

Speciation Has a Spatial Scale That Depends on Levels of Gene Flow

Yael Kisel* and Timothy G. Barraclough

Division of Biology and Natural Environment Research Council Centre for Population Biology, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, United Kingdom; and Jodrell Laboratory, Royal Botanic Gardens, Kew TW9 3DS, United Kingdom

Submitted February 23, 2009; Accepted September 24, 2009; Electronically published January 25, 2010

Dryad data: <http://datadryad.org/repo/handle/10255/dryad.887>.

ABSTRACT: Area is generally assumed to affect speciation rates, but work on the spatial context of speciation has focused mostly on patterns of range overlap between emerging species rather than on questions of geographical scale. A variety of geographical theories of speciation predict that the probability of speciation occurring within a given region should (1) increase with the size of the region and (2) increase as the spatial extent of intraspecific gene flow becomes smaller. Using a survey of speciation events on isolated oceanic islands for a broad range of taxa, we find evidence for both predictions. The probability of in situ speciation scales with island area in bats, carnivorous mammals, birds, flowering plants, lizards, butterflies and moths, and snails. Ferns are an exception to these findings, but they exhibit high frequencies of polyploid and hybrid speciation, which are expected to be scale independent. Furthermore, the minimum island size for speciation correlates across groups with the strength of intraspecific gene flow, as is estimated from a meta-analysis of published population genetic studies. These results indicate a general geographical model of speciation rates that are dependent on both area and gene flow. The spatial scale of population divergence is an important but neglected determinant of broad-scale diversity patterns.

Keywords: speciation, gene flow, oceanic islands, spatial scale.

Introduction

Although area is generally expected to affect speciation rates (MacArthur and Wilson 1967; Endler 1977; Rosenzweig 1995), most work on the spatial context of speciation has focused on patterns of range overlap between emerging species while ignoring questions of geographical scale (Mayr 1942; Butlin et al. 2008). Geographical theories of speciation predict that the probability of speciation occurring within a given region should (1) increase with the size of the region because of the greater opportunity for divergence within the region (MacArthur and Wilson

1967; Endler 1977; Rosenzweig 1995; Gavrillets and Vose 2005; Losos and Parent 2009) and (2) increase as the strength of gene flow decreases, for example, among organisms with shorter dispersal distances. Gene flow is the main process opposing population differentiation (Mayr 1963), and so the strength of gene flow between populations is expected to be an important determinant of the spatial scale at which genetic divergence and speciation can occur (Slatkin 1973, 1985; Doebeli and Dieckmann 2003). However, despite the potential of this body of theory to explain taxonomic and geographic variation in biodiversity (Ricklefs 2007), the extent to which the scale of speciation varies among taxa and the causes of such variation remain unknown.

Oceanic islands are useful for studying speciation because their well-defined boundaries and isolation make it easier to distinguish within-island (in situ) speciation from immigration in these regions than it is in continental regions. Several studies have used islands to explore the relationship between speciation rates and area. Diamond (1977) noted the lack of bird speciation in Pacific land masses that are smaller than New Zealand, but he also observed that insects, lizards, and ferns had diversified within smaller islands such as New Caledonia. Coyne and Price (2000) found no evidence worldwide for speciation in birds within oceanic islands that are smaller than 10,000 km², setting a lower bound for their minimum area for speciation. Losos and Schluter (2000) estimated the minimum area for speciation in Caribbean *Anolis* lizards to be 3,000 km², and they found that speciation rates increased linearly with island area above this limit. Similar relationships were found in cichlid fish in African lakes (Seehausen 2006) and *Bulimulus* snails in the Galapagos (Parent and Crespi 2006; Losos and Parent 2009), but with different minimum areas for speciation (in cichlids, <1 km²; in snails, 18.1 km²). Together with case studies of speciation on small islands (cichlids in crater lakes: Schliwien et al. 1994; Barluenga et al. 2006; palms on Lord Howe Island: Savolainen et al. 2006), these studies suggest

* Corresponding author; e-mail: y.kisel06@imperial.ac.uk.

that the spatial scale of speciation varies widely among taxa. However, only a few taxa have been investigated, and those have been located on different sets of islands; comparison of several taxa across a broad range of island sizes is needed to quantify taxonomic variation in the spatial scale of speciation and to identify its cause. To address this, we surveyed speciation events for a broad range of taxa on oceanic islands from around the world.

Islands also vary in many other factors that might affect rates of diversification (Carlquist 1974; Bauer 1988; Paulay 1994; Rosenzweig 1995). Even if a speciation-area relationship exists, it need not be the direct result of area; for instance, larger islands tend to have higher habitat diversity, which could foster higher rates of ecological speciation (Losos and Parent 2009). Island age might also affect diversification, either by increasing the time over which speciation can occur or through other effects related to the dynamics of island aging (Gillespie 2004; Emerson and Oromi 2005; Sequeira et al. 2008; discussed in Whittaker et al. 2009). In addition, the degree of isolation from other landmasses might affect speciation rates if lower colonization rates to more isolated islands leave more niches open to be filled by in situ speciation (Gillespie and Baldwin 2009). Our aim here is to use islands as a model to study the spatial scale of speciation rather than to explain island diversification in all its detail. However, because these other factors—especially habitat diversity and age—are likely to be correlated with island area, we included them in our analysis to be able to separate out the effects of area itself.

Another complication is the existence of archipelagos. As extreme examples of habitat fragmentation, archipelagos are expected to promote higher levels of diversification, especially for taxa that disperse well over land but not over water (Diamond 1977; Losos and Parent 2009). However, the degree to which the rate of speciation will be increased within an archipelago should depend on the dispersal ability of the taxon and the size of the water gaps between islands. For some taxa, barriers within islands may already be sufficiently strong isolating factors that rates of speciation are no higher in archipelagos than they are in single islands of comparable size. Therefore, we repeated our analyses, both including and excluding archipelagos, in order to test their effect on the probability of speciation.

Many traits of organisms and species have been hypothesized to affect rates of speciation (Jablonski 2008), but when viewed in a geographic context, dispersal ability is expected to be key. This has especially been argued in the specific context of oceanic islands (Diamond 1977; Paulay 1994; Ranker et al. 1994; Parent and Crespi 2006; Whittaker and Fernandez-Palacios 2007; Gillespie and Baldwin 2009; Givnish et al. 2009), where there are many examples of spectacular radiations of taxa with normally

poor dispersal abilities but a propensity for passive long-distance dispersal (e.g., weevils on Rapa [Paulay 1985] and snails on Bonin [Chiba 1999]). Diamond (1977) argued more specifically that dispersal ability might determine the threshold island area necessary for within-island speciation to take place, but this idea remains untested.

In this study, we used a comparative approach to measure the extent of variation in the spatial scale of speciation and to test the importance of gene flow in controlling this variation. To quantify the speciation-area relationship and the spatial scale of speciation, we surveyed the probability of in situ speciation on islands of different sizes for angiosperms, bats, birds, mammals of the order Carnivora, ferns, lizards, Macrolepidoptera (large butterflies and moths), and land snails. These taxa were chosen on the basis of the availability of required data (see “Material and Methods”), but they also represent a broad taxonomic range that varies in presumed dispersal ability. As a measure of the probability of in situ speciation on islands, we used the proportion of endemic lineages derived from single immigration events that have diverged within an island into two or more descendent species (Coyne and Price 2000; fig. 1). We also tested the relationship between the probability of in situ speciation and other island factors that could potentially confound our analysis of the speciation-area relationship.

After establishing the extent of variation in the spatial scale of speciation, we tested the importance of gene flow in setting the spatial scale of speciation by correlating the minimum area for speciation in each group with an independent measure of the strength of gene flow derived from the population genetic literature. To get comparable estimates of the strength of gene flow for the study taxa, F_{st} values were compiled from the molecular ecology literature for each taxon, along with a measure of the geographical scale of each study. F_{st} provides a measure of the genetic differentiation of populations within a species (0 = no differentiation, 1 = complete differentiation; Wright 1931) that should be robust to variation in the spatial arrangement of populations and the type of genetic marker used for analysis (Beaumont and Nichols 1996). It correlates strongly with broad-scale taxonomic variation in dispersal ability, and it does so more consistently than other population genetic measures such as number of migrants (Nm ; Bohonak 1999). When estimated from appropriate data—namely, from neutral loci unaffected by selection and from populations that are currently connected by gene flow and that have not undergone dramatic historical movements (Barton 2001)— F_{st} values provide a measure of the strength of gene flow that is comparable between species.

We found strong evidence for a general positive relationship between area and the probability of in situ spe-

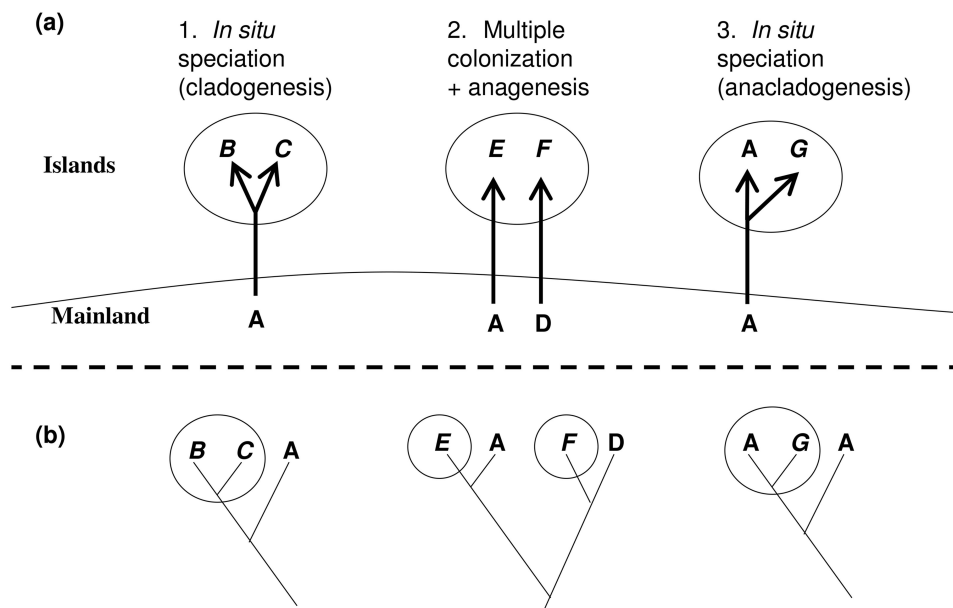


Figure 1: Patterns of diversification on islands. A and D are two species in the same genus, both of which are native to the mainland. B, C, E, F, and G (in italics) are endemic island species. *a*, Relationship between patterns of diversification and numbers of endemic species. (1) shows in situ, cladogenetic speciation (Stuessy et al. 2006) in which a mainland species reaches an island and subsequently splits within the island into two new species. (2) shows how multiple colonization followed by anagenetic change, with no diversification within the island, can create the same pattern of multiple species within one genus that are endemic to the same island (Coyné and Price 2000; Stuessy et al. 2006). (3) shows in situ, anacladogenetic speciation (Stuessy et al. 2006), in which a mainland species reaches an island and remains unchanged while budding off an endemic daughter species. *b*, Phylogenetic patterns resulting from these three modes of island diversification. Each phylogeny is presented underneath the diversification mode that produces it. Species native to islands are circled; tree nodes within the circles represent within-island speciation events.

ciation. Furthermore, we found that major taxa vary over more than six orders of magnitude in the minimum area required for speciation and that up to 76% of this variation can be explained by the variation between taxa in the average strength of gene flow. The results support the importance of area and gene flow in controlling speciation rates.

Material and Methods

Island Selection and Data Collection

Islands were selected on the basis of their level of isolation; we used only islands that are located at least 100 km away from any other landmass, including continents and other islands. We used only isolated islands in order to minimize the chance of continuing gene flow from outside populations of colonizing species and to minimize the chance of multiple colonizations leading to apparent speciation when no diversification has occurred within the island. Of the islands for which appropriate data were available, we excluded only Easter Island (Rapa Nui) because of its long history of habitat degradation and human-caused extinc-

tion before discovery by European taxonomists (Diamond 2005).

Isolated archipelagos were used as units of study when the longest leg in the minimum network connecting all of the islands was less than 100 km long, following the logic of Coyne and Price (2000). In all cases, the distance between islands was at least half the distance of the archipelago from the nearest other landmass. The area of each archipelago was calculated as the area of a 15–19-point minimum-spanning polygon enclosing all islands in the archipelago, using the Analyzing Digital Images software package (Pickle and Kirtley 2008). We repeated all analyses both with and without the inclusion of archipelagos. In the multiple-regression models that included archipelagos, we tested whether speciation is more likely within archipelagos than on single islands by including a term “ArchYN,” which was coded as a 0 (single island) or 1 (archipelago). Except where indicated, we present the results from analyses that included archipelagos.

Data on island area, elevation, isolation, and latitude were collected largely from the United Nations Environment Program (UNEP) online Island Directory database. Isolation was calculated following UNEP methods for “Iso-

Table 1: Area and speciation statistics, by taxonomic group

Group	Minimum area for speciation (km ²) ^a	No. islands with endemic species	No. islands with speciation events
Snails	.8	30 (17)	24 (12)
Angiosperms	15	32 (24)	21 (14)
Ferns	15	17 (11)	9 (5)
Birds	64 or 705	50 (33)	9 (1)
Lizards	108	27 (14)	10 (3)
Macrolepidoptera	141,200	6 (3)	1 (0)
Bats	416,400	14 (5)	2 (1)
Carnivora	587,713	2 (2)	1 (1)

Note: All island sizes >1 km² are shown rounded to the nearest square kilometer. Numbers in parentheses are from the data set that excluded archipelagos.

^a Estimated as the area of the smallest island with a speciation event: for snails, Nihoa; for angiosperms and ferns, Lord Howe; for birds, Norfolk (smaller value) or Tristan da Cunha; for lizards, Rodrigues; for Macrolepidoptera, Fiji; for bats, New Zealand; and for Carnivora, Madagascar.

lation Index” (Dahl 2004). Island data missing from the UNEP database were collected from primary literature or from government databases, where possible. Island statistics and references are given in table A1, which can be found in Dryad (<http://datadryad.org/repo/handle/10255/dryad.887>).

Island Species Data Collection

Study taxa were chosen on the basis of the availability of sufficient and comparable data, but also with the aim of representing a broad taxonomic range of plants and animals with presumed differences in dispersal ability. In the context of this study, “bats” refers to members of the order Chiroptera; “ferns” refers to those of class Filicopsida, phylum Pteridophyta, excluding Psilotopsida, Lycopsidea, and Equisetopsida; “lizards” refers to the order Squamata, excluding snakes and amphisbaenians; “snails” refers to terrestrial pulmonate snails in the orders Stylommatophora, Mesurethra, Heterurethra, and Sigmurethra; and “Macrolepidoptera” refers to butterflies and moths in the superfamilies Bombycoidea, Lasiocampoidea, Axioidea, Calliduloidea, Hedyloidea, Drepanoidea, Geometroidea, Hesperioidea, Mimalonoidea, Noctuidea, Papilionoidea, Sphingoidea, and Uranioidea.

Data were collected for all group/island combinations for which complete (to the best ability of the source authors) taxon lists, including endemism information, were found in the libraries of the Natural History Museum, London, and the Royal Botanic Gardens, Kew; in online sources (Avibase: Lepage 2008; Flora of Australia Online: Australian Biological Resources Study 2008); or in databases that were available to us (island lizards and carnivores worldwide: S. Meiri, unpublished data; endemic plants of selected islands: R. Salguero-Gomez, unpublished data; checklist of Mascarene plants: C. Thebaud, unpub-

lished data). The large scope of the study prohibited a comprehensive survey of recently published journal articles, and so we may not have found the most recent species lists in a few cases. Data were not available for all taxa on all islands included in the analysis (numbers of islands with data for each taxon given in table 1).

For each group/island combination where data were available, the names and endemism statuses of all native species were recorded according to the source. Therefore, in practice we used the species concepts held by the taxonomists who wrote the species lists. Species whose endemism statuses were in doubt were treated as nonendemic. The species lists we used are available upon request. Genera with apparent speciation events are listed in table A2, and sources for these species data are summarized in table A3 (both tables are available in Dryad).

Identification of Speciation Events

Following the method of Coyne and Price (2000), our measure of the probability of speciation within a given island for a given taxon is the proportion of endemic lineages derived from single immigration events that have diversified within an island into two or more descendent species. This approach controls for differences among islands in the number of colonizers because it divides the number of speciated lineages by the total number of lineages that colonized the island that could have speciated. We used a binary measure because our interest is in what controls the ability of lineages to speciate at all, rather than what controls the size of radiations.

We considered the number of genera with at least one endemic species to represent the number of endemic lineages and the number of genera with two or more endemic species to represent the number of lineages that have diversified in situ (Coyne and Price 2000; Stuessy et al. 2006).

Therefore, our measure of the probability of speciation on a particular island was the number of genera with two or more endemic species divided by the number of genera with one or more endemic species. We used only genera with endemic species (and not all native genera) in order to exclude lineages that have not been isolated enough from mainland populations or that have not been on an island long enough to speciate within that island. Because we used only genera with endemic species, we excluded the following islands and island groups, for which we found no record of endemic species in our study groups, from our data set: the Bounty Islands, Caroline Island, Cartier Island, Cocos (Keeling), Diego Garcia, the Gilbert Islands, the Hall Islands, Heard and McDonald, Niuaotupapu and Tafahi, the Prince Edward Islands, Tokelau, and Uvea.

Our measure ignores any cases of in situ speciation that have occurred through anacladogenesis (in which a daughter species diverges from an ancestral colonizing species that remains unchanged; fig. 1; Stuessy et al. 1990; Coyne and Price 2000) because the detailed morphological and genetic data required to identify such cases were not available for most genera. The rate of anacladogenetic speciation should vary with island characteristics in the same manner as the rate of cladogenetic speciation (which we are measuring), as they are both the result of essentially the same process of divergence. Thus, we expect that the inclusion of such speciation events would not qualitatively change our conclusions.

The minimum area for speciation of each taxon was estimated as the area of the smallest island or archipelago within which speciation has occurred. Statistics on speciation and endemism in each group are given in tables 1 and A3 (available in Dryad).

Adding Phylogenetic Information

Our method assumes that the chance of multiple endemic species within the same genus originating by multiple colonization events, rather than by in situ diversification, is rare (fig. 1). To validate this assumption, we searched the literature for phylogenetic information on the study genera. For each genus associated with a putative speciation event, searches were performed in TreeBase, National Center for Biotechnology Information GenBank Core-Nucleotide, and the Institute for Scientific Information (ISI) Web of Science to find published molecular phylogenies. All phylogenies that included more than one of the endemic species of a study genus with a putative speciation event on a particular island or archipelago were used. Cases of congeneric endemic species shown to not be each other's closest relatives were reclassified as multiple, non-speciated lineages (see Losos and Parent 2009). In a few

instances, published phylogenies indicated that multiple genera containing endemic species were all part of one larger endemic clade, and in this case, the genera involved were considered in the analyses to be a single speciated lineage. Also, in some cases, published phylogenies indicated that the endemic species in one genus were the result of multiple colonizations followed by multiple radiations; in such a case, the genus was treated as a grouping of multiple independent speciated lineages. Results of the phylogeny search are given in tables A2 and A4 (available in Dryad).

Statistical Analysis of the Speciation-Area Relationship

Speciation was treated as a binary response variable: each endemic lineage on each island has a value of 0 for speciation if it contains only one endemic species, and it has a value of 1 if it contains two or more endemic species. Overall regression models between this response variable and island area, considering all taxa together, were performed using the lme4 package in R, version 2.5 (Bates 2007; R Development Core Team 2007), for generalized linear mixed-effects models, using taxon as a random effect, a binomial error structure, and Laplace approximation for maximum likelihood estimates. Individual regression models for each taxon were performed in R using generalized linear models with a binomial error structure. R^2 values for all models were calculated with the formula $(SST - SSE)/SST$, where SSE is the deviance of the model and SST is the deviance of a null model (for mixed-effects models, consisting only of a different intercept for each taxon, which represents the mean probability of speciation over all islands).

We also constructed multiple-regression models to investigate the importance of the other island characteristics that might affect speciation probability. For both the individual-taxon models and the overall models considering all taxa together, we began with a maximal additive model including area, elevation as a proxy for habitat diversity in the absence of a direct measure (Ackerman et al. 2007), isolation from other landmasses, and whether the unit of study was a single island or an archipelago. We then calculated the Akaike Information Criterion corrected for small sample sizes (AIC_c; Burnham and Anderson 2002) for each submodel of this full model. For the overall models, we ran each submodel once with each of the possible random effect terms, which included one term indicating only different intercepts for each taxon (written as 1|group) and one term for each island environmental variable, indicating different slopes for each taxon with this variable (e.g., area|group, indicating different speciation-area slopes for each taxon). The best

models for each data set are listed in table A5 (available in Dryad).

The significance and importance of each predictor variable in the multiple regression models were evaluated using model averaging as described in Burnham and Anderson (2002). First, for each data set, the full set of additive models was generated. Then, the relative importance of each variable was calculated, on a scale from 0 to 1, as the sum of the Akaike weights of the models in which the variable appears; better models have larger Akaike weights, and a variable that contributes more to model fit (and as such is included in more of the best models) will thus have a higher relative-importance value. Parameter estimates and unconditional standard errors for each term were calculated by averaging over all models in which the variable appears, weighting values from individual models by the models' Akaike weights. A term was considered to be significant for a particular data set if the 95% confidence interval for its parameter estimate did not include 0.

Gene Flow Data

F_{st} data came primarily from the appendixes in Morjan and Rieseberg (2004), and they were supplemented by a search following the methods of Morjan and Rieseberg (2004) for TS = ((fern or pteridophyt* or snail or Lepidoptera* or Chiroptera or Carnivora or lizard) and ("gene flow" or Fst or Nm or Nem)) on ISI Web of Science. After this search, carnivore data were still lacking, and so data were added from another search, for TS = (carnivor* and "population structure"). In additional searches, only articles that presented an overall F_{st} value (as opposed to only pairwise F_{st} values) for variation between populations (rather than regions) were used. Two estimators of F_{st} , G_{st} (Nei 1973) and Φ_{st} (Excoffier et al. 1992; Excoffier 2001), were used when they were provided in the original sources instead of F_{st} .

All studies dealing with aquatic or marine species were excluded, as were studies of recent habitat fragmentation as a result of human activities, studies of clonality, or studies of hybridization between species or host races. We also excluded studies including historically isolated lineages, such as those separated by a major geographical barrier (e.g., *Myotis myotis* on either side of the Strait of Gibraltar; Castella et al. 2000), as they are often evidence of cryptic speciation, whereas we wanted estimates of gene flow within species. Only studies of wild populations of native organisms were used; recent introductions and crop pests were excluded. All F_{st} values derived from organelle markers (mitochondrial or chloroplast DNA) were excluded, as these reflect only female dispersal and not overall population patterns of genetic differentiation. One plant study in Morjan and Rieseberg's database (*Proteum*

glabrum; Morjan and Rieseberg 2004) had a negative F_{st} value, which was interpreted as $F_{st} = 0$ (Long 1986). Gene flow data used are summarized in table A6 (available in Dryad).

We checked for comparability of gene flow estimates from different studies. Studies using amplified fragment length polymorphisms (AFLP), random amplification of polymorphic DNA (RAPD), and inter-simple sequence repeats (iSSR) were found to have significantly higher F_{st} values than were studies using other marker types, even when correcting for taxon and the geographic scale of study (ANOVA, $F = 7.3$, $P < 2.2 \times 10^{-16}$), and so they were excluded from analysis. Isozymes were also excluded because they were used in only one study in our data set and because the comparability of F_{st} values derived from them with F_{st} values derived from other markers is unclear. The F_{st} values from studies using allozymes and from those using types of repeats (microsatellites or simple sequence repeats, minisatellites, tandem repeats) did not differ significantly (ANOVA, $F = 1.7$, $P = .192$), and so they were lumped for the final analysis. Studies using allozymes and such repeats also did not differ significantly in the number of populations (ANOVA, $F = 1.5$, $P = .229$) or the number of loci (ANOVA, $F = 0.60$, $P = .440$) used. There were small differences between the two marker types in the mean geographic scale of study, but these were inconsistent in sign between taxa and so cannot explain the pattern of taxon variation in F_{st} . F_{st} showed no overall relationship with the number of loci (linear regression, $P = .867$) or the number of populations (linear regression, $P = .885$) used in a study. Allele frequencies for individual markers were not consistently available, and thus their relationship with F_{st} could not be tested; however, they are not expected to be a major confounding factor.

As a measure of the geographic scale of each population genetic study, geographic range extent was measured as the greatest distance in kilometers between any two populations in the study, following the example of Bohonak (1999). We used the maximum distance between study populations, rather than the mean or the median, because we felt that this was a simpler and more practical measure that was still sufficient to resolve the wide range of geographic scales represented by the studies we used (whose maximum distances vary from 0.01 km to >14,000 km). In addition, use of alternative measures such as mean distance would be difficult because of the lack of detailed information in many of the sources. Our measure does not explicitly take into account the variation in species' geographic range sizes or in the spatial arrangement of sampled populations, but neither of these was easily quantified from the sources available to us, and neither is expected to have a strong independent effect on F_{st} (Beau-

mont and Nichols 1996). Distances were taken directly from papers if possible. Otherwise, they were calculated from population coordinates using the Vincenty formula (Veness 2008), measured from scaled maps given in the studies, or measured using Google Earth (ver. 4.3) if only place names were provided. Data for 23 taxa were excluded because the original reference either could not be found or did not contain sufficient information to calculate the geographic range extent.

Gene Flow Analyses

To test the effect of gene flow on the spatial scale of speciation, we correlated the natural log of the minimum island size for speciation in each study group with two summary measures of the spatial scale of gene flow. One potential problem with comparing F_{st} values is that F_{st} data tend to be measured at different spatial scales in the different taxa (fig. 2). Therefore, as summary measures, the mean F_{st} values were calculated for each group at two scales in turn: between 10 and 100 km and between 100 and 1,000 km. We chose these geographic scales for analysis because they correspond to the range of island sizes used in the speciation analysis (47 [73%] of 64 islands/island groups have maximum linear extents between 10 and 1,000 km) and because they are the only scales at which F_{st} data were available for all study taxa (except snails between 100 and 1,000 km). To control for the effect of outliers in the F_{st} data, we also tested the correlation of median F_{st} values with minimum island size for speciation.

For this analysis, the minimum island size for speciation for each study group was represented as the greatest distance between any two points of land within the island or archipelago. These extents were measured in Google Earth, version 4.3, and in the Analyzing Digital Images software package (Pickle and Kirtley 2008). Island linear extents were used instead of island areas in order for them to be directly comparable to the distances between populations used to represent the spatial scale of gene flow. Because of the uncertainty over bird speciation on Norfolk, we performed this analysis twice, once with Norfolk and once with Tristan da Cunha representing the smallest island with in situ bird speciation.

Results

Data Availability and Quality

We estimated speciation probabilities across a total of 64 islands, including 38 single islands and 26 archipelagos, taking into account 471 putative speciated genera. Phylogenies were available for 15% of these genera; an additional 15% of the genera were endemic to their island

and as such were most parsimoniously explained by in situ speciation. These data led us to exclude only seven of the putative speciated genera as being the result of multiple colonization (those seven genera were only ~5% of the genera for which phylogenetic or endemism information was available), confirming that the nonphylogenetic measure is a good measure of the number of genera that have speciated in situ and that it is little affected by multiple colonization events. Twelve genera were found to be part of a larger adaptive radiation already represented by another included genus, and so they were removed from the analysis. Three genera were found to be the result of two separate radiations, and one genus was found to be the result of three separate radiations; these were split into multiple speciated lineages accordingly. After taking all of this into account, our final data set included 457 speciated lineages. Phylogenetic data by genus are presented in table A2 and summarized in table A4 (available in Dryad).

Quantifying the Speciation-Area Relationship

Across taxa, on oceanic islands and archipelagos ranging in size from <1 (Nihoa) to >500,000 km² (Madagascar), there is a clear positive relationship between the probability of in situ speciation and island area ($P = 1.35 \times 10^{-5}$, $r^2 = 0.312$; when archipelagos are excluded: $P = 1.54 \times 10^{-4}$, $r^2 = 0.414$; fig. 3). The relationship between the probability of speciation and island area is significant in all taxa with sufficient data, except ferns (table 2). Bats and Macrolepidoptera had insufficient data to construct models for their speciation-area relationship, as they show evidence for speciation on only two islands and one island, respectively, but the data for both nevertheless support the same positive relationship seen in the other taxa: both are present on many small islands on which they have not speciated, while they have speciated on only the largest islands on which they are represented. In contrast, carnivores are rarely present at all on oceanic islands due to their poor dispersal over water, and there are endemic species on only two of the studied islands. Nonetheless, they also show evidence for speciation on only the largest island on which they are represented (Madagascar), and we predict that they would also show a positive speciation-area relationship if larger landmasses were considered. The lack of a speciation-area relationship in ferns, on the other hand, is not the result of a lack of data. Ferns have clearly speciated on small and large islands with similar probability, indicating that area is relatively unimportant in controlling their speciation.

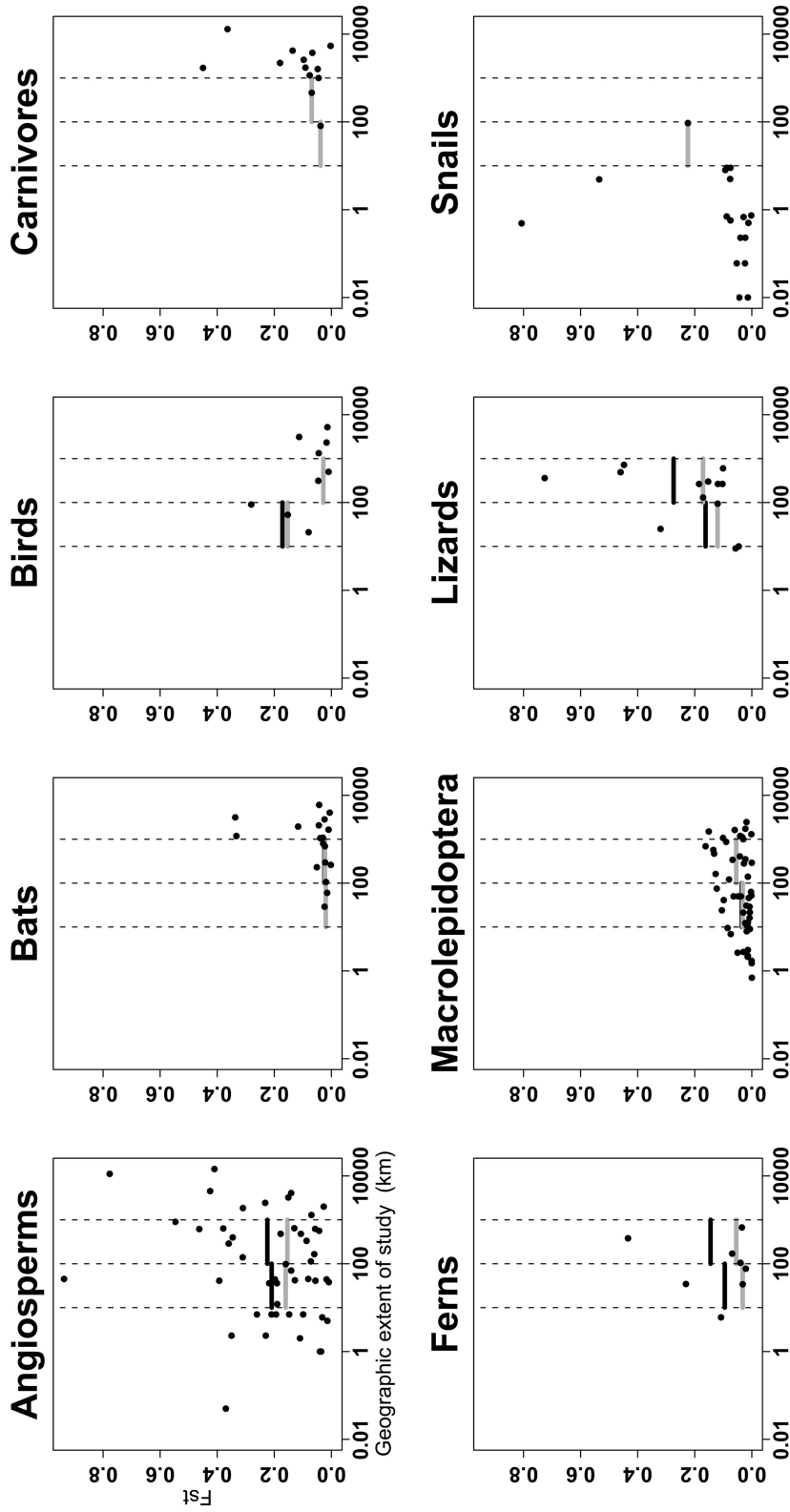


Figure 2: Relationship between F_{st} and the geographic extent of study, with F_{st} estimates taken from a review of the population genetic literature. Each point represents an F_{st} estimate from a single study, and only F_{st} values used in the analyses are shown. The geographic extent of study is the maximum distance between populations in one study. F_{st} values given are overall F_{st} values considering all populations in the study. Dashed lines delimit the two geographic scales that were used for the gene flow analysis. Solid black lines indicate the average F_{st} value for each geographic scale, and solid gray lines indicate the median F_{st} value. Where only a gray line is visible, the mean and median values are equal. No data were available for snails at the 100–1,000-km range.

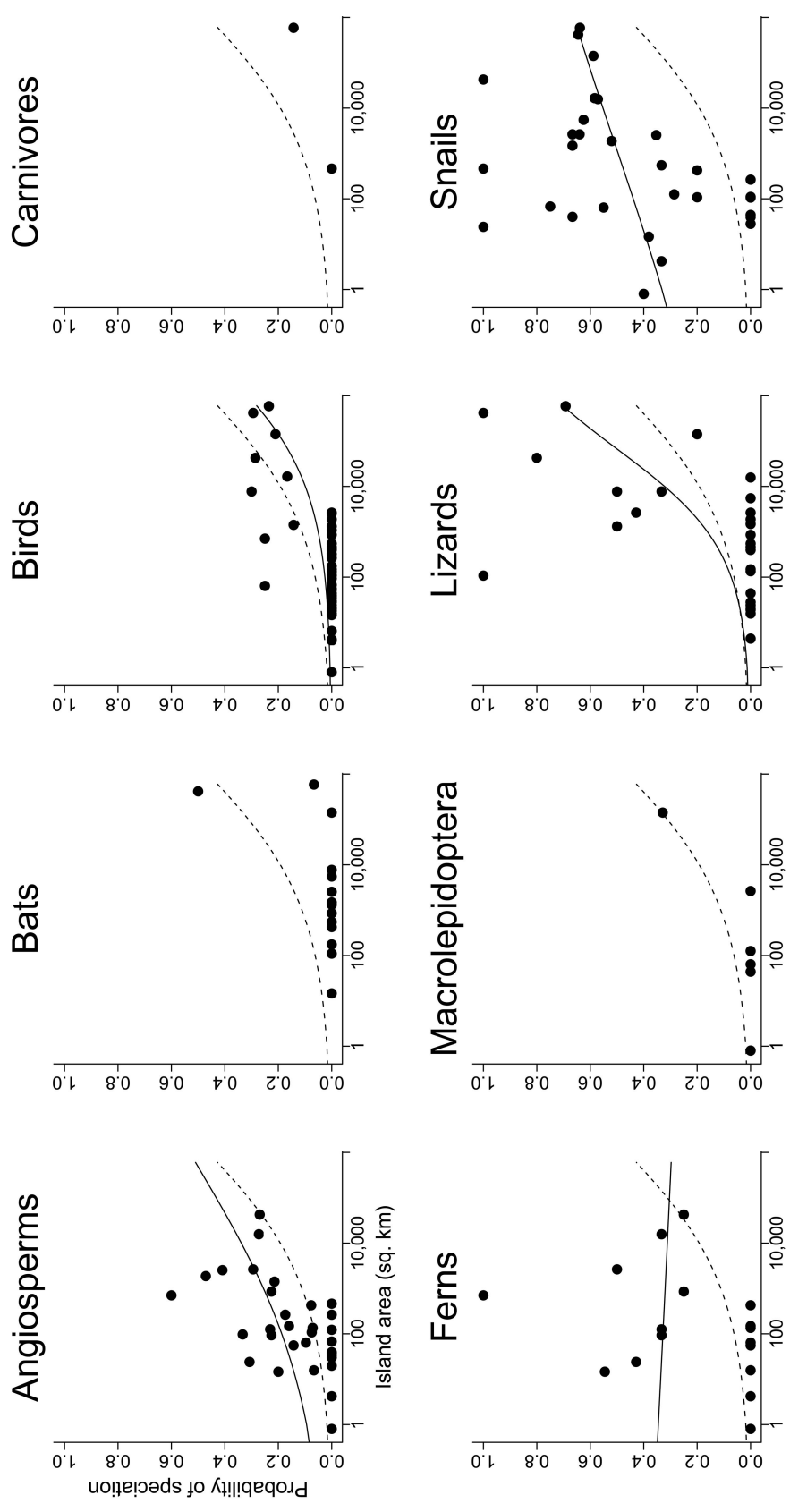


Figure 3: Relationship between the probability of speciation and area. Each point marks the percentage of lineages of a particular taxon, on a particular island, that have speciated. Solid lines mark the regression model for each taxon alone; dashed lines represent the overall model averaged over all taxa. No individual model line is given for bats, carnivores, or Macrolepidoptera, as these taxa had too few islands to model their speciation-area relationship. Lines are based on analyses including archipelagos.

Table 2: Area-only models for the probability of speciation

Taxon	<i>N</i>	<i>S</i>	<i>P</i> (speciation) model	<i>P</i> value for area	<i>r</i> ²
Overall model	64	38	$-4.00 + .279 \ln(\text{area}) + (\text{area} \text{group})$	1.35×10^{-5}	.312
Excluding archipelagos	40	20	$-5.89 + .420 \ln(\text{area}) + (\text{area} \text{group})$	1.54×10^{-4}	.414
Angiosperms	32	21	$-2.27 + .173 \ln(\text{area})$	1.75×10^{-6}	.277
Excluding archipelagos	24	14	$-3.14 + .350 \ln(\text{area})$	6.91×10^{-9}	.534
Birds	50	9	$-5.07 + .310 \ln(\text{area})$	1.66×10^{-5}	.580
Ferns	17	9	$-.641 - .0165 \ln(\text{area})$.882	.00158
Excluding archipelagos	11	5	$-.411 - .0844 \ln(\text{area})$.704	.0149
Lizards	27	10	$-4.35 + .391 \ln(\text{area})$	1.75×10^{-5}	.527
Excluding archipelagos	14	3	$-5.39 + .464 \ln(\text{area})$	8.6×10^{-4}	.732
Snails	30	24	$-.700 + .0988 \ln(\text{area})$.00234	.243
Excluding archipelagos	17	12	$-.810 + .0951 \ln(\text{area})$.0219	.216

Note: *N* = number of islands/island groups used to construct model. *S* = number of islands with speciation events. Parameter values are those given by the logistic models, and they produce predicted values that must be logit transformed to give an estimated probability of speciation. In the overall models, the “(area|group)” term is the random effect accounting for variation between our study taxa in the slope of the speciation-area relationship. Speciation probability was not modeled for bats, birds on single islands, Carnivora, or Macrolepidoptera because these groups had fewer than three islands with speciation events.

Measuring Minimum Areas for Speciation

The minimum area for speciation (estimated as the area of the smallest island or archipelago within which speciation has occurred) varies widely among taxa. Land snails have speciated within even the smallest island on which they have native species (Nihoa: 0.8 km²), whereas the only example of in situ speciation in Carnivora is on Madagascar (587,713.3 km²) and bats show no evidence of in situ speciation on any island except New Zealand (~416,400 km²) and Madagascar. Macrolepidoptera also appear to require large areas for speciation: the only island unit in which they show evidence of speciation is Fiji (141,200 km²). Angiosperms and lizards are intermediate in this regard, with minimum areas for speciation of 14.6 and 107.8 km², respectively (table 1; a more detailed summary of speciation events is in table A2 in Dryad). The situation in birds is unclear; even after genetic analysis, it is uncertain whether a putative speciation event on Norfolk Island (64 km²) is actually the result of multiple colonizations (Coyné and Price 2000). The next-smallest island unit within which bird speciation has potentially occurred is the Tristan da Cunha archipelago (705 km²), for which some evidence even supports a history of sympatric speciation within the smaller islands of the archipelago (Ryan et al. 2007; Grant and Grant 2009). The smallest single island with firm evidence for in situ bird speciation is Jamaica (11,400 km²; Coyné and Price 2000). Irrespective of the uncertainty for birds, it is evident that taxonomic variation in the spatial scale of speciation is great, spanning six orders of magnitude between snails and carnivorous mammals.

Testing the Importance of Area When Other Environmental Variables Are Included

The relationship with area is not an artifact of area’s correlation with another environmental variable. In our data set, island area is positively correlated with elevation (adjusted $r^2 = 0.478$; adjusted $r^2 = 0.365$ if archipelagos are excluded), island age (adjusted $r^2 = 0.111$; correlation is not significant if archipelagos are excluded), and whether an island is an archipelago (proportion of variance explained in ANOVA = 0.278), but in multiple regression models, both overall and for individual taxa, model-averaged parameter estimates indicate that island area is highly important and significant, independently of other island characteristics (best models are listed in table A5 in Dryad; model-averaged parameter estimates are given in tables 3 and 4).

In the overall models, area has a high relative-importance value (0.93 when archipelagos are included; 0.97 when archipelagos are excluded; relative-importance values are on a scale of 0–1), meaning that it is included in a high percentage of the best models. In addition, its parameter estimates are significantly greater than 0, supporting a positive speciation-area relationship. Isolation and elevation also have high relative-importance values and are also significant in the overall models, especially in the data set including archipelagos (table 3). Age and whether an island unit is an archipelago have low relative-importance values and are not significant in the overall models.

Area is also highly important and significant in most of the single-taxon models. It is the most important variable, and its parameter estimate is significantly greater than 0

Table 3: Model-averaged parameter estimates and relative-importance values for analyses that included archipelagos

Model	Area			Age			Elevation			Isolation			ArchYN	
	Relative-importance value	Estimate ± SE (CI)	Relative-importance value	Estimate ± SE (CI)	Relative-importance value	Estimate ± SE (CI)	Relative-importance value	Estimate ± SE (CI)	Relative-importance value	Estimate ± SE (CI)	Relative-importance value	Estimate ± SE (CI)	Estimate ± SE (CI)	Relative-importance value
Overall	.93	.20 ± .0044 (.19, .21)	.26	.0017 ± .0035 (-.0051, .0086)	.97	.022 ± 6.21 × 10⁻⁵ (.022, .022)	.97	.010 ± 1.29 × 10⁻⁵ (.010, .010)	.3	.010 ± 1.29 × 10⁻⁵ (.010, .010)	.3	.010 ± 1.29 × 10⁻⁵ (.010, .010)	.010 ± 1.29 × 10⁻⁵ (.010, .010)	.3 (-.077, .026)
Angiosperms	.99	.22 ± .0042 (.21, .23)	.20	.020 ± .016 (-.011, .051)	.76	.018 ± .00014 (.018, .018)	.87	.011 ± 2.86 × 10⁻⁵ (.011, .011)	.94	.011 ± 2.86 × 10⁻⁵ (.011, .011)	.94	.011 ± 2.86 × 10⁻⁵ (.011, .011)	-.66 ± .061 (-.78, -.54)	.94 (-.78, -.54)
Birds	.98	.32 ± .0086 (.30, .34)	.27	.040 ± .082 (-.12, .20)	.27	.0047 ± .00081 (.0031, .0063)	.26	-.00099 ± .00026 (-.0015, -.00048)	.79	-.00099 ± .00026 (-.0015, -.00048)	.79	-.00099 ± .00026 (-.0015, -.00048)	.79 ± .26 (.29, 1.29)	.79 ± .26 (.29, 1.29)
Ferns	.31	-.010 ± .018 (-.045, .025)	.31	.095 ± .16 (-.23, .42)	.34	.0074 ± .0020 (.0036, .011)	.33	.0019 ± .00017 (.0016, .0023)	.30	.0019 ± .00017 (.0016, .0023)	.30	.0019 ± .00017 (.0016, .0023)	.018 ± .60 (-1.1, 1.2)	.018 ± .60 (-1.1, 1.2)
Lizards	.86	.32 ± .017 (.29, .36)	.11	.015 ± .079 (-.14, .17)	.34	.017 ± .0027 (.012, .022)	.20	.00051 ± .00051 (-.00049, .0015)	.24	.00051 ± .00051 (-.00049, .0015)	.24	.00051 ± .00051 (-.00049, .0015)	.10 ± .44 (-.76, .97)	.10 ± .44 (-.76, .97)
Snails	.78	.074 ± .0020 (.070, .077)	.28	.024 ± .018 (-.011, .059)	.34	.0039 ± .00024 (.0034, .0044)	.21	1.76 × 10 ⁻⁵ ± 6.39 × 10 ⁻⁵ (-.00011, .00014)	.39	1.76 × 10 ⁻⁵ ± 6.39 × 10 ⁻⁵ (-.00011, .00014)	.39	1.76 × 10 ⁻⁵ ± 6.39 × 10 ⁻⁵ (-.00011, .00014)	.13 ± .11 (-.077, .34)	.13 ± .11 (-.077, .34)

Note: The parameter with the highest relative-importance value for each data set is highlighted in bold. Parameter estimates for significant variables are also highlighted in bold. Speciation probability was not modeled for bats, Carnivora, or Macrolepidoptera because these groups had fewer than three islands with speciation events. Age was available only for a subset of islands, and so parameter estimates for age come from regression models using this reduced subset. Parameter values for all other terms come from regression models using the full set of islands, unless age was found to be significant, in which case all parameter values were estimated using the reduced subset of islands with age data. CI = confidence interval.

Table 4: Model-averaged parameter estimates and relative-importance values for analyses that excluded archipelagos

Model	Area			Age			Elevation			Isolation		
	Relative-importance value	Estimate ± SE (CI)	Relative-importance value	Estimate ± SE (CI)	Relative-importance value	Estimate ± SE (CI)	Relative-importance value	Estimate ± SE (CI)	Relative-importance value	Estimate ± SE (CI)		
Overall	.97	.38 ± .016 (.35, .42)	.30	-.012 ± .023 (-.0587, .033)	.47	.011 ± .00045 (.010, .012)	.64	.0052 ± 2.83 × 10⁻⁵ (.0051, .0052)				
Angiosperms	.9998	.34 ± .0043 (.33, .35)	.12	-.010 ± .029 (-.067, .046)	.22	.0016 ± .00015 (.0013, .0019)	.33	.0021 ± 4.28 × 10⁻⁵ (.0020, .0022)				
Ferns	.37	-.032 ± .055 (-.14, .075)	.23	.010 ± .16 (-.31, .33)	.35	-.0020 ± .0025 (-.0069, .0030)	.35	1.1 × 10 ⁻⁴ ± .00018 (-.00024, .00047)				
Lizards	.16	-3.47 ± 1.6 × 10 ⁸ (-3.2 × 10 ⁸ , 3.2 × 10 ⁸)	.75	2.48 ± 4.35 × 10 ⁹ (-8.5 × 10 ⁹ , 8.53 × 10 ⁹)	.91	65.4 ± 1.6 × 10 ⁸ (-3.1 × 10 ⁸ , 3.1 × 10 ⁸)	.88	54.7 ± 1.1 × 10 ⁸ (-2.2 × 10 ⁸ , 2.2 × 10 ⁸)				
Snails	.40	.033 ± .0051 (.023, .043)	.42	.075 ± .024 (.028, .12)	.16	-.00048 ± .00047 (-.0014, .00044)	.14	-9.72 × 10 ⁻⁵ ± .00011 (-.00032, .00012)				

Note: The parameter with the highest relative-importance value for each data set is highlighted in bold. Parameter estimates for significant variables are also highlighted in bold. Speciation probability was not modeled for bats, birds, Carnivora, or Macrolepidoptera because these groups had fewer than three islands with speciation events. Age was available only for a subset of islands, and so parameter estimates for age come from regression models using this reduced subset. Parameter values for all other terms come from regression models using the full set of islands, unless age was found to be significant, in which case all parameter values were estimated using the reduced subset of islands with age data. CI = confidence interval.

(indicating a positive speciation-area relationship) for all single taxa except ferns in the data set that includes archipelagos and lizards in the data set that excludes archipelagos. It is also nonsignificant in ferns in the data set that excludes archipelagos. Elevation and isolation are also significant in most single-taxon models, although in all cases except that of lizards when archipelagos are excluded, they are much less important than area. The parameter estimate for isolation is positive in all taxa except birds; more distant islands tend to have a higher probability of speciation. Whether an island is an archipelago is important and significant in angiosperms and birds, while island age is relatively important and significant for only snails, and only when archipelagos are excluded. Over all models, island area is the most consistently important and significant island variable.

The Effect of Gene Flow

The minimum island size for speciation in each taxon correlates with the mean strength of gene flow for each taxon when gene flow is measured at the scale of 10–100 km (slope = -27.21 , $P = .00286$, adjusted $r^2 = 0.763$; fig. 4). Taxa that are able to speciate within smaller areas (indicated by a smaller minimum island size for speciation) are those with weaker gene flow (indicated by higher mean F_{st} values). At the scale of 100–1,000 km, at which snails are excluded due to lack of data, the same relationship is found, although it is marginally nonsignificant (slope = -16.51 , $P = .0695$, adjusted $r^2 = 0.417$). Widely overlapping 95% confidence intervals (at 10–100 km: -38.22 to -16.21 ; at 100–1,000 km: -30.56 to -2.46) indicate that there is no significant difference in slope between the two spatial scales. We found similar results using median F_{st} instead of mean F_{st} for each group (at 10–100 km: slope = -24.91 , $P = .0221$, adjusted $r^2 = 0.545$; at 100–1,000 km: slope = -20.40 , $P = .221$, adjusted $r^2 = 0.138$). Using Norfolk Island instead of Tristan da Cunha for the minimum area of speciation in birds also did not change the results (at 10–100 km: slope = -28.45 , $P = .00144$, adjusted $r^2 = 0.810$; at 100–1,000 km: slope = -14.69 , $P = .146$, adjusted $r^2 = 0.246$). When archipelagos are excluded, the same gene flow–minimum area relationship is again found at both spatial scales, although in this case the line is significant at 100–1,000 km (slope = -21.82 , $P = .0191$, adjusted $r^2 = 0.729$) but not at 10–100 km (slope = -24.02 , $P = .0917$, adjusted $r^2 = 0.358$). Again, widely overlapping confidence intervals for the slopes (at 10–100 km: -46.61 to -1.42 ; at 100–1,000 km: -33.07 to -10.56) indicate that there is no significant difference in the relationship between the two spatial scales.

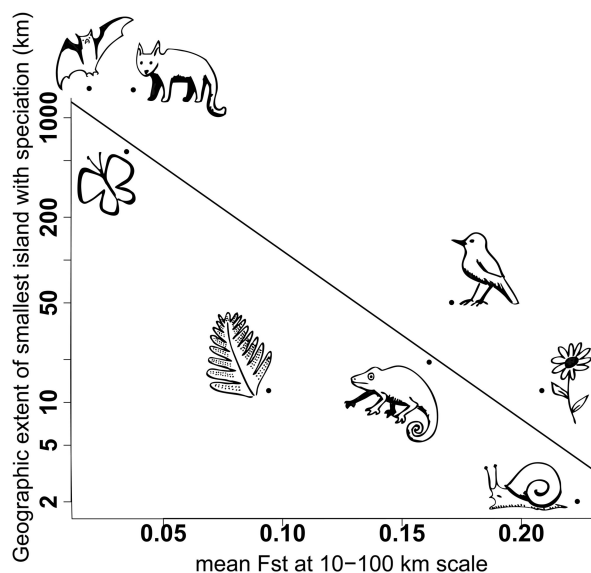


Figure 4: Minimum island size for speciation versus the average strength of gene flow when measured over geographic ranges of 10–100 km.

Discussion

Main Findings

Our results show that the speciation-area relationship, in which speciation is more likely and more frequent within larger areas, is a general pattern common to many groups of plants and animals. Ferns were the only group that showed no such relationship, perhaps because of their higher propensity for polyploid and hybrid speciation, the implications of which we discuss further below. The speciation-area relationship we found is not just a by-product of area's correlation with other island characteristics; island area was consistently important and significant in both overall and taxon-specific multivariate models, which also included island elevation (as a proxy for habitat diversity), age, isolation from other landmasses, and whether island units were archipelagos or single islands. Although all study taxa except ferns have in common a positive speciation-area relationship, they vary over six orders of magnitude in the minimum area required for speciation. Furthermore, this variation in the minimum area for speciation correlates with variation among taxa in the strength of gene flow. Taxa with higher rates of gene flow, measured at a common spatial scale, have a larger minimum area for speciation and a lower probability of speciation in any given area. This suggests that the population genetics of divergence directly control the incidence and rate of speciation: that there is a direct link

between microevolutionary and macroevolutionary processes.

The Effects of Other Island Characteristics on Speciation Probability

Although we found strong evidence for area as a major controller of speciation rates, this does not rule out roles for other environmental variables. In particular, isolation and elevation were also important and significant factors in the overall models and were important and significant in most of the individual-taxon models. The effect of elevation was always positive, as predicted if greater altitudinal variation increases the number of habitats and promotes greater ecological speciation (Ackerman et al. 2007; Losos and Parent 2009). In all cases except birds, the probability of in situ speciation increased with increasing isolation, consistent with predictions that lower colonization rates of distant islands should leave more niches available for speciation (Gillespie and Baldwin 2009). In birds, our measure of speciation probability increased on islands closer to other landmasses, which is unexpected according to our theoretical predictions but which might arise if some portion of inferred speciation events still represents multiple colonization (because colonization is expected to be greater on islands closer to other landmasses). However, isolation was the least important variable for birds, and it had a small effect on variation in our measure.

Interestingly, considering the great distinction that is usually made between single islands and archipelagos in island evolution theory, we found a significant and important effect of archipelagos only on speciation probability in birds and angiosperms. Furthermore, while the parameter estimate for birds is positive, as is expected if water gaps between islands act as additional dispersal barriers promoting speciation, the parameter estimate for angiosperms is negative, which is unexpected on theoretical grounds. The lack of significance and importance of the archipelago term in other taxa and in the overall models may indicate that the difference between water gaps and ecological barriers within islands in their strength as dispersal barriers is much greater for birds than for the other study taxa (Diamond 1977). For the other study taxa, barriers within islands may be strong enough that diversification within a heterogeneous island is comparable to diversification within an archipelago. Most importantly for our aims, the speciation-area relationship holds irrespective of whether archipelagos are included.

Broad comparative studies such as this one necessarily rely on surrogates and proxies for some underlying variables of interest, and so a lack of correlation in our study is not conclusive evidence against any environmental fac-

tor. Further work would particularly benefit from improved data on island ages; it is difficult to evaluate the biological relevance of ages taken from the geological literature (for instance, lava flows can sterilize an island some time after its actual origination and emergence; Whittaker et al. 2008), and ages are lacking for many islands and island groups.

The Spatial Scale of Speciation and Gene Flow

Consistent with the importance of gene flow in population genetics-based theories of speciation, estimates of the strength of gene flow explain up to 76% of the variation in the spatial scale of speciation across taxa. Taxa in which gene flow is weaker are able to speciate within smaller islands, suggesting that the strength of gene flow determines the spatial scale of speciation by controlling the minimum spatial extent at which differentiation of populations can occur. This result also accounts for the existence of thresholds in evolutionary species-area relationships (Losos and Schluter 2000): in situ speciation is expected to contribute significantly to local species richness only in areas large enough that gene flow does not prohibit population differentiation.

The main limitation for this analysis was the availability of gene flow data. Past studies have largely applied molecular markers to single-species questions, and meta-analyses like ours are necessarily posterior exercises limited by available data. While disparate studies are still comparable (Bohonak 1999; Morjan and Rieseberg 2004), targeted studies generating data for a set of species using a standardized sampling design would allow for more refined comparative analyses, including the use of more sophisticated measures of the spatial scale and strength of gene flow, such as the mean dispersal distance predicted from the slope of an isolation-by-distance (IBD) regression line for each species (Kinlan and Gaines 2003) or the *Sp* statistic (Vekemans and Hardy 2004). It would also be useful to have gene flow data for the specific genera and species for which island data were collected, instead of averaging over each major taxon (especially given the tendency of island species to evolve reduced dispersal ability; Carlquist 1974), but these data were not available in the literature.

Because of these constraints, we could compare only a limited number of different taxa at a relatively broad taxonomic scale while retaining enough information to provide reasonable sample sizes for estimating the study variables. Despite relatively low power, the result is robust for the sample available. The significance of the relationship varied, depending on the scale used (which determined whether snails were included) and whether archipelagos were included, but in an inconsistent way that reflected

low power rather than large changes in the underlying relationship. A significant relationship was also found by using an alternative measure that is closer to the underlying quantity of interest but less statistically robust than mean F_{st} (an estimate of the minimum scale at which neutral divergence is expected to occur within species of each major taxon; fig. 5).

Therefore, despite the above limitations and the relatively low power they entail, our results point to gene flow levels as a potentially important determinant of the spatial scale of speciation. It remains possible, however, that the relationship we found is the result of other confounding factors that vary between our study taxa in parallel with differences in gene flow. Incorporating more taxa, resolving the chosen taxa more finely, and generating more estimates of gene flow would be needed for more powerful tests of this hypothesis in the future.

The negative relationship we found between gene flow and the probability of speciation within a given area at first seems to contrast with the idea that either high (Eriksson and Bremer 1991; Owens et al. 1999; Phillimore et al. 2006) or intermediate (Price and Wagner 2004; Paulay and Meyer 2006) dispersal ability should lead to maximum diversification. These ideas are incompatible, however, only if every species is imagined to have a single value representing its dispersal ability. In reality, dispersal for any taxon is usually thought of as a leptokurtic probability function, with a long tail of infrequent long-distance dispersal events (Tilman and Kareiva 1997). Under this model, dispersal affects diversification in two different ways: shorter-distance dispersal within the species range maintains species cohesion, and rarer long-distance dispersal to new areas outside the species range allows the establishment of new, potentially isolated populations. By considering only those lineages that are able to reach oceanic islands, we intentionally focused on the effect of shorter-distance dispersal ability and controlled for long-distance dispersal, namely, colonization ability.

Evolutionary Explanations for the Observed Patterns

Several mechanisms could produce the speciation-area relationship that we observed (Gavrilets and Losos 2009). First, larger areas might offer more opportunity for geographical isolation, either by distance alone or via barriers to dispersal (MacArthur and Wilson 1967; Endler 1977; Rosenzweig 1995). Second, larger areas might encompass more habitat types, which could increase speciation rates through stronger divergent selection or by providing additional niche space allowing the coexistence of newly formed species (Losos and Parent 2009). We considered habitat variation in relation to elevation, but other unmeasured aspects of habitat variation might also scale with

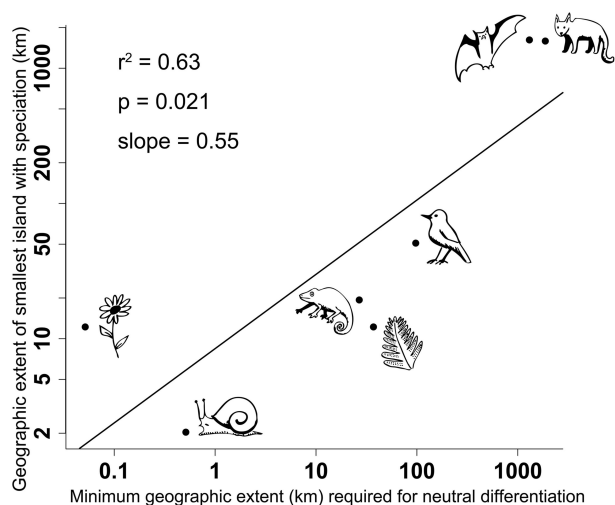


Figure 5: Results of an alternative gene flow analysis: the relationship between the minimum area for speciation and the spatial scale of neutral population differentiation. Minimum island size for speciation is plotted against the minimum geographic extent for each taxon at which gene flow has been observed to be weak enough to allow neutral genetic differentiation of populations ($F_{st} = 0.2$, corresponding to $Nm = 1$). This is estimated for each taxon by the geographic scale of the population genetic study with the smallest geographic scale and $F_{st} \geq 0.2$. Macrolepidoptera are excluded because none of the Macrolepidoptera studies in our population genetic data set have $F_{st} \geq 0.2$.

area. Third, larger areas can support larger population sizes, which might increase the rate of adaptive evolution by increasing the rate of origin of beneficial mutations upon which selection can act (Gavrilets and Vose 2005). Including data on population sizes might allow the third mechanism to be distinguished, but in the absence of such information we believe that the relationship between gene flow and the spatial scale of speciation is most consistent with speciation occurring through geographical isolation or ecological divergence into distinct, spatially structured habitats (Schluter 2001).

By providing the exception to the general pattern observed, ferns strengthen the support for these conclusions. Ferns are known to have a high incidence of speciation through hybridization and polyploidy (Wagner 1969; Otto and Whitton 2000), which are two major processes that allow speciation to occur in the face of gene flow (Berlocher 1998). In fact, of the two fern genera in our study with speciation events supported by published phylogenies, one is thought to have diversified through hybridization (Eastwood et al. 2004). In contrast, speciation as a result of hybridization and polyploidy is rare in animals and important but much less frequent in angiosperms (Otto and Whitton 2000). Thus, as expected if the speciation-area relationship is the result of gene flow–lim-

ited divergence, the group that most frequently speciates with continuing gene flow shows no significant speciation-area relationship. We conclude that pure sympatric speciation, namely, in the absence of any geographical isolation and in the presence of gene flow, appears to be infrequent in all taxa except ferns (see also Barraclough and Vogler 2000; Phillimore et al. 2008).

Extinction might also influence the relationship between diversification and area, most directly because extinction rates should be higher on smaller islands with smaller populations (MacArthur and Wilson 1967). For this reason, extinction has been used in the past to explain the relationship between island area and the number of single-island-endemic species (Mayr 1965). The effect of extinction on the speciation-area relationship cannot be tested with the type of data presented here; doing so would require studies of island taxa for which comprehensive fossil data are available and extinction rates can be estimated directly (perhaps birds; Steadman 2006). However, we believe that the association between decreased gene flow and increased diversification cannot be explained easily by extinction. There are some mechanisms, such as increased pathogen spread (Thrall et al. 2000) or swamping of local adaptation (Holt and Gomulkiewicz 2004), by which increased gene flow could increase the risk of extinction (and thereby decrease net diversification rate), but neither of these is a necessary outcome of increased gene flow. It is more usually expected that decreased gene flow should increase the risk of extinction, through either increased inbreeding (Lande 1988) or decreased recolonization rates within metapopulations (Gaggioti and Hanski 2004). Probability of speciation, on the other hand, is clearly predicted to increase with decreased gene flow. Therefore, we believe that it is more likely that the patterns we observe reflect differential rates of divergence and speciation, rather than an effect of extinction.

The Effects of Taxonomic Practice and Surveying Effort

As is assumed in most comparative studies of diversification, we assumed that entities named as species represent a similar level of evolutionary divergence across all taxa considered. If different taxa had been subjected to different taxonomic practices, our conclusions regarding scales at which speciation can occur could have been influenced. For instance, a taxon in which species are split more finely (so that they are equivalent to subspecies of other taxa) would be counted as being able to speciate within smaller islands. On the other hand, finer splitting, which would cause subspecies that are endemic to single islands to be elevated to species status, could lead to more cases of genera with only one endemic species on an island, and thus, lower calculated probabilities of speciation. As this

would not affect which genera are identified as having had speciation events, this would not affect the estimation of minimum areas for speciation but would change the slopes of the speciation-area relationships. In either case, it is unlikely that the differences in taxonomic practice among our study taxa are in the correct order (for instance, Lepidoptera lumped more than snails) to be solely responsible for the pattern of minimum areas for speciation we observe.

Data quality is likely to vary among islands and taxa as a result of differences in past surveying intensity. Total surveying effort has generally been greater for larger islands, but on small islands, less effort is necessary for complete description of their endemic species. Therefore, we do not believe that the chance of detecting whether a genus has speciated in situ is likely to vary systematically with island area.

Finally, our surveys of island characteristics, species lists, and phylogenies of study genera are not comprehensive, because of limitations on data availability. Future availability of appropriate data could perhaps alter observed patterns. However, we believe that our data include a high percentage of those data available and are complete enough to draw broad conclusions.

Implications for Evolutionary Studies of Diversity Patterns

Our results support a general geographical model of speciation in which area and gene flow interact via the spatial scale of speciation to control both speciation rates and resulting diversity patterns. Organisms with weaker gene flow are able to differentiate at finer spatial scales than are those with stronger gene flow, leading to increased speciation rates and higher taxonomic diversity within a given area. Variation among taxa in the strength of gene flow could be caused by several factors, including differences in dispersal ability, in the degree of habitat specificity (which controls which habitats will act as barriers to dispersal; Thorpe 1945), and in the strength of natural selection against between-population hybrids (whose survival is necessary for effective gene flow). The strength of selection against hybrids will depend on the rate of accumulation of genetic incompatibilities and the degree of local adaptation (Gavrilets 2004; Fuller 2008), both of which could vary systematically among taxa. Because the above model incorporates both species traits and geography, it should be useful for explaining both taxonomic and geographic variation in diversification rates and total diversity.

Furthermore, the strength of this model highlights more generally the potential of an evolutionary process-based framework for understanding speciation rates and higher-level patterns of species richness. Until now, macroevo-

lutionary studies have tested a diverse range of potential correlates of diversification, with mixed results and few general conclusions (for a review of factors tested, see Jablonski 2008). In particular, macroevolutionary studies that focused on organism traits, such as animal body size or plant woodiness, have generally found only weak correlations with diversification rates, explaining no more than 10%–25% of the observed variation in clade species richness, even using multivariate models (Phillimore et al. 2006). In contrast, there is stronger evidence for the link between population-level processes (including adaptive divergence, but also sexual selection and gene flow) and rates of speciation and diversification (e.g., Barraclough et al. 1995; Belliure et al. 2000; Stuart-Fox and Owens 2003; Funk et al. 2006; Seddon et al. 2008). These processes relate directly to the population genetic theory that forms the foundation of our understanding of speciation, and a framework based on these processes would be applicable to all organisms. Bridging the gap between population genetic theories of speciation and macroevolutionary approaches has great potential for improving our understanding of large-scale patterns of diversity.

Acknowledgments

We thank R. S. Gómez, S. Meiri, A. M. C. Santos, T. F. Stuessy, and C. Thebaud for access to their unpublished checklist data; R. A. Black and N. Cooper for helping us find bat and bird data; and T. J. Davies, J. Hortal, C. Lexer, L. McInnes, S. Meiri, A. B. Phillimore, A. Purvis, V. Savolainen, and two anonymous reviewers for comments on the manuscript. Y. K. was supported by a U.S. National Science Foundation Graduate Research Fellowship and a Deputy Rector's Award of Imperial College, London.

Literature Cited

- Ackerman, J. D., J. C. Trejo-Torres, and Y. Crespo-Chuy. 2007. Orchids of the West Indies: predictability of diversity and endemism. *Journal of Biogeography* 34:779–786.
- Australian Biological Resources Study, Canberra. 2008. Flora of Australia online: oceanic islands excluding Norfolk and Lord Howe Islands. <http://www.environment.gov.au/biodiversity/abrs/online-resources/flora/50/index.html>.
- Barluenga, M., K. Stolting, W. Salzburger, M. Muschick, and A. Meyer. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 439:719–723.
- Barraclough, T. G., and A. P. Vogler. 2000. Detecting the geographical pattern of speciation from species-level phylogenies. *American Naturalist* 155:419–434.
- Barraclough, T. G., P. H. Harvey, and S. Nee. 1995. Sexual selection and taxonomic diversity in passerine birds. *Proceedings of the Royal Society B: Biological Sciences* 259:211–215.
- Barton, N. H. 2001. The evolutionary consequences of gene flow and local adaptation: future approaches. Pages 329–340 in J. Clobert, E. Danchin, A. A. Dhondt, and J. D. Nichols, eds. *Dispersal*. Oxford University Press, Oxford.
- Bates, D. 2007. lme4: linear mixed-effects models using Eigen and R package version 0.99875-9.
- Bauer, A. M. 1988. Reptiles and the biogeographic interpretation of New Caledonia. *Tuatara* 30:39–50.
- Beaumont, M. A., and R. A. Nichols. 1996. Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences* 263:1619–1626.
- Belliure, J., G. Sorci, A. P. Møller, and J. Clobert. 2000. Dispersal distances predict subspecies richness in birds. *Journal of Evolutionary Biology* 13:480–487.
- Berlacher, S. H. 1998. Can sympatric speciation via host or habitat shift be proven from phylogenetic and biogeographic evidence? Pages 99–113 in D. J. Howard and S. H. Berlacher, eds. *Endless forms: species and speciation*. Oxford University Press, New York.
- Bohonak, A. J. 1999. Dispersal, gene flow, and population structure. *Quarterly Review of Biology* 74:21–45.
- Burnham, K. P., and D. R. Anderson. 2002. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, New York.
- Butlin, R. K., J. Galindo, and J. W. Grahame. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences* 363:2997–3007.
- Carlquist, S. J. 1974. *Island biology*. Columbia University Press, New York.
- Castella, V., M. Ruedi, L. Excoffier, C. Ibáñez, R. Arlettaz, and J. Hausser. 2000. Is the Gibraltar Strait a barrier to gene flow for the bat *Myotis myotis* (Chiroptera: Vespertilionidae)? *Molecular Ecology* 9:1761–1772.
- Chiba, S. 1999. Accelerated evolution of land snails *Mandarinina* in the oceanic Bonin Islands: evidence from mitochondrial DNA sequences. *Evolution* 53:460–471.
- Coyne, J. A., and T. D. Price. 2000. Little evidence for sympatric speciation in island birds. *Evolution* 54:2166–2171.
- Dahl, A. 2004. United Nations Environment Programme island directory. United Nations Environment Programme, Nairobi. <http://islands.unep.ch/isldir.htm>.
- Diamond, J. M. 1977. Continental and insular speciation in Pacific land birds. *Systematic Zoology* 26:263–268.
- . 2005. *Collapse: how societies choose to fail or succeed*. Viking, New York.
- Doebeli, M., and U. Dieckmann. 2003. Speciation along environmental gradients. *Nature* 421:259–264.
- Eastwood, A., Q. C. B. Cronk, J. C. Vogel, A. Hemp, and M. Gibby. 2004. Comparison of molecular and morphological data on St Helena: *Elaphoglossum*. *Plant Systematics and Evolution* 245:93–106.
- Emerson, B. C., and P. Oromi. 2005. Diversification of the forest beetle genus *Tarphius* on the Canary Islands, and the evolutionary origins of island endemics. *Evolution* 59:586–598.
- Endler, J. A. 1977. *Geographic variation, speciation, and clines*. Princeton University Press, Princeton, NJ.
- Eriksson, O., and B. Bremer. 1991. Fruit characteristics, life forms, and species richness in the plant family Rubiaceae. *American Naturalist* 138:751–761.
- Excoffier, L. 2001. Analysis of population subdivision. Pages 271–307 in D. J. Balding, M. Bishop, and C. Cannings, eds. *Handbook of statistical genetics*. Wiley, Chichester.

- Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131:479–491.
- Fuller, R. C. 2008. Genetic incompatibilities in killifish and the role of environment. *Evolution* 62:3056–3068.
- Funk, D. J., P. Nosil, and W. J. Etges. 2006. Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *Proceedings of the National Academy of Sciences of the USA* 103:3209–3213.
- Gaggiotti, O. E., and I. Hanski. 2004. Mechanisms of population extinction. Pages 337–366 in I. Hanski and O. E. Gaggiotti, eds. *Ecology, genetics, and evolution of metapopulations*. Elsevier, Amsterdam.
- Gavrilets, S. 2004. *Fitness landscapes and the origin of species*. Princeton University Press, Princeton, NJ.
- Gavrilets, S., and J. B. Losos. 2009. Adaptive radiation: contrasting theory with data. *Science* 323:732–737.
- Gavrilets, S., and A. Vose. 2005. Dynamic patterns of adaptive radiation. *Proceedings of the National Academy of Sciences of the USA* 102:18040–18045.
- Gillespie, R. 2004. Community assembly through adaptive radiation in Hawaiian spiders. *Science* 303:356–359.
- Gillespie, R. G., and B. G. Baldwin. 2009. Island biogeography of remote archipelagos: interplay between ecological and evolutionary processes. Pages 358–378 in J. B. Losos and R. E. Ricklefs, eds. *The theory of island biogeography at 40: impacts and prospects*. Princeton University Press, Princeton, NJ.
- Givnish, T. J., K. C. Millam, A. R. Mast, T. B. Paterson, T. J. Theim, A. L. Hipp, J. M. Henss, J. F. Smith, K. R. Wood, and K. J. Sytsma. 2009. Origin, adaptive radiation and diversification of the Hawaiian lobeliads (Asterales: Campanulaceae). *Proceedings of the Royal Society B: Biological Sciences* 276:407–416.
- Grant, P. R., and B. R. Grant. 2009. Sympatric speciation, immigration, and hybridization in island birds. Pages 326–357 in J. B. Losos and R. E. Ricklefs, eds. *The theory of island biogeography at 40: impacts and prospects*. Princeton University Press, Princeton, NJ.
- Holt, R. D., and R. Gomulkiewicz. 2004. Conservation implications of niche conservatism and evolution in heterogeneous environments. Pages 244–264 in R. Ferrière, U. Dieckmann, and D. Couvet, eds. *Evolutionary conservation biology*. Cambridge University Press, Cambridge.
- Jablonski, D. 2008. Species selection: theory and data. *Annual Review of Ecology, Evolution, and Systematics* 39:501–524.
- Kinlan, B. P., and S. D. Gaines. 2003. Propagule dispersal in marine and terrestrial environments: a community perspective. *Ecology* 84:2007–2020.
- Lande, R. 1988. Genetics and demography in biological conservation. *Science* 241:1455–1460.
- Lepage, D. 2008. Avibase: the world bird database. <http://www.bsc-eoc.org/avibase/avibase.jsp>.
- Long, J. C. 1986. The allelic correlation structure of Gainj-speaking and Kalam-speaking people. I. The estimation and interpretation of Wright's *F*-statistics. *Genetics* 112:629–647.
- Losos, J. B., and C. E. Parent. 2009. The speciation-area relationship. Pages 415–438 in J. B. Losos and R. E. Ricklefs, eds. *The theory of island biogeography at 40: impacts and prospects*. Princeton University Press, Princeton, NJ.
- Losos, J. B., and D. Schluter. 2000. Analysis of an evolutionary species-area relationship. *Nature* 408:847–850.
- MacArthur, R. H., and E. O. Wilson. 1967. *The theory of island biogeography*. Princeton University Press, Princeton, NJ.
- Mayr, E. 1942. *Systematics and the origin of species from the viewpoint of a zoologist*. Columbia University Press, New York.
- . 1963. *Animal species and evolution*. Belknap, Cambridge.
- . 1965. Avifauna: turnover on islands. *Science* 150:1587–1588.
- Morjan, C. L., and L. H. Rieseberg. 2004. How species evolve collectively: implications of gene flow and selection for the spread of advantageous alleles. *Molecular Ecology* 13:1341–1356.
- Nei, M. 1973. Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the USA* 70:3321–3323.
- Otto, S. P., and J. Whitton. 2000. Polyploid incidence and evolution. *Annual Review of Genetics* 34:401–437.
- Owens, I., P. Bennett, and P. Harvey. 1999. Species richness among birds: body size, life history, sexual selection or ecology? *Proceedings of the Royal Society B: Biological Sciences* 266:933–939.
- Parent, C. E., and B. J. Crespi. 2006. Sequential colonization and diversification of Galapagos endemic land snail genus *Bulimulus* (Gastropoda, Stylommatophora). *Evolution* 60:2311–2328.
- Paulay, G. 1985. Adaptive radiation on an isolated oceanic island: the Cryptorhynchinae (Curculionidae) of Rapa revisited. *Biological Journal of the Linnean Society* 26:95–187.
- . 1994. Biodiversity on oceanic islands: its origin and extinction. *American Zoologist* 34:134–144.
- Paulay, G., and C. Meyer. 2006. Dispersal and divergence across the greatest ocean region: do larvae matter? *Integrative and Comparative Biology* 46:269–281.
- Phillimore, A. B., R. P. Freckleton, C. D. L. Orme, and I. P. F. Owens. 2006. Ecology predicts large-scale patterns of phylogenetic diversification in birds. *American Naturalist* 168:220–229.
- Phillimore, A. B., C. D. L. Orme, G. H. Thomas, T. M. Blackburn, P. M. Bennett, K. J. Gaston, and I. P. F. Owens. 2008. Sympatric speciation in birds is rare: insights from range data and simulations. *American Naturalist* 171:646–657.
- Pickle, J., and J. Kirtley. 2008. *Analyzing digital images 2008*. Museum of Science, Boston.
- Price, J. P., and W. L. Wagner. 2004. Speciation in Hawaiian angiosperm lineages: cause, consequence, and mode. *Evolution* 58:2185–2200.
- Ranker, T., S. Floyd, and P. Trapp. 1994. Multiple colonizations of *Asplenium adiantum-nigrum* onto the Hawaiian archipelago. *Evolution* 48:1364–1370.
- R Development Core Team. 2007. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna.
- Ricklefs, R. E. 2007. History and diversity: explorations at the intersection of ecology and evolution. *American Naturalist* 170(suppl.):S56–S70.
- Rosenzweig, M. L. 1995. *Species diversity in space and time*. Cambridge University Press, Cambridge.
- Ryan, P. G., P. Bloomer, C. L. Moloney, T. J. Grant, and W. Delpont. 2007. Ecological speciation in South Atlantic island finches. *Science* 315:1420–1423.
- Savolainen, V., M. C. Anstett, C. Lexer, I. Hutton, J. J. Clarkson, M. V. Norup, M. P. Powell, D. Springate, N. Salamin, and W. J. Baker. 2006. Sympatric speciation in palms on an oceanic island. *Nature* 441:210–213.

- Schliewen, U. K., D. Tautz, and S. Paabo. 1994. Sympatric speciation suggested by monophyly of crater lake cichlids. *Nature* 368:629–632.
- Schluter, D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution* 16:372–380.
- Seddon, N., R. M. Merrill, and J. A. Tobias. 2008. Sexually selected traits predict patterns of species richness in a diverse clade of suboscine birds. *American Naturalist* 171:620–631.
- Seehausen, O. 2006. African cichlid fish: a model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences* 273:1987–1998.
- Sequeira, A., A. Lanteri, R. Albelo, S. Bhattacharya, and M. Sijapati. 2008. Colonization history, ecological shifts and diversification in the evolution of endemic Galapagos weevils. *Molecular Ecology* 17:1089–1107.
- Slatkin, M. 1973. Gene flow and selection in a cline. *Genetics* 75:733–756.
- . 1985. Gene flow in natural populations. *Annual Review of Ecology and Systematics* 16:393–430.
- Steadman, D. W. 2006. *Extinction and biogeography of tropical Pacific birds*. University of Chicago Press, Chicago.
- Stuart-Fox, D., and I. P. F. Owens. 2003. Species richness in agamid lizards: chance, body size, sexual selection or ecology? *Journal of Evolutionary Biology* 16:659–669.
- Stuessy, T. F., D. J. Crawford, and C. Marticorena. 1990. Patterns of phylogeny in the endemic vascular flora of the Juan Fernandez Islands, Chile. *Systematic Botany* 15:338–346.
- Stuessy, T. F., G. Jakubowsky, R. S. Gomez, M. Pfosser, P. M. Schluter, T. Fer, B. Y. Sun, and H. Kato. 2006. Anagenetic evolution in island plants. *Journal of Biogeography* 33:1259–1265.
- Thorpe, W. H. 1945. The evolutionary significance of habitat selection. *Journal of Animal Ecology* 14:67–70.
- Thrall, P. H., J. J. Burdon, and B. R. Murray. 2000. The metapopulation paradigm: a fragmented view of conservation biology. Pages 75–95 in A. G. Young and G. M. Clarke, eds. *Genetics, demography and viability of fragmented populations*. Cambridge University Press, Cambridge.
- Tilman, D., and P. M. Kareiva. 1997. *Spatial ecology: the role of space in population dynamics and interspecific interactions*. Princeton University Press, Princeton, NJ.
- Vekemans, X., and O. J. Hardy. 2004. New insights from fine-scale spatial genetic structure analyses in plant populations. *Molecular Ecology* 13:921–935.
- Veness, C. 2008. Geodesic distance between two latitude/longitude points using Vincenty ellipsoid formula in JavaScript. <http://www.movable-type.co.uk/scripts/latlong-vincenty.html>.
- Wagner, W. H. 1969. The role and taxonomic treatment of hybrids. *BioScience* 19:785–795.
- Whittaker, R. J., and J. M. Fernandez-Palacios. 2007. *Island biogeography*. Oxford University Press, Oxford.
- Whittaker, R. J., K. A. Triantis, and R. J. Ladle. 2008. A general dynamic theory of oceanic island biogeography. *Journal of Biogeography* 35:977–994.
- . 2009. A general dynamic theory of oceanic island biogeography: extending the MacArthur-Wilson theory to accommodate the rise and fall of volcanic islands. Pages 88–115 in J. B. Losos and R. E. Ricklefs, eds. *The theory of island biogeography at 40: impacts and prospects*. Princeton University Press, Princeton, NJ.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

Associate Editor: Kathleen Donohue
Editor: Mark A. McPeck



A survey of speciation events on oceanic islands shows that organisms that move around less can speciate within smaller areas. This general rule could help explain why taxonomic groups vary so much in diversity. Photograph by Ronit Weiss.