Syst. Biol. 67(2):181–194, 2018

DOI:10.1093/sysbio/syx071
Advance Access publication September 1, 2017

Issues and Perspectives in Species Delimitation using Phenotypic Data: Atlantean Evolution in Darwin's Finches

CARLOS DANIEL CADENA[1,*], FELIPE ZAPATA[2], AND IVÁN JIMÉNEZ[3]

[1]*Departamento de Ciencias Biológicas, Universidad de los Andes, Bogotá, Colombia;* [2]*Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, USA; and* [3]*Center for Conservation and Sustainable Development, Missouri Botanical Garden, St. Louis, MO, USA*
**Correspondence to be sent to: Departamento de Ciencias Biológicas, Universidad de los Andes, Carrera 1 no. 18A-10, Bogotá, Colombia; E-mail: ccadena@uniandes.edu.co.*

Received 5 April 2017; reviews returned 22 August 2017; accepted 30 August 2017
Associate Editor: John E. McCormack

Abstract.—Progress in the development and use of methods for species delimitation employing phenotypic data lags behind conceptual and practical advances in molecular genetic approaches. The basic evolutionary model underlying the use of phenotypic data to delimit species assumes random mating and quantitative polygenic traits, so that phenotypic distributions within a species should be approximately normal for individuals of the same sex and age. Accordingly, two or more distinct normal distributions of phenotypic traits suggest the existence of multiple species. In light of this model, we show that analytical approaches employed in taxonomic studies using phenotypic data are often compromised by three issues: 1) reliance on graphical analyses that convey little information on phenotype frequencies; 2) exclusion of characters potentially important for species delimitation following reduction of data dimensionality; and 3) use of measures of central tendency to evaluate phenotypic distinctiveness. We outline approaches to overcome these issues based on statistical developments related to normal mixture models (NMMs) and illustrate them empirically with a reanalysis of morphological data recently used to claim that there are no morphologically distinct species of Darwin's ground-finches (*Geospiza*). We found negligible support for this claim relative to taxonomic hypotheses recognizing multiple species. Although species limits among ground-finches merit further assessments using additional sources of information, our results bear implications for other areas of inquiry including speciation research: because ground-finches have likely speciated and are not trapped in a process of "Sisyphean" evolution as recently argued, they remain useful models to understand the evolutionary forces involved in speciation. Our work underscores the importance of statistical approaches grounded on appropriate evolutionary models for species delimitation. We discuss how NMMs offer new perspectives in the kind of inferences available to systematists, with significant repercussions on ideas about the phenotypic structure of biodiversity. [Morphology; normal mixture model; phenotype; principal components analysis; species limits; variable selection.]

Systematic biology seeks to discover and describe species, and to establish phylogenetic relationships among them and among clades at higher levels. Given these two main goals of the field, reviews published over a decade ago noted that the literature on theory and methods of phylogenetic inference and on theory of species concepts was extensive, whereas methods for delimiting species had received much less attention (Sites Jr. and Marshall 2003; Sites Jr. and Marshall 2004). Over the past few years, this imbalance has been partly overcome with considerable development, application, and integration of methods for species delimitation (Padial et al. 2010; Camargo and Sites Jr. 2013). Largely driven by increased availability of multilocus data sets brought about by advances in DNA sequencing technology, however, much recent progress has focused on probabilistic methods for analyses of molecular data (reviewed by Fujita et al. 2012; Carstens et al. 2013), whereas relatively little effort has been devoted to approaches using phenotypic data to delimit species (Wiens and Servedio 2000; Ezard et al. 2010; Guillot et al. 2012; Zapata and Jiménez 2012; Edwards and Knowles 2014; Solís-Lemus et al. 2014). Yet, because most fossil and living species have been discovered and named based on phenotypic distinctiveness (Luckow 1995; Mallet 2013; Miller 2016), and because genomic-based species delimitation approaches are no substitute for judicious assessments of other sources of information (Sukumaran and Knowles 2017), the theory and practice of delimiting species using phenotypic data remain central to modern systematics.

Although species descriptions employing phenotypic data are often nonquantitative and although systematists may often not be explicit about the rationale they follow to delimit species (Luckow 1995; McDade 1995; Sangster 2014; Allmon 2016), the use of objective criteria for species diagnosis based on phenotypic characters has a long tradition in taxonomy, rooted in evolutionary theory (Wiens and Servedio 2000; Zapata and Jiménez 2012; Futuyma 2013). The basic evolutionary model for the distribution of a continuous quantitative character within a species (Fisher 1918) assumes polygenic inheritance and random mating; under these assumptions, gene frequencies would be close to Hardy–Weinberg equilibrium, two or more loci would be near linkage equilibrium, and phenotypic variation among individuals of a single species would tend to be normally distributed (Templeton 2006). On the other hand, phenotypic variation may be best described by two or more distinct normal distributions (i.e., distributions differing in means, variances, or covariances); in this latter case, one may conclude that there is more than one species in a sample of individuals (Coyne and Orr 2004; Mallet 2008). This conclusion is granted under the assumption that parameter differences between normal distributions do not reflect genetic

polymorphisms (e.g., sex-related variation), ontogenetic variation, or phenotypic plasticity. Therefore, differences between normal distributions caused by few loci of large effect (e.g., Smith 1993) or largely driven by environmental factors (e.g., Moczek and Emlen 1999) do not constitute evidence of more than one species. While this Fisherian model readily applies to (polygenic) continuous traits and not to qualitative traits, we note that phenotypic variation commonly regarded as qualitative is actually continuous (Stevens 1991), including variation in the shape (e.g., Leaché et al. 2009) and color (e.g., McKay et al. 2014) of morphological structures.

We stress that distinct phenotypic distributions may represent evidence of species boundaries given a variety of species definitions (sensu de Queiroz 1998). For instance, distinct phenotypic distributions may serve as a species criterion (i.e., as a standard to judge whether a group of organisms qualifies as a species) under species definitions that emphasize phenotypic and genotypic clusters (e.g., Ezard et al. 2010), interbreeding (e.g., Mayr 1992), phenotypic cohesion (e.g., Bond and Stockman 2008), or diagnosability (e.g., Crisp and Weston 1993). The same is true under species definitions that are alternative descriptions of the general lineage concept of de Queiroz (1998), including the evolutionary species definitions of Simpson (1951) and Wiley (1978). Therefore, the Fisherian model described above serves as a conceptual basis to infer species limits given diverse views regarding the nature of species.

Despite the long tradition of the basic model for species delimitation based on quantitative phenotypic characters, statistical tools for its formal application to empirical data were fairly limited until recently. Procedures allowing one to fit combinations of normal distributions to phenotypic variation among specimens, without *a priori* knowledge of species limits, were initially developed in the late XIX century (Pearson 1894). However, practical application only became possible following computational advances in the 1970s (i.e., the expectation–maximization algorithm; McLachlan and Peel 2000) and software development from the late XX century into the present (e.g., Fraley and Raftery 2002; Fraley et al. 2012). Because these statistical approaches entered the literature on species delimitation only a few years ago (Ezard et al. 2010; see also Hausdorf and Hennig 2010 for an application to molecular data), it is not surprising that even recent studies do not employ them when analyzing phenotypic data to delimit species. Instead, systematists frequently infer species limits examining phenotypic variation based on visual inspection of scatter plots defined by a few axes that account for most phenotypic variance, often derived from principal components analysis (PCA). In addition, systematists often delimit species based on differences between groups of specimens in the central tendency of phenotypes. This is true of work on living plants and animals (reviewed by Rieseberg et al. 2006), as well as in studies of extinct taxa in the fossil record (reviewed by Allmon 2016).

Here, we show that the way in which analytical approaches are commonly employed to examine phenotypic data in taxonomic studies is often inadequate in light of the evolutionary model underlying species delimitation described above. It follows that if species delimited by inadequate statistical approaches are used as units for subsequent analyses, then any mistakes may carry on and influence views in other areas of inquiry, such as speciation research. Focusing on Darwin's finches from the Galapagos Islands, an iconic group for the study of natural selection, speciation, and adaptive radiation (Lack 1947; Bowman 1961; Grant 1999; Grant and Grant 2008; Grant and Grant 2014), we provide an example of how employing statistical approaches explicitly related to the basic evolutionary model underlying the use of phenotypic data in species delimitation may enhance assessments of species limits and thus our understanding of evolutionary processes.

Sisyphean Evolution in Darwin's Finches?

Among Darwin's finches, the many studies of ground-finches in the genus *Geospiza* have been especially productive in terms of insights into species formation and the role of geographic isolation, natural selection, and hybridization in microevolutionary processes that may scale up to macroevolutionary patterns (reviewed by Grant 1999; Grant and Grant 2008; Grant and Grant 2014). There has been considerable disagreement in the literature about the number of species in the group (reviewed by McKay and Zink 2015), but most modern taxonomic treatments have recognized six species of ground-finches (Lack 1947; Rising et al. 2011). However, based on genomic evidence (Lamichhaney et al. 2015) and some vocal and behavioral data, three subspecies were recently elevated to species rank, bringing the total number of recognized species to nine (Remsen Jr. et al. 2017).

In a provocative recent paper, McKay and Zink (2015) offered an intriguing alternative perspective on the taxonomy and evolution of ground-finches (see also Zink 2002). These authors boldly argued that morphological evidence for the existence of multiple species of *Geospiza* is lacking and they presented the iconoclastic argument that different phenotypes should be considered transient ecomorphs within a single species. Furthermore, according to these authors, ground-finches are an appropriate model to study forces involved in geographic variation and local adaptation, but not to demonstrate the workings of speciation because in their view speciation in the group has not occurred. Instead, incipient speciation has been repeatedly stalled or reversed owing to shifting conditions affecting the strength and direction of natural selection and to ongoing gene flow (McKay and Zink 2015). Because speciation is initiated but never completed, McKay and Zink (2015) described evolution in ground-finches as "Sisyphean"

in reference to Sisyphus, a character in Greek mythology condemned by the gods to ceaselessly push a boulder up a mountain, only to watch it roll back down, repeating this task eternally. Because of its originality in challenging "entrenched orthodoxy regarding speciation in Darwin's Finches," the study by McKay and Zink (2015) was duly recognized with an award by a major ornithological organization (Cooper Ornithological Society 2016).

A central premise of the arguments by McKay and Zink (2015) was their assertion that phenotypic discontinuities do not exist among recognized species of ground-finches (contra Lack 1947; Grant et al. 1985). This claim seems particularly important in light of recent work indicating that some pairs of species (i.e., *Geospiza fortis* vs. *G. acutirostris*, *G. scandens* vs. *G propinqua*; *G. magnirostris* vs. *G. conirostris*) appear not to be clearly distinguishable from each other with genomic data (Fig. 1b in Lamichhaney et al. 2015). Although McKay and Zink (2015) rightly noted that given currently available phenotypic data "the real test of species limits is determining the extent to which specimens form multiple morphological clusters when *a priori* specimen identifications are ignored," they did not formally conduct such a test. Instead, their approach illustrates three problematic issues in analyses of phenotypic data for species delimitation. In the next section, we describe these issues and outline possible solutions afforded by statistical tools directly related to the basic evolutionary model underlying the use of phenotypic data in species delimitation. We then implement these solutions in a reanalysis of the morphological data on *Geospiza* ground-finches to revisit the question of whether morphological evidence supports the hypothesis that there are several species in the group.

Three Frequent Issues in Analyses of Phenotypic Data for Species Delimitation

Graphical analyses may convey little information on phenotype frequencies crucial to assess evidence for multiple species

Many species delimitation studies rely on visual inspection of bivariate (rarely trivariate) scatter plots of phenotypic space to detect discontinuities and thus define phenotypic groups (e.g., Fig. 1 in McKay and Zink 2015). These scatter plots may offer only limited insight into the structure of character variation because visual cluttering and record overplotting hinder perception of phenotype frequencies crucial to identify groups (McLachlan 2004). We illustrate this problem with a hypothetical example in which specimens from a given locality seem to reveal no phenotypic discontinuities, with intermediate phenotypes across the range of

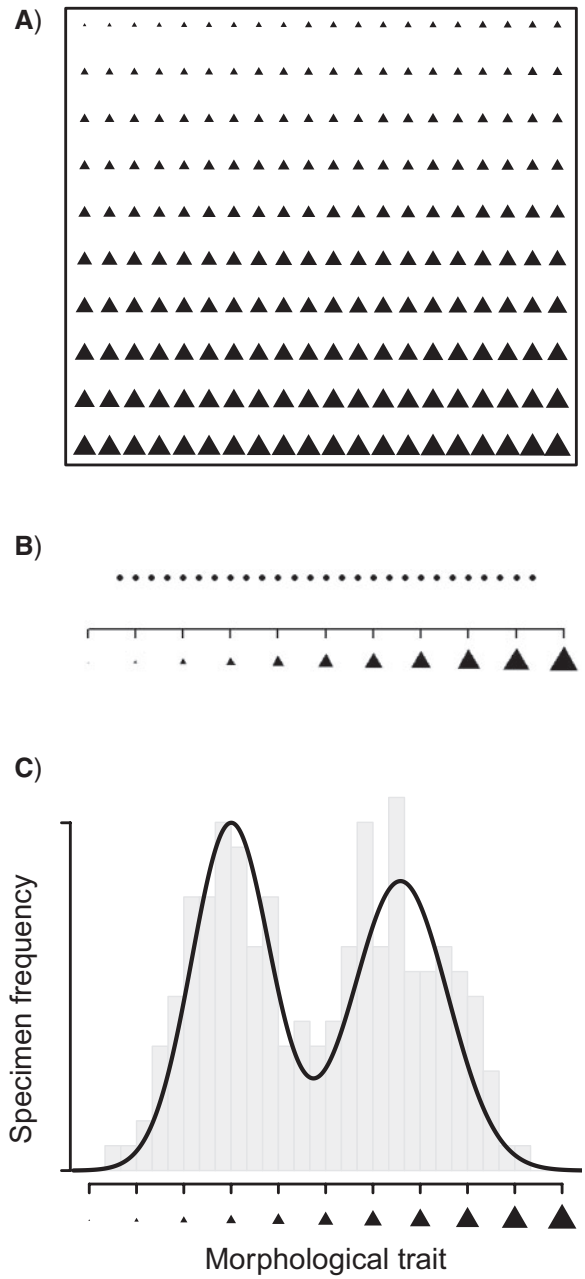

Figure 1. Visual inspection of phenotypic data may yield limited insight regarding species limits. A) Sample of 200 museum specimens (triangles) arranged according to a morphological phenotype (triangle size), from small in the upper left to large in the lower right. The specimens appear to form a smooth gradient with no morphological gaps. B) Plot of specimen measurements along a single continuous axis representing the size of the morphological trait in (A). At the resolution of the measurements, extreme values in the sample seem to be gradually connected by intermediate phenotypes throughout. Thus, there seem to be no obvious morphological gap, suggesting the specimens correspond to a single variable species. C) Two distinct normal distributions are revealed by examining the frequency (gray bars) of specimen phenotypes in the sample, suggesting that specimens may correspond to two species. In fact, the sample was drawn from a mixture of two normal distributions (continuous black lines).

variation (Fig. 1a); accordingly, a univariate scatter plot fails to reveal evidence of distinct phenotypic groups (Fig. 1b). The problem with scatter plots concealing crucial information (also common in 2D and 3D scatter plots) is revealed by a histogram of phenotype frequencies employing the same data, which reveals two distinct normal distributions (Fig. 1c). Following the model for species delimitation based on continuous phenotypic characters described above, this histogram suggests the existence of two species.

Graphical analysis of phenotype frequencies (e.g., Fig. 1c) may be effective to detect groups when few characters are relevant (but see McLachlan and Peel 2000, page 9). However, it may be difficult to detect distinct normal distributions in phenotypic spaces defined by more than two dimensions, where complex covariance structures are likely (McLachlan 2004). Moreover, graphical analysis may suggest the existence of several distinct distributions when, in fact, all variation derives from a single normal distribution (Day 1969; McLachlan and Peel 2000, page 17; Supplementary Material Appendix 1 at http://dx.doi.org/10.5061/dryad.9gh90). In general, then, detection of phenotypic groups exclusively based on graphical analysis is potentially highly subjective and difficult to replicate, or, as stated by Pearson (1894) over a century ago: "To throw the solution on the judgment of the eye in examining the graphical results is, I feel certain, quite futile." Therefore, graphical analysis of phenotype frequencies is a useful but limited tool for species delimitation.

Recent statistical developments allow systematists to go beyond graphical analysis by using normal mixture models (NMMs, McLachlan and Peel 2000) as a formal approach to test for the existence of distinct species based on multivariate phenotypic data (Ezard et al. 2010; Guillot et al. 2012; Edwards and Knowles 2014; Kleindorfer et al. 2014). These models conceptualize phenotypic variation as a combination (i.e., a mixture) of distinct normal distributions; a mixture may include one or more distinct normal distributions, representing the hypothesis of one or more species, respectively. The parameters of a NMM specifying a particular hypothesis include the means and variance–covariance matrices describing the Gaussian phenotypic distribution of each species. These parameters can be estimated using maximum likelihood from data on phenotypic measurements, without *a priori* knowledge of species limits, employing the expectation–maximization algorithm (McLachlan and Krishnan 2008). Comparison of empirical support among models representing different hypotheses is often based on the Bayesian Information Criterion (BIC; Schwarz 1978), which evaluates the likelihood of each model while adjusting for model complexity (Fraley and Raftery 2002).

Reduction of dimensionality via PCA may exclude important characters for species delimitation

Species delimitation studies often begin analyses by reducing the dimensionality of phenotypic space, typically via principal component analysis (PCA) or related procedures (McLachlan 2004; Ezard et al. 2010), and then focusing attention on few principal components that account for most of the variation in the data. For example, McKay and Zink (2015) focused on three principal components explaining 99% of the variation in six morphological characters of *Geospiza* ground-finches (see their Fig. 1). This use of PCA and related procedures in taxonomy was suggested decades ago (Sneath and Sokal 1973) and is still prescribed nowadays (e.g., Ezard et al. 2010). However, there is no reason to believe that principal components accounting for most of the variation in a data set are most useful for group discrimination (Chang 1983).

To illustrate the problem of reducing dimensionality to the principal components accounting for most of the variation, we use a hypothetical example based on two phenotypically distinct species, each represented by a bivariate normal distribution (Fig. 2a). The first principal component of the mixture of these two distributions explains >99% of the variation and, yet, it is useless to distinguish the two species (Fig. 2b). In contrast, the second principal component accounts for <1% of the variation and readily discriminates species (Fig. 2c). This example is bivariate for simplicity, but the statistical principle applies to mixtures of two normal distributions in any number of dimensions (Chang 1983). We stress that the problem at hand is not rotation of the data using PCA or related procedures, because such rotation may serve a number of useful purposes; rather, the problem is using the amount of phenotypic variance explained by each principal component as a proxy for its usefulness to distinguish phenotypic groups (Chang 1983).

Although alternatives to PCA and related approaches for dimensionality reduction should be regularly considered in analysis aiming to detect groups in multivariate space (McLachlan and Peel 2000; McLachlan 2004), they are rarely implemented in species delimitation studies. For example, one may reduce dimensionality based on *a priori* considerations about which set of characters may be best to diagnose particular species and then use those characters in analyses based on NMMs. In particular, when *a priori* information about specific traits separating species is available (e.g., original species descriptions), one should favor analyzing variation in such traits; far from being circular (McKay and Zink 2015), it is only natural that one should critically examine evidence for species limits precisely in the dimensions in which such limits are hypothesized to exist (see also Remsen Jr. 2010; Patten and Remsen Jr. 2017). Alternatively, one may use methods that aim to find the set of variables (phenotypic traits) that best discriminates groups in a NMM, with no *a priori* information about groups (Raftery and Dean 2006; Maugis et al. 2009a; Maugis et al. 2009b).

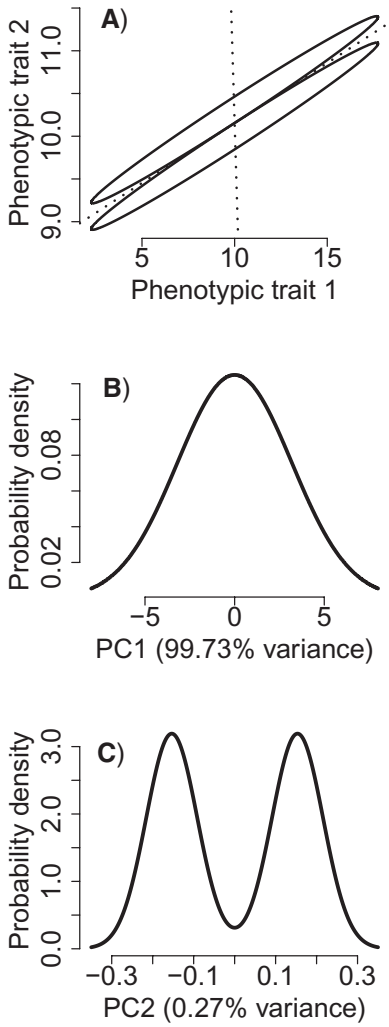

FIGURE 2. The usefulness of principal components to discriminate species is not necessarily proportional to the total phenotypic variance they explain. A) Hypothetical example of two distinct species in the space defined by two phenotypic traits. Each species is described by a bivariate normal distribution, shown as ellipses covering 95% of the individuals of each species. Dotted lines represent the two principal component axes of the normal mixture of the two species. Note that the two principal components are orthogonal, forming a right angle that may not be apparent due to the magnification of the ordinate relative to the abscissa. B) Probability density of individuals of the two species along the first principal component (PC1). This axis is useless to discriminate the phenotypic distributions of the two species, despite the fact that it explains 99.73% of the variance. C) Probability density of individuals of the two species along the second principal component (PC2). The phenotypes of the two species can be readily distinguished along this axis, even though it explains only 0.27% of the variance. Because systematists often discard phenotypic axes accounting for small fractions of the total variance, they may miss crucial phenotypic evidence for species limits.

Differences in central tendency are not evidence of distinct phenotypic distributions

Distinct normal distributions in quantitative characters constitute evidence for the existence of distinct species, but differences in central tendency between groups of individuals defined *a priori* do not. This issue has been pointed out previously (e.g., Mayr et al. 1953; Luckow 1995; Patten and Unitt 2002), but seems to be ignored when statistical procedures to investigate differences in central tendency (e.g., *t*-tests, analysis of variance, Cohen's *d*) are advanced as potentially valid tools to evaluate species boundaries (e.g., Simpson 1951; Henderson 2006; Tobias et al. 2010). McKay and Zink (2015, page 695 and their Fig. 2) have done as much by suggesting that statistical differences in average phenotypes between allopatric island populations of ground-finches could be equated to distinct morphological groups which, in turn, would have to be recognized as species.

Because this issue appears to commonly afflict assessments of species limits between allopatric forms (e.g., Tobias et al. 2010), we illustrate it with a hypothetical example likely encountered in many studies of species delimitation: spatially separated populations of a species exhibiting geographic variation in phenotype. Following the analysis of McKay and Zink (2015; their Fig. 2), we imagine researchers sampling individuals of a species in two islands and then documenting significant differences in the central tendency of phenotypes between the two island populations (Fig. 3a). This difference in central tendency, however, does not constitute evidence that phenotypic variation is best described by two (or more) distinct normal distributions. In fact, the relevant NMM analysis indicates that a single normal distribution best explains phenotypic variation across individuals from the two island populations (Fig. 3b), consistent with the Fisherian model for a single species. Therefore, there is no evidence for more than one species in the sample of specimens regardless of differences in average phenotypes. This illustration focuses on *a priori* groups of specimens defined by geography (i.e., spatially separated populations), but the issue may affect comparisons involving groups of specimens defined by time (i.e., allochronic populations; Simpson 1951) or by any other criterion. In general, phenotypic variation conforming to a single normal distribution can be arbitrarily split into parts that differ in central tendency (Supplementary Material Appendix 2 available on Dryad). Therefore, on their own, differences in central tendency between groups of specimens cannot be regarded as evidence for distinct normal distributions.

The solution to the above problem is simple: do not treat phenotypic differences in central tendency as evidence for the existence of distinct phenotypic groups and, therefore, distinct species. No matter how statistically significant, even very large effect sizes are not germane in light of the basic model for species delimitation based on quantitative phenotypic characters. In light of this model, the focus of analysis should be on determining the number of normal distributions needed to describe phenotypic variation among specimens, as well as on estimating the parameters of those distributions (e.g., means and

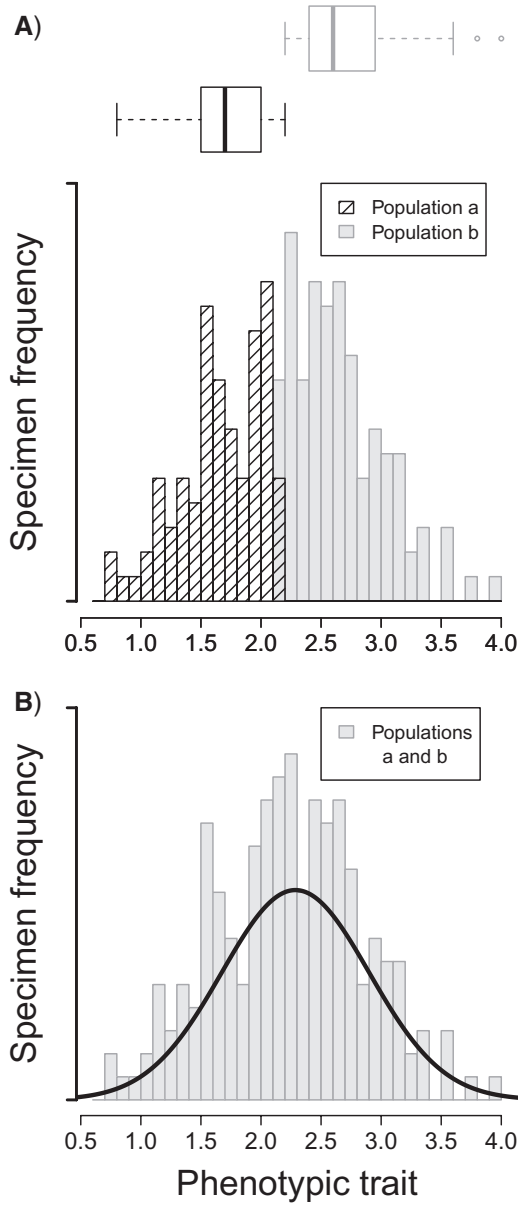

FIGURE 3. Differences in central tendency are not evidence of distinct phenotypic distributions. A) Specimens from two island populations a and b (striped and gray histogram bars, respectively) differ markedly in the central tendency of a phenotypic trait (abscissa), as described by boxplots on top. The boxplots show the median (solid thick line), the interquartile range (box), whiskers extending to the most extreme values within $\times 1.5$ interquartile ranges from the box, and outliers. The means of the two groups are 1.717 and 2.707, and their standard deviations are 0.3528 and 0.3867, respectively. The difference between means is statistically significant (0.99, 95% CI: 0.89−1.09, *t*-test *P*-value $< 2 \times 10^{-16}$). Cohen's $d = 2.66$, which is generally regarded as a large effect size. B) Despite the difference in central tendency, a single normal distribution (continuous black line) describes phenotypic variation across specimens from both islands (gray bars) better than two normal distributions. In particular, empirical support for a normal mixture model assuming that populations a and b constitute two distinct normal distributions is substantially lower than that for a model specifying a single normal distribution: a difference of 10 in BIC. Thus, in light of the basic model for species delimitation based on quantitative phenotypic characters, there are no grounds to suggest that the specimens represent two distinct species despite marked differences in central tendency.

variance–covariance matrices). Indeed, strong evidence may exist for more than one distinct normal distribution in the absence of differences in central tendency (Hennig 2010), suggesting biologically meaningful differences between species in the variance of phenotypic traits (Supplementary Material Appendix 2 available on Dryad). As explained above, NMMs are a useful tool to test for distinct normal distributions.

Are There Phenotypically Distinct Groups of Ground-Finches?

We examined phenotypic variation among *Geospiza* ground-finches by analyzing data from six morphological measurements of museum specimens (wing length, tail length, tarsus length, bill length, bill width, and bill depth) taken on adult males by H. S. Swarth for his monographic revision of the birds of the Galapagos (Swarth 1931). These were the same data employed by McKay and Zink (2015), which we use here with permission from the California Academy of Sciences; our sample sizes differ from those of the earlier study (486 vs. 501 male individuals) because we excluded a few individuals that were duplicated in the original data set. The data we employed and the R code used to conduct the analyses described below are available as Supplementary Material Appendices 3 and 4 available on Dryad, respectively. We emphasize that the purpose of our analysis was not to establish species limits among ground-finches nor to provide species diagnoses. Rather, we sought to revisit the question of whether the morphological data analyzed by McKay and Zink (2015) are consistent with the existence of multiple species and thereby illustrate an empirical application of approaches involving variable selection and NMMs for species delimitation. As we argue below, the morphological data we analyzed are one of multiple sources of evidence that researchers may use for subsequent assessments of species limits under a variety of species definitions (sensu de Queiroz 1998).

We asked how many distinct groups of ground-finches exist in the Galapagos using morphological data from specimens collected across the archipelago (total 18 islands). To define the morphological space for this analysis, we followed McKay and Zink (2015) and used PCA on the covariance matrix of log-transformed data. Rather than examining evidence for species limits using only the first three principal components accounting for >99% of the variation (McKay and Zink 2015), we used the R package clustvarsel (Scrucca and Raftery 2004) to reduce the dimensionality of the data by selecting the set of principal components most useful for group discrimination in NMMs, without *a priori* information about groups (Raftery and Dean 2006; Maugis et al. 2009a, 2009b). We used the R package mclust 5.0 (Scrucca et al. 2016) to fit a wide range of NMMs. At one extreme, NMMs assuming one morphological group represented the Sisyphean evolution hypothesis that there is a single species of ground-finch

(McKay and Zink 2015). Toward the opposite end, NMMs assuming up to 30 distinct morphological groups represented hypotheses alluded to by McKay and Zink (2015) when they suggested ground-finches may comprise "dozens of cluster species," or "1 or 6 or 30 species" (page 695). We also fitted NMMs specifying the six (Lack 1947) or nine (Lamichhaney et al. 2015; Remsen Jr. et al. 2017) species recognized by alternative taxonomic treatments of *Geospiza*, using the original specimen identifications in Swarth's data updated to reflect changes in nomenclature. We used the Bayesian Information Criterion (BIC; Schwarz 1978) to measure empirical support for different NMMs (Fraley and Raftery 2002) and thereby explicitly evaluated the hypothesis that there is only one species of ground-finch (McKay and Zink 2015) relative to hypotheses that there are several species in the group (Lack 1947; Lamichhaney et al. 2015; Remsen Jr. et al. 2017).

We found the first four principal components to be most useful for group discrimination; NMMs ignoring the fourth principal component, although it explained only 0.6% of the morphological variance, had substantially less empirical support (ΔBIC $\geqslant$ 55) than those including it. Therefore, in contrast to McKay and Zink (2015), we did not discard the fourth principal component for analysis. The models specifying seven and eight distinct morphological groups of ground-finches received the strongest support (ΔBIC $\leqslant$ 1.26). Support for all other models was considerably lower (ΔBIC in all cases >20; Fig. 4). In turn, the model with the lowest support represented the Sisyphean evolution hypothesis proposing no distinct morphological groups of ground-finches (i.e., that there is a single group; McKay and Zink 2015), which had a 500 BIC difference to the second-worse model and >821 BIC difference to the two best models. Support for hypotheses consistent with the alternative scenarios that there might be "dozens of cluster species" or 30 species (McKay and Zink 2015, page 695) was also poor. Relative to the best models, models specifying groupings consistent with taxonomy recognizing six or nine species were weakly supported (Fig. 4), considering differences in BIC scores greater than six are typically regarded as strong or very strong evidence against models with lower support (Kass and Raftery 1995). In sum, the data provided poor empirical support for the hypothesis that ground-finches consist of only one species (McKay and Zink 2015) and strongly supported hypotheses of several (but not dozens or 30; McKay and Zink 2015) morphologically distinct groups (Fig. 5, Supplementary Material Fig. 1 available on Dryad). However, those groups did not exactly align with existing taxonomic treatments of *Geospiza* (Fig. 6, Supplementary Material Fig. 2 available on Dryad).

Despite their comparatively low empirical support, models specifying six or nine morphological groups according to taxonomic treatments of *Geospiza* (Lack 1947; Remsen Jr. et al. 2017) were partially consistent with the best models (Fig. 6 and Supplementary Material Fig. 2 available on Dryad). For example, in the best models, all specimens of two of the nine currently

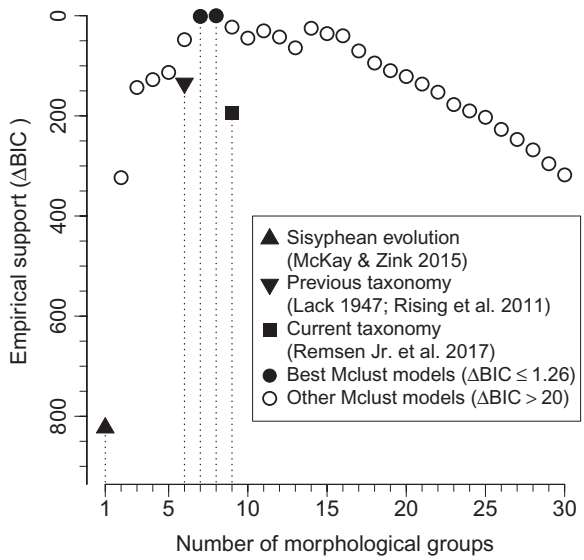

FIGURE 4. Analysis of morphological data strongly supported hypotheses that there are multiple distinct groups of *Geospiza* ground-finches. The plot shows the empirical support (ordinate) for normal mixture models assuming 1–30 distinct morphological groups (abscissa), and for the two models specifying groupings of specimens reflecting taxonomic treatments recognizing six species (Lack 1947; Rising et al. 2011) or nine species (Remsen Jr. et al. 2017). Empirical support was measured as difference in BIC relative to the best model (ΔBIC). The two models with highest empirical support assumed seven and eight distinct morphological groups. Empirical support for the model corresponding to the Sisyphean evolution hypothesis positing there is a single species of ground-finch (i.e., a single morphological group; McKay and Zink 2015) was negligible (ΔBIC >820).

recognized species (*G. scandens*, *G. septentrionalis*) were assigned to two respective morphological groups which included few or no specimens of other species (Fig. 6; Supplementary Material Fig. 2 available on Dryad). Discrepancies between our analysis and current taxonomy were most evident in cases such as those of 1) *G. propinqua*, *G. conirostris*, and *G. fortis*, which were assigned to three, three (or two) and four morphological groups, respectively, or 2) *G. fuliginosa* and *G. acutirostris*, in which all specimens were assigned to the same group to the exclusion of nearly all specimens of other species. In addition, some morphological groups included specimens of multiple species (e.g., morphological Group 1 contained specimens identified as *G. propinqua*, *G. fortis*, *G. magnirostris*, and *G. conirostris*). Part of the lack of agreement between the morphological groups we detected and groups recognized by taxonomy may be accounted for by considering that species may be told apart by phenotypic characters different from those we considered. For example, *G. fuliginosa* and *G. acutirostris* are indistinguishable in our analysis, but are distinct given subtle differences in bill profile and marked differences in songs (Grant and Grant 2008); this illustrates the importance of careful selection of traits to be included in studies of species delimitation using phenotypic data including analyses involving NMMs. Likewise, some of the discrepancies with current taxonomy (Remsen Jr. et al. 2017) involved cases in which

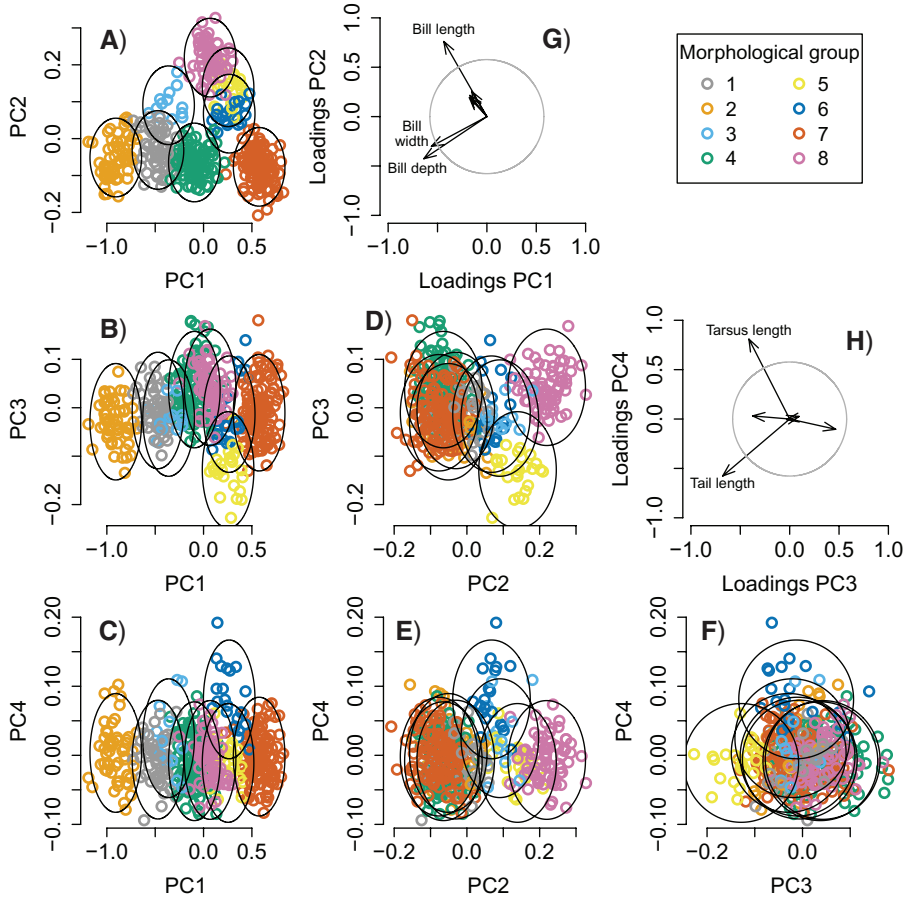

FIGURE 5. Eight morphological groups of *Geospiza* ground-finches identified by one of the best normal mixture models (NMMs). Panels (A–F) show the groups in the space defined by the four principal components most useful for group discrimination (PC1–PC4). Colored symbols represent specimens assigned to different morphological groups and ellipses show 95% high-density regions for normal distributions representing each morphological group. Arrows in G and H display the contribution of measured morphological traits to each principal component, gauged by the loadings of each trait on each principal component (i.e., elements of normalized eigenvectors). Circles show the length of arrows expected if all six traits contributed equally to bidimensional principal component spaces; arrows exceeding this expectation contribute most significantly and are labeled. PC1 and PC2 reflect general aspects of beak size and shape (G), with Group 2 having long, deep, and wide beaks, Group 7 having short, shallow, and narrow beaks, and morphological Group 8 having long, shallow, and narrow beaks. PC3 and PC4 reflect aspects of tail and tarsus length (H), with Group 5 having a relatively long tail, and Group 6 having a relatively long tarsus. PC4 is particularly useful to distinguish Group 6 despite explaining only 0.6% of the total variance. The morphological distribution of groups in the other well supported NMM is fairly similar to the one shown here, the main difference being that Group 3 is merged into Groups 1 and 8 (Supplementary Material Fig. 1 available on Dryad).

species delimitation was not based on morphology, but rather resulted from recent genomic analyses revealing that phenotypically similar populations are distantly related (Lamichhaney et al. 2015). This likely explains why our analysis did not fully discriminate some species pairs in the morphological space we examined (*G. conirostris* vs. *G. propinqua* and *G. difficils* vs. *G. septentrionalis*), although they may be more distinct in other phenotypic spaces including bill profile and song as well as behavior (Grant et al. 2000; Grant and Grant 2002). Also, we assumed that specimen identifications in the data set we analyzed were faultless; thus, part of the apparent mismatch between morphological groups detected in our analyses and taxonomy may reflect identification errors. Evaluating this possibility would require detailed examinations of individual specimens beyond the scope of our work.

Geographic context is an important consideration in assessments of species limits using phenotypic traits. Under a wide range of species definitions (sensu de Queiroz 1998), distinct phenotypic groups among sympatric individuals are readily accepted as evidence for the existence of distinct species (Mayr 1992; Mallet 2008). However, distinct phenotypic groups corresponding to nonsympatric populations may be less readily accepted as evidence of distinct species under species definitions that emphasize "intrinsic" over "extrinsic" barriers to gene exchange (Harrison 1998) because such groups may reflect ephemeral within-species differentiation due to geographic isolation or local adaptation (Zapata and Jiménez 2012). The morphological groups of ground-finches we detected (Fig. 5, Supplementary Material Fig. 1 available on Dryad) cannot be interpreted to reflect within-species,

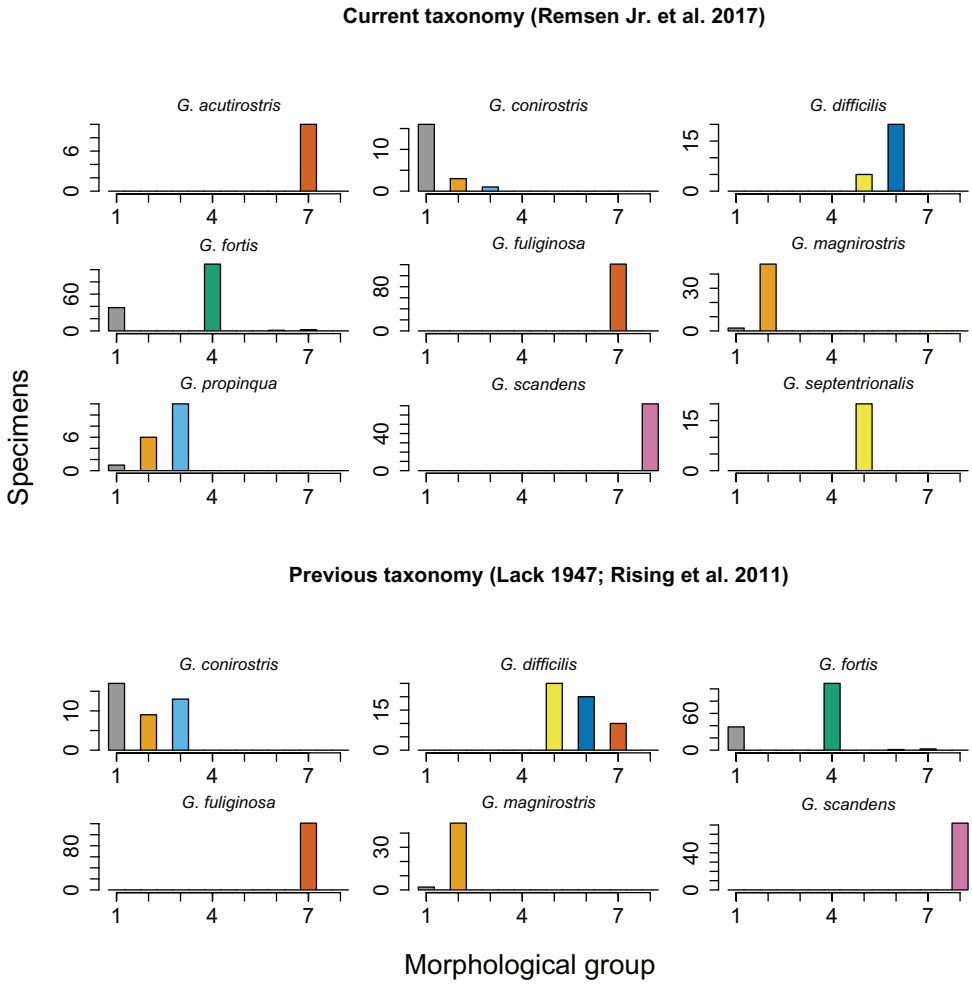

FIGURE 6. Eight morphological groups of *Geospiza* ground-finches in the Galapagos Archipelago identified by one of the best normal mixture models partially correspond to the nine species recognized by current taxonomy (Remsen Jr. et al. 2017) and to the six species recognized by previous taxonomy (Lack 1947; Rising et al. 2011). Each histogram shows, for each recognized species, the number of specimens assigned to each of the eight morphological groups. Groups are colored according to the scheme in Fig. 5.

TABLE 1. Number of islands in the Galapagos Archipelago where each of the eight morphological groups of *Geospiza* ground-finches identified by one of the best normal mixture models were found to occur (diagonal) and co-occur with other groups (off diagonal)

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6	Group 7	Group 8
Group 1	8	4	2	4	2	1	6	4
Group 2	—	9	2	4	4	3	7	5
Group 3	—	—	3	0	1	0	1	0
Group 4	—	—	—	11	2	3	10	7
Group 5	—	—	—	—	4	2	2	2
Group 6	—	—	—	—	—	3	3	3
Group 7	—	—	—	—	—	—	14	3
Group 8	—	—	—	—	—	—	—	9

Note: All groups co-occurred with each other in at least one island, except for cases involving Group 3, which did not co-occur with three other groups. Note, however, that Group 3 was not recovered as distinct in the other best model, which identified only seven groups (Supplementary Table 1 available on Dryad).

among-island variation because these groups occurred on multiple islands and were sympatric with other groups; all of the morphological groups identified in the best NMMs were widely distributed across the Galapagos Archipelago (median = 8.5 or 9.0, range 3–14 islands per group; Table 1 and Supplementary Material Table 1 available on Dryad) and most islands harbored several groups (up to six in Santiago and seven in Santa Cruz; Fig. 7 and Supplementary Material Fig. 3 available on Dryad). Importantly, almost all

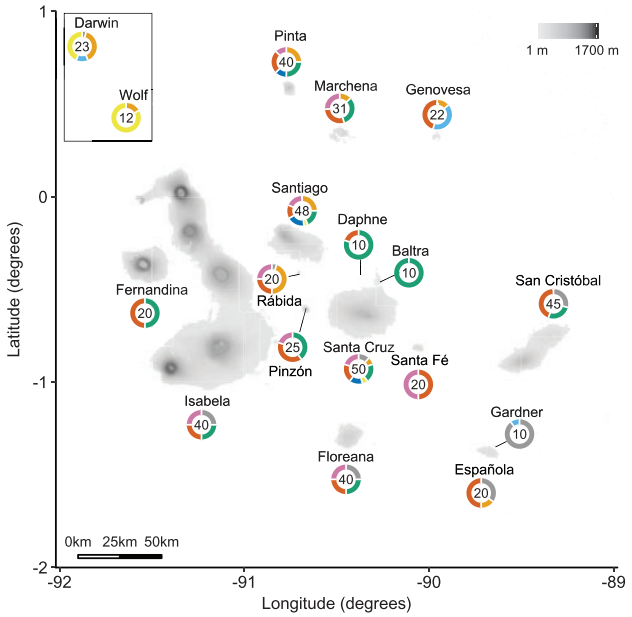

FIGURE 7. Eight distinct morphological groups of *Geospiza* ground-finches identified by one of the best normal mixture models have broad geographic distributions across the Galapagos Archipelago. For each island, numbers indicate individuals included in the analysis and ringplots depict the fraction of such individuals assigned to each morphological group following the color scheme in Fig. 5. The existence of distinct morphological groups in potential sympatry within islands (e.g., >4 groups in Santa Cruz, Santiago, and Pinta) suggests that such groups are unlikely to reflect within-species differentiation due to geographic isolation or local adaptation.

morphological groups co-occurred with each other in at least one island; the only exception was morphological Group 3 in one of the models, which co-occurred with four out of the other seven groups (Table 1 and Supplementary Material Table 1 available on Dryad). McKay and Zink (2015) indicated that different morphs of ground-finches exist within islands and argued that if such morphs were treated as species, then one would need to recognize dozens of species in the group; our analysis suggests this is not the case given the occurrence of all morphological groups in multiple islands.

At this point we note that because the specimens we analyzed were collected several decades ago (Swarth 1931), they may not faithfully reflect patterns in morphological variation nor the geographic distributions of morphological groups in the present. This is because over the past century, ground-finch populations have experienced a few colonization and extinction events, changes in the degree of morphological differentiation among populations due to natural and human-mediated hybridization, and bouts of selection in shifting directions over multiple generations in association with environmental variation in space and time (Harris 1973; De León et al. 2011; Grant and Grant 2014). Thus, we refrain from additional discussions about species limits involving comparisons of historical morphological data with contemporary evidence (e.g., genomics; Lamichhaney et al. 2015). Nonetheless, our analyses serve to demonstrate that statistically distinct morphological groups of ground-finches existed in the past, and we strongly suspect they still exist in the present. Accordingly, we suggest that the burden of proof for systematists proposing to lump ground-finches into a single species based on morphological data is on showing that distinct groups do not exist.

Conclusions; or, Atlantean Evolution in Darwin's Finches

Our reanalysis of morphological data pointed strongly to the existence of several groups of phenotypically distinct *Geospiza* ground-finches based only on six linear morphological measurements. In addition, we found evidence of distinct phenotypes in geographic scenarios (i.e., sympatry within islands) where one should not expect them if populations had not achieved evolutionary independence. Specifically, because the variation in quantitative morphological traits we examined is polygenic (Grant and Grant 1989; Grant and Grant 1994; Lamichhaney et al. 2015) and not caused by differences in sex or age (we restricted analyses to adult males), the existence of distinct phenotypic groups in areas where populations come into contact implies there are likely several species of ground-finches. Therefore, we contend that ground-finches are not an example of Sisyphean evolution (McKay and Zink 2015), a term that could well apply to other systems in nature (Seehausen 2006; Nosil et al. 2009; Rudman and Schluter 2016). Instead, evolutionary forces maintaining populations of ground-finches apart are likely in place, just as in Greek mythology Atlas prevents the merging of the Earth and the sky with his shoulders. Ground-finches thus likely represent an example of what one might call "Atlantean evolution." One, of course, does not need a new term to refer to speciation, but thinking of Atlas brings to mind *atlas*, a collection of maps, which reminds one of the central role of geography in speciation (Price 2008) and in the basic model underlying species delimitation based on phenotypic variation.

The question of exactly how many species of Darwin's ground-finches are there remains open and requires further attention to morphology, including careful scrutiny of discrepancies between morphological variation and taxonomy (e.g., Fig. 6). In addition, morphological variation should be further examined in light of biological factors including additional phenotypic characters, ecological niches, mating behavior, population dynamics, and patterns of genetic and genomic variation among populations (Grant 1999; Huber et al. 2007; Grant and Grant 2008; Farrington et al. 2014; Grant and Grant 2014; Lamichhaney et al. 2015; McKay and Zink 2015). Fruitful discussions about species limits in the group would likely start by addressing some of the additional thought-provoking arguments advanced by McKay and Zink (2015) that we did not touch on and which are beyond the scope of our work (e.g., the extent to which morphological

TABLE 2. Summary of three frequent issues in analysis of phenotypic data for species delimitation, possible solutions afforded by normal mixture models (NMMs), and associated challenges

Issues	Potential pitfalls	Proposed solutions	Some challenges
1. Graphical analyses may convey little information on phenotype frequencies	Failure to detect distinct phenotypic groups (Fig. 1)	Use NMMs to model phenotypic variation according to evolutionary theory	Improve approaches to estimate the number of distinct normal distributions (or "components") in NMMs
	Infer the existence of several phenotypic groups when, in fact, only one exists (Supplementary Material Appendices 1 available on Dryad)		
	Assessment can be highly subjective and difficult to replicate, particularly in analyses including > 2 phenotypic traits		
2. Reduction of dimensionality via PCA (or related approaches) may exclude important characters	Failure to detect distinct phenotypic groups by considering only principal components that explain a high proportion of the overall phenotypic variance (Fig. 2)	Use variable-selection approaches to reduce the number of dimensions (i.e., traits or principal components) to those that best discriminate groups in NMMs, using no *a priori* group information	Improve variable selection techniques for NMMs
		Reduce dimensionality to traits thought useful for distinguishing phenotypic groups, based on *a priori* considerations, and use those traits in NMMs	
3. Differences in central tendency are not evidence of distinct phenotypic distributions	Misinterpret differences in mean phenotypes as evidence of distinct groups (Fig. 3, Supplementary Material Appendix 1 available on Dryad)	Use NMMs to determine the number of distinct normal distributions (i.e., phenotypic groups) and estimate the parameters describing such distributions (means and variance–covariance matrices)	

groups are stable lineages over time or the evidence for the existence of distinct gene pools). Any such discussions, however, as well as discussions over species delimitation in other organisms, should bear in mind that phenotypic evidence for species limits is best assessed using statistical approaches appropriately grounded on evolutionary models.

OUTLOOK

We have described three issues that commonly affect species delimitation using phenotypic data, potential pitfalls that may derive from not considering such issues, and some proposed solutions (Table 2). However, we also highlight that the approaches used here to analyze phenotypic data for species delimitation are not free of problems. Issues such as estimation of the number of groups in NMMs (McLachlan and Peel 2000; McLachlan and Rathnayake 2014) or how to select variables for NMM analyses of multidimensional data sets (Poon et al. 2013) are critical areas of active research in statistics in which progress remains to be made (Table 2). Despite these issues, however, we argue that the statistical tools we used are appropriate because they are directly related to the basic evolutionary model underlying species delimitation using phenotypic data (Fisher 1918). Moreover, these tools allow systematists to go beyond fairly limited graphical analysis, and to break free from problems resulting from reduction of dimensionality using PCA or related approaches and from comparisons of measurements of central tendency. The value of embracing approaches with a solid theoretical basis despite limitations in their implementation in systematics is clear considering other developments in the field in which theory predated robust methodologies that subsequently blossomed. Such developments include the use of statistical methods to study species limits among fossil populations (Newell 1956), the application of probabilistic models to infer phylogenetic trees (Felsenstein 1981), time-calibration of molecular phylogenies (Kishino and Hasegawa 1990), and the estimation of species trees from gene trees (Maddison 1997).

Practical approaches to fit NMMs without *a priori* information about species limits offer a fresh perspective in inferences available to systematists and bring into question conventional reliance on two criteria often employed to establish species limits: lack of overlap in

phenotypic ranges and gaps in phenotypic distributions. In the absence of NMMs, it seemed reasonable to argue that species limits should be based on fixed phenotypic differences because continuous variation could only be subdivided using subjective criteria (Cracraft 1989; Davis 1997). Accordingly, overlap of phenotypic ranges has been conventionally stressed as a criterion to suggest samples of individuals are conspecific (e.g., Simpson 1951; Davis and Heywood 1963; Zink 2002; McKay and Zink 2015). However, under the framework offered by NMMs, overlap in phenotypic ranges is not relevant for species delimitation for two reasons. First, one may find strong empirical support for models in which the phenotypic ranges of distinct normal distributions overlap (e.g., Figs. 1 and 5), indicating that range overlap does not imply absence of evidence of species limits. Second, because phenotypic variation conforming to a single normal distribution can be arbitrarily split into parts with nonoverlapping ranges (Supplementary Material Appendix 1 available on Dryad), absence of range overlap does not imply strong empirical support for models with more than a single species.

In addition to lack of overlap of phenotypic ranges, gaps in phenotypic distributions have been conventionally used as a species criterion (Mallet 2013). Such gaps may be defined as phenotypic regions with low frequency of individuals and therefore do not necessarily imply lack of overlap in phenotypic ranges; indeed, a gap may exist between two distinct phenotypic distributions that overlap in their extremes (e.g., Fig. 1). Although true phenotypic gaps (along with multimodality in phenotypic distributions) are sufficient to suggest species boundaries (Zapata and Jiménez 2012; Mallet 2013; but see Day 1969; McLachlan and Peel 2000, page 17), they are not necessary to demonstrate such boundaries exist because NMMs specifying more than one species may be strongly supported in the absence of phenotypic gaps. An example of support for more than one normal distribution in the absence of phenotypic gaps was provided at the inception of NMMs: Karl Pearson inferred two groups among specimens of the shore crab (*Carcinus maenas*) from the Bay of Naples, even though the mixture of the groups was not bimodal and therefore they were not separated by a gap (Pearson 1894). Moreover, Pearson examined the possibility of inferring the existence of groups with different phenotypic variances but identical phenotypic means, which are by definition not separated by a gap (Supplementary Material Appendix 2 available on Dryad).

To conclude, we note that the criteria for species delimitation discussed above are relevant in the context of ideas about the reality of species. In particular, it has been argued that if the hypothesis that species are real entities in nature is correct, then biological diversity should be a patchwork of phenotypic clusters delineated by gaps (Coyne and Orr 2004; Barraclough and Humphreys 2015). This prediction, however, would not necessarily follow from the hypothesis that species are real if, as we argue, phenotypically distinct species need not be separated by gaps. In other words, species may be real, phenotypically distinct entities in nature even if phenotypic gaps are not major elements structuring biological diversity. Because statistical approaches related to NMMs now allow systematists to make unprecedented formal inferences about the existence of species even in the absence of phenotypic gaps, they constitute particularly useful tools to describe the structure of biological diversity, a necessary step to understand the evolutionary processes that generated it.

Supplementary Material

Data available from the Dryad Digital Repository: http://dx.doi.org/10.5061/dryad.9gh90.

Acknowledgements

We thank the California Academy of Sciences (Jack Dumbacher) for allowing us to use and publish morphological data from H. S. Swarth's archive. Bailey McKay kindly provided tabulated data. We thank Peter Grant, James Mallet, Bailey McKay, Van Remsen Jr., Sara Ruane, Elizabeth Spriggs, Peter Stevens, members of C.D. Cadena's laboratory group, anonymous reviewers, and associate editor John McCormack for discussion and helpful comments on the manuscript. C.D. Cadena thanks the Universidad de los Andes for allowing a sabbatical leave that enabled him to complete this project.

References

Allmon W.D. 2016. Studying species in the fossil record: a review and recommendations for a more unified approach. In: Allmon W.D., Yacobucci M.M., editors. Species and speciation in the fossil record. Chicago (IL): University of Chicago Press. p. 59–120.

Barraclough T.G., Humphreys A.M. 2015. The evolutionary reality of species and higher taxa in plants: a survey of post-modern opinion and evidence. New Phytologist 207:291–296.

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.

Bowman R. 1961. Morphological differentiation and adaptation in the Galapagos finches. University of California Publications in Zoology 58:1–302.

Camargo A., Sites J.W. Jr. 2013. Species delimitation: a decade after the renaissance. In: Pavlinov IY editor. The Species Problem: Ongoing Issues. InTech. p. 225–247; doi:10.5772/3313.

Carstens B., Pelletier T.A., Reid N.M., Satler J.D. 2013. How to fail at species delimitation. Mol. Ecol. 22:4369–4383.

Chang W.-C. 1983. On using principal components before separating a mixture of two multivariate normal distributions. Appl. Stat. 32:267–275.

Cooper Ornithological Society. 2016. Katma Award 2015, to Bailey McKay and Robert Zink. Condor 118:209–210.

Coyne J.A., Orr H.A. 2004. Speciation. Sunderland (MA): Sinauer Associates.

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.A.,

editors. Speciation and its consequences. Sunderland (MA): Sinauer Associates. p. 28–59.
Crisp M.D., Weston P.H. 1993. Geographic and ontogenetic variation in morphology of Australian waratahs (*Telopea*: Proteaceae). Syst. Biol. 42:49–76.
Davis J.I. 1997. Evolution, evidence and the role of species concepts in phylogenetics. Syst. Botany 22:373–403.
Davis P.H., Heywood V. 1963. Principles of angiosperm taxonomy. Princeton (NJ): D. van Nostrand Co.
Day N.E. 1969. Estimating the components of a mixture of normal distributions. Biometrika 56:463–474.
De León L.F., Raeymaekers J.A.M., Bermingham E., Podos J., Herrel A., Hendry A.P. 2011. Exploring possible human influences on the evolution of Darwin's finches. Evolution 65:2258–2272.
de Queiroz K. 1998. The general lineage concept of species, species criteria, and the process of speciation. In: Howard D.J., Berlocher S.H., editors. Endless forms: species and speciation. New York: Oxford University Press. p. 57–75.
Edwards D.L., Knowles L.L. 2014. Species detection and individual assignment in species delimitation: can integrative data increase efficacy? Proc. R. Soc. Lond. B 281:20132765.
Ezard T.H.G., Pearson P.N., Purvis A. 2010. Algorithmic approaches to aid species' delimitation in multidimensional morphospace. BMC Evol. Biol. 10:175.
Farrington H.L., Lawson L.P., Clark C.M., Petren K. 2014. The evolutionary history of Darwin's finches: speciation, gene flow, and introgression in a fragmented landscape. Evolution 68:2932–2944.
Felsenstein J. 1981. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. 17:368–376.
Fisher R.A. 1918. The correlation between relatives on the supposition of Mendelian inheritance. Trans. Roy. Soc. Edinburgh 52:399–433.
Fraley C., Raftery A.E. 2002. Model-based clustering, discriminant analysis, and density estimation. J. Am. Stat. Assoc. 97:611–631.
Fraley C., Raftery A.E., Murphy T.B., Scrucca L. 2012. mclust Version 4 for R: normal mixture modeling for model-based clustering, classification, and density estimation. Technical Report No. 597. Department of Statistics, University of Washington.
Fujita M.W., Leaché A.D., Burbrink F.T., McGuire J.A., Moritz C. 2012. Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol. Evol. 27:480–488.
Futuyma D.J. 2013. Evolution. 3rd ed. Sunderland (MA): Sinauer Associates.
Grant B.R., Grant P.R. 1989. Evolutionary dynamics of a natural population. The Large Cactus Finch of the Galápagos. Chicago (IL): Unversity of Chicago Press.
Grant B.R., Grant P.R. 2002. Simulating secondary contact in allopatric speciation: an empirical test of premating isolation. Biol. J. Linnean Soc. 76:545–556.
Grant P.R. 1999. Ecology and evolution of Darwin's finches. Princeton (NJ): Princeton University Press.
Grant P.R., Abbott I., Schluter D., Curry R.L., Abbott L.K. 1985. Variation in the size and shape of Darwin's finches. Biol. J. Linnean Soc., 25:1–39.
Grant P.R., Grant B.R. 1994. Phenotypic and genetic effects of hybridization in Darwin's finches. Evolution 48:297–316.
Grant P.R., Grant B.R. 2008. How and why species multiply: the radiation of Darwin's finches. Princeton (NJ): Princeton University Press.
Grant P.R., Grant B.R. 2014. 40 years of evolution: Darwin's Finches on Daphne Major Island. Princeton (NJ): Princeton University Press.
Grant P.R., Grant B.R., Petren K. 2000. The allopatric phase of speciation: the sharp-beaked ground finch (*Geospiza difficilis*) on the Galápagos islands. Biol. J. Linnean Soc. 69:287–317.
Guillot G., Renaud S., Ledevin R., Michaux J., Claude J. 2012. A unifying model for the analysis of phenotypic, genetic, and geographic data. Syst. Biol. 61:897–911.
Harris M.P. 1973. The Galápagos avifauna. Condor 75:265–278.
Harrison R.G. 1998. Linking evolutionary pattern and process: the relevance of species concepts for the study of speciation. In: Howard D.J., Berlocher S.H., editors. Endless forms: species and speciation. New York: Oxford University Press. p. 19–31.
Hausdorf B., Hennig C. 2010. Species delimitation using dominant and codominant multilocus markers. Syst. Biol. 59:491–503.
Henderson A. 2006. Traditional morphometrics in plant systematics and its role in palm systematics. Bot. J. Linnean Soc. 151: 103–111.
Hennig C. 2010. Methods for merging Gaussian mixture components. Adv. Data Anal. Classif. 4:3–34.
Huber S.K., De León L.F., Hendry A.P., Bermingham E., Podos J. 2007. Reproductive isolation of sympatric morphs in a population of Darwin's finches. Proc. R. Soc. Lond. B 274:1709–1714.
Kass R.E., Raftery A.E. 1995. Bayes factors. J. Am. Stat. Assoc., 90:773–795.
Kishino H., Hasegawa M. 1990. Converting distance to time: application to human evolution. Methods Enzymol. 183:550–570.
Kleindorfer S., O'Connor J.A., Dudaniec R.Y., Myers S.A., Robertson J., Sulloway F.J. 2014. Species collapse via hybridization in Darwin's tree finches. Am. Naturalist 183:325–341.
Lack D. 1947. Darwin's Finches. New York: Cambridge University Press.
Lamichhaney S, Berglund J, Almén MS, Maqbool K, Grabherr M, Martinez-Barrio A, Promerová M, Rubin C-J, Wang C, Zamani N, Grant BR, Grant PR, Webster MT, Andersson L. 2015. Evolution of Darwin's finches and their beaks revealed by genome sequencing. Nature 518:371–375.
Leaché A.D., Koo M.S., Spencer C.L., Papenfuss T.J., Fisher R.N., McGuire J.A. 2009. Quantifying ecological, morphological, and genetic variation to delimit species in the coast horned lizard species complex (*Phrynosoma*). Proc. Nat. Acad. Sci. USA 106: 12418–12423.
Luckow M. 1995. Species concepts: assumptions, methods, and applications. Syst. Botany 20:589–518.
Maddison W.P. 1997. Gene trees in species trees. Syst. Biol. 463: 523–536.
Mallet J. 2008. Hybridization, ecological races and the nature of species: empirical evidence for the ease of speciation. Phil. Trans. R. Soc. Lond. B 363:2971–2986.
Mallet J. 2013. Concepts of species. In: Levin S.A., editor. Encyclopedia of biodiversity. Vol. 6. Waltham (MA): Academic Press. p. 679–691.
Maugis C., Celeux G., Martin-Magniette M.-L. 2009a. Variable selection for clustering with Gaussian mixture models. Biometrics 65:701–709.
Maugis C., Celeux G., Martin-Magniette M.-L. 2009b. Variable selection in model-based clustering: a general variable role modeling. Comput. Stat. Data Anal. 53:3872–3882.
Mayr E. 1992. A local flora and the biological species concept. Am. J. Botany 79:222–238.
Mayr E., Linsley E.G., Usinger R.L. 1953. Methods and principles of systematic zoology. New York: McGraw-Hill.
McDade L.A. 1995. Species concepts and problems in practice: insights from botanical monographs. Syst. Botany 20:606–622.
McKay B.D., Mays H.L. Jr, Yao C.-T., Wan D., Higuchi H., Nishiumi I. 2014. Incorporating color into integrative taxonomy: analysis of the Varied Tit (*Sittiparus varius*) complex in East Asia. Syst. Biol. 63:505–517.
McKay B.D., Zink R.M. 2015. Sisyphean evolution in Darwin's finches. Biol. Rev. 90:689–698.
McLachlan G., Krishnan T. 2008. The EM algorithm and extensions. 2nd ed. Hoboken (NJ): John Wiley and Sons (Series in probability and statistics).
McLachlan G, Peel D.A. 2000. Finite mixture models. Hoboken (NJ): John Wiley and Sons (Series in probability and statistics).
McLachlan G.J. 2004. Discriminant analysis and statistical pattern recognition. Hoboken (NJ): John Wiley and Sons (Series in probability and statistics).
McLachlan G.J., Rathnayake S. 2014. On the number of components in a Gaussian mixture model. WIREs Data Min. Knowl. Discov. 4:341–355.
Miller W. 2016. The species problem: concepts, conflicts, and patterns preserved in the fossil record. In: Allmon W.D., Yacobucci M.M., editors. Species and speciation in the fossil record. Chicago (IL): University of Chicago Press. p. 28–58.
Moczek A.P., Emlen D.J. 1999. Proximate determination of male horn dimorphismin the beetle *Onthophagus taurus* (Coleoptera: Scarabaeidae). J. Evol. Biol. 12:27–37.

Newell N.D. 1956. Fossil populations. In: Sylvester-Bradley P.C., editor. The species concept in paleontology: a symposium. London (UK): London Systematics Association. p. 63–82.
Nosil P., Harmon L.J., Seehausen O. 2009. Ecological explanations for (incomplete) speciation. Trends Ecol. Evol. 24:145–156.
Padial J.M., Miralles A., De la Riva I., Vences M. 2010. The integrative future of taxonomy. Front. Zool. 7:216.
Patten M.A., Remsen J.V. Jr. 2017. Complementary roles of phenotype and genotype in subspecies delimitation. J. Heredity 108: 462–464.
Patten M.A, Unitt P. 2002. Diagnosability versus mean differences of sage sparrow subspecies. Auk 119:26–35.
Pearson K. 1894. Contributions to the theory of mathematical evolution. Philos. Trans. R. Soc. Lond. A, 185:71–110.
Poon L.K.M., Zhang N.L., Liu T., Liu A.H. 2013. Model-based clustering of high-dimensional data: variable selection versus facet determination. Int. J. Approx. Reason. 54:196–215.
Price T. 2008. Speciation in birds. Greenwood Village (CO): Roberts and Company Publishers.
Raftery A.E., Dean N. 2006. Variable selection for model-based clustering. J. Am. Stat. Assoc. 101:168–178.
Remsen J.V. Jr. 2010. Subspecies as a meaningful taxonomic rank in avian classification. Ornithol. Monogr. 67:62–78.
Remsen JV, Jr., Areta JI, Cadena CD, Claramunt S, Jaramillo A, Pacheco JF, Pérez-Emán J, Robbins MB, Stiles FG, Stotz DF, Zimmer KJ. 2017. A classification of the bird species of South America, version 28 April, 2017. American Ornithologists' Union. Available from: URL http://www.museum.lsu.edu/~Remsen/SACCBaseline.htm.
Rieseberg L.H., Wood T.E., Baack E.J. 2006. The nature of plant species. Nature 440:524–527.
Rising J.D., Jaramillo A., Copete J.L., Ryan P.G., Madge S.C. 2011. Family Emberizidae (Buntings and New World Sparrows). In: del Hoyo J., Elliott A., Christie D.A., editors. Handbook of the Birds of the World. Vol. 16. Tanagers to New World Blackbirds. Barcelona: Lynx Edicions. p. 428–683.
Rudman S.M., Schluter D. 2016. Ecological impacts of reverse speciation in threespine stickleback. Curr. Biol. 26:490–495.
Sangster G. 2014. The application of species criteria in avian taxonomy and its implications for the debate over species concepts. Biol. Rev. 89:199–214.
Schwarz G. 1978. Estimating the dimension of a model. Ann. Stat. 6: 461–464.
Scrucca L., Fop M., Murphy T.B., Raftery A.E. 2016. mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. R Journal. 8:289–317.
Scrucca L, Raftery A.E. 2004. clustvarsel: a package implementing variable selection for model-based clustering in R. Preprint available on arXiv https://arxiv.org/abs/1411.0606.
Seehausen O. 2006. Conservation: losing biodiversity by reverse speciation. Curr. Biol. 16:R334–R337.
Simpson G.G. 1951. The species concept. Evolution 5:285–298.
Sites J.W. Jr, Marshall J.C. 2003. Delimiting species: a Renaissance issue in systematic biology. Trends Ecol. Evol. 18:462–470.
Sites J.W. Jr, Marshall J.C. 2004. Operational criteria for delimiting species. Ann. Rev. Ecol. Evol. Syst. 35:199–227.
Smith T.B. 1993. Disruptive selection and the genetic basis of bill size polymorphism in the African finch *Pyrenestes*. Nature 363: 618–620.
Sneath P.H.A., Sokal R.R. 1973. Numerical taxonomy. San Francisco (CA): W. H. Freeman and Co.
Solís-Lemus C., Knowles L.L., Ané C. 2014. Bayesian species delimitation combining multiple genes and traits in a unified framework. Evolution 69:492–507.
Stevens P.F. 1991. Character states, morphological variation, and phylogenetic analysis: a review. Syst. Botany 16:553–583.
Sukumaran J., Knowles L.L. 2017. Multispecies coalescent delimits structure, not species. Proc. Nat. Acad. Sci. USA 114: 1607–1612.
Swarth H.S. 1931. The avifauna of the Galápagos Islands. Occas. Pap. Calif. Acad. Sci. 18:1–299.
Templeton A.R. 2006. Population genetics and microevolutionary theory. Hoboken (NJ): John Wiley & Sons.
Tobias J.A., Seddon N., Spottiswoode C.N., Pilgrim J.D., Fishpool L.D.C., Collar N.J. 2010. Quantitative criteria for species delimitation. Ibis 152:724–746.
Wiens J.J., Servedio M.R. 2000. Species delimitation in systematics: inferring diagnostic differences between species. Proc. R. Soc. Lond. B 267:631–636.
Wiley E.O. 1978. The evolutionary species concept reconsidered. Syst. Zool. 27:17–26.
Zapata F., Jiménez I. 2012. Species delimitation: inferring gaps in morphology across geography. Syst. Biol. 61:179–194.
Zink R.M. 2002. A new perspective on the evolutionary history of Darwin's finches. Auk 119:864–871.