies are to be widely applied in animal models.

It remains an open question, however, whether natural populations in general will have sufficient variance in relatedness to facilitate the pedigree-free animal model approach or whether our study population is unusual in this respect. Our population was an isolated island population with low to moderate levels of genetic polymorphism, and our experimental design ensured that we had a sample richer in full siblings than one might expect at random. Nevertheless, if future studies do reveal substantial variation in relatedness in other wild populations, then our analyses suggest that the pedigree-free animal model may provide an opportunity to study the evolutionary genetics of wild populations, particularly if its use is complemented by the development of increasingly powerful molecular markers, more sophisticated analytical methods for reconstructing genealogies from marker data, and the subsequent dimension reduction of the resulting relatedness matrix.

All work was conducted under permits obtained from The Australian Bird and Bat Banding Scheme, the Environment Protection Agency of Queensland and the Animal Ethics Committee of the University of Queensland.

We thank the Heron Island Research Station, P & O Heron Island Resorts, Queensland Marine Parks, NERC Molecular Genetics Facility (Sheffield), S. Chenoweth, D. Dawson, C. Edwards, J. Hadfield, K. Henderson, W. Hill, J. Kikkawa, L. Kruuk, A. Krupa, M. Losiak, B. Maher, F. Manson, K. Ritland, S. Robinson-Wolrath, S. Scott, B. Sheldon, S. Twigg, B. Walsh, J. Wang and C. Wiley for help and discussion. This work was supported through funding from the Australian Research Council and the Natural Environment Research Council (UK).

REFERENCES

- Andrew, R. L., Peakall, R., Wallis, I. R., Wood, J. T., Knight, E. J. & Foley, W. J. 2005 Marker-based quantitative genetics in the wild: the heritability and genetic correlation of chemical defenses in eucalyptus. *Genetics* **171**, 1989–1998. (doi:10.1534/genetics.105.042952)
- Blows, M. W., Chenoweth, S. F. & Hine, E. 2004 Orientation of the genetic variance–covariance matrix and the fitness surface for multiple male sexually selected traits. *Am. Nat.* **163**, 329–340. (doi:10.1086/381941)
- Brommer, J. E., Merila, J., Sheldon, B. C. & Gustafsson, L. 2005 Natural selection and genetic variation for reproductive reaction norms in a wild bird population. *Evolution* **59**, 1362–1371.
- Charmanter, A. & Réale, D. 2005 How do misassigned paternities affect the estimation of heritability in the wild? *Mol. Ecol.* **14**, 2839–2850. (doi:10.1111/j.1365-294X.2005.02619.x)
- Charmanter, A., Perrins, C., McCleery, R. H. & Sheldon, B. C. 2006a Evolutionary response to selection on clutch size in a long-term study of the mute swan. *Am. Nat.* **167**, 453–465. (doi:10.1086/499378)
- Charmanter, A., Perrins, C., McCleery, R. H. & Sheldon, B. C. 2006b Quantitative genetics of age at reproduction in wild swans: support for antagonistic pleiotropy models of senescence. *Proc. Natl Acad. Sci. USA* **103**, 6587–6592. (doi:10.1073/pnas.0511123103)
- Clegg, S. M. & Owens, I. P. F. 2002 The ‘island rule’ in birds: medium body size and its ecological explanation. *Proc. R. Soc. B* **269**, 1359–1365. (doi:10.1098/rspb.2002.2024)
- Clegg, S. M., Degnan, S. M., Kikkawa, J., Moritz, C., Estoup, A. & Owens, I. P. F. 2002a Genetic consequences of sequential founder events by an island-colonizing bird. *Proc. Natl Acad. Sci. USA* **99**, 8127–8132. (doi:10.1073/pnas.102583399)
- Clegg, S. M., Degnan, S. M., Moritz, C., Estoup, A., Kikkawa, J. & Owens, I. P. F. 2002b Microevolution in island forms: the roles of drift and directional selection in morphological divergence of a passerine bird. *Evolution* **56**, 2090–2099.
- Coltman, D. W. 2005 Testing marker-based estimates of heritability in the wild. *Mol. Ecol.* **14**, 2593–2599. (doi:10.1111/j.1365-294X.2005.02600.x)
- Coltman, D. W., O’Donoghue, P., Jorgenson, J. T., Hogg, J. T., Strobeck, C. & Festa-Bianchet, M. 2003 Undesirable evolutionary consequences of trophy hunting. *Nature* **426**, 655–658. (doi:10.1038/nature02177)
- Csilléry, K., Johnson, T., Beraldi, D., Clutton-Brock, T., Coltman, D., Hansson, B., Spong, G. & Pemberton, J. M. 2006 Performance of marker-based relatedness estimators in natural populations of outbred vertebrates. *Genetics* **173**, 2091–2101. (doi:10.1534/genetics.106.057331)
- Degnan, S. M., Robertson, B. C., Clegg, S. M. & Moritz, C. C. 1999 Microsatellite primers for studies of gene flow and mating systems in white-eyes (*Zosterops*). *Mol. Ecol.* **8**, 159–160. (doi:10.1046/j.1365-294X.1999.00799.x)

- Dieringer, D. & Schlötterer, C. 2003 Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite datasets. *Mol. Ecol. Notes* **3**, 167–169. (doi:10.1046/j.1471-8286.2003.00351.x)
- Falconer, D. S. & Mackay, T. F. C. 1996 *Introduction to quantitative genetics*. Harlow, UK: Longman Group.
- Fernández, J. & Toro, M. A. 2006 A new method to estimate relatedness from molecular markers. *Mol. Ecol.* **15**, 1657–1667. (doi:10.1111/j.1365-294X.2006.02873.x)
- Frentiu, F. D. 2004 Adaptive morphological evolution in an island bird (*Zosterops lateralis chlorocephalus*): a quantitative genetic approach. Unpublished PhD thesis. University of Queensland, Australia.
- Frentiu, F. D., Lange, C. L., Burke, T. & Owens, I. P. F. 2003 Isolation of microsatellite loci in the Capricorn silveryeye, *Zosterops lateralis chlorocephalus* (Aves: Zosteropidae). *Mol. Ecol. Notes* **3**, 462–464. (doi:10.1046/j.1471-8286.2003.00484.x)
- Frentiu, F. D., Clegg, S. M., Blows, M. W. & Owens, I. P. F. 2007 Large body size in an island-dwelling bird: a microevolutionary analysis. *J. Evol. Biol.* **20**, 639–649. (doi:10.1111/j.1420-9101.2006.01242.x)
- Garant, D. & Kruuk, L. E. B. 2005 How to use molecular marker data to measure evolutionary parameters in wild populations. *Mol. Ecol.* **14**, 1843–1859. (doi:10.1111/j.1365-294X.2005.02561.x)
- Garant, D., Kruuk, L. E. B., McCleery, R. H. & Sheldon, B. C. 2004 Evolution in a changing environment: a case study with great tit fledging mass. *Am. Nat.* **164**, E115–E129. (doi:10.1086/424764)
- Garant, D., Kruuk, L. E. B., Wilkin, T. A., McCleery, R. H. & Sheldon, B. C. 2005 Evolution driven by differential dispersal within a wild bird population. *Nature* **433**, 60–65. (doi:10.1038/nature03051)
- Gilmour, A. R., Gogel, B. J., Cullis, B. R., Welham, S. J. & Thompson, R. 2002 *AsReml User Guide Release 1.0*. Hemel Hempstead, UK: VSN International Ltd.
- Gleeson, D. J., Blows, M. W. & Owens, I. P. F. 2005 Genetic covariance between indices of body condition and immunocompetence in a passerine bird. *BMC Evol. Biol.* **5**, 61. (doi:10.1186/1471-2148-5-61)
- Hadfield, J. D. & Owens, I. P. F. 2006 Strong environmental determination of a carotenoid-based plumage trait is not mediated by carotenoid availability. *J. Evol. Biol.* **19**, 1104–1114. (doi:10.1111/j.1420-9101.2006.01095.x)
- Hadfield, J. D., Burgess, M. D., Lord, A., Phillimore, A. B., Clegg, S. M. & Owens, I. P. F. 2006 Direct versus indirect sexual selection: genetic basis of colour, size and recruitment in a wild bird. *Proc. R. Soc. B* **273**, 1347–1353. (doi:10.1098/rspb.2005.3459)
- Hadfield, J. D., Nutall, A., Osorio, D. & Owens, I. P. F. 2007 Testing the phenotypic gambit: phenotypic, genetic and environmental correlations of colour. *J. Evol. Biol.* **20**, 549–557. (doi:10.1111/j.1420-9101.2006.01262.x)
- Hardy, O. J. & Vekemans, X. 2002 SPAGeDI: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Mol. Ecol. Notes* **2**, 618–620. (doi:10.1046/j.1471-8286.2002.00305.x)
- Jamieson, A. & Taylor, S. C. S. 1997 Comparisons of three probability formulae for parentage exclusion. *Anim. Genet.* **28**, 397–400. (doi:10.1111/j.1365-2052.1997.00186.x)
- Kikkawa, J. & Wilson, J. M. 1983 Breeding and dominance among the Heron Island silveryeyes *Zosterops lateralis chlorocephala*. *Emu* **83**, 181–198.
- Klaper, R., Ritland, K., Mousseau, T. A. & Hunter, M. D. 2001 Heritability of phenolics in *Quercus laevis* inferred using molecular markers. *J. Hered.* **92**, 421–426. (doi:10.1093/jhered/92.5.421)
- Kruuk, L. E. B. 2004 Estimating genetic parameters in natural populations using the ‘animal model’. *Phil. Trans. R. Soc. B* **359**, 873–890. (doi:10.1098/rstb.2003.1437)
- Kruuk, L. E. B., Clutton-Brock, T. H., Slate, J., Pemberton, J. M., Brotherstone, S. & Guinness, F. E. 2000 Heritability of fitness in a wild mammal population. *Proc. Natl Acad. Sci. USA* **97**, 698–703. (doi:10.1073/pnas.97.2.698)
- Kruuk, L. E. B., Merilä, J. & Sheldon, B. C. 2001 Phenotypic selection on a heritable size trait revisited. *Am. Nat.* **158**, 557–571. (doi:10.1086/323585)
- Kruuk, L. E. B., Slate, J., Pemberton, J. M., Brotherstone, S., Guinness, F. & Clutton-Brock, T. 2002 Antler size in red deer: heritability and selection but no evolution. *Evolution* **56**, 1683–1695.
- Lynch, M. & Walsh, B. 1998 *Genetic and the analysis of quantitative traits*. Sunderland, MA: Sinauer.
- Lynch, M. & Ritland, K. 1999 Estimation of pairwise relatedness with molecular markers. *Genetics* **152**, 1753–1766.
- MacColl, A. D. C. & Hatchwell, B. J. 2003 Heritability of parental effort in a passerine bird. *Evolution* **57**, 2191–2195.
- McCleery, R. H., Pettifor, R. A., Armbruster, P., Meyer, K., Sheldon, B. C. & Perrins, C. M. 2004 Components of variance underlying fitness in a natural population of the great tit *Parus major*. *Am. Nat.* **164**, E62–E72. (doi:10.1086/422660)
- Merilä, J. & Sheldon, B. C. 2000 Lifetime reproductive success and heritability in nature. *Am. Nat.* **155**, 301–310. (doi:10.1086/303330)
- Merilä, J., Kruuk, L. E. B. & Sheldon, B. C. 2001 Cryptic evolution in a wild bird population. *Nature* **412**, 76–79. (doi:10.1038/35083580)
- Mezey, J. G. & Houle, D. 2005 The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. *Evolution* **59**, 1027–1038.
- Moore, A. J. & Kruuk, P. F. 2002 Quantitative genetic analysis of natural populations. *Nat. Rev. Genet.* **3**, 971–978. (doi:10.1038/nrg951)
- Mousseau, T. A., Ritland, K. & Heath, D. D. 1998 A novel method for estimating heritability using molecular markers. *Heredity* **80**, 218–224. (doi:10.1046/j.1365-2540.1998.00269.x)
- Nicholls, J. A., Double, M. C., Rowell, D. M. & Magrath, R. D. 2000 The evolution of cooperative and pair breeding in thornbills *Acanthiza* (Pardalotidae). *J. Avian Biol.* **31**, 165–176. (doi:10.1034/j.1600-048X.2000.310208.x)
- Owens, I. P. F. 2006 Where is behavioural ecology going? *Trends Ecol. Evol.* **21**, 356–361. (doi:10.1016/j.tree.2006.03.014)
- Queller, D. C. & Goodnight, K. F. 1989 Estimating relatedness using genetic markers. *Evolution* **43**, 258–275. (doi:10.2307/2409206)
- Ritland, K. 1996 Marker-based method for inferences about quantitative inheritance in natural populations. *Evolution* **50**, 1062–1073. (doi:10.2307/2410647)
- Ritland, K. 2000a Detecting inheritance with inferred relatedness in nature. In *Adaptive genetic variation in the wild* (eds T. A. Mousseau, B. Sinervo & J. Endler), pp. 187–199. Oxford, UK: Oxford University Press.
- Ritland, K. 2000b Marker-inferred relatedness as a tool for detecting heritability in nature. *Mol. Ecol.* **9**, 1195–1204. (doi:10.1046/j.1365-294x.2000.00971.x)
- Ritland, K. 2004 *MARK—marker inferred relatedness and quantitative inheritance program*. Vancouver, Canada: University of British Columbia.
- Ritland, K. & Ritland, C. 1996 Inferences about quantitative inheritance based on natural population structure in the yellow monkeyflower, *Mimulus guttatus*. *Evolution* **50**, 1074–1082. (doi:10.2307/2410648)
- Robertson, B. C., Degnan, S. M., Kikkawa, J. & Moritz, C. C. 2001 Genetic monogamy in the absence of paternity

- guards: the Capricorn silvereve, *Zosterops lateralis chlorocephalus*, on Heron Island. *Behav. Ecol.* **12**, 666–673. (doi:10.1093/beheco/12.6.666)
- Robinson-Wolrath, S. I. & Owens, I. P. F. 2003 Large size in an island-dwelling bird: intraspecific competition and the dominance hypothesis. *J. Evol. Biol.* **16**, 1106–1114. (doi:10.1046/j.1420-9101.2003.00615.x)
- Rodríguez-Ramilo, S. T., Toro, M. A., Caballero, A. & Fernández, J. 2007 The accuracy of a heritability estimator using molecular information. *Conserv. Genet.* **8**, 1189–1198. (doi:10.1007/s10592-006-9273-z)
- Rönnegård, L. & Carlborg, Ö. 2007 Separation of base allele and sampling term effects gives new insights in variance component QTL analysis. *BMC Genet.* **8**, 1. (doi:10.1186/1471-2156-8-1)
- Rönnegård, L., Mischenko, K., Holmgren, S. & Carlborg, Ö. 2007 Increasing the efficiency of variance component quantitative trait loci analysis by using reduced-rank identity-by-descent matrices. *Genetics* **176**, 1935–1938. (doi:10.1534/genetics.107.071977)
- SAS 2001 *SAS/STAT software*, release 8.2. Cary, NC: SAS Institute, Inc.
- Scott, S. N., Clegg, S. M., Blomberg, S. P., Kikkawa, J. & Owens, I. P. F. 2003 Morphological shifts in island-dwelling birds: the roles of generalist foraging and niche expansion. *Evolution* **57**, 2147–2156.
- Shikano, T. 2005 Marker-based estimation of heritability for body color variation in Japanese flounder *Paralichthys olivaceus*. *Aquaculture* **249**, 95–105. (doi:10.1016/j.aquaculture.2005.03.023)
- Thomas, S. C. 2005 The estimation of genetic relationships using molecular markers and their efficiency in estimating heritability in natural populations. *Phil. Trans. R. Soc. B* **360**, 1457–1467. (doi:10.1098/rstb.2005.1675)
- Thomas, S. C. & Hill, W. G. 2000 Estimating quantitative genetic parameters using sibships reconstructed from marker data. *Genetics* **155**, 1961–1972.
- Thomas, S. C., Pemberton, J. M. & Hill, W. G. 2000 Estimating variance components in natural populations using inferred relationships. *Heredity* **84**, 427–436. (doi:10.1046/j.1365-2540.2000.00681.x)
- Thomas, S. C., Coltman, D. W. & Pemberton, J. M. 2002 The use of marker-based relationship information to estimate the heritability of body weight in a natural population: a cautionary tale. *J. Evol. Biol.* **15**, 92–99. (doi:10.1046/j.1420-9101.2002.00372.x)
- van Kleunen, M. & Ritland, K. 2005 Estimating heritabilities and genetic correlations with marker-based methods: an experimental test in *Mimulus guttatus*. *J. Hered.* **96**, 368–375. (doi:10.1093/jhered/esi039)
- Van Oosterhout, C., Hutchinson, W. F., Wills, D. P. M. & Shipley, P. 2004 MICROCHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Mol. Ecol. Notes* **4**, 535–538. (doi:10.1111/j.1471-8286.2004.00684.x)
- Wang, J. L. 2002 An estimator for pairwise relatedness using molecular markers. *Genetics* **160**, 1203–1215.
- Wang, J. 2004a Estimating pairwise relatedness from dominant genetic markers. *Mol. Ecol.* **13**, 3169–3178. (doi:10.1111/j.1365-294X.2004.02298.x)
- Wang, J. L. 2004b Sibship reconstruction from genetic data with typing errors. *Genetics* **166**, 1963–1979. (doi:10.1534/genetics.166.4.1963)
- Wilson, A. J., McDonald, G., Moghadam, H. K., Herbing, C. M. & Ferguson, M. M. 2003 Marker-assisted estimation of quantitative genetic parameters in rainbow trout, *Oncorhynchus mykiss*. *Genet. Res.* **81**, 145–156. (doi:10.1017/S0016672302006055)