

Species Delimitation: Inferring Gaps in Morphology across Geography

FELIPE ZAPATA^{1,2,4,*} AND IVÁN JIMÉNEZ³

¹Department of Biology, University of Missouri—St. Louis, One University Boulevard, St. Louis, MO 63121, USA; ²Graduate Studies Program, Division of Science and Conservation, Missouri Botanical Garden, PO Box 299, St. Louis, MO 63166, USA; ³Center for Conservation and Sustainable Development, Division of Science and Conservation, Missouri Botanical Garden, PO Box 299, St. Louis, MO 63166, USA; and ⁴Present address: Department of Integrative Biology, University of California, 3060 Valley Life Sciences Building, Berkeley, CA 94720-3140, USA;

*Correspondence to be sent to: Department of Integrative Biology, University of California, 3060 Valley Life Sciences Building, Berkeley, CA 94720-3140, USA; E-mail: felipe.zapata@berkeley.edu.

Felipe Zapata and Iván Jiménez have contributed equally to this work.

Received 4 August 2010; reviews returned 3 November 2010; accepted 9 May 2011

Associate Editor: Richard Glor

Abstract.—Species are commonly delimited on the basis of gaps in patterns of morphological variation, but there seems to be little recent work on methods to objectively assess such gaps. Here, we introduce a statistical approach that uses measurements of continuous morphological characters and geographic variation in those characters to (i) measure the strength of the evidence for the existence of a gap in morphological variation between two hypothesized species and (ii) examine if a gap in morphological variation between two hypothesized species can be explained by an alternative hypothesis of geographic variation within a species. This approach is based on recent developments in analyses of multivariate normal mixtures, estimates of multivariate tolerance regions, and principal coordinates of neighboring matrices. We demonstrate the application of the approach by examining previously proposed hypotheses of species limits in the plant genus *Escalonia*. We discuss the main features of the method, including potential limitations, in relation to other approaches that use gaps in morphological variation as a criterion for species delimitation. The method we propose can help strengthen the link between the theory and practice of species delimitation by increasing the transparency and consistency of taxonomic decisions based on morphology, thus contributing to integrative approaches for species delimitation that consider morphological and geographic data on an equal footing with other kinds of information. [Geographic variation; morphological variation; normal mixture; operational criteria; species limits; tolerance limit.]

Conventionally, species are recognized by seemingly discontinuous, nonoverlapping patterns of variation in characters of individuals from geographically circumscribed populations (Mayr 1942; Davis and Heywood 1963). However, identifying and delimiting species has been a controversial endeavor in systematics. Disagreements stem largely from the desire of biologists to provide an objective definition of species (Mayr 1982). Debates range from philosophical arguments as to what species are (ontology, e.g., Mishler and Donoghue 1982; Wilson 1999; Stamos 2003) to discussions on operational methods to delimit species (epistemology, e.g., Howard and Berlocher 1998; Sites and Marshall 2004). Underlying the controversy is the notion that understanding species boundaries is essential for elucidating the processes driving the origin and maintenance of biological diversity (Coyne and Orr 2004), as well as for basic and applied research in other areas of biology (Gaston 2000; Agapow et al. 2004; Padial and De La Riva 2006; Light et al. 2008). de Queiroz (1998, 2005) suggested that recognizing species as separately evolving segments of metapopulation-level lineages provides a context for integrating information from different sources for inferring lineage separation and thus for evaluating hypotheses about species boundaries. This pluralistic and unifying framework fosters the use and development of a wide range of methods for inferring species limits (de Queiroz 2007) including non-tree- and tree-based approaches (Sites and Marshall 2004). Non-tree-based approaches focus on gene flow assessments, such as crossability experiments to evaluate the level

of pre- and postzygotic reproductive isolation (e.g., Ackermann et al. 2008). Tree-based approaches focus on properties of phylogenetic trees (Sites and Marshall 2004) and include the use of morphological or molecular data to evaluate reciprocal monophyly (e.g., Wiens and Penkrot 2002; Lehtonen 2008) and coalescent theory to model the relationship between gene trees and species history (e.g., Knowles and Carstens 2007; Cummings et al. 2008). Recent studies have combined non-tree- and tree-based approaches, using georeferenced occurrence records and spatially explicit environmental information to evaluate demographic exchangeability and gene flow between populations (e.g., Raxworthy et al. 2007; Stockman and Bond 2007; Bond and Stockman 2008). However, non-tree-based approaches to delimit species using morphological variation have received little attention recently (Wiens 2007, but see Wiens and Servedio 2000) despite substantial past interest (e.g., Sneath and Sokal 1973). This is surprising because many, perhaps most, species are still delimited based on statements made by museum/herbarium-based systematists about patterns of morphological variation (Luckow 1995; Futuyma 1998; Wiens and Servedio 2000).

How do systematists use information on morphological variation to decide what to call a species? Often, seemingly nonoverlapping patterns of morphological variation are used as a criterion for inferring species boundaries (Davis and Nixon 1992; Wiens and Servedio 2000; Sites and Marshall 2004). This operational criterion—gaps in the pattern of morphological variation—is based on the idea that morphological

discontinuities suggest that some evolutionary force (e.g., lack of gene flow, natural selection) may prevent two distinct lineages from homogenizing (Wiens and Servedio 2000; Wiens and Penkrot 2002; Coyne and Orr 2004; Mallet 2008). Here, we focus on two issues that systematists contend with when using gaps in morphological variation as a criterion for inferring species limits.

Foremost, systematists need to weigh the strength of the evidence indicating a gap in morphological variation (or lack thereof). However, there seems to be little current work on how to do so. Various kinds of cluster analyses have been proposed to examine species limits using morphological data (Sneath and Sokal 1973), but algorithms used to define clusters lack a clear relationship to the operational criterion of a morphological gap (Dunn and Everitt 1982, p. 101–104) and respective underlying theory (Mayr 1992; de Queiroz and Good 1997; Wiens and Penkrot 2002). Clustering algorithms assume a hierarchical structure even though morphological variation at and below the species level is unlikely to be hierarchical and thus force data into groups irrespective of whether such groups exist in nature (Crisp and Weston 1993; de Queiroz and Good 1997). Some systematists have also tried to examine species limits using statistics that evaluate differences in central tendency (e.g., analysis of variance, discriminant analysis) among sets of samples (Luckow 1995; Henderson 2006); however, a difference in central tendency between samples is not the same as a gap between samples. In univariate or multivariate space, two samples may differ in central tendency in the absence of a gap separating two modes (McLachlan and Peel 2000; Ray and Lindsay 2005). Because the operational criterion of interest is a gap in morphology (see above), and not differences in central tendency (Luckow 1995, see also Patten and Unitt 2002), statistical tests focusing on measures of central tendency would not seem germane. Recently, a way to measure the strength of the evidence for the existence of a gap in qualitative or discrete characters between pairs of hypothesized species was developed based on statistical tolerance regions (Wiens and Servedio 2000). Here, we extend this approach to the case of quantitative continuous morphological characters using recent developments in statistics related to the topography of multivariate normal mixtures (Ray and Lindsay 2005) and estimates of multivariate tolerance regions (Krishnamoorthy and Mathew 1999; Krishnamoorthy and Mondal 2006).

A second issue that systematists confront is that the operational criterion of a morphological gap may require, contingent on species definitions (*sensu* de Queiroz 1998), that populations of the hypothesized species are sympatric (Mayr 1963; Futuyma 1998; Coyne and Orr 2004; Mallet 2008). Accordingly, systematists who try to evaluate species limits among nonsympatric populations may want to assess whether a morphological discontinuity interpreted as a species boundary can be explained by an alternative hypothesis of geographic variation within a single species (de Queiroz

and Good 1997; de Queiroz 2007). Ideally, the alternative hypothesis would identify the factors thought to shape geographic variation in morphology and would be contrasted against a hypothesis of a species boundary using spatial data on such factors. In practice, often little is known about the factors determining morphological variation and their spatial variation, and systematists resort to variables derived from spatial coordinates to examine geographic patterns of morphological variation. For instance, Mantel tests (Mantel 1967; Sokal 1979; de Queiroz and Good 1997) can be used to model simple large-scale spatial trends, akin to those predicted in neutral traits according to models of isolation by distance (Wright 1943, 1946, 1969). Aside from isolation by distance, environmental differences can determine spatial variation in morphology within a species (e.g., local adaptation) potentially resulting in spatial patterns more complex than large-scale spatial trends, which are difficult to describe using Mantel approaches or trend surface polynomial regression (Gabriel and Sokal 1969; Legendre and Legendre 1998). Here, we show how recent advances in spatial ecology, particularly the development of spatial analysis by means of principal coordinates of neighbor matrices (Borcard and Legendre 2002; Dray et al. 2006; Griffith and Peres-Neto 2006), can be used to model complex spatial patterns of morphological variation, like those that can arise from spatially structured evolutionary forces acting on a single species and thus provide a powerful alternative hypothesis of geographic variation in morphology that does not require a species boundary.

Below, we first describe the conceptual basis of the approach that systematists can use to deal with both issues stated above. Next, we demonstrate the application of this approach by using herbarium data to evaluate previously proposed hypotheses (Sleumer 1968) about species limits in the plant genus *Escallonia* L.f. (Escalloniaceae), a widespread group of neotropical shrubs and trees.

THE CONCEPTUAL BASIS

We emphasize that the approach we describe here is not designed to formulate hypotheses of species limits. Such hypotheses can be derived from previous taxonomic studies, as in our example data set (see below), or generated from the evaluation of other operational criteria (e.g., molecular phylogenetics or artificial crosses). Our method is designed to help systematists weigh the strength of the evidence supporting a previously proposed hypothesis about a species boundary. Thus, we consider taxa designated as species to be hypotheses of species as biological entities in nature (Hey et al. 2003; Wheeler 2004; Henderson 2005). Accordingly, we refer to taxa designated as species as “hypothesized species” and to the limits of those taxa as “hypothesis of species limits” or “hypothesis of species boundaries.”

Inferring Gaps in Morphology

Wiens and Servedio (2000) developed a statistical approach to evaluate hypotheses of species limits based on qualitative or discrete characters. Their approach was based on theory that emphasizes pairwise comparisons among hypothesized species, similar in this sense to approaches based on assessments of crossability (e.g., Rieseberg 2000; Dettman et al. 2003; Ackermann et al. 2008) that are necessarily based on pairwise comparisons. Here, we extend the approach developed by Wiens and Servedio (2000) to the case of quantitative continuous characters. In this case, we will assume that phenotypic variation arises from the effect of several genes and that there is random mating among the individuals in a geographic locality that belong to a single species. Under these assumptions, morphological variation is reasonably described by a normal distribution (Sokal and Rohlf 1995; Templeton 2006). Conversely, the hypothesis that there are two species in a geographic locality is supported when the distribution of morphological variation has two (or more, see below) modes (local maxima), so that there is a gap separating two approximately normal distributions (Futuyma 1998). Like the criterion based on gaps in the distribution of qualitative or discrete characters, the criterion based on gaps in the distribution of continuous characters assumes that bimodal (or multimodal) distributions do not result from polymorphisms, ontogenetic variation or phenotypic plasticity. Based on these assumptions, a hypothesis of species limits can be evaluated by assessing the number of modes in the distribution of continuous morphological characters of two hypothesized species, as we explain below.

To assess the number of modes in a univariate or multivariate morphological space X defined by measurements of two hypothesized species, one can estimate the probability density function (pdf) of the mixture of two normal distributions describing the morphological variation implied by the hypothesized species limits (McLachlan and Peel 2000):

$$f(X) = \pi_1 \phi(X; \mu_1, \Sigma_1) + \pi_2 \phi(X; \mu_2, \Sigma_2), \quad (1)$$

where π_1 and π_2 are the mixing proportions of the distributions describing the morphological variation of each of the two hypothesized species, $\phi(X; \mu_1, \Sigma_1)$ is the pdf of the distribution describing the morphological variation of one species with mean μ_1 and variance Σ_1 and $\phi(X; \mu_2, \Sigma_2)$ is the pdf of the distribution describing the morphological variation of the other species with mean μ_2 and variance Σ_2 . For our purposes, both hypothesized species should be weighted equally, so $\pi_1 = \pi_2 = 0.5$. Hereafter, we use carets to denote sample-based estimates of population parameters. For instance, $\hat{f}(X)$, $\hat{\phi}(X; \hat{\mu}_1, \hat{\Sigma}_1)$ and $\hat{\phi}(X; \hat{\mu}_2, \hat{\Sigma}_2)$ denote estimates of $f(X)$, $\phi(X; \mu_1, \Sigma_1)$ and $\phi(X; \mu_2, \Sigma_2)$, respectively.

When X is univariate, a mixture of two normal distributions can have one or two modes (Robertson and Fryer 1969), and evidence for bimodality (i.e., two local maxima) can be detected by visually inspecting a

plot of $\hat{f}(X)$, as illustrated by a numerical example of $f(X)$ (Fig. 1a), random samples from $\phi(X; \mu_1, \Sigma_1)$ and $\phi(X; \mu_2, \Sigma_2)$ (Fig. 1b) and the respective $\hat{f}(X)$ (Fig. 1c). Bimodality in $\hat{f}(X)$ can be regarded as a necessary, but not sufficient, condition to support the hypothesis that a sample of morphological measurements represents two species. Although bimodality implies that two local maxima are separated by a local minimum, the latter being the gap in morphological variation that suggests a species boundary, it does not directly inform how morphologically distinct two hypothesized species are. Following Wiens and Servedio (2000), one may select a priori a frequency cutoff below which overlap of phenotypic values between hypothesized species is considered rare enough as to suggest evolutionary isolation. Then, univariate tolerance regions for normal distributions (Natrella 1963) can be used to determine the degree of phenotypic overlap between hypothesized species. By example, the value \hat{a} in Figure 1b estimates with confidence $\gamma = 0.95$ the upper limit of the one-tailed tolerance region encompassing a proportion $\beta = 0.9$ of the hypothesized species represented by $\hat{\phi}(X; \hat{\mu}_1, \hat{\Sigma}_1)$. Likewise, the value \hat{b} in Figure 1b estimates with confidence $\gamma = 0.95$ the lower limit of the one-tailed tolerance region encompassing a proportion $\beta = 0.9$ of the hypothesized species represented by $\hat{\phi}(X; \hat{\mu}_2, \hat{\Sigma}_2)$. The confidence level γ is the frequency with which estimated tolerance regions encompass a proportion $\geq \beta$ of the sampled population: If we were to draw a large number of random samples from a given statistical population ($\phi(X; \mu_1, \Sigma_1)$ or $\phi(X; \mu_2, \Sigma_2)$), a proportion γ of those random samples would yield estimated tolerance regions that contain a proportion $\geq \beta$ of the population from which the samples were drawn (Fig. 1d). For a given confidence level γ , no overlap between tolerance regions would indicate that there is a gap between $\phi(X; \mu_1, \Sigma_1)$ and $\phi(X; \mu_2, \Sigma_2)$, so that proportions $\geq \beta$ of the two distributions do not overlap (Fig. 1b). If the overlap does not exceed the a priori frequency cutoff, then the results would support the hypothesis that the sample of morphological measurements represents two species.

In contrast to the case above in which X is univariate, a mixture of two multivariate normal distributions can have more than two modes (Ray and Lindsay 2005). Detecting and characterizing evidence for more than one mode in multivariate space X involves inspecting a plot of $\hat{f}(X)$ along the ridgeline manifold. The ridgeline manifold is a curve guaranteed to include all the critical points (maxima, minima, and saddles) of $\hat{f}(X)$ (Fig. 2; Ray and Lindsay 2005). Points along the ridgeline manifold correspond to values of variable α that range from zero at the multivariate mean of one hypothesized species to one at multivariate mean of the other (Fig. 2c,d). Visual inspection can reveal local maxima and minima of $\hat{f}(X)$ along the ridgeline manifold (Fig. 2f). For the same reason described above, more than one mode in $\hat{f}(X)$ along the ridgeline manifold

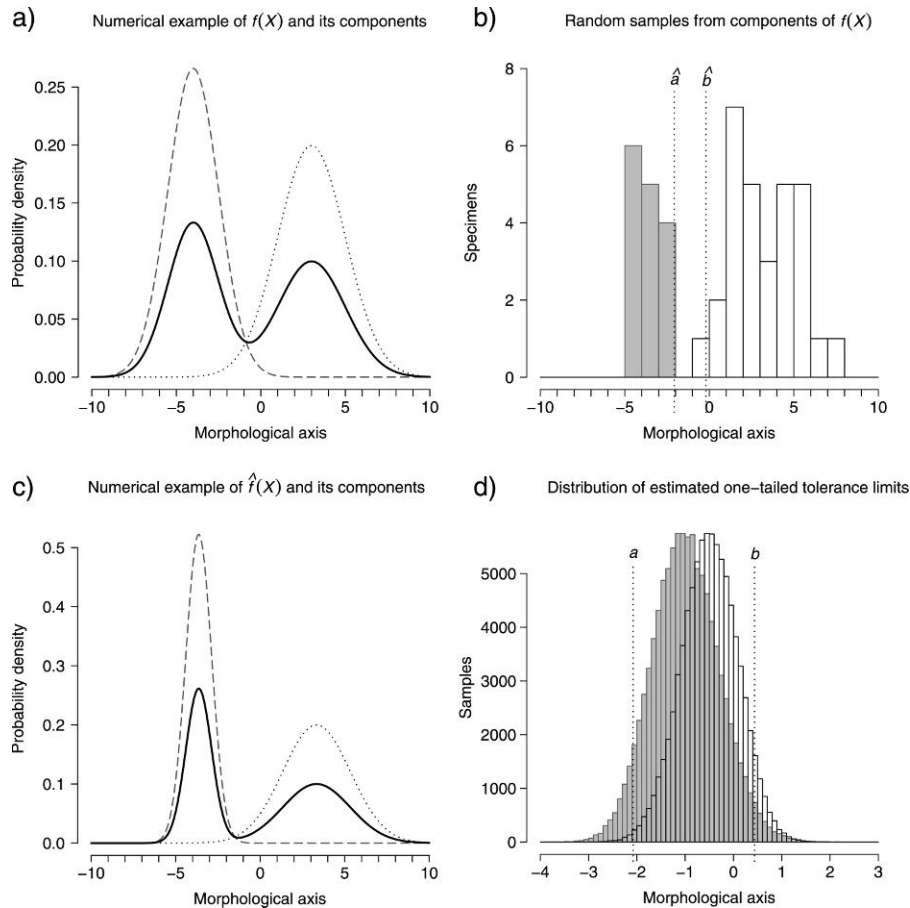


FIGURE 1. Numerical example of univariate normal mixtures and tolerance regions. a) pdf for the mixture of two distributions of a morphological variable (continuous line) and the corresponding statistical populations describing hypothesized species with parameters chosen for illustrative purposes ($\phi(X; \mu_1 = -4, \Sigma_1 = 2.25)$ dashed line, $\phi(X; \mu_2 = 3, \Sigma_2 = 4)$ dotted lines). b) Simulated random samples from each of the two statistical populations in (a). Gray bars: $\hat{\mu}_1 = -3.638$, $\hat{\Sigma}_1 = 0.583$, $n = 15$; white bars: $\hat{\mu}_2 = 3.331$, $\hat{\Sigma}_2 = 3.987$, $n = 30$. Values \hat{a} and \hat{b} estimate with confidence $\gamma = 0.95$ the limits of one-tailed tolerance regions encompassing a proportion $\beta = 0.9$ of the hypothesized species represented by gray and white bars, respectively. c) pdf of the normal mixture (continuous line) and its components (dashed and dotted lines) estimated from the samples in (b). d) Distribution of 100,000 estimated tolerance limits for each of the two populations in (a). Each limit was calculated for $\gamma = 0.95$ and $\beta = 0.9$, from simulated samples as those in (b). A proportion 0.947 ($\sim \gamma$) of the estimated limits for $\phi(X; \mu_1 = -4, \Sigma_1 = 2.25)$ (gray bars) were equal to or higher than a , the actual (as opposed to estimated) upper limit of the interval containing a proportion = 0.9 of the population. Likewise, a proportion 0.946 ($\sim \gamma$) of the estimated limits for $\phi(X; \mu_2 = 3, \Sigma_2 = 4)$ (white bars) were equal to or lower than b , the respective actual lower limit.

can be regarded as a necessary, but not sufficient, condition to support the hypothesis that the sample of morphological measurements represents two species. The frequency cutoff approach of Wiens and Servidio (2000) can also be applied when X is multivariate by examining overlap of ellipsoidal tolerance regions for multivariate normal distributions (Krishnamoorthy and Mathew 1999; Krishnamoorthy and Mondal 2006). Given a confidence level γ , one can calculate various ellipsoids defining tolerance regions for one of the distributions, for example, $\hat{\phi}(X; \hat{\mu}_1, \hat{\Sigma}_1)$, using various β values. Each of these ellipsoids shares a single point along the ridgeline manifold (corresponding to a single value of α) with an ellipsoid defining a tolerance region for the other distribution, $\hat{\phi}(X; \hat{\mu}_2, \hat{\Sigma}_2)$ (Fig. 3a, Ray and Lindsay 2005). A plot of the proportions β covered by these ellipsoids for several points along the ridgeline

manifold (corresponding to several values of α) reveals the estimated phenotypic overlap between the hypothesized species (Fig. 3b). If the overlap does not exceed the frequency cutoff selected a priori, then there would be empirical support for the hypothesis that the sample of morphological measurements represents two species.

Gaps versus Geographic Variation in Morphology

Phenotypic discontinuities can result from geographic differentiation within a single species (e.g., McGowen et al. 2001; Stenström et al. 2002; Olson et al. 2004). Therefore, under some species definitions (sensu de Queiroz 1998), it is often useful to examine whether a morphological discontinuity can be explained solely by geographic variation within a lineage, without having to invoke a boundary between species (de Queiroz and

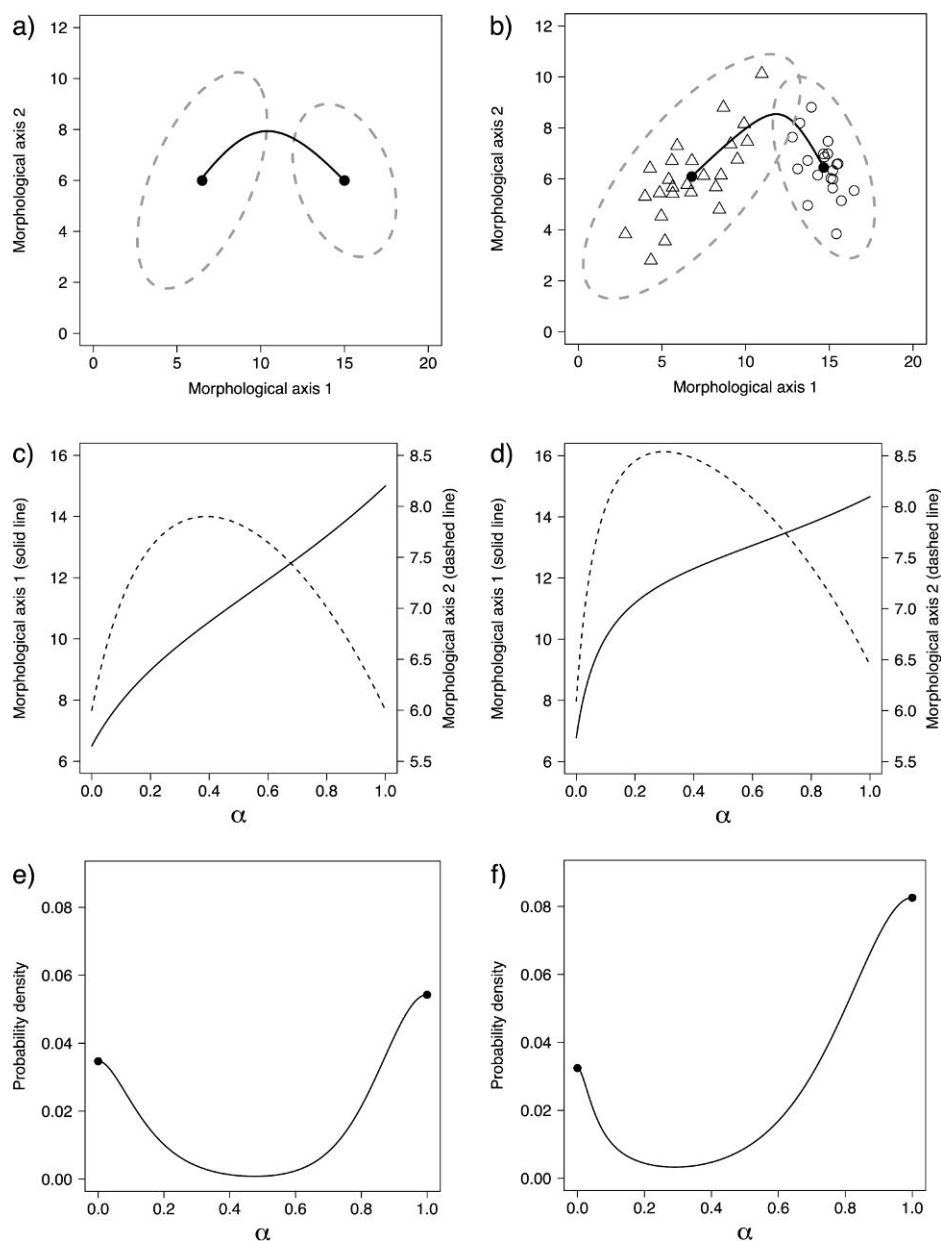


FIGURE 2. Numerical example of multivariate normal mixtures and tolerance regions. a) Two multivariate normal distributions describing hypothesized species with parameters chosen for illustrative purposes $\left(\phi \left(X; \mu_1 = \begin{bmatrix} 6.5 \\ 6 \end{bmatrix}, \Sigma_1 = \begin{bmatrix} 2.5 & 1.5 \\ 1.5 & 3.0 \end{bmatrix} \right), \phi \left(X; \mu_2 = \begin{bmatrix} 15 \\ 6 \end{bmatrix}, \Sigma_2 = \begin{bmatrix} 1.6 & -0.5 \\ -0.5 & 1.5 \end{bmatrix} \right) \right)$ are represented by ellipses delineating regions including a proportion to 0.95 of each statistical population and the respective multivariate (closed circles). The black continuous line that joins both multivariate means is the ridgeline manifold. b) Simulated random samples from each of the statistical populations in (a). Triangles: $\hat{\mu}_1 = \begin{bmatrix} 6.78 \\ 6.09 \end{bmatrix}, \hat{\Sigma}_1 = \begin{bmatrix} 4.79 & 2.55 \\ 2.55 & 2.62 \end{bmatrix}, n = 25$; circles: $\hat{\mu}_2 = \begin{bmatrix} 14.66 \\ 6.45 \end{bmatrix}, \hat{\Sigma}_2 = \begin{bmatrix} 0.95 & -0.55 \\ -0.55 & 1.30 \end{bmatrix}, n = 20$ Estimated (as opposed to actual) multivariate means are shown as closed circles and the estimated ridgeline manifold as a black continuous line. The dashed lines delimit tolerance regions for proportion $\beta = 0.9$ and confidence level $\gamma = 0.95$. c) Points along the ridgeline manifold in (a) correspond to values of variable α that range from 0 to 1 and map onto morphological space as shown by the two ordinates representing morphological Axes 1 and 2. d) Relationship between α and morphological Axes 1 and 2 as in (c) but based on estimated values from the random samples in (b). e) pdf evaluated at various points corresponding to values of α along the ridgeline manifold in (a). f) Estimated pdf evaluated at various points corresponding to values of α along the ridgeline manifold in (b).

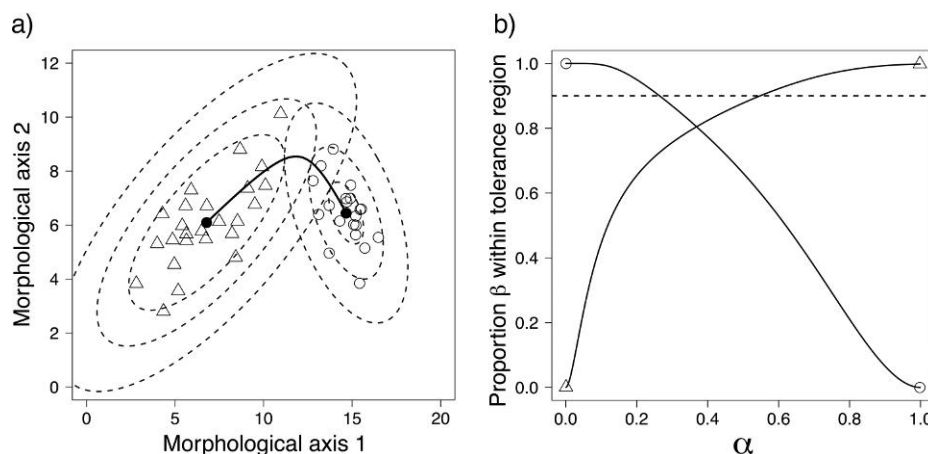


FIGURE 3. Numerical example of multivariate normal mixtures and tolerance regions. a) Estimated multivariate means (closed circles) and ridgeline manifold (black continuous line) based on simulated random samples (same as in Fig. 2b) from the two hypothesized species in Figure 2a. Dashed lines delimit three elliptical tolerance regions for each sample. These six tolerance regions form three pairs, each pair sharing a single point in the ridgeline manifold (corresponding to a single value of α). b) The continuous lines show, for different values of α along the ridgeline manifold (in the abscissa), the estimated proportion β (in the ordinate) covered with confidence $\gamma = 0.95$ by the two elliptical tolerance regions sharing a single point in the ridgeline manifold. Open triangles mark the ends of the line representing one hypothesized species and open circles for the other. The horizontal dotted line marks proportion $\beta = 0.9$.

Good 1997; de Queiroz 2007). To this end, geographic variation in morphology can be modeled using principal coordinates of neighbor matrices (Borcard and Legendre 2002; Dray et al. 2006; Griffith and Peres-Neto 2006). This approach uses the geographic coordinates of the collection localities of individual specimens on which morphological variables were measured to construct a matrix of geographic distances. Principal coordinate analysis of this distance matrix yields eigenvectors, known as spatial eigenvectors, which are useful to describe complex patterns of spatial variation in morphology, and can be used as explanatory variables in regression models to directly confront a hypothesis of species limits against the alternative hypothesis of geographic variation within a species, as explained next.

Multiple regression models can be used when morphological space X is univariate. A first step in this analysis involves accounting for spatial linear trends in morphology by regressing the morphological variable on the geographic coordinates of the specimens in the sample (see Borcard and Legendre 2002; Borcard et al. 2004). When a statistically significant spatial linear trend is detected in this first step, the response variable in the following step would be the residuals yielded after extracting the spatial linear trend. Otherwise, the response variable in the following step would be the original morphological variable. The next step involves a second multiple regression in which spatial eigenvectors enter as explanatory variables to determine whether variation in morphology can be modeled as geographic variation within a single species (Fig. 4). A dummy variable (coding one of the hypothesized species as zero and the other as one (Draper and Smith 1998)) also enters as explanatory variable to determine whether a significant amount of variation in morphology can be explained by the hypothesis of a species boundary (Fig. 4).

Further independent variables in this second multiple regression are interaction terms between each spatial eigenvector and the dummy variable, constructed by multiplying each spatial eigenvector by the dummy variable. These interaction terms also represent the hypothesis of a species boundary. For instance, the interaction term between the dummy variable and the first spatial eigenvector represents the idea that the relationship between the continuous morphological character (the response variable) and the first spatial eigenvector differs between both hypothesized species. The hypothesis of species limits would be supported if the analysis yields at least one statistically significant regression coefficient for the dummy variable or any of its interactions with spatial eigenvectors. In that case, one can conclude that variables representing the hypothesis of a species limit (the dummy variable or any of its interactions) accounted for variation in morphology that was not captured by spatial variables (spatial eigenvectors) representing an alternative hypothesis of geographic variation within a species. This result would considerably strengthen the evidence supporting the hypothesis of species limits because the spatial eigenvectors have remarkable power to model a wide range of spatial structures (Borcard and Legendre 2002; Dray et al. 2006; Griffith and Peres-Neto 2006).

When X is multivariate, redundancy analysis (RDA, Rao 1964) can be used to evaluate a hypothesis of species limits against the alternative hypothesis of geographic variation within a species. RDA is an extension of multiple regression that allows regressing several response variables (instead of just one response variable in multiple regression) on several independent variables (ter Braak 1995; Legendre and Legendre 1998). The analysis proceeds in two steps that are similar to those used when X is univariate. First, one runs an RDA in which

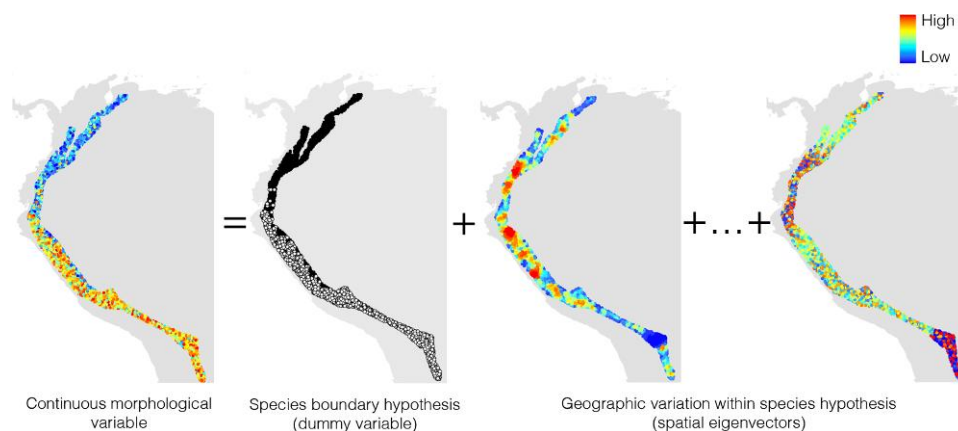


FIGURE 4. Illustration of how a multiple linear regression can be used to determine whether a gap in morphology can be explained by a hypothesis of a species boundary or by a hypothesis of geographic variation within a single species. The map on the left of the equal sign shows the geographic pattern of morphological variation measured on specimens of two hypothesized species. The hypothesis of a species boundary assigns the specimens measured to one (white circles) or the other (black circles) hypothesized species. The hypothesis of geographic variation in morphology within a single species is modeled with spatial eigenvectors derived from a principle coordinate analysis of a matrix of geographic distances of the collection localities of the specimens measured. Only two spatial eigenvectors are shown here, corresponding to eigenvalues of rank 1 (left) and 1001 (right). This figure is available in black and white in print and in color at *Systematic Biology* online.

the multivariate morphological measurements are the response variables and the geographic coordinates of the specimens in the sample are the only independent variables. If a statistically significant spatial linear trend is detected in this first step, the response variables in the next step would be the (multivariate) residuals around the linear trend, otherwise the response variables would be the original multivariate morphological measurements. The second step in the analysis involves an RDA in which the spatial eigenvectors are independent variables representing the hypothesis of geographic variation within a species, and as before, a dummy variable and its interactions with spatial eigenvectors represent the hypothesis of species limits. A statistically significant coefficient for any of the two latter independent variables provides support for the hypothesis of a species limit and indicates that a multivariate morphological discontinuity hypothesized to delimit two species cannot be entirely explained by a hypothesis of geographic variation within a species.

EMPIRICAL EXAMPLES

Study System

To illustrate the application of our approach, we used 2 examples from an ongoing systematic study on the evolution and diversification of the plant genus *Escallonia*. The complete data set examined here consists of 6 of the 39 taxa designated as species in the only available monograph of the genus (Sleumer 1968), taken here as hypotheses of species in nature: *Escallonia micrantha*, *Escallonia millegrana*, *Escallonia discolor*, *Escallonia piurensis*, *Escallonia resinosa*, and *Escallonia schreiteri*. Example A includes *Escallonia millegrana* and *Escallonia micrantha*, a pair of morphologically similar taxa for which species limits are uncertain. Molecular phylogenies

based on nuclear loci show that individuals sampled from both these taxa form a monophyletic group well isolated from all other taxa within *Escallonia* (Zapata 2010). Example B includes a group of four taxa that are morphologically similar in reproductive characters (i.e., inflorescence type and petal, ovary, and calyx shapes) but differ in vegetative morphology (i.e., leaf shape). All individuals sampled from these four taxa belong to a single well-supported clade in the aforementioned molecular phylogenies (Zapata 2010). Examples A and B illustrate different outcomes that may result from using the methods developed here.

Materials and Methods

We used a total of 85 herbarium specimens to represent the geographic range of morphological and ecological variation in the six taxa included in Examples A and B (Fig. 5). Specimens were assigned to each taxon after detailed studies of morphology, using a dichotomous key (Sleumer 1968), and carrying out comparative analyses using collections available from more than 17 herbaria including many of the same specimens that Sleumer studied. For each specimen, latitude–longitude coordinates were obtained from herbarium labels, high resolution maps, and online gazetteers (Guralnick et al. 2006). We characterized variation in 6 vegetative and 21 reproductive continuous characters (online Appendices 2–4 available from <http://www.sysbio.oxfordjournals.org/>), selected after careful study of specimens and literature review (Kausel 1953; Sleumer 1968), that describe the form of leaves (e.g., petiole length, lamina length, and width), inflorescences (length of main axis and first secondary axis), and flowers (e.g., ovary length and width, calyx tube length, calyx lobe length and width, and petal

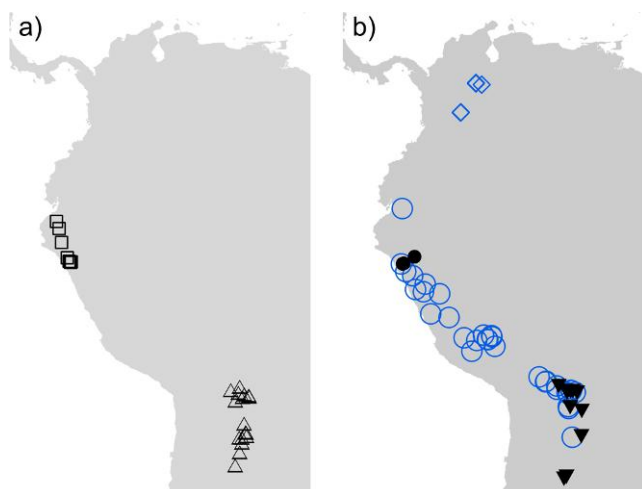


FIGURE 5. Outline of Northwestern South America and collection localities for the herbarium specimens included in the morphological samples of hypothesized *Escallonia* species. a) Example A: \square = *Escallonia micrantha*, Δ = *Escallonia millegrana*. b) Example B: \diamond = *Escallonia discolor*, \bullet = *Escallonia piurensis*, \circ = *Escallonia resinosa*, \blacktriangledown = *Escallonia schreiteri*. This figure is available in black and white in print and in color at *Systematic Biology* online.

length and width). We measured only mature leaves and flowers in all specimens. Vegetative characters were measured using a standard metric ruler on dried specimens, and floral characters were measured using a digital caliper (Digimatic CD-6" CS; Mitutoyo Japan) or a micrometer eyepiece on a stereoscopic dissecting microscope (SMZ645; Nikon USA) on flowers that were previously rehydrated. All measurements were recorded from three different structures for each specimen whenever possible and then averaged for each specimen.

For each of the two examples (A and B), we used measurements of continuous characters to derive orthogonal morphological axes using principal component analysis (PCA) on a correlation matrix. The first two resulting principal components, PC1 and PC2, captured 48.8% and 59.6% of the morphological variation in the samples for Examples A and B, respectively, and we used them to define the multivariate morphological space X for further analyses. In particular, we used the multivariate spaces defined by PC1 and PC2 to estimate the ridgeline manifold (Fig. 2b) and corresponding pdf (Fig. 2f) for each pair of taxa (i.e., hypothesized species). When the pdf along the ridgeline manifold did not exhibit more than one mode, we stopped the analyses because the first necessary condition to support the hypothesis of a species boundary was not met (see Inferring Gaps in Morphology section). Otherwise, we performed further analyses to estimate how morphologically distinct the two taxa were. In particular, we calculated a series of ellipsoid tolerance regions for each taxon (Fig. 3a) and estimated the proportions β covered by each of these ellipsoids (Fig. 3b) using a fixed confidence level $\gamma = 0.95$. Following Wiens and Servedio (2000), we used 0.1 as an a priori frequency cutoff below which overlap of

phenotypic values between hypothesized species was considered rare enough to suggest lineage isolation. Therefore, if the overlap of tolerance regions for two taxa calculated with $\gamma = 0.95$ did not exceed the frequency cutoff, we considered the two taxa sufficiently distinct to support the hypothesis of a species limit.

When taxa were morphologically distinct according to the frequency cutoff described above, we used RDA to ask whether the morphological discontinuity could be explained by a hypothesis of geographic variation within a single species. First, the values of PC1 and PC2 defining X were evaluated for spatial linear trends using the geographic coordinates of each specimen as explanatory variables (see Borcard et al. 2004). After detrending, we used the multivariate residuals as response variables in a second RDA. Among the explanatory variables in the second RDA, we used the spatial eigenvectors associated with positive eigenvalues (see Borcard and Legendre 2002; Dray et al. 2006; Griffith and Peres-Neto 2006) to represent the hypothesis of geographic variation within a single species. We used the geographic coordinates of the specimens to derive the spatial eigenvectors following the description in Dormann et al. (2007). All other explanatory variables in the second RDA represented the hypothesis of a species limit, including a dummy variable coding for each putative species and its interactions with the spatial eigenvectors. Only explanatory variables that were not highly correlated were retained in the RDA models. We assessed the statistical significance of the regression coefficient for each explanatory variable through permutation (Legendre and Legendre 1998). We used a standard model simplification procedure (Crawley 2002) to find a minimal adequate model by removing nonsignificant terms that did not increase model explanatory power. We measured model explanatory power as the "adjusted redundancy statistic," which is the RDA equivalent of the adjusted regression coefficient of determination (Peres-Neto et al. 2006). All statistical analyses were carried out in the R environment (R Development Core Team 2009) using the packages ellipse (Murdoch and Chow 2007), labdsv (Roberts 2007), mvtnorm (Genz et al. 2009), spdep (Bivand et al. 2009), vegan (Oksanen et al. 2009), and our scripts (online Appendices 5 and 6).

Results

Example A.—We measured 11 specimens assigned to *E. micrantha* and 17 assigned to *E. millegrana*. These samples separated clearly in multivariate morphological space (Fig. 6a). A plot of the estimated pdf along the ridgeline manifold revealed two local maxima, one at $\alpha = 0$, the estimated mean of the distribution describing *E. millegrana*, and the other at $\alpha = 1$, the estimated mean of the distribution describing *E. micrantha* (Fig. 7a). This implied bimodality in $\hat{f}(X)$, and thus, it fulfilled the first necessary condition to suggest a morphological gap separating this pair of taxa (hypothesized

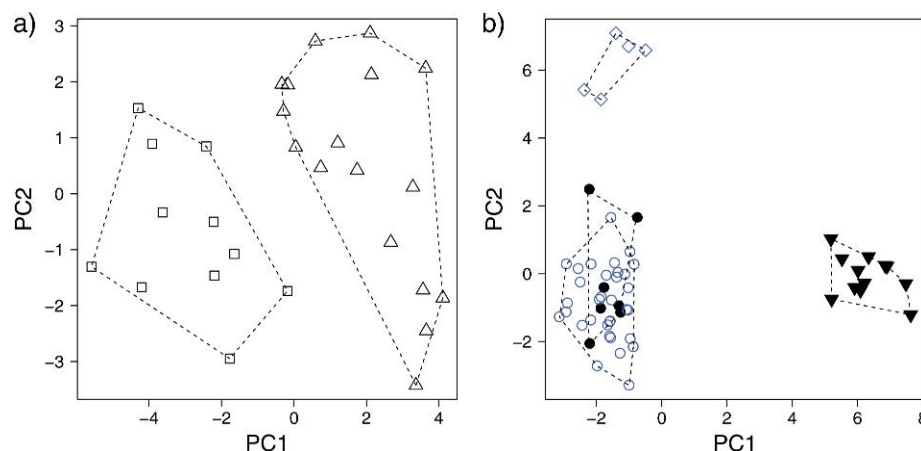


FIGURE 6. PCA describing the pattern of morphological variation among hypothesized *Escallonia* species. a) Example A, \square = *Escallonia micrantha*, Δ = *Escallonia millegrana*. b) Example B: \diamond = *Escallonia discolor*, \bullet = *Escallonia piurensis*, \circ = *Escallonia resinosa*, \blacktriangledown = *Escallonia schreiteri*. The dashed polygons correspond to the minimum convex hulls for the samples of each hypothesized species. This figure is available in black and white in print and in color at *Systematic Biology* online.

species). However, tolerance regions covering proportions $\beta < 0.9$ of the distributions of both taxa overlapped (Fig. 7b), failing to support the hypothesis of a species boundary according to the a priori frequency cutoff. As we did not find enough evidence in these data to support the idea that *E. micrantha* and *E. millegrana* were separated by a gap in the morphological space defined by PC1 and PC2, we conducted no further analyses because it was not necessary to confront the hypothesis of a species boundary against an alternative hypothesis of geographic variation within a species.

Example B.—Only 7 specimens were available for *E. piurensis* and 5 for *E. discolor*, whereas 12 and 35 specimens were available for *E. schreiteri* and *E. resinosa*, respectively. Specimens assigned to *E. discolor* and *E. schreiteri* separated clearly from all other taxa in the morphological space defined by PC1 and PC2, whereas

specimens assigned to *E. piurensis* and *E. resinosa* did not (Fig. 6b).

The plots of $\hat{f}(X)$ along the ridgeline manifold for *E. schreiteri* and each of the other taxa were all bimodal (Fig. 8a, c, e), suggesting a morphological gap between *E. schreiteri* and all other three taxa. The corresponding plots of the proportions covered by tolerance regions revealed that the phenotypic overlap was smaller than the frequency cutoff in all cases (Fig. 8b, d, f), indicating that *E. schreiteri* was sufficiently distinct from all other three taxa to support the hypothesis of a species limit. Therefore, we evaluated this hypothesis against an alternative of geographic variation within a species. Although the pattern of morphological variation in the pairwise comparison between *E. schreiteri* and *E. resinosa* was spatially structured and partly explained by spatial eigenvectors, a dummy variable representing

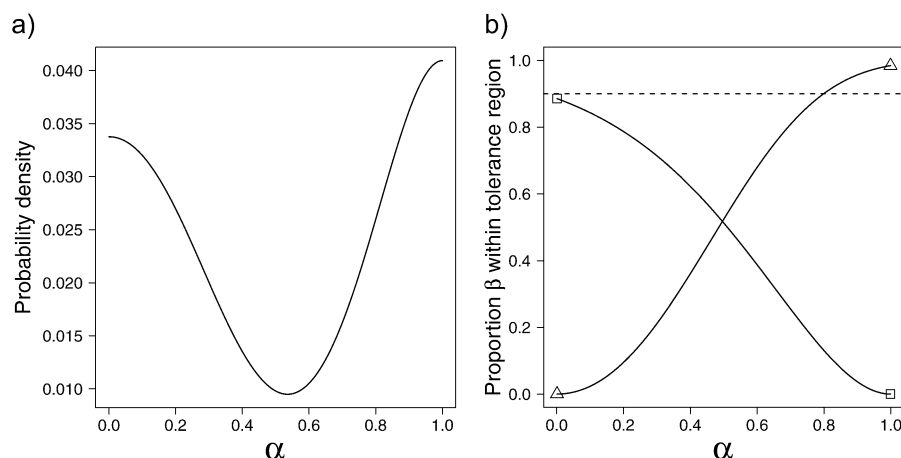
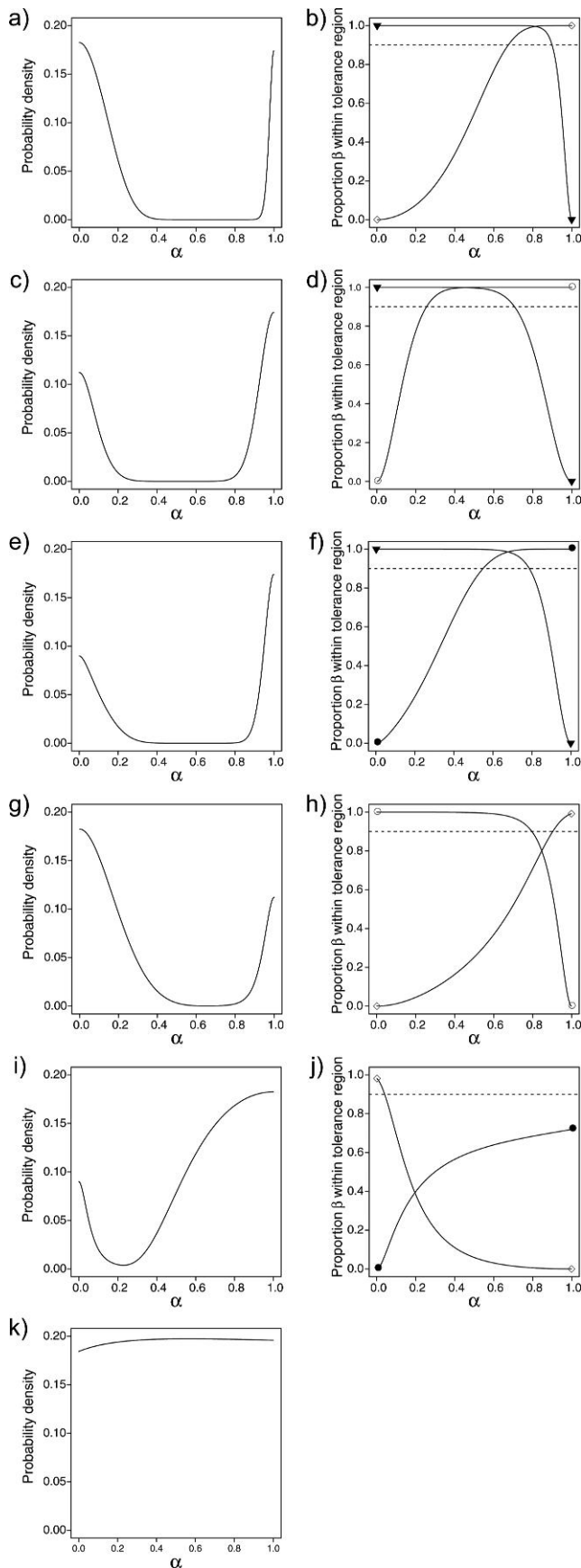


FIGURE 7. Inference of morphological gap between *Escallonia* species in Example A. a) Estimated pdf, $\hat{f}(X)$, evaluated for various points along the ridgeline manifold (α) for the mixture of two normal distributions of morphological variation implied by the hypothesized species limit; b) estimated proportion β (in the ordinate) covered by the elliptical tolerance regions ($\gamma = 0.95$) sharing a single point at α in the ridgeline manifold (in the abscissa). Symbols mark the end of lines representing each species; \square = *Escallonia micrantha*, Δ = *Escallonia millegrana*. The horizontal dotted line marks proportion $\beta = 0.9$.



a species boundary accounted for variation in morphology that was not captured by spatial variables (Table 1). In contrast, the morphological gaps separating *E. schreiteri* from *E. discolor* and *E. schreiteri* from *E. piurensis* were explained by geographic differentiation among allopatric populations (Fig. 5) within a single species. The relationship between PC1 and PC2 and the variables representing the hypotheses of species limits were not statistically significant for either of the latter pairwise comparisons (Table 1). Furthermore, the pattern of morphological variation in both pairwise comparisons corresponded to large-scale spatial trends (RDA for detrending yielded significant regression coefficients, not shown) rather than to more complex spatially structured patterns (Table 1, nonsignificant regression coefficients for spatial eigenvectors), suggesting simple geographic gradients in morphology in these taxa.

The plots of $\hat{f}(X)$ along the ridgeline manifold for *E. discolor* and *E. resinosa* and *E. discolor* and *E. piurensis* were both bimodal (Fig. 8g, i), consistent with the idea that a morphological gap separates each of these pairs of taxa. However, for each of these pairwise comparisons, tolerance regions covering proportions $\beta < 0.9$ overlapped (Fig. 8h, j), failing to support the hypothesis of a species boundary. As we did not find enough evidence in these data to support the idea that *E. discolor*, *E. resinosa*, and *E. piurensis* were separated by morphological gaps, it was not necessary to conduct further analyses to evaluate the hypothesis of geographic variation within a species.

Finally, the plot along the ridgeline manifold for *E. piurensis* and *E. resinosa* was unimodal (Fig. 8k), thus not meeting the first necessary condition to support the hypothesis of a species boundary. Therefore, we conducted no further analyses for this pairwise comparison.

DISCUSSION

Numerous statistical methods have been developed recently to help systematists weigh the strength of the evidence related to different operational criteria to infer species boundaries (e.g., Wiens and Servedio 2000; Wiens and Penkrot 2002; Sites and Marshall 2004; Pons et al. 2006; Knowles and Carstens 2007;

FIGURE 8. Inference of morphological gaps among *Escallonia* species in Example B. Each panel in the left column corresponds to the estimated pdf, $\hat{f}(X)$, evaluated for various points along the ridge-line manifold (α) for the mixture of two normal distributions of morphological variation implied by the hypothesized species limits. Each panel in the right column corresponds to the estimated proportion β (in the ordinate) covered by the elliptical tolerance regions ($\gamma = 0.95$) sharing a single point at α in the ridgeline manifold (in the abscissa). In these latter panels, symbols mark the end of lines representing each species and the horizontal dotted line marks proportion $\beta = 0.9$. a–b) *Escallonia discolor* (\diamond) and *Escallonia schreiteri* (\blacktriangledown), c–d) *Escallonia resinosa* (\circ) and *E. schreiteri* (\blacktriangledown), e–f) *Escallonia piurensis* (\bullet) and *E. discolor* (\diamond), g–h) *E. discolor* (\diamond) and *E. resinosa* (\circ), i–j) *E. piurensis* (\bullet) and *E. discolor* (\diamond), k) *E. piurensis* and *E. resinosa*.

TABLE 1. Results from RDA after accounting for spatial linear trend for three taxa (hypothesized species) pairs

Taxa pair and adjusted redundancy statistic	Source of variation	df	Variance	F	P
<i>Escallonia schreiteri</i> and <i>Escallonia resinosa</i> $Ra^2 = 0.88$ ($Ra^2 = 0.89$)	SEV1	1	0.017 (0.016)	0.640 (0.684)	0.48 (0.43)
	SEV2	1	0.462 (0.463)	18.150 (19.147)	0.01 (0.01)
	SEV3	1	0.410 (0.407)	16.000 (17.114)	0.01 (0.01)
	SEV4	1	0.022 (0.022)	0.852 (0.9145)	0.37 (0.34)
	SEV5	1	0.016 (0.016)	0.612 (0.657)	0.34 (0.40)
	SEV6	1	0.091 (0.091)	3.570 (3.822)	0.03 (0.06)
	SEV7	1	0.004	0.154	0.76
	Dum.	1	7.950 (7.926)	311.264 (333.089)	0.01 (0.01)
	SEV1 * Dum.	1	0.000	0.040	0.93
	SEV2 * Dum.	1	0.016	0.626	0.43
	SEV3 * Dum.	1	0.001	0.054	0.93
	SEV4 * Dum.	1	0.022 (0.022)	0.852 (0.9145)	0.37 (0.34)
	SEV5 * Dum.	1	0.022	0.880	0.35
	SEV7 * Dum.	1	0.043	1.710	0.16
	Residual	30 (37)	0.766 (0.880)		
<i>E. schreiteri</i> and <i>Escallonia discolor</i> $Ra^2 = 0$	SEV1	1	0.036	0.250	0.73
	Dum.	1	0.160	1.080	0.33
	SEV1 * Dum.	1	0.004	0.003	0.97
	Residual	13	1.933		
<i>E. schreiteri</i> and <i>Escallonia piurensis</i> $Ra^2 = 0$	SEV1	1	0.002	0.017	0.99
	Dum.	1	0.002	0.019	0.96
	SEV1 * Dum.	1	0.085	0.753	0.40
	Residual	15	1.693		

Note: Taxa names and adjusted redundancy statistic (Ra^2) for each RDA are shown in the first column. For each source of variation listed in the second column, the subsequent columns provide degrees of freedom (d.f.), variance, the F statistic, and the significance of the respective regression coefficient (P). Throughout the table values obtained after model simplification are shown in parentheses. Spatial eigenvectors (SEV) are listed in descending order according to their respective eigenvalues from SEV1 to SEVn. The dummy variable (Dum.) and its interactions with spatial eigenvectors (e.g., SEV1*Dum.) are listed afterwards. Rows corresponding to statistically significant ($P < 0.05$) regression coefficients are shown in bold font.

Raxworthy et al. 2007; O'Meara 2010). However, there is little recent work on how to measure the strength of the evidence related to the operational criterion of nonoverlapping patterns of variation in quantitative continuous morphological characters (Wiens 2007). Here, we presented a method that uses morphological and geographic data to formally assess whether there is enough evidence to indicate that a morphological gap supports a hypothesis of a species boundary and whether such a gap could be explained by an alternative hypothesis of geographic variation within a single species. This method, hereafter GCG (Gaps in Continuous characters across Geography), is based on recent developments in analyses of multivariate normal mixtures (Ray and Lindsay 2005), estimates of multivariate tolerance regions (Krishnamoorthy and Mathew 1999; Krishnamoorthy and Mondal 2006), and principal coordinates of neighboring matrices (Borcard and Legendre 2002; Borcard et al. 2004). We illustrated the application of GCG by using herbarium data to evaluate previously proposed hypotheses (Sleumer 1968) about species limits within two groups of morphologically similar taxa, both forming well-supported clades in molecular phylogenies of the neotropical plant genus *Escallonia* (Zapata 2010). Analysis of these empirical examples shows that GCG is

readily applied to real-world data of the kind typically available to museum/herbarium-based systematists. Further work on groups where species boundaries are well-known a priori will be needed to assess the performance of GCG relative to alternatives.

Below, we discuss properties of GCG that may be important determinants of its performance and use those properties as a basis of comparison with other non-tree-based approaches (sensu Sites and Marshall 2004) that emphasize discontinuities in morphological variation (i.e., gaps) as a criterion to infer species limits: population aggregation analysis (Davis and Nixon 1992), hereafter PAA, and estimates of tolerance regions for discrete or qualitative characters (Wiens and Servedio 2000), hereafter GD (Gaps in Discrete characters). We do not further discuss approaches based on clustering analyses or statistical tests of differences in central tendency because they do not seem to be clearly related to the operational criterion of a gap in morphological variation (see introduction). Likewise, we do not discuss tree-based approaches (sensu Sites and Marshall 2004) to delimit species using morphological data (Baum and Donoghue 1995, Wiens and Penkrot 2002) because they are based on fundamentally different criteria and underlying theory (Sites and Marshall 2004) and thus are outside the scope of this paper.

In principle, hypotheses of species limits can be supported or rejected (Hey et al. 2003, Wheeler 2004, Henderson 2006); however, how a single empirical study employing GCG can support or reject a given hypothesis depends on the extent to which various assumptions are reasonably met, as well as other aspects of the study design and system. Like PAA and GD, GCG assumes that morphological discontinuities presented as evidence supporting hypotheses of species limits do not result from polymorphisms, ontogenetic variation, or phenotypic plasticity. GCG, but not PAA or GD, assumes that morphological variation is reasonably described by a normal distribution. When PAA, GD, or GCG fail to support a particular hypothesis, it is possible to argue that important morphological characters supporting the hypothesis were not included in the analysis. This argument, however, would not apply to hypotheses of species limits explicitly stated in terms of particular morphological characters, as may be the case for taxa described in formal statements about measures of dispersion of different characters (Wheeler 2004; Henderson 2005).

An important property of GCG is that it can be effectively applied to morphological spaces with any number of dimensions. For simplicity, we illustrated the implementation of this method using examples based on analyses in bivariate morphological spaces (the first two axes from PCA). However, all steps in the method (i.e., estimating the pdf along the ridgeline manifold, the tolerance regions and coefficients of the RDA) can be carried out using data in more dimensions. It is worth noting that the ridgeline manifold of the multivariate normal mixture of two taxa (i.e., two hypothesized species) is always one dimensional, regardless of the number of dimensions of the morphological space (Ray and Lindsay 2005). This implies a major reduction of dimensionality that readily allows visual inspection of key features of high-dimensional morphological spaces. Specifically, inspecting the pdf along the one-dimensional manifold (Fig. 2f) reveals gaps and modes that are otherwise difficult to visualize. Furthermore, using the one-dimensional manifold to compare the proportions covered by tolerance regions (Fig. 3b) simplifies the otherwise difficult task of assessing by visual inspection the overlap between high-dimensional ellipsoidal tolerance regions. Importantly, GCG accounts for within-taxa correlations between morphological dimensions (through covariance terms), unlike PAA and GD, and does not assume independence of the variables used to describe multivariate morphological space, in contrast to GD.

The distinction between morphological gaps among samples (i.e., groups of specimens representing particular hypothesized species) and morphological gaps among the populations from which the samples were derived is nontrivial and central to species delimitation (Wiens and Servedio 2000). In GCG, as well as in GD, support for a hypothesis of a species boundary is gauged according to the overlap of tolerance regions inferred from samples. The degree of overlap between

tolerance regions is an estimate of morphological gaps between the populations from which the samples were derived. In contrast, in PAA species limits are drawn by gauging overlap of intervals defined by extreme sample values. We believe that focusing on tolerance regions is desirable because they attempt to describe the properties of the populations from which the samples were drawn. Hypothesis about species limits are about such populations (Hey et al. 2003; Wheeler 2004; Henderson 2005), not about samples.

Calculation of tolerance regions to infer morphological distinctiveness, as proposed here following GD, assumes that the observed sample is a random draw from a statistical population representing a hypothesized species. This assumption means that the probability of any individual being included in the sample is equal across all individuals in the statistical population and that the inclusion of one individual in the sample does not influence the probability of including any other individual (Zar 1999). The samples used in systematic studies are generally not random samples of hypothesized species. This is particularly exacerbated in museum/herbarium-based studies where species are often recognized from limited samples, which themselves frequently cover a small part of the hypothesized species' total geographic range and phenotypic variability (Davis and Heywood 1963; Henderson 2006). Even though PAA does not resort to formal inferences about statistical populations based on sample estimates (see above), error in estimates of species limits produced by this method can also result from nonrandom sampling of individuals and populations. Problems may arise, for instance, when bias geographic sampling fails to secure specimens from a zone of intergradation that bridge what may be mistakenly perceived as a morphological gap between allopatric populations (Davis and Nixon 1992). The mismatch between assumed and actual sampling strategies is not only an issue for methods emphasizing discontinuities in morphological variation as a criterion to infer species limits but is also common to virtually all approaches to infer species limits (see also Manel et al. 2003; Pons et al. 2006; Bond and Stockman 2008; Lohse 2009; Papadopoulou et al. 2009). Implementing formal sampling designs across large geographical scales may often be difficult, and future simulation work may help us understand the relationships between properties of different methods to delimit species and the procedures by which samples of hypothesized species are obtained. In the mean time, the mismatch between assumed and actual sampling strategies mandates caution in the interpretation of approaches for species delimitation.

Difficulties in the interpretation of results can arise when an analysis based on GCG fails to support a given hypothesis because often it is not possible to equate absence of evidence with evidence of absence, an issue common to all attempts to evaluate scientific hypotheses using statistical inferences (Gotelli and Ellison 2004). For instance, it is possible to interpret absence of evidence for a morphological gap as a lack of statistical

power due to insufficient data. Eliminating this possibility may be difficult because there is not a way to formally determine what constitutes a large enough sample size without knowing the population parameters defining the distribution mixture corresponding to a pair of hypothesized species (i.e., the population means, variances and covariances). In this context, there is a noteworthy difference between overlap of tolerance regions, as used by GCG and GD, and overlap of intervals defined by extreme sample values used in PAA. Given random sampling of a population (e.g., $\phi(X; \mu_1, \Sigma_1)$) and fixed values of β and γ , tolerance regions become narrower as sample size increases (online Appendix 7) because increased sample size results in increased precision in the estimate of the population distribution (i.e., increased precision of $\hat{\phi}(X; \hat{\mu}_1, \hat{\Sigma}_1)$). This implies higher statistical power to detect morphological discontinuities. In consequence, as sample size increases, GCG as well as GD may tend to recognize more morphological discontinuities. This contrasts with the tendency of the interval defined by extreme sample values to increase as sample size increases (Zar 1999; online Appendix 7) and the respective tendency of PAA to recognize less morphological discontinuities as sample size increases (Davis and Nixon 1992). Strictly speaking, these differences among methods occur under random sampling of statistical populations representing hypothesized species. However, random sampling is not common in systematic studies (see above).

PAA was explicitly designed as a method for the identification of phylogenetic species (sensu Nixon and Wheeler 1990, see Davis and Nixon 1992). In contrast, GCG and GD were designed to gauge the strength of the evidence indicating a discontinuity in morphological variation and, as such, they are not necessarily tied to any particular species definition (sensu de Queiroz 1998) because a gap in morphology can provide evidence of a species boundary under different species definitions, potentially yielding different results (Crisp and Weston 1993; Luckow 1995). For instance, when two taxa are considered sufficiently distinct to support the hypothesis of a species limit based on nonoverlapping tolerance regions calculated for some a priori values of γ and β , but morphological distinctiveness can be explained by geographic differentiation among allopatric populations, the phylogenetic (Cracraft 1983, 1989; Nixon and Wheeler 1990) and biological (Mayr 1942) species definitions may lead to different conclusions about species limits (Cracraft 1992). Under the phylogenetic species definition, allopatric taxa that are morphologically distinct (such as taxa pairs *E. schreiteri*–*E. discolor* and *E. schreiteri*–*E. piurensis* in our empirical Example B) would be considered different species, whereas under the biological species definition, it is possible to regard allopatric taxa that are morphologically distinct as geographic variants of the same species. Thus, to yield consistent conclusions about species limits, results from the approach we presented here need to be interpreted in the light of an explicit species definition.

Beyond particular species definitions, recognizing species as separately evolving segments of metapopulation-level lineages (de Queiroz 1998; de Queiroz 2005; de Queiroz 2007) implies that multiple operational criteria can be used to infer lineage separation, but no single one of them is necessary. Thus, a hypothesis of species limits that is not supported by evidence of morphological discontinuities can, nonetheless, correspond to a separation between lineages in nature, such as with morphologically convergent species and/or morphologically cryptic species (Whittall et al. 2004). Such hypotheses could be supported by other operational criteria, including pre- and postzygotic reproductive isolation, or differential exclusivity of gene genealogies. Therefore, support for a hypothesis of species limits may often be more easily interpreted than the absence of evidence supporting such hypothesis. This asymmetry might seem to be a problem if one requires that an inference from a single empirical study should be able to falsify a hypothesis of species boundaries. However, hypothesis-driven research does not need to adopt such a strict requirement. Support (or lack thereof) for a given scientific hypothesis can be judged in the light of multiple statistical inferences derived from several empirical studies, interpreted considering the nuances related to samples size and other aspects of study design (Bernays and Wege 1987; Hilborn and Mangel 1997; Gotelli and Ellison 2004 but see Krebs 2000). This approach recognizes that a single empirical study does not evaluate any hypothesis in isolation but rather bundles containing the hypothesis of interest along with background assumptions (Quine 1951; Duhem 1954).

The above issues regarding the relationship between GCG and species limits decisions involve major controversies about the ontology of species and the nature of the scientific method that are largely outside the scope of the present paper. Our focus here was on presenting an approach, explicitly based on theory about species limits, to weigh the strength of the evidence indicating a gap in morphological variation and assess whether such gap can be interpreted as geographic variation within a single species rather than as a species boundary. We hope that this approach will help to address concerns raised by several systematists (Davis and Heywood 1963; McDade 1995; Luckow 1995; Stevens 2000; Henderson 2006; Wiens 2007) regarding current practices to delimit species based on morphology. In particular, GCG can help strengthen the link between the theory and practice of species delimitation by increasing the transparency and consistency of taxonomic decisions based on morphology. There might be inevitable disagreements regarding decisions such as the relevance of a particular species definition (sensu de Queiroz 1998) or the a priori selection of a frequency cutoff below which overlap of phenotypic values between hypothesized species is considered rare enough as to suggest evolutionary isolation (Wiens and Servedio 2000). However, once those decisions are adopted, and given a set of explicit assumptions about the distribution of

morphological variation and sampling of such a distribution, application of GCG to a particular data set provides an objective assessment of the strength of the evidence indicating whether a morphological discontinuity supports a particular hypothesis about species limits.

SUPPLEMENTARY MATERIAL

Supplementary material, including data files and/or online-only appendices, can be found at <http://www.sysbio.oxfordjournals.org/>.

FUNDING

This work was supported in part by U.S. National Science Foundation (grant number OISE-0738118 to Peter F. Stevens and F.Z.); American Society of Plant Taxonomists, Idea Wild; and the Whitney R. Harris World Ecology Center at the University of Missouri–St. Louis.

ACKNOWLEDGMENTS

We especially thank P.F. Stevens for insightful discussions on species delimitation. We thank D.W. Stevens and C.M. Taylor for discussions regarding species-level taxonomy. P.F. Stevens, E.A. Kellogg, K.M. Olsen, P.C. Hoch, and C.D. Cadena provided helpful comments on drafts of this manuscript. We also thank R. DeBry, R. Glor, J. Sites, and two anonymous reviewers for constructive comments on earlier versions of this manuscript.

REFERENCES

- Ackermann M., Achatz M., Weigend M. 2008. Hybridization and crossability in *Caiophora* (Loasaceae subfam. Loasoideae): are interfertile species and inbred populations results of a recent radiation? *Am. J. Bot.* 95:1109–1121.
- Agapow P.M., Bininda-Emonds O.R.P., Crandall K.A., Gittleman J.L., Mace G.M., Marshall J.C., Purvis A. 2004. The impact of species concept on biodiversity studies. *Q. Rev. Biol.* 79:161–179.
- Baum D.A., Donoghue M.J. 1995. Choosing among alternative “phylogenetic” species concepts. *Syst. Bot.* 20:560–573.
- Bernays E.A., Wege, P. 1987. Significance levels of inferential statistics and their interpretations: a lesson from feeding deterrent experiments. *Ann. Entomol. Soc. Am.* 80: 9–11.
- Bivand R., Anselin L., Assunção R., Berke O., Bernat A., Carvalho M., Chun Y., Christensen B., Dormann C., Dray S., Halbersma R., Krainski E., Lewin-Koh N., Li H., Ma J., Millo G., Mueller W., Ono H., Peres-Neto P.R., Reder M., Tiefelsdorf M., Yu D. 2009. *spdep*: spatial dependence: weighting schemes, statistics and models. R package version 0.4-50. Available from: <http://CRAN.R-project.org/package=spdep>.
- Bond J.E., Stockman A.K. 2008. An integrative method for delimiting cohesion species: finding the population-species interface in a group of Californian trapdoor spiders with extreme genetic divergence and geographic structuring. *Syst. Biol.* 57:628–646.
- Borcard D., Legendre P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol. Modell.* 153:51–68.
- Borcard D., Legendre P., Avois-Jacquet C., Tuomisto H. 2004. Dissecting the spatial structure of ecological data at multiple scales. *Ecology*. 85:1826–1832.
- Coyne J.A., Orr H.A. 2004. *Speciation*. Sunderland (MA): Sinauer.
- Cracraft J. 1983. Species concepts and speciation analysis. *Curr. Ornithol.* 1:159–187.
- Cracraft J. 1989. Speciation and its ontology: the empirical consequences of alternative species concepts for understanding patterns and processes of differentiation. In: Otte D., Endler J., editors. *Speciation and its consequences*. Sunderland (MA): Sinauer. p. 28–59.
- Cracraft J. 1992. The species of the birds-of-paradise (Paradisaeidae): applying the phylogenetic species concept to a complex pattern of diversification. *Cladistics*. 8:1–43.
- Crawley M.J. 2002. *Statistical computing: an introduction to data analysis using S-Plus*. New York: John Wiley and Sons.
- Crisp M.D., Weston P.H. 1993. Geographic and ontogenetic variation in morphology of Australian Waraths (Telopea: Proteaceae). *Syst. Biol.* 42:49–57.
- Cummings M., Neel M., Shaw K. 2008. A genealogical approach to quantifying lineage divergence. *Evolution*. 62:2411–2422.
- Davis J.I., Nixon K.C. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. *Syst. Biol.* 41:421–435.
- Davis P.H., Heywood V. 1963. *Principles of angiosperm taxonomy*. Princeton (NJ): D. van Nostrand Co.
- de Queiroz K. 1998. The general lineage concept of species, species criteria, and the process of speciation: a conceptual unification and terminological recommendations. In: Howard D.J., Berlocher S.H., editors. *Endless forms: species and speciation*. New York: Oxford University Press. p. 57–75.
- de Queiroz K. 2005. A unified concept of species and its consequences for the future of taxonomy. *Proc. Calif. Acad. Sci.* 56:196–215.
- de Queiroz K. 2007. Species concepts and species delimitation. *Syst. Biol.* 56:879–886.
- de Queiroz K., Good D.A. 1997. Phenetic clustering in biology: a critique. *Q. Rev. Biol.* 72:3–30.
- Dettman J.R., Jacobson D.J., Turner E., Pringle A., Taylor J.W. 2003. Reproductive isolation and phylogenetic divergence in *Neurospora*: comparing methods of species recognition in a model eukaryote. *Evolution*. 57:2721–2741.
- Dormann C.F., McPherson J.M., Araújo M.B., Bivand R., Bollinger J., Carl G., Davies R.G., Hirzel A., Jetz W., Kissling W.D., Kühn I., Ohlemüller R., Peres-Neto P.R., Reineking B., Schröder B., Schurr F.M., Wilson R. 2007. Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*. 30:609–628.
- Draper N.R., Smith H. 1998. *Applied regression analysis*. 3rd ed. New York: John Wiley and Sons.
- Dray S., Legendre P., Peres-Neto P.R. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecol. Modell.* 196:483–493.
- Duhem P. 1954. *The aim and structure of physical theory*. Translation by Phillip Wiener. Princeton (NJ): Princeton University Press.
- Dunn G., Everitt B.S. 1982. *An introduction to mathematical taxonomy*. Cambridge (UK): Cambridge University Press.
- Futuyma D.J. 1998. *Evolutionary Biology*. Sunderland (MA): Sinauer.
- Gabriel K.R., Sokal R.R. 1969. A new statistical approach to geographic variation analysis. *Syst. Zool.* 18:259–278.
- Gaston K.J. 2000. Global patterns in biodiversity. *Nature*. 405:220–227.
- Genz A., Bretz F., Miwa T., Mi X., Leisch F., Scheipl F., Hothorn T. 2009. *mvtnorm*: multivariate normal and t distributions. R package version 0.9-7. Available from: <http://CRAN.R-project.org/package=mvtnorm>.
- Griffith D.A., Peres-Neto P.R. 2006. Spatial modelling in ecology: the flexibility of eigenfunction spatial analyses. *Ecology*. 87:2603–2613.
- Gotelli, N.J., Ellison, A. M. 2004. *A primer of ecological statistics*. Sunderland (MA): Sinauer.
- Guralnick R., Wiczkorek J., Beaman R., Hijmans R., Group B.W. 2006. *BioGeomancer*: automated georeferencing to map the world’s biodiversity data. *PLoS Biol* 4:e381.
- Henderson A. 2005. The methods of herbarium taxonomy. *Syst. Bot.* 30: 456–469.
- Henderson A. 2006. Traditional morphometrics in plant systematics and its role in palm systematics. *Bot. J. Linn. Soc.* 151:103–111.
- Hey J., Waples R.S., Arnold M.L., Butlin R.K., Harrison R.G. 2003. Understanding and confronting species uncertainty in biology and conservation. *Trends Ecol. Evol.* 18:597–603.

- Hilborn R., Mangel M. 1997. The ecological detective: confronting models with data. Princeton (NJ): Princeton University Press.
- Howard D.J., Berlocher S.H., editors. 1998. Endless forms: species and speciation. Oxford: Oxford University Press.
- Kausel E. 1953. Revisión del género "Escallonia" en Chile. *Darwiniana*. 10:169–255.
- Knowles L.L., Carstens B.C. 2007. Delimiting species without monophyletic gene trees. *Syst. Biol.* 56:887–895.
- Krebs C.J. 2000. Hypothesis testing in ecology. In: Boitani L., Fuller T.K., editors. Research techniques in animal ecology. New York: Columbia University Press. p. 1–14.
- Krishnamoorthy K., Mathew T. 1999. Comparison of approximation methods for computing tolerance factors for a multivariate normal population. *Technometrics*. 41:234–249.
- Krishnamoorthy K., Mondal S. 2006. Improved tolerance factors for multivariate normal distributions. *Comm. Stat. Sim. Comp.* 35: 461–478.
- Legendre P., Legendre L. 1998. Numerical ecology. 2nd English ed. Amsterdam: Elsevier Science BV.
- Lehtonen S. 2008. An integrative approach to species delimitation in Echinodorus (Alismataceae) and the description of two new species. *New Bull.* 63:525–563.
- Light J.E., Touns M.A., Reed D.L. 2008. What's in a name: the taxonomic status of human head and body lice. *Mol. Phylogenet. Evol.* 47:1203–1216.
- Lohse K. 2009. Can mtDNA barcodes be used to delimit species? A response to Pons et al. (2006). *Syst. Biol.* 58:439–442.
- Luckow M. 1995. Species concepts: assumptions, methods and applications. *Syst. Bot.* 20:589–605.
- Mallet J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. *Proc. R. Soc. Lond. B.* 363:2971–2986.
- Manel S., Schwartz M.K., Luikart G., Taberlet P. 2003. Landscape genetics: combining landscape ecology and populations genetics. *Trends Ecol. Evol.* 18:189–197.
- Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–220.
- Mayr E. 1942. Systematics and the origin of species. New York: Columbia University Press.
- Mayr E. 1963. Animal species and evolution. Cambridge (MA): Harvard University Press.
- Mayr E. 1982. The growth of the biological thought. Cambridge (MA): Harvard University Press.
- Mayr E. 1992. A local flora and the biological species concept. *Am. J. Bot.* 79:222–238.
- McDade L.A. 1995. Species concepts and problems in practice: insights from botanical monographs. *Syst. Bot.* 20:606–622.
- McGowen M.H., Wiltshire R.J.E., Potts B.M., Vaillancourt R.E. 2001. The origin of Eucalyptus vernicosa, a unique shrub eucalyptus. *Biol. J. Linn. Soc.* 74:397–405.
- McLachlan G., Peel D.A. 2000. Finite mixture models. New York: John Wiley and Sons.
- Mishler B.D., Donoghue M.J. 1982. Species concepts: a case for pluralism. *Syst. Zool.* 31:491–503.
- Murdoch D., Chow E.D. (portion to R by Frias Celayeta, J.M.). 2007. ellipse: functions for drawing ellipses and ellipses-like confidence regions. R. package version 0.3–5. Available from: <http://CRAN.R-project.org/package=ellipse>.
- Natrella M.G. 1963. Experimental Statistics. Washington (DC): U.S. Dept. of Commerce, National Bureau of Standards, U.S. Govt.
- Nixon K.C., Wheeler Q.D. 1990. An amplification of the phylogenetic species concept. *Cladistics*. 6:211–223.
- Oksanen J., Kindt R., Legendre P., O'Hara B., Simpson G.L., Solymos P., Stevens M.H.H., Wagner H. 2009. vegan: community ecology package R package version 1.15–4. Available from: <http://CRAN.R-project.org/package=vegan>.
- Olson L.E., Goodman S.M., Yoder A.D. 2004. Illumination of cryptic species boundaries in long-tailed shrew tenrecs (Mammalia: Tenrecidae; Microgale), with new insights into geographic variation and distributional constraints. *Biol. J. Linn. Soc.* 83:1–22.
- O'Meara B.C. 2010. New heuristic methods for joint species delimitation and species tree inference. *Syst. Biol.* 59:59–73.
- Padial J., De La Riva I. 2006. Taxonomic inflation and the stability of species lists: the perils of ostrich's behavior. *Syst. Biol.* 55:859–867.
- Papadopoulou A., Monaghan M.T., Barraclough T.G., Vogler A.P. 2009. Sampling error does not invalidate the Yule-coalescent model for species delimitation. A response to Lohse (2009). *Syst. Biol.* 58: 442–444.
- Patten M.A., Unitt P. 2002. Diagnosability versus mean differences of Sage Sparrow subspecies. *Auk*. 119:26–35.
- Peres-Neto P.R., Legendre P., Dray S., Borcard D. 2006. Variation partitioning of species data matrices: estimation and comparison of fractions. *Ecology*. 87:2614–2625.
- Pons J., Barraclough T.G., Gomez-Zurita J., Cardoso A., Duran D., Hazell S., Kamoun S., Sumlin W., Vogler A.P. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst. Biol.* 55:595–609.
- Quine, W.V. 1951. Two Dogmas of Empiricism. *Philos. Rev.* 60:20–43.
- R Development Core Team. 2009. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: <http://www.R-project.org>.
- Rao C.R. 1964. The use and interpretation of principal component analysis in applied research. *Sankhya A.* 26:329–358.
- Raxworthy C., Ingram C., Rabibisoa N., Pearson R. 2007. Applications of ecological niche modeling for species delimitation: a review and empirical evaluation using day geckos (Phelsuma) from Madagascar. *Syst. Biol.* 56:907–923.
- Ray S., Lindsay B. 2005. The topography of multivariate normal mixtures. *Ann. Stat.* 33:2042–2065.
- Rieseberg L.H. 2000. Crossing relationships among ancient and experimental sunflower hybrid lineages. *Evolution*. 54:859–865.
- Roberts D.W. 2007. labdsv: ordination and multivariate analysis for ecology. R package version 1.3–1. Available from: <http://ecology.msu.montana.edu/labdsv/R>.
- Robertson C.A., Fryer J.G. 1969. Some descriptive properties of normal mixtures. *Skand. Aktuarietidskr.* 52:137–146.
- Sites J.W., Marshall J.C. 2004. Operational criteria for delimiting species. *Annu. Rev. Ecol. Evol. Syst.* 35:199–227.
- Sleumer H.O. 1968. Die Gattung Escallonia (Saxifragaceae). *Verhandelingen der Koninklijke Nederlandse Akademie van Wetenschappen, Afd. Natuurkunde*. 58:1–146.
- Sneath P.H.A., Sokal R.R. 1973. Numerical taxonomy: the principles and practice of numerical classification. San Francisco (CA): W.H. Freeman and Co.
- Sokal R.R. 1979. Testing statistical significance of geographic variation patterns. *Syst. Zool.* 28:227–232.
- Sokal R.R., Rohlf F.J. 1995. Biometry: the principles and practice of statistics in biological research. New York: W. H. Freeman and Co.
- Stamos D.N. 2003. The species problem: biological species, ontology and the metaphysics of biology. Oxford: Lexington Books.
- Stenström A., Jónsdóttir I.S., Augner M. 2002. Genetic and environmental effects on morphology in clonal sedges in the Eurasian Arctic. *Am. J. Bot.* 89:1410–1421.
- Stevens P.F. 2000. Botanical systematics 1950–2000: Change, progress, or both? *Taxon*. 49:635–659.
- Stockman A.K., Bond J.E. 2007. Delimiting cohesion species: extreme population structuring and the role of ecological interchangeability. *Mol. Ecol.* 16:3374–3392.
- Templeton A.R. 2006. Population genetics and microevolutionary theory. Hoboken (NJ): John Wiley and Sons.
- ter Braak C.J.F. 1995. Ordination. In: Jongman R.H.G., ter Braak C.J.F., van Tongeren O.F.R., editors. Data analysis in community and landscape ecology. Cambridge (UK): Cambridge University Press. p. 91–173.
- Wheeler Q.D. 2004. Taxonomic triage and the poverty of phylogeny. *Proc. R. Soc. Lond. B.* 359:571–583.
- Whittall J.B., Hellquist C.B., Schneider E.L., Hodges S.A. 2004. Cryptic species in an endangered pondweed community (Potamogeton, Potamogetonaceae) revealed by AFLP markers. *Am. J. Bot.* 91: 2022–2029.
- Wiens J.J. 2007. Species delimitation: new approaches for discovering diversity. *Syst. Biol.* 56:875–878.
- Wiens J.J., Penkrot T.A. 2002. Delimiting species using DNA and morphological variation and discordant species limits in spiny lizards (Sceloporus). *Syst. Biol.* 51:69–91.

- Wiens J.J., Servedio M.R. 2000. Species delimitation in systematics: inferring diagnostic differences between species. *Proc. R. Soc. Lond. B.* 267:631–636.
- Wilson R.A. 1999. *Species: new interdisciplinary essays*. Cambridge (MA): MIT Press.
- Wright S. 1943. Isolation by distance. *Genetics*. 28:114–138.
- Wright S. 1946. Isolation by distance under diverse system of mating. *Genetics*. 31:39–59.
- Wright S. 1969. *Evolution and the genetics of populations*. Vol. IV: variability within and among natural populations. Chicago (IL): Chicago University Press.
- Zapata F. 2010. *Phylogenetics and diversification of Escallonia (Escalloniaceae)* [Ph.D. dissertation]. St. Louis (MO): University of Missouri.
- Zar J.I. 1999. *Biostatistical analysis*. Upper Saddle River (NJ): Prentice Hall.