11

Coalescent-based analysis of demography: applications to biogeography on Sulawesi

Ben J. Evans

11.1 Introduction

If we were to travel back in time 18 000 years to Southeast Asia, we would encounter a radically different physical landscape. We would of course find a much more subtle impact of humans then compared to now, but perhaps equally striking would be a completely different extent of land and sea. During the last glacial period and the ones that preceded it, large regions that are now ocean were above sea level. It was possible during these intermittent but protracted periods to travel over land from mainland Asia to the islands of Borneo, Sumatra and Java, or from Australia to the island of New Guinea. These land connections account for faunal affinities between islands in the western part of Southeast Asia (the 'Sunda Region') and the eastern part of Southeast Asia (the 'Sahul Region'). Between these regions is Wallacea (Sulawesi, the Moluccan Archipelago and the Lesser Sunda Islands), which includes islands that were never connected by dry land to mainland Asia or to Australia. Abrupt transitions in flora and fauna in this region prompted biogeographers to demarcate the interface between the Asian and Australian faunal zones with boundaries such as Wallace's Line (Fig 11.1a; Wallace 1863, Huxley 1868, Lydekker 1896, George 1981).
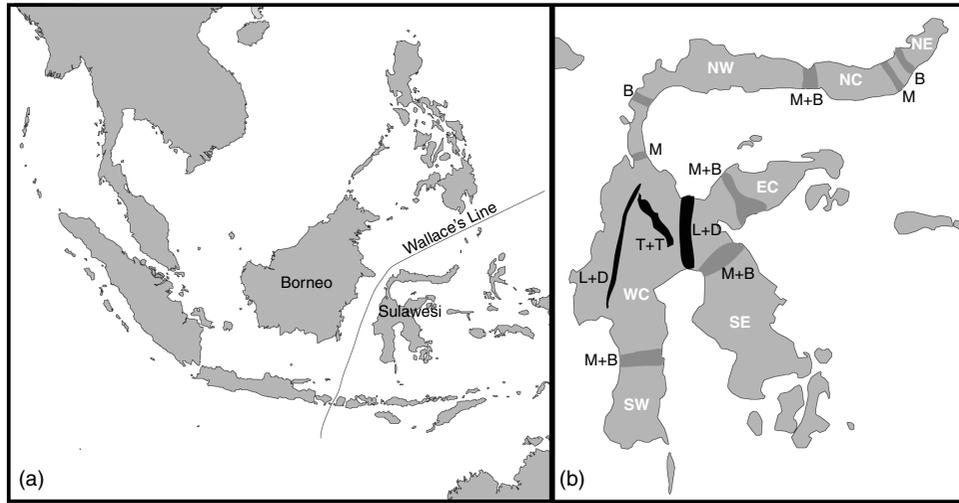
Figure 11.1 Geographical locations of major biogeographical boundaries in the Indonesian Archipelago and on Sulawesi. (a) Position of Wallace's Line relative to Borneo and Sulawesi. (b) Locations of geographically coincident boundaries of differentiated lineages in more than one taxon. In white font are the Southwest (SW), Southeast (SE), West-central (WC), East-central (EC), Northwest (NW), North central (NC) and Northeast (NE) areas of endemism defined by Sulawesi endemics from the genus *Macaca* (M) and *Bufo* (B) with boundaries in grey. Many boundaries of these lineages are shared (M + B). Other shared and probably more ancient boundaries (in black) are apparent in the genera *Limnonectes* and *Draco* (L + D) and the genera *Tarsius* and *Thoopterus* (T + T).

Wallacea is both a biogeographical transition zone and also a site of remarkable species endemism. Sulawesi Island, the largest in Wallacea, is arguably also the most complex in terms of geological and biogeographical history. Sulawesi is about 174 000 km^2, which is slightly bigger than the state of New York and about twice as large as Portugal. Forged from the accretion of at least four major palaeoislands (Hall 2001, this volume, Chapter 3), this island was subject to periodic marine inundation in multiple regions, such as the Tempe Depression at the base of the southwest peninsula and the Gorontalo Depression in the middle of the north peninsula. The ecology of Sulawesi is remarkably variable and different portions experience distinct rainfall cycles and have different climates and soil substrates (Whitten et al. 2002). Being an island composed largely of peninsulas, overland distances between different parts of Sulawesi can be quite large (up to ~1300 km), so geographic distance has the potential to contribute to genetic differentiation (Wright 1943). It is no surprise, therefore, that substantial population structure and/or high species diversity exists in many of Sulawesi's endemic terrestrial fauna (van der Vecht 1953, Musser 1987, Bridle et al. 2001, Evans et al. 2003a, b, c, 2008, Larson et al. 2005, McGuire et al. 2007, Brown et al. 2010, Musser et al. 2010).

What is surprising, however, is the extent to which biogeographical patterns coincide across evolutionary lineages on Sulawesi. While no two lineages can have identical evolutionary histories, physical aspects of a habitat such as the margins of land and water or altitudinal and climatic gradients can lead to biogeographical similarities among co-distributed (sympatric) species in terms of, for example, the geographical distribution of distinct populations, divergence time between populations, or the extent of migration between populations (Avise 2000). Recent studies of terrestrial vertebrates on Sulawesi have recovered similar patterns in multiple, distantly related groups that suggest a shared evolutionary impact of the ecological and geological factors.

11.1.1 Biogeographical patterns on Sulawesi

One shared pattern is exemplified by the Sulawesi macaques (genus *Macaca*) and the Celebes toad (*Bufo celebensis*). In both groups the geographic distributions of differentiated populations include three differentiated populations/species on the north peninsula, one on each of the other peninsulas (southwest, southeast, and east central), and one in the centre (Fig 11.1b; Evans et al. 1999, 2001, 2003b, c, 2008). The geographic location of the margins of differentiated areas of endemism is essentially identical in Sulawesi macaques and Celebes toads at the Gorontalo Depression between the northwest and north central area of endemism (AOE), at the Tempe Depression between the southwest and west-central AOE, and at the Bongka River between the northern margin between the west-central and east-central AOE (Fig 11.1b). In both groups, differentiation between the north-central and the northeast AOE is low compared to pairwise comparisons between other areas of endemism (Evans et al. 2003b, c, 2008). Exceptions to this general pattern of similar patterns of differentiation occur on the margin between the west-central and northwest AOE, which in toads is situated north of the margin in macaques, and also between the north-central and the northeast AOE, which in toads is east of the margin in macaques, at least for mitochondrial DNA.

Another shared pattern of differentiation is exemplified by flying lizards (genus *Draco*) and fanged frogs (genus *Limnonectes*; Fig 11.1b; Iskandar and Tjan 1996, Evans et al. 2003a, McGuire et al. 2007, Setiadi et al. 2010). Some of these species exhibit differentiation similar to some key features to the macaques and toads (for example, across the Gorontalo and Tempe Depression), but species ranges differ in key aspects that may reflect a longer residence on Sulawesi. Species of flying lizards and fanged frogs on the North peninsula both have ranges that extend down the west side of the West central AOE (Evans et al. 2003a, McGuire et al. 2007, Setiadi et al. 2011). Both groups also have a species or multiple species whose range spans the east central, southeast and east portion of the West central AOE (East of Lake Poso) (Evans et al. 2003a, McGuire et al. 2007, Setiadi et al. 2011).

A third shared pattern is substantial differentiation across the Palu–Koro fault that runs from the city of Poso at the base of the north peninsula to the Sea of Bone, which lies between the southwest and southeast peninsulas. The boundary between differentiated populations roughly corresponds to this fault in tarsiers (genus *Tarsius*) and in the bat *Thoopterus nigrescens* (Fig 11.1b; Campbell et al. 2007, Merker et al. 2009). This pattern is not found in Sulawesi macaques or Celebes toads and may reflect an underlying cause that pre-dates the arrival of these groups to Sulawesi. Analysis of tarsier vocalisations suggests that in addition to deep divergences across the Palu–Koro fault, areas of endemism exist in tarsiers that share some features with Sulawesi macaques and Celebes toads (Shekelle et al. 1997, Shekelle and Leksono 2004).

Perhaps not surprisingly, there is incomplete information about divergence, population structure and geographic ranges of most of Sulawesi's vertebrates, including its large vertebrates. Moreover, some species appear to share some features of these general biogeographical patterns but not others. For example, squirrels of the genus *Prosciurillus* have species distributions that bear similarities to the first and second distributional patterns described above (Musser et al. 2010). In particular, *P. leucomus* occurs throughout the north peninsula, *P. topapuensis* and *P. weberi* occur in the west central portion of Sulawesi, and a fossil record exists for another species in the southwest peninsula (Musser et al. 2010). *Prosciurillus alstoni* occurs in the southeast and west-central portion of Sulawesi but also in the northern region of West central Sulawesi (Musser et al. 2010) – this distribution is not shared with any of the species discussed above. Further sampling is needed to iron out the precise distributions of these squirrel species and their parasites, and the degree to which these distributions correspond with other Sulawesi endemics (Musser et al. 2010). Likewise, babirusas (genus *Babyrousa*) – an endemic pig of Sulawesi – from North Sulawesi appear to be morphologically divergent from populations on the Togian, Buru and Sula islands, but the margins of differentiation and the ranges of differentiated forms are not known (Meijaard and Groves 2002). Interestingly, a babirusa fossil of an extinct southwestern peninsula population appears distinct from the North Sulawesi population (Meijaard and Groves 2002). Another example is the anoa (genus *Bubalus*), the endemic dwarf buffalo of Sulawesi. There are reports of morphological and cytological variation, but limited sampling prevents the conclusive determination of whether multiple divergent lineages are co-distributed (sympatric) or not (Burton et al. 2005). In a survey by Burton and colleagues (2005) individuals that were morphologically distinguishable on the basis of horn shape, body size, pelage colour and tail length were sampled higher than and lower than 1000 m above sea level, suggesting that variation in Sulawesi anoas may not be associated with altitudinal gradients (that is, that there is not a 'highland' and 'lowland' anoa even though there may be multiple species of anoa). In the Celebes warty pig, *Sus celebensis*, mitochondrial DNA

from the north peninsula is diverged from that of the rest of Sulawesi (Larson et al. 2005), but it is still unclear whether there exists further population subdivision on Sulawesi.

Some vertebrate groups, such as fanged frogs (Evans et al. 2003a, Setiadi et al. 2011) and shrews (genus *Crocidura*, Ruedi et al. 1998), have multiple sympatric forms in different parts of Sulawesi. This offers the opportunity to test whether sympatric forms are more or less closely related than non-sympatric forms. For example, after accounting for the effect of shared evolutionary history, variation in morphology and life history among sympatric fanged frogs on Sulawesi suggests the action of natural selection (Setiadi et al. 2011). This supports an adaptive radiation of these frogs on Sulawesi to occupy ecological niches that are not available on the Philippines – an archipelago that supports a higher diversity of frog families than Sulawesi (Setiadi et al. 2011). Undoubtedly, of course, many species have biogeographical histories that are not similar to these rough patterns (Bridle et al. 2004). One study of endangered Sulawesi tortoises (*Indotestudo forstenii*), for example, did not find substantial genetic differences among captive individuals, although the exact geographic provenance of these samples is unknown (Ives et al. 2008). However, if we ignore patterns that are not universally shared by all species, conservation decisions will have little biological basis (Evans et al. 2004). Patterns of differentiation shared by multiple (but not all) species therefore provide a useful guide for conservation prioritisation on Sulawesi. Arguably, we should target multiple areas of genetic endemism on different portions of this island, especially regions that currently do not have large national parks (Evans et al. 2003c, 2008). According to a census in 2000, about 15 million people live on Sulawesi. Habitat alteration is occurring at an alarming pace, including within protected areas (Bickford et al. 2007), and there is urgent need for further conservation measures.

11.1.2 Coalescence, divergence population genetics and isolation-migration models

Molecular polymorphism is influenced by demography, ancestry, mutation, genetic drift and natural selection. The evolutionary history of molecular polymorphism can be represented by a gene tree (a genealogy) that traces the evolutionary history of different alleles back in time to a single ancestral allele. This 'reverse evolution' backwards in time from descendant to ancestor is known as coalescence. Theory has been developed that decouples the mutational process that generates molecular polymorphism, and the coalescent genealogical process that accounts for its evolutionary history (Hudson 1991). This theory has been applied to a number of problems in evolutionary genetics including inferences concerning natural selection, recombination, phylogenetics and migration (Nielsen and Wakeley 2001, Nordborg 2001, Edwards et al. 2007).

Speciation occurs when an ancestral lineage diverges into two or more descendant lineages. Speciation could result from a physical barrier to migration that emerges within the ancestral lineage, in which case all parts of the genome begin diverging at the same time except those portions that are polymorphic in the ancestral lineage. Divergence of ancestral polymorphisms by definition begins before the time of speciation. Alternatively, speciation could occur in a stepwise fashion wherein incompatibilities between descendant lineages (genetic, morphological, behavioural) encoded by different parts of the genome accumulate over time. In the second scenario, reproductive isolation of many different parts of the genome occurs at different times (de Queiroz 1998, Wu 2001). Study of these alternative scenarios using multilocus polymorphism data, is facilitated by coalescent methods that permit comparison of demographic models that have and that do not have migration after divergence, an approach called divergence population genetics (Kliman et al. 2000). A useful feature of these methods is that realistic evolutionary phenomena can be incorporated into demographic models including intra- and interlocus recombination, biparental or uniparental inheritance, variation among loci in the rate of mutation, changes in population size, gene flow among populations and speciation without gene flow (Hey and Nielsen 2004, Becquet and Przeworski 2007).

Multiple approaches have been developed to study divergence population genetics (Wakeley and Hey 1997, Kliman et al. 2000, Becquet and Przeworski 2007, Hey and Nielsen 2004, Leman et al. 2005, Putnam et al. 2007). This study uses one of them, implemented by the program MIMAR (Becquet and Przeworski 2007), to explore how violations of model assumptions could affect estimation of parameters of demographic models (Fig 11.2a). Typically the null model (hereafter the isolation model) is one in which a single ancestral population diverges into two descendant populations with no subsequent gene flow. The alternative model (hereafter the isolation–migration model) permits ongoing gene flow (symmetrical or asymmetrical) between the descendant populations. Estimated parameters include the mutation parameters for the ancestral and both descendant populations ($\theta = 4N_e\mu$ for diploid biparentally inherited loci), the divergence time (τ), and (for the alternative model only) the magnitude of gene flow in terms of effective number of migrants per generation in each direction (m_{12} and m_{21}, respectively). Instead of directly analysing sequence data, MIMAR uses four summary statistics (s_1, s_2, ss and sf) that are affected by the parameters of the model (that is, θ_1, θ_2, θ_A, τ, m_{12} and m_{21}). The summary statistics concern the distribution among populations of derived mutations, categorising them as being either fixed in one population or the other (sf), polymorphic in one population but not the other (s_1) or vice versa (s_2), or as polymorphisms shared by both populations (ss) (Wakeley and Hey 1997). MIMAR uses a Markov chain Monte Carlo approach based on genealogical coalescence to estimate the posterior probabilities of the model parameters (detailed in
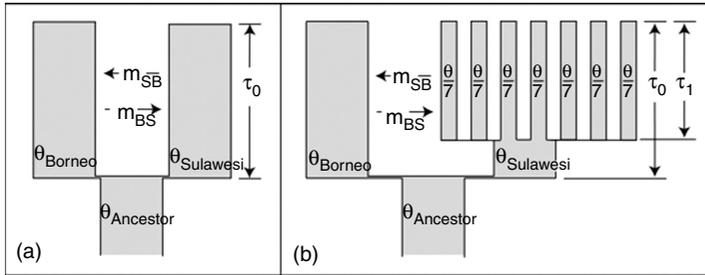
Figure 11.2 Demographic parameters used for simulations and estimated by MIMAR analysis. All simulations were performed without migration between Sulawesi and Borneo ($m_{SB} = m_{BS} = 0$) and the simulations were analysed with models without or with asymmetric migration between Sulawesi and Borneo. (a) Control simulations were performed with essentially no population structure on Borneo or on Sulawesi ($\delta = 0.001$, see text). (b) Simulations were also performed with a model with asymmetric population structure ($\delta = 0.5$) in which the Sulawesi lineage split into 7 demes at time $\tau_1 = 0.5 * \tau_0$ after divergence with Borneo.

Becquet and Przeworski 2007). The aim of this analysis is essentially to assess how frequently coalescent simulations with alternative parameter values generate data with summary statistics similar to those of the observed data.

11.1.3 Violations of models used in divergence population genetics

For computational reasons, models used to study divergence population genetics make simplifying assumptions concerning the evolutionary history of the species being studied – assumptions such as constant population size of each of the three lineages (the ancestral and both descendant lineages), constant rates of migration between the descendant lineages and no population structure in the ancestral or descendant lineages. It is therefore of interest to understand the impact of making these assumptions in situations when there is reason to suspect that they are not true.

Recently, two studies explored the impact of violations of assumptions of the standard models used in divergence population genetics. Becquet and Przeworski (2009) tested the impact of violating assumptions due to ancestral population structure, early gene flow immediately after speciation but not later on, and secondary contact – no gene flow immediately after speciation followed by gene flow after a period of reproductive isolation. Using two approaches in divergence population genetics – MIMAR (Becquet and Przeworski 2007) and IM (Hey and Nielsen 2004), Becquet and Przeworski identified conditions in which biologically plausible violations of model assumptions could lead to biased estimation of model parameters. For example, if the ancestral population is not panmictic, IM and MIMAR both tended to overestimate the ancestral effective population size (Becquet and

Przeworski 2009). Becquet and Przeworski (2009) also found that violation of the assumption of constant rate of migration led to an inappropriate inference of no migration if migration after divergence occurred only in an early phase. Strasburg and Rieseberg (2010) examined other violations including symmetrical population structure in both descendant populations, linkage among loci and gene flow from an unsampled species, using the programs IM and IM_A (Hey and Nielsen 2004, 2007). Strasburg and Rieseberg concluded that estimates of model parameter were generally robust to the violation they considered. For example, analysis of simulations with symmetrical population structure generated 90% highest posterior densities for model parameters that included the true parameter value about as frequently as the control (Strasburg and Rieseberg 2010).

Here I have explored another violation of the standard model used in divergence population genetic analysis that is related to studies of dispersal to the Indonesian island of Sulawesi. I explored how 'asymmetric population structure' – when one of the descendant populations develops substantial population structure while the other remains essentially panmictic – affects conclusions drawn from divergence population genetics analysis using the program MIMAR. This is relevant to studies of dispersal for Sulawesi biogeography because many groups appear to have more substantial population structure on Sulawesi compared to similarly sized portions of Borneo.

11.2 Methods

In an attempt to make the results of these simulations relevant to at least one dataset, I performed simulations to mimic roughly a recent analysis of macaque monkeys from Borneo and Sulawesi by Evans et al. (2010). This includes using the following features: (1) a single dispersal of a macaque ancestor to Sulawesi τ_0 generations ago counting back in time from the present followed by (2) the Sulawesi population being subdivided into seven (equally sized) demes which subsequently exchange no migration from the present going back in time for τ_1 generations, where $\tau_1 < \tau_0$ (Fig 11.2). Simulations matched the dataset of Evans et al. (2010) in terms of the number of loci (29), number of silent sites per locus (for an average of 459 silent sites per locus), the number of autosomal, X-chromosome and Y-chromosome loci (14 autosomal, 14 X-chromosome, 1 Y-chromosome), inheritance and mutation rate scalars of each locus, and the number of individuals sequenced for each locus for each *Macaca* species in the study, which included *M. nemestrina*, *M. nigra*, *M. nigrescens*, *M. hecki*, *M. tonkeana*, *M. maura* and *M. ochreata* (Table 11.1). Mitochondrial DNA sequences were not included by Evans et al. (2010) to avoid polarisation of polymorphisms at this rapidly evolving locus. Additionally, parameter values of the simulations roughly approximate estimated

Table 11.1 Data and sampling from Evans et al. (2010) including gene acronym (Name), number of silent sites sequenced (bp), mutation rate scalar (m scalar), number of samples from each of eight species or populations, and counts of 4 types of segregating sites (s₁, s₂, s_S, s_F) used in MIMAR analysis (see MIMAR documentation for details).

Gene	bp	μ scalar	M. nigra	M. nigrescens	M. hecki	M. tonkeana (West)	M. tonkeana (East)	M. maura	M. ochreata	M. nemestrina (Borneo)	S_1	S_2	S_s	S_f
ASIP	439	1	2	0	4	6	2	0	2	18	6	3	0	0
ATXN10	423	1.7	2	2	2	6	2	2	2	12	6	4	0	0
GPR15	138	0.9	2	2	4	6	2	2	2	16	2	2	0	0
IRBP	452	0.8	2	2	4	6	0	2	2	8	7	4	0	1
KFL10	133	1.3	2	2	4	6	2	0	2	16	3	2	1	0
PDYN	113	0.8	2	2	4	6	2	0	2	18	0	2	0	0
TRIM22	179	1	2	0	2	0	2	0	2	8	3	0	0	0
AFP	751	1	2	0	2	4	0	0	2	16	7	8	1	1
APOE	583	1.2	0	0	2	4	2	0	2	18	7	2	0	6
B2M	644	0.7	2	2	2	4	2	0	2	18	1	9	0	6
beta	376	0.9	2	2	4	4	2	2	2	24	3	5	0	0
CCL2	862	1.2	2	2	4	4	2	0	2	16	15	21	1	0
NRAMP	421	0.6	2	2	4	6	0	2	2	18	7	3	0	0
TTR	858	1.2	0	0	2	2	2	0	2	8	9	13	0	2
AMLEX	454	0.2	1	1	2	3	1	1	1	15	1	0	0	0

CXORF15	506	0.8	1	1	2	2	1	1	1	11	1	1	0	1
DBX	462	1	1	1	2	3	1	1	1	15	4	1	1	1
EFiAX	457	0.9	1	1	0	1	0	0	1	12	2	0	0	1
NLGN4X	212	2	1	1	2	3	1	1	1	15	0	4	1	0
PRX	314	2	1	1	2	3	1	1	1	15	3	9	1	0
RPMX	495	0.5	1	1	2	3	1	1	1	11	3	0	0	0
SMCX	385	0.7	1	1	2	2	1	1	1	12	0	1	0	0
SOX3	209	1.5	1	0	1	3	1	1	0	11	0	0	0	0
TBL1X	1357	1.2	1	1	2	3	1	1	1	12	12	15	1	0
TMSB4X	418	0.6	1	1	2	3	1	1	1	11	0	1	0	0
USP9X	187	0.9	1	1	2	3	1	1	1	14	1	1	0	1
UTX	186	0.3	1	1	2	3	1	1	1	15	0	0	0	0
ZFX	188	0.6	1	1	2	3	1	1	1	13	0	0	0	0
Y	1104	1.6	1	1	2	3	1	1	1	8	15	0	0	4

values for macaques from Sulawesi and Borneo, including divergence time, θ_{Borneo}, $\theta_{Sulawesi}$ and $\theta_{Ancestor}$ (see below).

Coalescent simulations were performed using ms (Hudson 2002). Custom PERL scripts were used to concatenate these simulations and convert them into input files for MIMAR analysis. Following Becquet and Przeworski (2009), simulations and MIMAR analysis were performed without intralocus recombination. The model I considered had five parameters including three mutation parameters $\theta_{Ancestor}$, θ_{Borneo}, $\theta_{Sulawesi}$, the divergence time between the descendant lineages on Borneo and Sulawesi (τ_0) and the time after colonization that the Sulawesi population diverged simultaneously into seven lineages that did not subsequently exchange migrants until the present (τ_1). For all simulations, $\theta_{Ancestor} = \theta_{Borneo} = \theta_{Sulawesi} = 0.002$ (per site) which is similar to results from macaque monkeys on Borneo (Evans et al. 2010). τ_0 was set to $4.4N_e$ generations, a value chosen because it falls within the 95% confidence interval estimated by Evans et al. (2010).

To incorporate mutation rate heterogeneity and differences in effective population size of autosomal DNA, the X-chromosome and the Y-chromosome, θ for each locus was multiplied by a mutation rate scalar (Table 11.1) and by an inheritance scalar. Mutation rate scalars were estimated by Evans et al. (2010) using an outgroup (baboons). Inheritance scalars for each locus were estimated by maximum likelihood (Evans et al. 2010) using a demographic model with unequal proportion of males and females due to sex-specific variation in reproductive success. The maximum likelihood scalars for aDNA, xDNA and yDNA were 1.00, 0.87 and 0.18, respectively, instead of the ideal expectation of 1.00, 0.75 and 0.25. These maximum likelihood estimates are consistent with field observations of macaque societies that suggest that there is higher variation in reproductive success in males compared to females (Dittus 1975, de Ruiter et al. 1992, Keane et al. 1997, Van Noordwijk and Van Schaik 2002, Widdig et al. 2004).

In order to simulate population structure on Sulawesi, the simulations included division of one descendant lineage into seven subpopulations at time τ_1, where $\tau_1 = \delta\tau_0$ and δ is equal to 0.001 or 0.5. When $\delta = 0.001$, the MIMAR model is essentially not violated so this was treated as a control simulation. Twenty simulations were performed for each value of δ, and all of them were analysed under models with and without asymmetric migration parameters between Borneo and Sulawesi. Each model was analysed twice using a different random seed for MIMAR using the sharcnet computer cluster (www.sharcnet.ca).

For each simulation, model comparison was performed using a goodness of fit test (Becquet and Przeworski 2007). This involves comparison of the summary statistics used in MIMAR analysis and also summary statistics not used in the analysis, including F_{ST} (Hudson et al. 1992), the average pairwise number of nucleotide differences for Sulawesi and Borneo, and Tajima's D (Tajima 1989) for Sulawesi and Borneo. Summary statistics were calculated from the simulations using PERL

scripts and the isolation–migration model was considered better if one or more of the test statistics fit the data significantly worse in the isolation model (Fig 11.2; Becquet and Przeworski 2007, 2009).

11.3 Results

The data analysed in this study were generated using simulations with no migration between any of the descendant populations. At least three sources of variation could contribute to differences between the actual and inferred model parameter values. First, even though each simulation was performed with fixed parameter values ($\theta_{Ancestor}$, θ_{Borneo}, $\theta_{Sulawesi}$, τ_0, τ_1), stochastic differences among simulations in the coalescent will impact the posterior distribution of the parameter values recovered by MIMAR analysis. Second, the simulations with population structure on Sulawesi (that is, where $\delta = 0.5$) are expected to have a larger value for $\theta_{Sulawesi}$ because population subdivision with fixed deme size increases effective population size compared to a panmictic population with no subdivision (Nei and Takahata 1993). Third, variation among loci in mutation rates (encapsulated with a mutation rate scalar) could affect parameter estimates, although this effect should be similar in the control ($\delta = 0.001$) and test ($\delta = 0.5$) simulations. And fourth, any biases in the estimation of the posterior distribution of parameter values by MIMAR could also contribute differences in parameter estimates recovered from simulations.

11.3.1 No violations of the isolation model

As a first step, simulations were analysed with essentially panmictic populations in both of the descendant lineages that arose from the ancestral lineage (where $\delta = 0.001$). These simulations were analysed with an isolation model (no migration following divergence of the descendant lineages) and with the isolation–migration model (asymmetric migration after divergence). Goodness of fit tests indicated that in no case was the isolation–migration model preferred over the isolation model (data not shown). The mode of the posterior distributions of $\theta_{Sulawesi}$ and θ_{Borneo} was similar in both models and was close to the actual value of 0.002 in most simulations (Fig 11.3a). In analysis of the control simulations ($\delta = 0.001$) using the isolation model, 90% confidence intervals did not include the actual parameter value twice for $\theta_{Sulawesi}$, twice for θ_{Borneo}, five times for τ_0, and three times for $\theta_{Ancestor}$. In all cases the 90% confidence intervals derived from the same simulation included the actual value for more than one of these parameters. In analysis of the control simulations ($\delta = 0.001$) using the isolation–migration model, 90% confidence intervals did not include the actual parameter value twice for $\theta_{Sulawesi}$, twice for θ_{Borneo}, three times for τ_0, and four times for $\theta_{Ancestor}$. In analyses with the isolation model
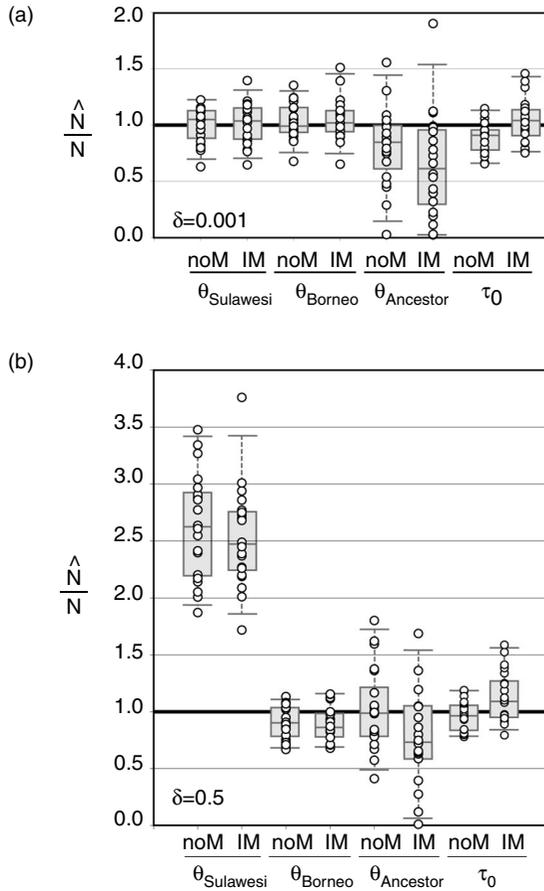
Figure 11.3 Parameter estimates recovered from MIMAR analysis of simulations (a) without ($\delta = 0.001$) and (b) with ($\delta = 0.5$) population structure on Sulawesi using a model without (noM) or with (IM) asymmetric migration between Sulawesi and Borneo. The *y*-axis is the ratio of the mode of the posterior distribution compared to the actual value used in the simulation ($\hat{N}/N$). Box and whiskers plots indicate the first and third quartiles, the medians, and the 95 percentiles of the modes of the posterior distributions.

and with the isolation–migration model, the dispersion of the estimates of $\theta_{Ancestor}$ was greater than the dispersion of estimates of $\theta_{Sulawesi}$ and θ_{Borneo} (Fig 11.3a).

11.3.2 Asymmetric population structure: a violation of the isolation model

When simulations generated with asymmetric population subdivision in one descendant lineage ($\delta = 0.5$) were analysed using the isolation model, parameter estimates were fairly robust to violations of model assumptions. Similar to the

control simulations, in no instance did goodness-of-fit tests prefer the isolation–migration model over the isolation model (data not shown). When these simulations were analysed using the isolation model, 90% confidence intervals did not include the actual parameter value three times for θ_{Borneo}, three times for τ_0, and once for $\theta_{Ancestor}$. The 90% confidence interval did not include the actual value for θ_{Borneo} and for τ_0 for analysis of one simulation. When these simulations were analysed using the isolation–migration model, 90% confidence intervals did not include the actual parameter value three times for θ_{Borneo}, once for τ_0, and twice for $\theta_{Ancestor}$, and these inaccuracies were all from analyses of different simulations. As predicted by theory, estimates of $\theta_{Sulawesi}$ were higher when $\delta = 0.5$ (Fig 11.3). In the simulations where $\delta = 0.5$, the actual parameter value for $\theta_{Sulawesi}$ was lower than the 90% confidence interval in all of the analyses with the isolation model and also in all analyses with the isolation–migration model (Fig 11.3b).

11.4 Discussion

A population at mutation–drift equilibrium has a roughly constant level of polymorphism over time because variation is introduced by mutation at the same rate that it is removed by genetic drift. Divergence population genetics explores systems that are not in mutation–drift equilibrium because population size has changed, and because the system as a whole is no longer panmictic (even though each descendant population is panmictic under model assumptions). This study used simulations and a coalescent-based analysis – MIMAR (Becquet and Przeworski 2007) – to explore how violation of a demographic model used in divergence population genetics could impact inferences concerning non-equilibrium systems based on multi-locus DNA sequence data. Using the biogeographical history of macaque monkeys on the Southeast Asian islands of Sulawesi and Borneo as an example, data were simulated in which one descendant population on Sulawesi acquired population structure, whereas another on Borneo remained panmictic. Analysis of these simulations is of interest because multiple species on Sulawesi suggest that population structure on this island is quite common. In particular, multiple distantly related lineages such as toads and monkeys, tarsiers and bats, and flying lizards and fanged frogs, have similar geographic distributions of substantially differentiated populations on Sulawesi. The degree to which patterns are shared across species is the culmination of multiple variables including the timing of initial colonisation, dynamic abiotic factors such as marine inundation, ecological transitions and heterogeneity, and species-specific phenomena such as adaptation, ecological tolerance, genetic drift and demography.

Overall MIMAR analyses of the simulations performed here suggest that this model-based approach to studying biogeography is robust to extreme model

violation by asymmetrical population structure in descendant populations. While dispersion of the parameter estimate for the ancestral polymorphism parameter $\theta_{Ancestor}$ tended to be larger than that of the polymorphism parameters for the extant populations ($\theta_{Sulawesi}$ and θ_{Borneo}), this difference was not substantially affected by asymmetric population structure in the descendant populations (Fig 11.3). Asymmetric population structure did not lead to an inappropriate inference of ongoing migration between Borneo and Sulawesi, and estimates of the timing of divergence of populations on Sulawesi and Borneo appeared to not be substantially biased. This finding is similar to those reported for symmetrical population structure by Strasburg and Rieseberg (2010). An unsurprising caveat to these conclusions is that asymmetric population structure did inflate estimates of the polymorphism parameter of the structured lineage ($\theta_{Sulawesi}$), a result that is predicted by population genetic theory. Becquet and Przeworski (2009) recovered a similar result for $\theta_{Ancestor}$ in their study of a structured ancestral population. Similar to the findings of Becquet and Przeworski (2007), control simulations (without population structure on Sulawesi) produced broader posterior distributions for $\theta_{Ancestor}$ than for $\theta_{Sulawesi}$ or θ_{Borneo} (Fig 11.3).

In MIMAR, inferences about migration and ancestral population sizes are based on variance in coalescence times across loci (Becquet and Przeworski 2007, 2009). Larger ancestral populations are expected to have more variation among loci in coalescence times because of ancestral polymorphism at the moment of speciation. Migration between descendant lineages is also expected to increase variance in coalescence times across loci compared to a model with no migration. In an isolation model, all loci must coalesce at or before the time of speciation. Theory predicts that after ~9 to $12N_e$ generations, where N_e is the effective population size of each descendant population, greater than 95% of the loci should be reciprocally monophyletic in two diverged populations (Hudson and Coyne 2002). In the simulations analysed here, divergence time was $4.4N_e$ generations, so presumably many of the loci are not reciprocally monophyletic in each of the simulated descendant populations. This is also suggested by the observed data from Evans et al. (2010), on which the simulations were based, which have relatively few fixed differences between Sulawesi and Borneo (Table 11.1). These data are instead characterised by a preponderance of derived segregating sites that are found in only one or the other island, although relatively few polymorphisms are shared by populations on both islands.

In these simulations, population subdivision increases the overall effective population size (Nei and Takahata 1993). Consequently, alleles shared with the other descendant population (that is, the simulated panmictic population representing individuals from Borneo) due to ancestral polymorphism are less likely to be lost because of drift on Sulawesi. The net effect of population subdivision, therefore, could be to increase the variance in coalescence times across loci. This in

turn could contribute to an incorrect signature of migration or errors in estimates of $\theta_{Ancestor}$. However, this does not appear to be the case for the parameter values considered here, and the recent analysis of macaque migration between Borneo and Sulawesi also did not recover significant support for migration (Evans et al. 2010). Instead of accounting for this variance with migration or a larger $\theta_{Ancestor}$, MIMAR correctly inflated the estimate of $\theta_{Sulawesi}$. This is undoubtedly a consequence of the large number of Sulawesi-specific polymorphisms compared to the number of polymorphisms shared between Borneo and Sulawesi in the simulations. The impact of asymmetric population structure could be more substantial if θ_{Borneo} were larger, if $(\tau_0 - \tau_1)$ were smaller, and/or if τ_0 were smaller, and this is an interesting avenue for further work. Also relevant to the conclusions of Evans et al. (2010) but not considered here is an assessment of what level of migration could be missed by coalescent-based analysis using isolation and isolation-migration models – that is, whether there could be an impact of asymmetric population structure on the incidence of false negatives for ongoing migration.

11.4.1 Future directions

When and from where did ancestors of Sulawesi endemics colonise this island? To what degree do patterns of differentiation and timings of population subdivision correspond among species? Did closely related Sulawesi endemics evolve *in situ*? We have only begun to address these questions and many exciting studies lie ahead. Model-based approaches offer a promising strategy to help us identify biogeographical trends and exceptions using molecular data. An obvious application of this research is for conservation planning on Sulawesi. While not all species will match areas of endemism identified in species discussed here, the available data argue strongly for a 'distributed' approach to conservation on Sulawesi wherein different portions of this island are independently targeted for management, as opposed to targeting a single ecosystem or region.

Acknowledgements

References

Avise, J. C. (2000). *Phylogeography. The History and Formation of Species.* Cambridge, MA: Harvard University Press.

Becquet, C. and Przeworski, M. (2007). A new approach to estimate parameters of speciation models with application to apes. *Genome Research*, **17**, 1505–19.

Becquet, C. and Przeworski, M. (2009). Learning about modes of speciation by computational approaches. *Evolution*, **63**, 2547–62.

Bickford, D. P., Supriatna, J., Andayani, N. et al. (2007). *Indonesia's Protected Areas Need More Protection: Suggestions from Island Examples*. Cambridge: Cambridge University Press.

Bridle, J. R., Garn, A., Monk, K. A. and Butlin, R. K. (2001). Speciation in *Chitaura* grasshoppers (Acrididae: Oxyinae) on the island of Sulawesi: colour patterns, morphology and contact zones. *Biological Journal of the Linnean Society*, **72**, 373–90.

Bridle, J. R., Pedro, P. M. and Butlin, R. K. (2004). Habitat fragmentation and biodiversity: testing for the evolutionary effects of refugia. *Evolution*, **58**, 1394–6.

Brown, R. M., Linkem, C. W., Siler, C. D. et al. (2010). Phylogeography and historical demography of *Polypedates leucomystax* in the islands of Indonesia and the Philippines: Evidence for recent human-mediated range expansion? *Molecular Phylogenetics and Evolution*, **57**, 598–619.

Burton, J. A., Hedges, S. and Mustari, A. H. (2005). The taxonomic status, distribution and conservation of the lowland anoa *Bubalus depressicornis* and mountain anoa *Bubalus quarlesi*. *Mammal Review*, **35**, 25–50.

Campbell, P., Putnam, A. S., Bonney, C. et al. (2007). Contrasting patterns of genetic differentiation between endemic and widespread species of fruit bats (Chiroptera: Pteropodidae) in Sulawesi, Indonesia. *Molecular*

Phylogenetics and Evolution, **44**, 474–82.

de Queiroz, K. (1998). The general lineage concept of species, species criteria, and the process of speciation. In *Endless Forms: Species and Speciation*, ed. D. Howard and S. Berlocher. New York: Oxford Press, pp. 57–75.

de Ruiter, J. R., Scheffrahn, W., Trommelen, G. J. J. M. et al. (1992). Male social rank and reproductive success in wild long-tailed macaques. In *Paternity in primates: genetic tests and theories*. ed. R. D. Martin, A. F. Dixon and E. J. Wickings. Basel, Switzerland: Karger, pp. 175–191.

Dittus, W. (1975). Population dynamics of the toque macaque, *Macaca sinica*. In *Socioecology and Psychology of Primates*. ed. R. H. Tuttle. The Hague: Mounton, pp. 125–52.

Edwards, S. V., Lie, L. and Pearl, D. K. (2007). High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences*, **104**, 5936–41.

Evans, B. J., Morales, J. C., Supriatna, J. and Melnick, D. J. (1999). Origin of the Sulawesi macaques (Cercopithecidae, *Macaca*) as inferred from a mitochondrial DNA phylogeny. *Biological Journal of the Linnean Society*, **66**, 539–60.

Evans, B. J., Supriatna, J. and Melnick, D. J. (2001). Hybridization and population genetics of two macaque species in Sulawesi, Indonesia. *Evolution*, **55**, 1685–702.

Evans, B. J., Brown, R. M., McGuire, J. A. et al. (2003a). Phylogenetics of fanged frogs (Anura; Ranidae; *Limnonectes*): testing biogeographical hypotheses at the Asian–Australian faunal zone interface. *Systematic Biology*, **52**, 794–819.

Evans, B. J., Supriatna, J., Andayani, N. and Melnick, D. J. (2003b). Diversification of Sulawesi macaque monkeys: decoupled evolution of mitochondrial and autosomal DNA. *Evolution*, **57**, 1931–46.

Evans, B. J., Supriatna, J., Andayani, N. et al. (2003c). Monkeys and toads define areas of endemism on Sulawesi. *Evolution*, **57**, 1436–43.

Evans, B. J., Cannatella, D. C. and Melnick, D. J. (2004). Understanding the origins of areas of endemism in phylogeographic analyses: a reply to Bridle et al. *Evolution*, **58**, 1397–1400.

Evans, B. J., McGuire, J. A., Brown, R. M., Andayani, N. and Supriatna, J. (2008). A coalescent framework for comparing alternative models of population structure with genetic data: evolution of Celebes toads. *Biology Letters*, **4**, 430–3.

Evans, B. J., Pin, L., Melnick, D. J. and Wright, S. I. (2010). Sex-linked inheritance in macaque monkeys: implications for effective population size and dispersal to Sulawesi. *Genetics*, **185**, 923–37.

George, W. (1981). Wallace and his line. In *Wallace's Line and Plate Tectonics*, ed. T. C. Whitmore. Oxford: Clarendon Press, pp. 3–8.

Hall, R. (2001). Cenozoic reconstructions of SE Asia and the SW Pacific: changing patterns of land and sea. In *Faunal and Floral Migrations and Evolution in SE Asia–Australia*, ed. I. Metcalfe, J. Smith, M. Morwood and I. Davidson. Lisse, The Netherlands: Swets and Zeitlinger Publishers, pp. 35–56.

Hey, J. and Nielsen, R. (2004). Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics*, **167**, 747–60.

Hey, J. and Nielsen, R. (2007). Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences*, **104**, 2785–90.

Hudson, R. R. (1991). Gene genealogies and the coalescent process. *Oxford Surveys in Evolutionary Biology*. **7**, 1–44.

Hudson, R. R. (2002). Generating samples under a Wright–Fisher neutral model of genetic variation. *Bioinformatics*, **18**, 337–8.

Hudson, R. R. and Coyne, J. A. (2002). Mathematical consequences of the genealogical species concept. *Evolution*, **56**, 1557–65.

Hudson, R. R., Slatkin, M. and Maddison, W. P. (1992). Estimation of levels of gene flow from DNA sequence data. *Genetics*, **132**, 583–9.

Huxley, T. H. (1868). On the classification and distribution of the Alectoromorphae and Heteromorphae. *Proceedings of the Zoological Society of London*, **1868**, 294–319.

Iskandar, D. T. and Tjan, K. N. (1996). The amphibians and reptiles of Sulawesi, with notes of the distribution and chromosomal number of frogs. In *Proceedings of the First International Conference on Eastern Indonesian-Australian Vertebrate Fauna. Manado, Indonesia, Nov 22–26*, ed. D. J. Kitchner and A. Suyanto, pp. 39–46.

Ives, I., Spinks, P. Q. and Shaffer, H. B. (2008). Morphological and genetic variation in the endangered Sulawesi tortoise *Indotestudo forstenii*; evidence of distinct lineages? *Conservation Genetics*, **9**, 709–713.

Keane, B., Dittus, W. P. and Melnick, D. J. (1997). Paternity assessment in wild

groups of toque macaques *Macaca sinica* at Polonnaruwa, Sri Lanka using molecular markers. *Molecular Ecology*, **6**, 267–82.

Kliman, R. M., Andolfatto, P., Coyne, J. A. et al. (2000). The population genetics of the origin and divergence of the *Drosophila simulans* complex species. *Genetics*, **156**, 1913–31.

Larson, G., Dobney, K., Albarella, U. et al. (2005). Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, **3007**, 1618–21.

Leman, S. C., Chen, Y., Stajich, J. E., Noor, M. A. F. and Uyenoyama, M. (2005). Likelihoods from summary statistics: recent divergence between species. *Genetics*, **171**, 1419–36.

Lydekker, R. (1896). *A Geographical History of Mammals.* Cambridge: Cambridge University Press.

McGuire, J. A., Brown, R. M., Mumpuni, Riyanto, A. and Andayani, N. (2007). The flying lizards of the *Draco lineatus* group (Squamata: Iguania: Agamidae): a taxonomic revision with descriptions of two new species. *Herpetological Monographs*, **21**, 179–212.

Meijaard, E. and Groves, C. (2002). Upgrading three subspecies of babirusa (*Babyrousa* sp.) to full species level. *Asian Wild Pig News*, **2**, 33–9.

Merker, S., Driller, C., Pertwitasari-Farajallah, D., Pamungkas, J. and Zischler, H. (2009). Elucidating geological and biological processes underlying the diversification of Sulawesi tarsiers. *Proceedings of the National Academy of Sciences*, **106**, 8459–64.

Musser, G. G. (1987). The mammals of Sulawesi. In *Biogeographical Evolution of the Malay Archipelago*, ed.

T. C. Whitmore. Oxford: Clarendon Press, pp. 73–93.

Musser, G. G., Durden, L. A., Holden, M. E. and Light, J. E. (2010). Systematic review of endemic Sulawesi squirrels (Rodentia, Sciuridae), with descriptions of new species of associated sucking lice (Insecta, Anoplura), and phylogenetic and zoogeographic assessments of sciurid lice. *Bulletin of the American Museum of Natural History*, **339**, 1–160.

Nei, M. and Takahata, N. (1993). Effective population size, genetic diversity and coalescence time in subdivided populations. *Journal of Molecular Evolution*, **37**, 240–44.

Nielsen, R. and Wakeley, J. (2001). Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–96.

Nordborg, M. (2001). Coalescent theory. In *Handbook of Statistical Genetics*, ed. D. Balding, M. Bishop and C. Cannings. Chichester, UK: Wiley, pp. 1–37.

Putnam, A. S., Scriber, J. M. and Andolfatto, P. (2007). Discordant divergence times among Z-chromosome regions between two ecologically distinct swallowtail butterfly species. *Evolution*, **61**, 912–27.

Ruedi, M., Auberson, M. and Savolainen, V. (1998). Biogeography of Sulawesian shrews: testing for their origin with a parametric bootstrap on molecular data. *Molecular Phylogenetics and Evolution*, **9**, 567–71.

Setiadi, M. I., McGuire, J. A., Brown, R. M. et al. (2011). Adaptive radiation and ecological opportunity in Sulawesi and Philippine fanged frogs (Limnonectes). *American Naturalist*, **178**(2), 221–40.

Shekelle, M. and Leksono, S. M. (2004). Rencana konservasi di Pulau Sulawesi:

dengan menggunakan tarsius sebagai 'flagship taxon'. *Biota*, **9**, 1–10.

Shekelle, M., Leksono, S. M., Ichwan, L. L. S. and Masala, Y. (1997). The natural history of the tarsiers of North and Central Sulawesi. *Sulawesi Primate Newsletter*, **4**, 3–8.

Strasburg, J. L. and Rieseberg, L. H. (2010). How robust are 'isolation with migration' analyses to violations of the IM model? A simulation study. *Molecular Biology and Evolution*, **27**, 297–310.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–95.

Van Noordwijk, M. A. and Van Schaik, C. P. (2002). Career moves: transfer and rank challenge decisions by male long-tailed macaques. *Behavior*, **138**, 359–95.

van der Vecht, J. (1953). The carpenter bees (*Xylocopa* Latr.) of Celebes. *Idea*, **9**, 57–9.

Wakeley, J. and Hey, J. (1997). Estimating ancestral population parameters. *Genetics*, **145**, 847–55.

Wallace, A. R. (1863). On the physical geography of the Malay Archipelago. *Journal of the Royal Geographical Society*, **1863**, 217–34.

Whitten, T., Henderson, G. S. and Mustafa, M. (2002). *The Ecology of Sulawesi*. Hong Kong: Periplus Editions Ltd.

Widdig, A., Bercovich, F. B., Streich, W. J., Sauermann, U., Nürnberg, P. and Krawczak, M. (2004). A longitudinal analysis of reproductive skew in male rhesus macaques. *Proceedings of the Royal Society of London B*, **271**, 819–26.

Wright, S. (1943). Isolation by distance. *Genetics*, **28**, 114–38.

Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–65.