Ⓔ

Process-Based Models of Species Distributions and the Mid-Domain Effect

Sean R. Connolly[*]

Centre for Coral Reef Biodiversity, School of Marine Biology and Aquaculture, James Cook University, Townsville, Queensland 4811, Australia

ABSTRACT: Null models that place species ranges at random within a bounded geographical domain produce hump-shaped species richness gradients (the "mid-domain effect," or MDE). However, there is debate about the extent to which these models are a suitable null expectation for effects of environmental gradients on species richness. Here, I present a process-based framework for modeling species distributions within a bounded geographical domain. Analysis of null models consistent with the mid-domain hypothesis shows that MDEs are indeed likely to be ubiquitous consequences of geographical domain boundaries. Comparing the probability distributions of range locations for the process-based and randomization-based models reveals that randomization models probably overestimate the contribution of MDEs to empirical patterns of species richness, but it also indicates that other testable predictions from randomization models are likely to be robust. I also show how this process-based framework can be extended beyond null models to incorporate effects of environmental gradients within the domain. This study provides a first step toward an ecological theory of species distributions in geographical space that can incorporate both "geometric constraints" and effects of environmental gradients, and it shows how such a theory can inform our understanding of species richness gradients in nature.

Keywords: macroecology, mid-domain effect (MDE), geographic range, species richness, latitudinal gradient.

Understanding the causes of large-scale gradients in species richness has been a core enterprise of ecology since its inception. Numerous hypotheses for these gradients have been proposed, including physical environmental variables

* E-mail: sean.connolly@jcu.edu.au.

that increase the probability of species origination, invasion, or persistence, as well as differences in the time over which species have accumulated (for recent reviews, see Chown and Gaston 2000; Willig et al. 2003). In a seminal article, Colwell and Hurtt (1994) proposed a radically different kind of explanation, that the overlap of species ranges, and thus species richness, should increase toward the center of a geographical domain, even in the absence of environmental gradients. This hypothesis was inspired by the finding that species ranges, when placed at random within geographical domains, tend to produce a "mid-domain effect" (MDE)—a quasi-parabolic gradient in species richness, peaking in the center of a geographical domain (Colwell and Hurtt 1994; Colwell and Lees 2000). Subsequently, a variety of null models have been used to assess how much observed species richness gradients can be explained by mid-domain effects (for reviews, see Zapata et al. 2003; Colwell et al. 2004). Moreover, assessments of the effects of environmental factors on species richness (e.g., energy, area) have been reassessed in light of the MDE (Jetz and Rahbek 2002; Sanders 2002; Connolly et al. 2003), as well as other hypotheses about the causes of species richness gradients (e.g., Rapoport's rule; Lyons and Willig 1997; Hughes et al. 2002).

Null models are controversial, and mid-domain models are no exception (Zapata et al. 2003; Colwell et al. 2004). This controversy is due to the nonbiological nature of many null models, which makes both model formulation and the interpretation of model fit more problematic than for explicitly process-oriented models (Roughgarden 1983). For mid-domain models, the key question is how well randomizations of species ranges approximate the effects of geographical domain boundaries on species' distributions. Colwell and Hurtt (1994) reasoned that environmental conditions vary but that species' responses to environmental conditions would be sufficiently individualistic that, in the aggregate, no part of the domain would be more hospitable to species than any other part. From this, they conjectured that species' responses to geographically variable environmental conditions would tend to produce species distributions that were largely inde-

pendent of one another and in which one particular range location was as likely as any other.

Mid-domain models typically either randomize both range size and location (hereafter termed "fully randomized" models; Colwell and Hurtt 1994; Bokma et al. 2001; Grytnes 2003) or they randomize the locations of observed range sizes (Piñeda and Caswell 1998; Jetz and Rahbek 2001). In general, quantitative agreement between fully randomized models and empirical species richness gradients is poor, even when the empirical gradient is hump-shaped (for reviews, see Zapata et al. 2003; Colwell et al. 2004). This poor fit occurs because the model always predicts the same frequency distribution of range sizes relative to domain size (Arita 2004). In other words, fully randomized models assume that the frequency distribution of range sizes is entirely a property of the domain, independent of species' dispersal abilities or the breadth or narrowness of their environmental tolerances. On the other hand, randomization models that use observed range sizes tend to agree more closely with empirical data (e.g., Jetz and Rahbek 2001; Connolly et al. 2003). Using observed range sizes in a null model implicitly assumes that range size is determined entirely by species' intrinsic characteristics (e.g., dispersal characteristics, breadth of environmental tolerances). In nature, however, the distribution of environmental conditions, as well as species' intrinsic properties, plays a role in the determination of range size. Consequently, these models may inadvertently "smuggle in" effects of environmental gradients and overestimate the importance of MDE (Zapata et al. 2003).

Clearly, there are problems with both of these assumptions about how the frequency distribution of range sizes is determined. As a result, there is no consensus about whether geometric boundary constraints induce MDEs in nature, and if so, how important they are. Indeed, some have argued that the species richness gradients produced by null models are statistical artifacts that do not actually follow from biological assumptions consistent with a null expectation of no environmental gradients (Bokma et al. 2001; Hawkins and Diniz-Filho 2002; Zapata et al. 2003; but see Grytnes 2003). One way to resolve this problem is to extend mid-domain theory with more process-based models that allow species to differ in their breadth of environmental tolerances, whose assumptions are explicitly consistent with the mid-domain hypothesis and that do not incorporate features of observed data in a way that might smuggle in the effects of environmental gradients.

In this article, I undertake this extension of mid-domain theory by developing a general model of species distributions within a bounded geographical domain. Within this framework, I formulate specific null models for which the parameters governing the probability that a region of the domain contains suitable habitat are constant within a bounded geographical domain. I analyze these models to determine, first, whether MDEs are produced, and second, if they are, how well randomization models approximate them. I also show how this framework can be very naturally extended beyond the null hypothesis to formulate models that include effects of environmental gradients as well as domain boundaries. Such models allow us to quantify how environmental gradients cause species distributions to deviate from the patterns expected under MDEs alone.

Randomization Models

I compare predictions from the process-based null models with two randomization-based null models that are emerging as the preferred models in mid-domain analyses (Colwell et al. 2004). Both models randomly place observed range sizes within a bounded geographical domain (here normalized to extend from 0 to 1), but they do so in different ways. In the first model (here termed the "range shuffling model"), all geometrically feasible range locations are equally likely (Piñeda and Caswell 1998; Lees et al. 1999), so the probability that a species with range size r overlaps location x on the domain is

$$\Pr(x|r) = \begin{cases} 1 & \text{if } 1 - r \leq x \leq r \\ \dfrac{r}{1-r} & \text{if } r \leq x \leq 1 - r \\ \dfrac{1-x}{1-r} & \text{if } x > \max(r, 1-r) \\ \dfrac{x}{1-r} & \text{if } x < \min(r, 1-r) \end{cases} \tag{1}$$

(cf. Lees et al. 1999). Fully randomized models are special cases of this model; they follow equation (1), but as noted earlier, they also specify the frequency distribution of range sizes (Colwell and Hurtt 1994; Willig and Lyons 1998; Arita 2004). In the second model (the "spreading dye" model of Jetz and Rahbek 2001), a point of origin is chosen at random from within the domain, and the range then randomly spreads out from that point. An analytical expression for a one-dimensional version of this model is

$$\Pr(x|r) = \begin{cases} 1 & \text{if } 1 - r \leq x \leq r \\ r & \text{if } r \leq x \leq 1 - r \\ x + \dfrac{r}{2} & \text{if } x < \min\{r, 1-r\} \\ 1 - x + \dfrac{r}{2} & \text{if } x > \max\{r, 1-r\} \end{cases} \tag{2}$$

(for a full derivation, see app. A in the online edition of

the *American Naturalist*). Despite their differences, both models predict that all possible interior range locations (i.e., ranges not abutting domain boundaries) are equally likely; however, in the spreading dye model, species are more likely to abut a domain boundary (see app. A for details). Consequently, this model will always exhibit a shallower MDE than the range shuffling model.

To compare these models with the process-based models, I calculated predictions for the probability that a species' range overlaps location x for each randomization model from

$$\Pr(x) = \int_{r=0}^{1} \Pr(x|r)h(r)\,dr, \qquad (3)$$

where $\Pr(x|r)$ is the probability that a species of range size r overlaps location x, obtained from equation (1) or (2), as appropriate, and $h(r)$ is the probability distribution of range size obtained from the relevant process model (eq. [6], below). This is equivalent to simulating a set of species distributions according to the process model, then randomizing the locations of the "observed" (i.e., simulated) ranges.

General Model

To maintain consistency with the randomization models, the theory presented here models species as distributed within a one-dimensional, bounded domain whose length is normalized to unity. In each model, the location of a species' peak conditions is determined by a probability distribution $\omega(y)$, and its range extends outward until it encounters a species-specific range limit or a domain boundary, whichever occurs first. With these assumptions, the probability that a species' range overlaps location x is given by

$$\Pr(x) = \int_{y=0}^{1} \omega(y)\,\Pr(x|y)\,dy, \qquad (4)$$

where $\Pr(x|y)$ is the probability that a species encounters no distributional limit before reaching x, given that its peak conditions are at location y. In words, equation (4) states that the probability that a species range overlaps location x is the probability that its peak conditions are at location y and it does not encounter a range limit before extending to location x, summed over all possible values of y.

Similarly, an expression for the probability distribution

for range location (lower range endpoint at x_L, upper endpoint at x_U) can be written

$$
f(x_L, x_U) = \begin{cases}
\displaystyle\int_{y=0}^{1} \Pr(0|y)\Pr(1|y)\omega(y)\,dy & 0 = x_L, x_U = 1 \\[2ex]
\displaystyle\int_{y=x_L}^{1} g(x_L|y)\Pr(1|y)\omega(y)\,dy & 0 < x_L, x_U = 1 \\[2ex]
\displaystyle\int_{y=0}^{x_U} \Pr(0|y)g(x_U|y)\omega(y)\,dy & 0 = x_L, x_U < 1 \\[2ex]
\displaystyle\int_{y=x_L}^{x_U} g(x_L|y)g(x_U|y)\omega(y)\,dy & 0 < x_L, x_U < 1
\end{cases},
$$

$$(5)$$

where $g(x|y)$ is the probability density for the encounter of a distributional limit at x, given peak conditions at y. Thus, the first term is the (discrete) probability that a species encounters no distributional limit before encountering the upper and lower domain boundaries (i.e., it is pandemic). The second term is the probability density for a species that extends all the way to the upper domain boundary but encounters a species-specific distributional limit at x_L before reaching the lower domain boundary. The third term represents the converse: a species that abuts the lower domain boundary but encounters an upper range limit at x_U before reaching the upper boundary. The final term is the probability density for a species that encounters a lower limit before reaching the lower domain boundary and an upper limit before reaching the upper domain boundary. Equation (5) can be integrated to obtain a probability distribution for range size, r:

$$
h(r) = \begin{cases}
f(0, 1) & r = 1 \\[2ex]
\displaystyle f(0, r) + f(1-r, 1) + \int_{x_L=0}^{1-r} f(x_L, x_L + r)\,dx_L & r < 1
\end{cases}.
$$

$$(6)$$

This expression was used in equation (3) to generate the range shuffling and spreading dye predictions for comparison with the process-based null models presented below.

Poisson Model

An ideal starting point for a model of species distributions is to assume that species encounter distributional limits according to a Poisson process. There are three reasons for this. First, a null model allowing species to differ in their environmental tolerances can be formulated as parsimoniously as possible: with a single parameter. Second, assumptions of the Poisson model can be readily relaxed (see "Alternative Null Models"), allowing us to assess the robustness of model results. Third, it allows us to move beyond the null model and incorporate effects of environmental gradients in a straightforward way. Specifically, under a Poisson process, the probability that a barrier to range expansion is encountered in an infinitesimally small interval $(x, x + dx)$ is

$$\Pr[\text{barrier in } (x, x + dx)] = \lambda(x)dx, \qquad (7)$$

where $\lambda(x)$ is the stochastic rate at which barriers to range expansion are encountered at x. In general, we might expect $\lambda(x)$ to vary as a function of x (e.g., as a function of the quality of the environment at x or the steepness of gradients in environmental conditions at x). For this general model, the probability that a species' range does not encounter a range limit between its peak conditions at y and some other location x follows from the theory of inhomogeneous Poisson processes:

$$\Pr(x|y) = \begin{cases} \exp\left(-\displaystyle\int_{z=y}^{x} \lambda(z)dz\right) & y < x \\[2em] \exp\left(-\displaystyle\int_{z=x}^{y} \lambda(z)dz\right) & y > x \end{cases}. \qquad (8a)$$

Similarly, the probability density for the encounter of distributional limits is

$$g(x|y) = \begin{cases} \lambda(x)\exp\left(-\displaystyle\int_{z=y}^{x} \lambda(z)dz\right) & y < x \\[2em] \lambda(x)\exp\left(-\displaystyle\int_{z=x}^{y} \lambda(z)dz\right) & y > x \end{cases}. \qquad (8b)$$

If environmental suitability results from the interaction of many factors distributed independently in the domain, consistent with a null hypothesis that environmental gra-

dients do not influence species distributions (sensu Colwell and Hurtt 1994; Willig and Lyons 1998), then the probability distribution of peak conditions should be approximately uniform ($\omega(x) = 1$), and the stochastic rate at which species encounter range limits should be approximately constant across the domain ($\lambda(x) = \lambda$). With these constraints, equation (8a) can be combined with equation (4) to obtain the species richness gradient:

$$p(x) = \int_{y=0}^{x} e^{-\lambda(x-y)}dy + \int_{y=x}^{1} e^{-\lambda(y-x)}dy$$

$$= \frac{2 - e^{-\lambda x} - e^{-\lambda(1-x)}}{\lambda}. \qquad (9)$$

This model exhibits a quasi-parabolic mid-domain effect: the species richness gradient is steepest at domain boundaries, and it flattens as it approaches a maximum at the mid-domain (fig. 1A). This result is robust to the Poisson process assumption: for the general class of null models in which $\Pr(x|y)$ depends only on the distance between x and y (i.e., not on their specific locations), there is always a quasi-parabolic MDE (see app. B in the online edition of the *American Naturalist*).

Combining equations (5), (8a), and (8b) and noting that $r = x_U - x_L$, we can obtain the probability distribution of range endpoints:

$$f(x_L, x_R) = \begin{cases} e^{-\lambda} & 0 = x_L, x_U = 1 \\ \lambda r e^{-\lambda r} & 0 < x_L, x_U = 1 \\ \lambda r e^{-\lambda r} & 0 = x_L, x_U < 1 \\ \lambda^2 r e^{-\lambda r} & 0 < x_L, x_U < 1 \end{cases}. \qquad (10a\text{–}10d)$$

Like the MDE itself, this distribution exhibits a feature that is common to randomization-based null models. In particular, equation (10d) is independent of range location. In other words, conditional on range extent, r, all interior range locations (i.e., range locations other than those abutting a range boundary) are equally likely. However, the probability that a species' range abuts one or both domain boundaries is always higher for the process-based null model than for either of the randomization models; consequently, its MDE is always overestimated by the randomization models, particularly for intermediate λ (fig. 1B; see app. C in the online edition of the *American Naturalist*).

Alternative Null Models

Weibull Model. The Poisson process follows from the "law of rare events" (see, e.g., Taylor and Karlin 1994) and
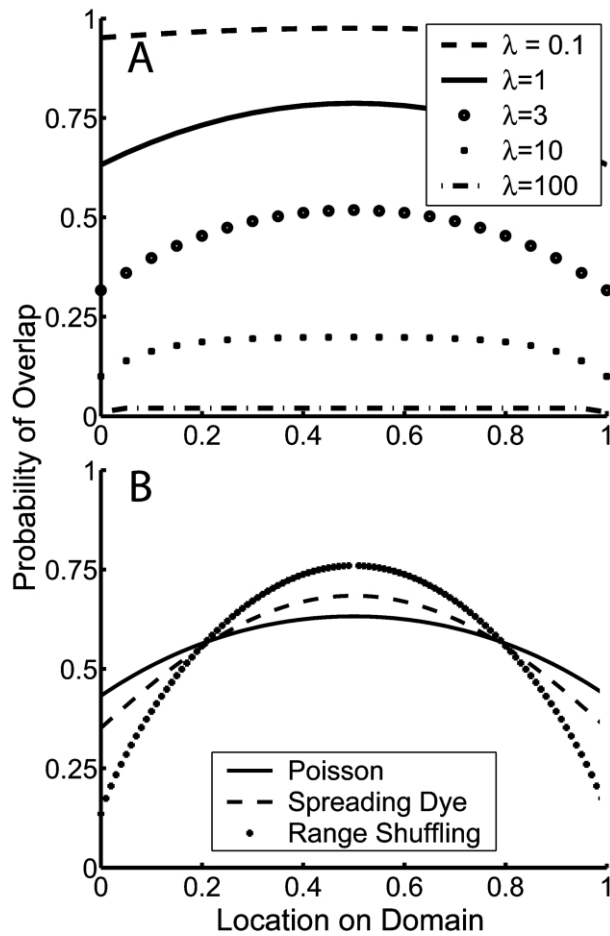
Figure 1: *A*, Probability of range overlap versus location on the domain for the Poisson model (eq. [6]) for a variety of values of the range limitation parameter, λ. Higher values of λ correspond to higher rates of encounter of range endpoints and thus smaller ranges. *B*, Representative comparison of the difference between the Poisson model's mid-domain effect and those of the two null models for an equivalent probability distribution of range extents as per eq. (3) (λ = 2).

entails two key assumptions: rarity and independence. The rarity assumption—that the probability of multiple events in one infinitesimally small interval is negligibly small— is clearly satisfied here because distributional limits will only be encountered once on either side of a species' peak conditions. The independence assumption implies that the probability of encountering a species-specific range limit at some location within the domain depends only on environmental conditions at that location (i.e., it is independent of how far a species has already successfully extended its range). In nature, this latter assumption may be violated; for instance, the probability of encountering a barrier to range expansion may be low near the location of peak conditions and then increase with increasing dis-

tance from that location. The Weibull distribution captures this form of distance-dependent stochastic process (Cox and Oakes 1984). Therefore, to assess the robustness of the results from the Poisson model, I developed a second model for which the distance from species' peak conditions to their distributional limits were drawn from a Weibull distribution.

For the Weibull model, the probability that a species' range extends from the location of its peak conditions to encompass location x is given by

$$\Pr(x|y) = \begin{cases} \exp[-\lambda^\gamma (x - y)^\gamma] & y < x, \\ \exp[-\lambda^\gamma (y - x)^\gamma] & y > x, \end{cases} \quad (11a)$$

and the probability density for the encounter of distributional limits is

$$g(x|y) = \begin{cases} \lambda^\gamma \gamma (x - y)^{\gamma-1} \exp[-\lambda^\gamma (x - y)^\gamma] & y < x \\ \lambda^\gamma \gamma (y - x)^{\gamma-1} \exp[-\lambda^\gamma (y - x)^\gamma] & y > x, \end{cases} \quad (11b)$$

where γ is the so-called shape parameter of the Weibull distribution. When $\gamma = 1$, this model is equivalent to the Poisson model. When $\gamma > 1$, then the stochastic rate at which range limits are encountered increases with increasing distance from the location of peak conditions (with larger γ corresponding to a more steeply increasing rate). As with the Poisson model (and consistent with the general proof in app. B), this model exhibits a quasi-parabolic MDE. As γ increases, the shape of the species richness gradient shifts so that species richness increases more sharply from domain boundaries toward an increasingly flat plateau through the mid-domain (fig. 2*A*). Despite this variability in exact shape, the MDE of the Weibull model, like that of the Poisson model, is always shallower than that of either randomization model (fig. 2*B*). Also consistent with the Poisson model, numerical integration of the Weibull model's analogue of equation (10d) indicates that, conditional on range size, all possible interior range locations (i.e., range locations other than those abutting domain boundaries) are equally likely.

Patchy Distribution Model. In the preceding models, all of the domain that lies between a species' distributional limits is occupied. However, in nature, many species consist of multiple geographically separated subpopulations connected by dispersal. This may occur because only a portion of the domain between potential distributional limits is habitable (i.e., patches of suitable habitat are separated by intervening regions of unsuitable habitat), or it may also occur as a result of the interaction between local population extinction and recolonization (i.e., unoccupied regions are habitable, but populations are not established
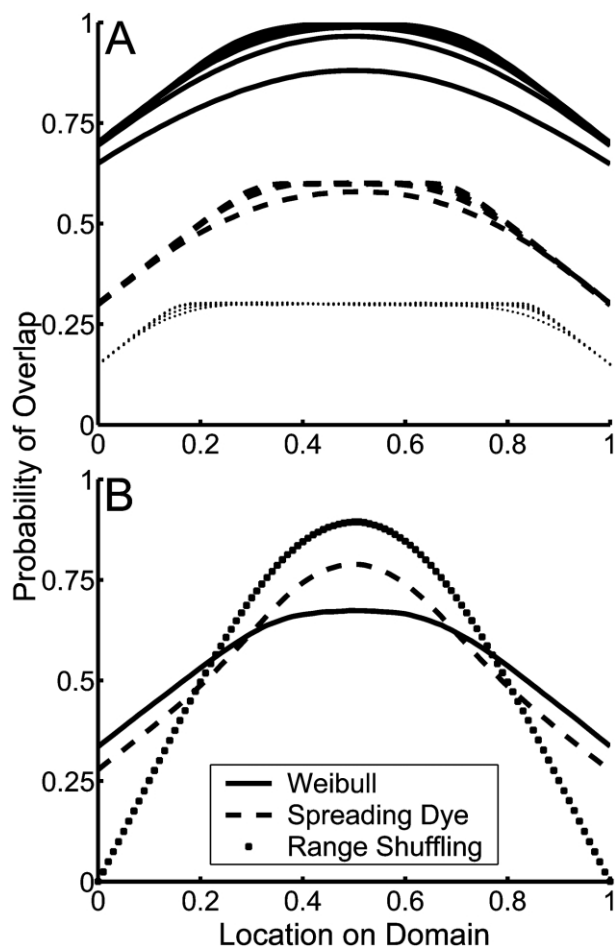
Figure 2: *A*, Probability of range overlap versus location on the domain for the Weibull model for a variety of values of the parameters λ and γ. For each line type, values of γ range from 2 (*lowermost line*) to 10 (*uppermost line*), with λ varied so that the mean of the Weibull distribution is constant (at 0.7, 0.3, and 0.15 for the *solid*, *dashed*, and *dotted lines*, respectively). *B*, Representative comparisons of the difference between the Weibull model's mid-domain effect and that of the two null models for an equivalent probability distribution of range sizes as per eq. (3) ($\lambda = 0.36$; $\gamma = 5$). Numerical investigations indicate that this result—shallower species richness gradients in the Weibull model than the randomization models—is consistent across parameter values.

there). To determine how this influences species richness gradients, I relaxed the range contiguity assumption by allowing species to consist of multiple discrete populations in between their potential range limits (which are still encountered at rate λ, as in the Poisson null model). Specifically, transitions from occupied to unoccupied habitat, and vice versa, occur at Poisson rates α and β, respectively (and consequently, the sizes of occupied patches and distances between them are exponentially distributed with means $1/\alpha$ and $1/\beta$, respectively). The derivation of species

richness gradients and probability distributions of range endpoints for this model is somewhat complicated (for details, see app. D in the online edition of the *American Naturalist*).

Based on patterns in range overlap, this model produces a broad range of MDEs, depending on how patchy species distributions are (fig. 3*A*–3*C*). MDEs are most pronounced when three conditions are met: range limitation is weak (λ on the order of 10^1 or smaller; fig. 3*A*), there are few—but more than one—patches of occupied habitat (β on the order of 10^0–10^1; fig. 3*B*), and patches of occupied habitat are small (α is large, on the order of 10^2 or greater; fig. 3*C*). However, because species in this model do not necessarily occupy all habitat between their potential range limits, range overlap (the probability that a species' range encompasses a location) and local presence-absence (the probability that a species occupies a location) are not equivalent. If, instead of range overlap, we consider local occupancy, a different picture emerges (fig. 3*D*–3*F*). Even for parameter values that produce very peaked MDEs from the standpoint of range overlap (i.e., regional species richness), MDEs in local occupancy (i.e., local species richness) are consistently shallow, similar in shape to—or in some cases, flatter than—the MDEs of the Poisson and Weibull models.

In contrast to the previous models, the patchy distribution model does not predict that, conditional on range size, all possible interior range locations are equally likely. Rather, because species may not occupy habitat adjacent to their potential distributional limits, range endpoints that would have abutted a domain boundary in the Poisson model are "smeared" into the domain so that range locations near domain boundaries are more likely than locations near the mid-domain (fig. 4; app. D). The extent to which this occurs depends on how frequently patches of occupied habitat recur within a species' distributional limits (i.e., on β) regardless of the other model parameters. In general, the effect is either quite small or confined to a narrow region near domain boundaries. The exception occurs when species ranges are highly noncontiguous (β is on the order of 10^1; occupied habitat patches are, on average, separated by distances ≈ 1/10 of the domain).

Discussion

These results support the hypothesis that the effects of domain boundaries alone will induce mid-domain effects, but they indicate that the height of these MDEs is constrained. In particular, for the Poisson and Weibull models, the maximum difference in species richness between the mid-domain and domain boundaries is approximately twofold (figs. 1, 2). Exceptions occur when species range over a few small patches of occupied habitat distributed
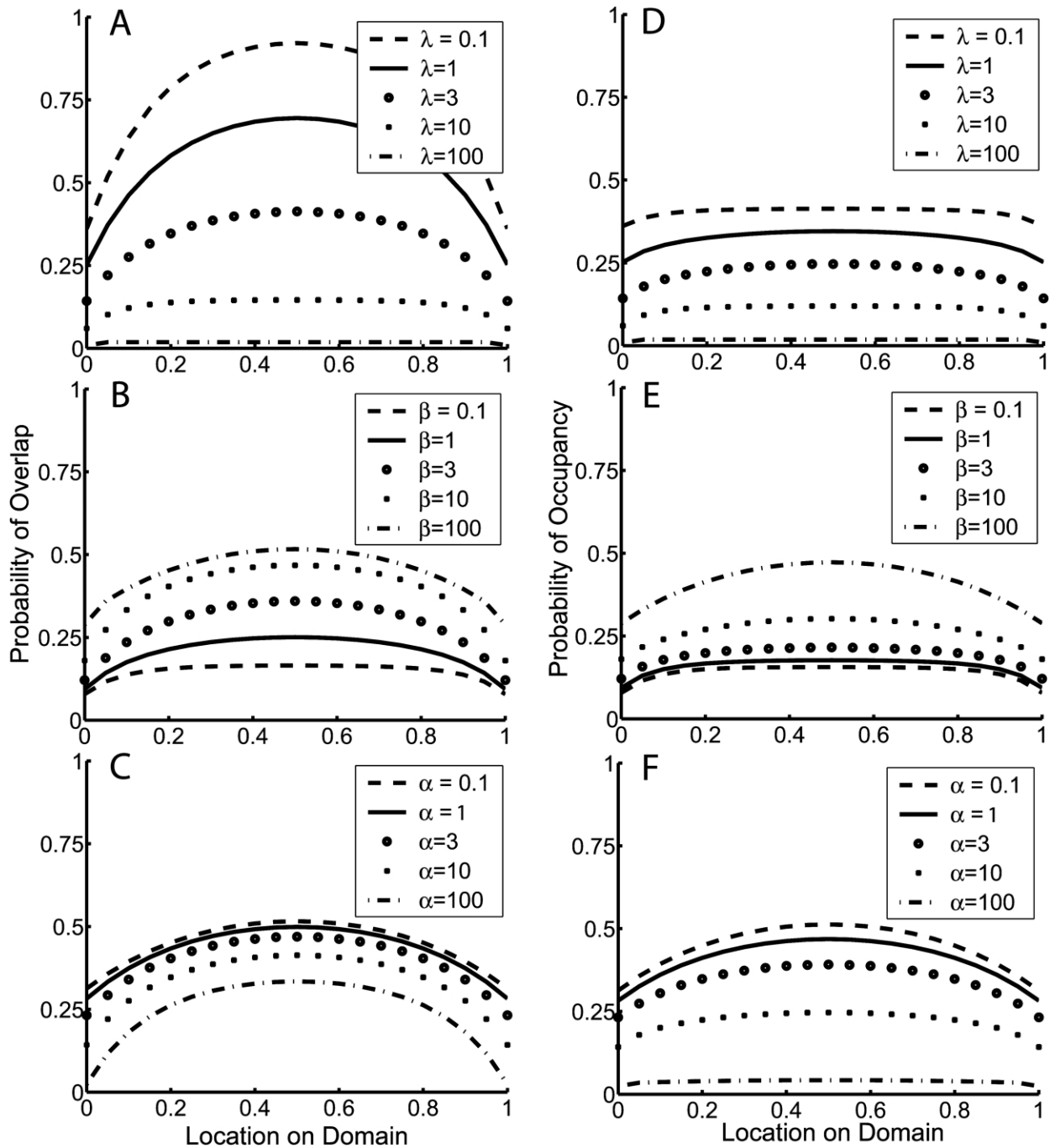
Figure 3: Probability of range overlap (*A–C*) and occupancy (*D–F*) versus location on the domain for the patchy distribution model, illustrating the effects of the three model parameters. *A, D,* Effect of varying range limitation, λ ($\alpha = 10$; $\beta = 5$). *B, E,* Effect of varying the distances between patches of occupied habitat, β ($\lambda = 3$; $\alpha = 10$). *C, F,* Effect of varying patchiness of occupied habitat, α ($\lambda = 3$; $\beta = 5$).

widely within the domain, in which case there is a pronounced MDE in range overlap but not local habitat occupancy (fig. 3). However, it is unlikely that this accounts for pronounced species richness gradients in nature: the contrasting patterns in range overlap versus local occupancy produced by the model imply that local species richness should remain fairly constant over very large gradients in regional species richness, whereas empirical data
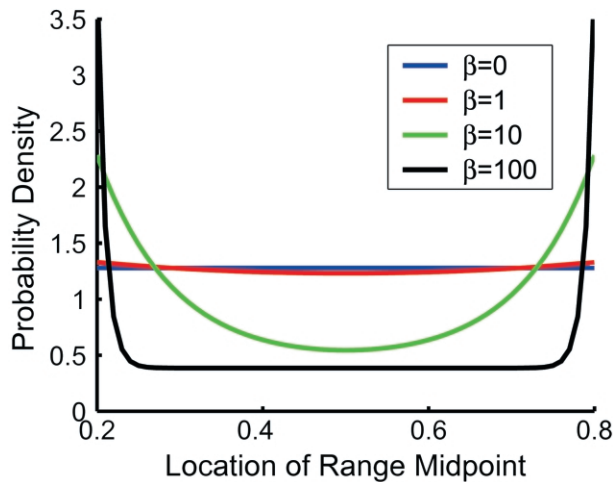
Figure 4: Probability density for alternative interior range locations, conditional on range size (eq. [D5c]), illustrated here for a range size of 0.4 (i.e., a range encompassing 40% of the domain) and a range of values of the habitat recurrence parameter, β. Note that because $r = 0.4$, only range midpoints between $x = 0.2$ and $x = 0.8$ are geometrically feasible. A horizontal line indicates a uniform distribution: all interior range locations (i.e., ranges not abutting a domain boundary) are equally likely. Bowl-shaped lines indicate that ranges are more likely to be near one or the other domain boundary than centered within the domain. Numerical investigation indicates that the effect of changes in β on the probability density is consistent for different combinations of λ and α. Therefore, only one case ($\lambda = 1$ and $\alpha = 10$) is shown here for illustration.

consistently show strong increases in local richness with regional richness (e.g., Ricklefs 1987; Cornell and Karlson 1997; Karlson et al. 2004; Witman et al. 2004).

It is likely that a strong constraint on the height of MDEs, such as that identified here, will prove to be a common feature of process-based null models, indicating that the extent to which boundary constraints can induce MDEs in nature will be limited. In particular, a constraint on the height of MDEs is robust to relaxation of the Poisson process and (for local species richness) the contiguous range assumptions (figs. 2, 3*D*–3*F*). Moreover, there is an intuitive explanation for this result. For species richness to have a high peak in the mid-domain, species must have broad environmental tolerances (i.e., encounter unsuitable conditions rarely), so that many species extend their distributions to encompass the mid-domain. However, in the absence of gradients in the quality of environmental conditions, these same broad tolerances increase the probability that species will extend all the way to domain boundaries, leading to relatively high species richness adjacent to those boundaries. This finding raises questions about the interpretation of statistical concordance between randomization models and observed patterns in species richness. This concordance arises because fully randomized

and range shuffling models, like many empirical richness gradients, generally exhibit a gradual decline from high values of species richness to zero or near zero at domain boundaries (e.g., Willig and Lyons 1998; Lees et al. 1999; Koleff and Gaston 2001; Ellison 2002; Connolly et al. 2003). However, if the magnitude of MDEs is constrained, then comparing observed versus randomized species richness gradients will overestimate the variation in species richness that can be explained by MDE.

Although the theory presented here raises questions about the interpretation of statistical concordance between observed species richness gradients and the predictions of randomization models, it also identifies an alternative way to use randomizations to conduct more robust tests of MDE. Except for species distributions that are highly non-contiguous, process-based null models and randomization models consistently agree on one prediction: conditional on range size, the geographical distribution of interior range locations is either uniform or nearly so. In other words, if a species richness gradient is principally due to MDE, then species ranges should be approximately randomly (i.e., uniformly) distributed among interior range locations, with a possible accumulation of species at or very near domain boundaries. This prediction is eminently testable, using data available from most data sets on species distributions (for an example, see Connolly et al. 2003).

Although testing for and quantifying deviations from a null model can be informative in and of itself, there is a need to understand whether and how MDEs may interact with the effects of environmental gradients to produce species richness gradients (Colwell et al. 2004). One way forward is to construct models that explicitly incorporate different assumptions about how environmental gradients and domain boundaries influence species distributions and then to compare these models with equivalent null models that omit these gradients. The general model presented here (eqq. [4]–[6]) is readily adapted for that purpose. For instance, consider an extension of the Poisson model (eqq. [8a], [8b]) that allows for symmetric variation in $\lambda(x)$ across the domain:

$$\lambda(x) = \begin{cases} \dfrac{a}{x^b} & 0 < x < 0.5 \\[2mm] \dfrac{a}{(1-x)^b} & 0.5 < x < 1 \end{cases}. \tag{12}$$

Equation (12) allows a broad range of gradients in environmental conditions (fig. 5*A*). It includes the null model as a limiting case: as b approaches zero, $\lambda(x)$ remains fairly constant through the interior of the domain and then increases rapidly toward infinity as domain boundaries are
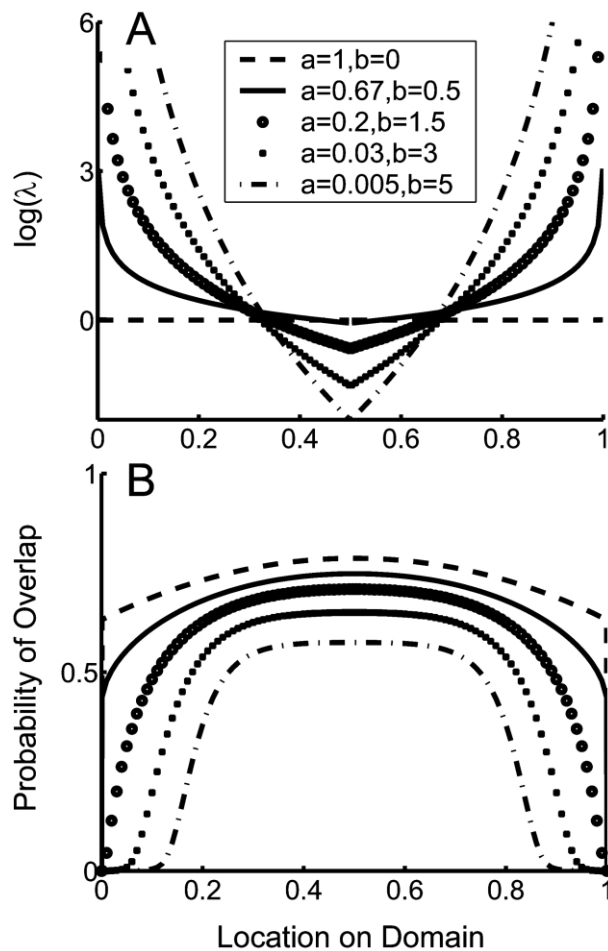
based or randomization-based null models, and it is qualitatively similar to that of some latitudinal gradients in species richness, which deviate from the quasi-parabolic shape of null models in the same qualitative way (e.g., New World bats and marsupials: Willig and Lyons 1998; New World parrots: Koleff and Gaston 2001; Indo-Pacific reef fishes and corals: Connolly et al. 2003). It also resembles the pattern of decline at the upper end of many altitudinal gradients (e.g., Andean plants: Kessler 2001; Himalayan plants: Grytnes and Vetaas 2002; North American ants: Sanders 2002). More strikingly, parameter values that lead to this kind of species richness gradient also produce a distinctive probability distribution of range locations that is markedly different from the uniform distributions produced by null models. In particular, large ranges are disproportionately centered near the mid-domain, but small ranges are centered away from the mid-domain, in the region where species richness exhibits its steep decline (fig. 6). Interestingly, the latitudinal ranges of Indo-Pacific reef fishes and corals also exhibit this peculiar geographical distribution (Connolly et al. 2003). This concordance suggests that even simple environmental gradient models can shed light on how and why real distributions of species ranges deviate from the assumptions of null models. Again, only this one study (to my knowledge) has conducted empirical analyses of this kind, but clearly, further theoretical and empirical work on the distributions of range locations in geographical space has the
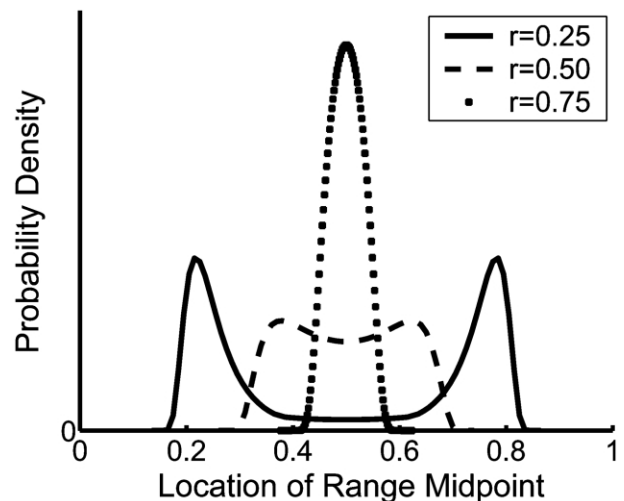
Figure 5: Effect of geographical gradients in the range limitation parameter, $\lambda(x)$, on species richness gradients (eq. [12]). *A*, Gradients in range limitation, $\lambda(x)$. *B*, Probability of range overlap versus location on the domain predicted by the gradients in $\lambda(x)$. The legend in *A* applies to both figures. Note that the dashed line is equivalent to the null model with $\lambda = 1$.

approached. As *b* increases, $\lambda(x)$ varies more gradually across the domain.

In specific empirical contexts, it may be particularly informative to model $\lambda(x)$ as an explicit function of different environmental variables such as habitat type, productivity, or temperature. However, even the simple model presented above produces some differences with the corresponding null model that are illuminating (fig. 5*B*). In particular, as the environmental gradient steepens, species richness exhibits an increasingly bell-like shape: there is a broad, shallow gradient in species richness in the mid-domain, then a steep decline followed by a tailing off of species richness as domain boundaries are approached. This tailing-off pattern is not produced by either process-

Figure 6: Probability distribution of range locations associated with parameters that generate bell-shaped distributions, showing the shift from a bimodal distribution at small range sizes to a unimodal distribution at large range sizes ($a = 0.03$; $b = 3$; cf. fig. 5). Numerical investigations indicate that the same pattern is generated for other combinations of parameter values that also produce bell-shaped species richness gradients.

potential to shed considerable light on how environmental gradients shape species distributions in nature.

One of the powerful features of mid-domain models is that they can be assessed against much more stringent conditions than, for example, regression models: they predict not only the shape of a species richness gradient but also the bivariate distribution of range sizes and locations that gives rise to those gradients (cf. "strong" vs. "weak" tests of McGill [2003]). The theoretical framework presented here preserves that important feature of null models (eq. [5]), but it also allows us to make the same kinds of predictions for models incorporating environmental gradients (e.g., fig. 6). Consequently, null models and environmental gradient models can be formulated for specific ecological contexts and explicitly fitted to empirical data. In this article, results have been expressed in terms of probability distributions on a species-by-species basis (e.g., probability of range overlap) because species differ in their breadths of environmental tolerances (and, for "non-null" models, their sensitivity to environmental gradients). Quantitative fits of models to data will require accounting for this heterogeneity in a parsimonious way, but methods exist for precisely these kinds of extensions (e.g., mixture models; Norris and Pollock 1996). A second consideration is nonindependence of environmental tolerances among species (e.g., closely related species may tolerate similar conditions), but again, the effects of nonindependence can be readily accounted for in parameter estimation (Buckland et al. 1993; Anderson et al. 1994).

The mid-domain hypothesis has prompted ecological biogeographers to recognize that species richness at a particular place depends not only on environmental quality at that location but on the broadscale distribution of environmental conditions that determines the locations and sizes of species ranges. Unfortunately, ecology has lacked a formal theory of species distributions in geographical space that allows us to model species ranges as a function of the distribution of environmental conditions that promote or inhibit the origination or persistence of species. This contrasts with many other areas of ecology, which are increasingly adopting approaches that confront process-oriented models directly with empirical data (e.g., behavioral and population ecology: Hilborn and Mangel 1997; some aspects of macroecology: e.g., Hubbell 2001; Engen et al. 2002; Gaston and He 2002). The present contribution offers a way forward, allowing formulation of null models as well as alternatives that incorporate environmental gradients. Using such process-based approaches to analyze species richness patterns, as a complement to prevailing regression and randomization approaches, is feasible and worth further exploration.

Literature Cited

Anderson, D. R., K. P. Burnham, and G. C. White. 1994. AIC model selection in overdispersed capture-recapture data. Ecology 75: 1780–1793.

Arita, H. T. 2004. Range size in mid-domain models of species diversity. Journal of Theoretical Biology 232:119–126.

Bokma, F., J. Bokma, and M. Mönkkönen. 2001. Random processes and geographic species richness patterns: why so few species in the north? Ecography 24:43–49.

Buckland, S. T., D. R. Anderson, K. P. Burnham, and J. L. Laake. 1993. Distance sampling: estimating abundance of biological populations. Chapman & Hall, London.

Chown, S. L., and K. J. Gaston. 2000. Areas, cradles, and museums: the latitudinal gradient in species richness. Trends in Ecology & Evolution 15:311–315.

Colwell, R. K., and G. C. Hurtt. 1994. Nonbiological gradients in species richness and a spurious Rapoport effect. American Naturalist 144:570–595.

Colwell, R. K., and D. C. Lees. 2000. The mid-domain effect: geometric constraints on the geography of species richness. Trends in Ecology & Evolution 15:70–76.

Colwell, R. K., N. Gotelli, and C. Rahbek. 2004. The mid-domain effect and species richness patterns: what have we learned so far? American Naturalist 163:E1–E23.

Connolly, S. R., D. R. Bellwood, and T. P. Hughes. 2003. Indo-Pacific biodiversity of coral reefs: deviations from a mid-domain model. Ecology 84:2178–2190.

Cornell, H. V., and R. H. Karlson. 1997. Local and regional processes as controls of species richness. Pages 250–268 *in* D. Tilman and P. Kareiva, eds. Spatial ecology: the role of space in population dynamics and interspecific interactions. Princeton University Press, Princeton, NJ.

Cox, D. R., and D. Oakes. 1984. Analysis of survival data. Chapman & Hall, London.

Ellison, A. M. 2002. Macroecology of mangroves: large-scale patterns and processes in tropical coastal forests. Trees: Structure and Function 16:181–194.

Engen, S., R. Lande, T. Walla, and P. J. deVries. 2002. Analyzing spatial structure of communities using the two-dimensional Poisson lognormal species abundance model. American Naturalist 160: 60–73.

Gaston, K. J., and F. He. 2002. The distribution of species range size: a stochastic process. Proceedings of the Royal Society of London B 269:1079–1086.

Grytnes, J. A. 2003. Ecological interpretations of the mid-domain effect. Ecology Letters 6:883–888.

Grytnes, J. A., and O. R. Vetaas. 2002. Species richness and altitude: a comparison between null models and interpolated plant species richness along the Himalayan altitudinal gradient, Nepal. American Naturalist 159:294–304.

Hawkins, B. A., and J. A. F. Diniz-Filho. 2002. The mid-domain effect cannot explain the diversity gradient of Nearctic birds. Global Ecology and Biogeography 11:419–426.

Hilborn, R., and M. Mangel. 1997. The ecological detective: confronting models with data. Princeton University Press, Princeton, NJ.

Hubbell, S. P. 2001. The unified neutral theory of biodiversity and biogeography. Princeton University Press, Princeton, NJ.

Hughes, T. P., D. R. Bellwood, and S. R. Connolly. 2002. Biodiversity hotspots, centers of endemicity, and the conservation of coral reefs. Ecology Letters 5:775–784.

Jetz, W., and C. Rahbek. 2001. Geometric constraints explain much of the species richness pattern in African birds. Proceedings of the National Academy of Sciences of the USA 98:5661–5666.

———. 2002. Geographic range size and determinants of avian species richness. Science 297:1548–1551.

Karlson, R. H., H. V. Cornell, and T. P. Hughes. 2004. Coral communities are regionally enriched along an oceanic biodiversity gradient. Nature 429:867–870.

Kessler, M. 2001. Patterns of diversity and range size of selected plant groups along an elevational transect in the Bolivian Andes. Biodiversity and Conservation 10:1897–1921.

Koleff, P., and K. J. Gaston. 2001. Latitudinal gradients in diversity: real patterns and random models. Ecography 24:341–351.

Lees, D. C., C. Kremen, and L. Andriamampianina. 1999. A null model for species richness gradients: bounded range overlap of butterflies and other rainforest endemics in Madagascar. Biological Journal of the Linnean Society 67:529–584.

Lyons, S. K., and M. L. Willig. 1997. Latitudinal patterns of range size: methodological concerns and empirical evaluations for New World bats and marsupials. Oikos 79:568–580.

McGill, B. J. 2003. Strong and weak tests of macroecological theory. Oikos 102:679–685.

Norris, J. L., and K. H. Pollock. 1996. Nonparametric MLE under two closed capture-recapture models with heterogeneity. Biometrics 52:639–649.

Piñeda, J., and H. Caswell. 1998. Bathymetric species-diversity patterns and boundary constraints on vertical range distributions. Deep Sea Research II: Topical Studies in Oceanography 45:83–101.

Ricklefs, R. E. 1987. Community diversity: relative roles of local and regional processes. Science 235:167–171.

Roughgarden, J. 1983. Competition and theory in community ecology. American Naturalist 122:583–601.

Sanders, N. J. 2002. Elevational gradients in ant species richness: area, geometry, and Rapoport's Rule. Ecography 25:25–32.

Taylor, H. M. and S. Karlin. 1994. An introduction to stochastic modeling. Academic Press, San Diego, CA.

Willig, M. R., and S. K. Lyons. 1998. An analytical model of latitudinal gradients of species richness with an empirical test for marsupials and bats in the New World. Oikos 81:93–98.

Willig, M. R., D. M. Kaufman, and R. D. Stevens. 2003. Latitudinal gradients of biodiversity: pattern, process, scale, and synthesis. Annual Review of Ecology and Systematics 34:273–309.

Witman, J., R. J. Etter, and F. Smith. 2004. The relationship between regional and local species diversity in marine benthic communities: a global perspective. Proceedings of the National Academy of Sciences of the USA 101:15664–15669.

Zapata, F. A., K. J. Gaston, and S. L. Chown. 2003. Mid-domain models of species richness gradients: assumptions, methods, and evidence. Journal of Animal Ecology 72:677–690.