

Species-Level Heritability Reaffirmed: A Comment on “On the Heritability of Geographic Range Sizes”

Gene Hunt,^{1,*} Kaustuv Roy,^{1,†} and David Jablonski^{2,‡}

1. Section of Ecology, Behavior, and Evolution, Division of Biological Sciences, University of California, San Diego, La Jolla, California 92093-0116;

2. Department of Geophysical Sciences, University of Chicago, Chicago, Illinois 60637

Submitted February 25, 2004; Accepted June 24, 2004;

Electronically published May 31, 2005

Online enhancements: appendixes.

Keywords: heritability, geographic range size.

Geographic range has attracted attention as an important macroevolutionary parameter because it has been argued to be a species-level trait involved in higher-level sorting processes (Jablonski 1986, 1987, 2000; Brown 1995; Grantam 1995; Gould 2002; Okasha 2003). Previous studies have found that geographic ranges of pairs of related taxa are more similar than expected by chance (i.e., geographic range is “heritable”) in Cretaceous marine mollusks (Jablonski 1987), herbaceous plants (Ricklefs and Latham 1992; Qian and Ricklefs 2004), and North American birds and mammals (Brown 1995, pp. 193–194). However, Webb and Gaston (2003) recently argued that the analytical methods used by Jablonski (1987; and presumably other studies that used similar methods) are inappropriate and overestimate the correlation between ranges of closely related taxa. Using a new method, they concluded that range sizes of closely related species of both Cretaceous gastropods and living birds are not more similar than expected by chance.

Webb and Gaston’s main concern was the appropriateness of standard statistical methods for analyzing geographic range size distributions, which are almost always right skewed (Willis 1922; Brown 1995; Gaston 1998). This

skewness violates the assumptions of parametric statistics commonly used to measure heritability, such as parent-offspring regressions and sib-sib product-moment correlations (Falconer and Mackay 1996). Jablonski (1987) recognized this problem and used nonparametric statistics, but Webb and Gaston suggested that even nonparametric methods could be affected by the skewed nature of the data and hence developed a new method for analyzing range size heritability.

Although Webb and Gaston have contributed valuable empirical data to the question of range size heritability, we show here that their method is flawed in that it implicitly assumes that range sizes are uniformly distributed. When range sizes have their characteristic right skew, Webb and Gaston’s method generally fails to detect heritability even when range sizes of closely related species are strongly correlated. We also reassess Jablonski’s (1987) Cretaceous mollusk data and find strong support for his conclusion of significantly heritable geographic range size in these animals. Moreover, we note that Webb and Gaston’s own analyses suggest that range size is also weakly but significantly heritable in modern bird species.

Webb and Gaston’s Method

In order to assess the similarity of geographic ranges of closely related species, Webb and Gaston (2003) defined a metric of range size asymmetry, *Asy*. For a pair of closely related species (sister species or ancestor-descendant pairs), *Asy* is defined as

$$Asy = 1 - \frac{S}{B},$$

where *S* is the smaller and *B* the larger of the two geographic ranges; *Asy* is a dissimilarity metric that is equal to 0 when the pair of species have ranges of equal size and approaches 1 as the difference between the two range sizes increases. To determine the significance of observed *Asy* values, Webb and Gaston compared them to the expected distribution of *Asy* if the smaller geographic range (*S*) were a random fraction of the larger range (*B*). Because range

* E-mail: eghunt@ucsd.edu.

† E-mail: kroy@ucsd.edu.

‡ E-mail: djablons@midway.uchicago.edu.

sizes were measured in whole numbers, this approach allows S to be any integer from 1 to B with equal probability. Under this assumption, the expected value of Asy is a simple function of B : $E(Asy) = (B - 1)/(2B)$. If the observed values of Asy are generally less than the expected Asy , one would conclude that geographic range sizes in closely related species are significantly similar and thus show evidence of species-level heritability.

The key assumption that Webb and Gaston (2003) made is that the smaller range size (S) is a “random fraction” of the larger range (B). In other words, conditional on a particular value of B , the distribution of S is uniform from 1 to B . However, under the null hypothesis that ancestor and descendant ranges are drawn independently from the distribution of range sizes, the distribution of S for any B will in fact mirror the overall probability distribution of range sizes, truncated at B (fig. 1A; this proposition is demonstrated explicitly in app. A in the online edition of the *American Naturalist*). Thus, S can be expected to be uniformly distributed from 1 to B only when range sizes themselves are uniformly distributed. This condition is clearly not met in the data sets analyzed by Webb and Gaston, both of which show the strongly right-skewed distributions that are characteristic of geographic range size data (Willis 1922; Brown 1995; Gaston 1998; fig. 1B). Hence, instead of testing the null hypothesis of no correlation between closely related species, Webb and Gaston’s procedure essentially tests a composite null hypothesis that geographic ranges are not heritable and that they are uniformly distributed. As a result, test results may be determined by the shape of the frequency distribution of range sizes and not their heritability. Note that our concerns about Webb and Gaston’s method are not focused on the use of Asy , which is a straightforward distance metric. Although its advantages and disadvantages relative to other similarity/dissimilarity metrics are unknown at present, it may be a reasonable way to measure the difference between pairs of ranges. However, interpreting observed Asy values requires knowing the expected distribution of this statistic in the absence of range size heritability. The generation of this null distribution is the primary concern of this comment.

What If Geographic Ranges Are Not Uniformly Distributed?

To evaluate Webb and Gaston’s (2003) method, we applied it to simulated data generated under different assumptions about the distribution of range sizes. We considered three different range size distributions: uniform; right skewed, drawn from an exponential distribution; and symmetric, drawn from a normal distribution. Our use of the exponential distribution to model right-skewed distributions is

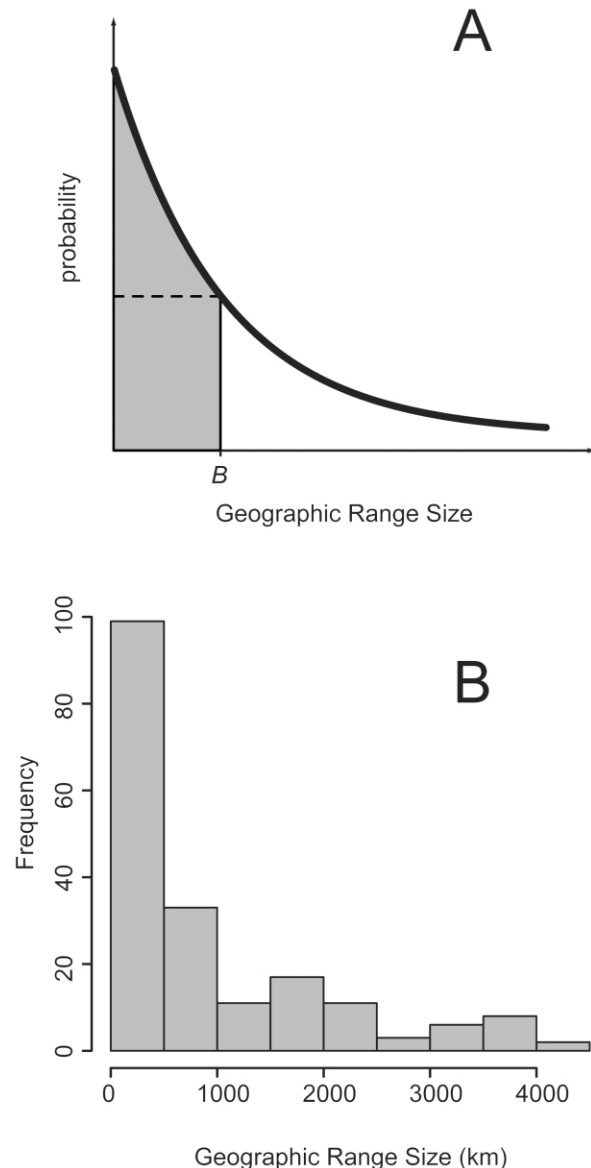


Figure 1: A, Hypothetical distribution of geographic range sizes shown as the thick solid line. For two range sizes drawn from this distribution, B is the larger of the pair. Conditional on this value of B , the distribution of the smaller range size will mirror the probability distribution of range sizes, truncated at B (shaded region). In contrast, Webb and Gaston’s (2003) approach assumes that the smaller range is a random fraction of B (dashed line), regardless of how range size is distributed in the taxa under study. B, Empirical frequency distribution of geographic range sizes of Cretaceous gastropods ($n = 95$) from Jablonski (1987). Note the strong right skew typical of range size distributions.

based on the observation that empirical distributions of range sizes are typically very strongly right skewed on an arithmetic scale, so much so that the smallest size class is almost always the modal class (Gaston 1998). The expo-

nential distribution shares this quality, and range sizes of Cretaceous gastropods (fig. 1) can be approximated quite well by an exponential distribution. Using other distributions (including lognormal and gamma distributions) to model strongly right-skewed data yielded qualitatively similar results.

In all simulations, range sizes of ancestors and descendants (or sister species) were drawn independently from the appropriate distribution. Thus, true heritability is 0 in all cases. In order to match the data and methods of Webb and Gaston, we rounded all ranges to the nearest integer and scaled them to range from 1 to 100 (no important differences result if this step is omitted). For each simulated data set, we calculated the values of B and Asy and compared them to the expectation derived by Webb and Gaston.

Representative simulated geographic ranges under all three scenarios are shown in figure 2, along with the resulting plots of Asy with respect to B . Each plot shows $n = 103$ pairs of geographic ranges, the sample size of Webb and Gaston's bird data set. When geographic ranges are uniformly distributed, Webb and Gaston's method performs as expected: the observed values of Asy are distributed symmetrically around the expected Asy (fig. 2B). In this particular realization, 49 values of Asy are greater than and 54 are lower than the expected Asy , a result that is consistent with points having no preferred tendency to fall above or below the expectation ($P = .69$, exact binomial test).

The results are, however, very different when geographic ranges are exponentially distributed (fig. 2D, 2E). Under these conditions, there is a strong tendency for observed values of Asy to exceed the expectation of Webb and Gaston. For the realization shown, 69 values of Asy exceed the expectation, 32 are less than the expectation, and two points are equal to their expectation (fig. 2). Results as imbalanced as these are very unlikely unless the null distribution of Asy differs from the expectation derived by Webb and Gaston ($P = .003$, exact binomial test of the probability of observing at least 69 of 102 values greater than the expectation, given equal probabilities of being greater than or less than the expectation). Geographic range sizes tend to be more asymmetric than the Webb and Gaston expectation because, for a given value of B , most of the observed values of S are clustered at the small end of the range size distribution simply because most ranges are small. This clustering of small S values produces a greater difference between S and B (and thus a higher Asy value) than would be seen if range sizes were evenly distributed.

When geographic ranges are normally distributed, the opposite pattern emerges: most of the observed values of Asy are less than the expectation (fig. 2G, 2H). Using Webb

and Gaston's test, one would mistakenly conclude that the lower than expected difference between the ranges of sister species constitutes significant evidence of geographic range heritability (27 and 76 values of Asy are greater and less than the expectation, respectively; $P < .00001$, exact binomial test). Because most of the weight is at the center of the distribution, Asy values tend to be low, and therefore values of S are disproportionately close to B , relative to the uniform assumption.

The specific realizations shown in the first two columns of figure 2 are typical of those produced in these simulations. The right-hand column of figure 2 shows the distribution of deviations from the Webb and Gaston expectation for 10,000 simulated pairs of ranges for each distribution. Deviations were measured as the observed Asy minus the expected Asy , divided by the expected Asy . If Asy is overestimated, deviations will be shifted to values greater than 0; if Asy is underestimated, deviations will tend to be less than 0. Thus Asy is roughly unbiased when geographic ranges are uniformly distributed (fig. 2C), yields values systematically higher than the expectation when ranges are exponentially distributed (fig. 2F), and yields values systematically lower than the expectation when ranges are normally distributed (fig. 2I). The tendency for right skewness to inflate Asy values accounts for Webb and Gaston's otherwise counterintuitive finding that closely related species of birds or gastropods tend to have range sizes that are significantly dissimilar, even though rank correlations in both data sets are significantly positive. This interpretation is further supported by the presence of a consistent and robust signal suggesting significant heritability in the gastropod data of Jablonski (1987), as discussed below.

Reassessing Range Size Heritability in Cretaceous Gastropods

Webb and Gaston's (2003) main concern was that the statistical methods used by Jablonski (1987) overestimated the correlation of geographic range size between ancestor-descendant pairs. Webb and Gaston suggested that two aspects of Jablonski's data may have caused the spurious finding of significant heritability. First, they noted that range sizes are strongly skewed, and in particular there were many ancestor-descendant pairs in which both species have very small ranges. Second, they suggested that the outliers may have somehow unduly influenced the inferred similarity between ancestor and descendant range sizes. Here we consider both of these claims and reevaluate the gastropod data set of Jablonski (1987) to test the robustness of the original results. The data set analyzed here is the same as that in Jablonski (1987) with minor updates to the range sizes of three species pairs; these updates have

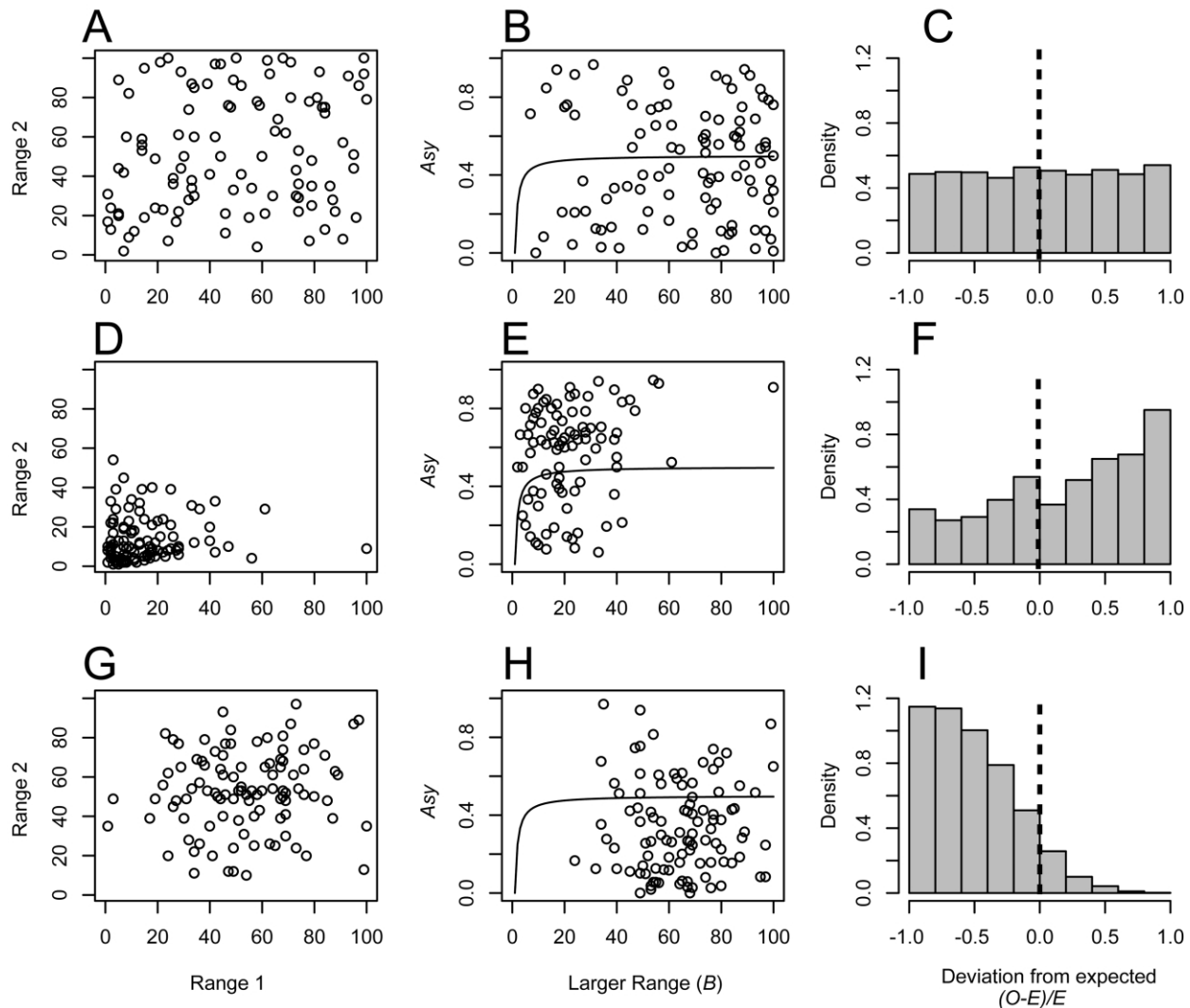


Figure 2: Application of Webb and Gaston's (2003) method to simulated data sets. Each row corresponds to a different distribution of range sizes. In the first row (A–C), range sizes are drawn from a uniform distribution; in the middle row (D–F), range sizes were drawn from a right-skewed (exponential) distribution; in the bottom row (G–I), range sizes were drawn from a normal distribution. The left column (A, D, G) plots the original range sizes for $n = 103$ simulated ancestral and descendant species pairs. The middle column (B, E, H) plots the Asy of each species pair with respect to the value of the larger geographic range (B). The curve on each of these plots shows the expected value of Asy according to Webb and Gaston. The right column (C, F, I) shows the distribution of the deviation of observed Asy values from the expectation of Webb and Gaston for 10,000 random draws from each distribution. These deviations were calculated as the observed Asy minus the expected Asy, divided by the expected Asy. These deviations will average 0 (shown as a dotted line) if the Webb and Gaston expected value of Asy is an unbiased predictor of observed Asy values.

negligible influence on the results. The raw data are presented in appendix B in the online edition of the *American Naturalist* (table B1).

It is straightforward to show that skewness, in itself, has no impact on Jablonski's (1987) result. This can be demonstrated by comparing the observed Spearman's rank correlation coefficient (r_s) for these data with the distribution of r_s obtained by randomizing the range sizes of

the ancestral (or descendant) species. Shuffling the order of ancestral range sizes destroys any similarity due to heritability by scrambling ancestor-descendant pairs while preserving the skewness of the original data set. The observed r_s for the Jablonski gastropod data set is 0.61, which is greater than any of 10,000 rank correlation coefficients generated from the shuffled data ($P < .0001$; fig. 3). In addition, it is also worth noting that similar randomiza-

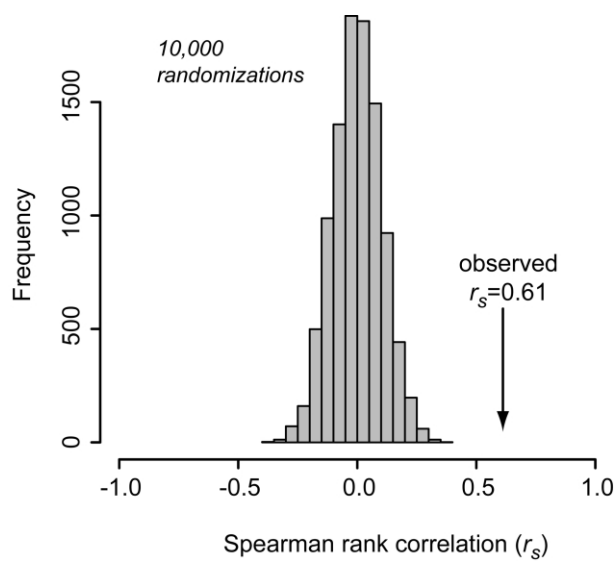


Figure 3: Significance of rank correlation in geographic ranges of ancestor-descendant pairs of Cretaceous gastropods assessed using randomization. The histogram shows the distribution of Spearman rank correlations (r_s) generated by shuffling the geographic ranges of ancestors 10,000 times. This shuffling destroys any heritability by separating ancestor-descendant pairs but preserves other aspects of the empirical distribution of range sizes. The arrow shows the observed value of r_s , which is larger than any of the 10,000 randomized values.

tions reported by Webb and Gaston (2003, p. 557) for 103 avian sister species pairs suggest that range sizes of sister species are more similar than would be expected by chance, although the pattern is weaker than in Cretaceous gastropods ($r_s = 0.26$, $P < .01$).

These tests are examples of randomization or permutation tests (Sokal and Rohlf 1995, p. 803; Manly 1997), which are useful for generating null distributions without making assumptions about the distribution from which the data are drawn. Note that despite the nonnormality of range sizes, the null distribution of r_s is centered around 0 (fig. 3), indicating that r_s does not tend to overestimate similarity even when the data are strongly right skewed. The randomization-based P values for r_s presented in this comment are nearly identical to those obtained from statistical tables with r_s adjusted for tied observations (no correction for ties is needed for randomization tests because the null distribution is based on data with exactly the same number of ties as the observed data).

If one uses a similar randomization routine to generate the null distribution for Asy , one finds that the median Asy of Jablonski's gastropod data is in fact lower than any of 10,000 shuffled data sets (i.e., data sets in which the order of the ancestral range sizes is randomly permuted; median $Asy = 0.61$, $P < .0001$). Thus, the Asy metric,

when compared to an appropriate null distribution, also gives very strong evidence that range size is heritable in Cretaceous gastropods. Because separate Asy values are produced for each species pair rather than a single value characterizing the whole data set, the additional step of taking the median Asy is required to perform a statistical test (one could also use the mean Asy , which, in our case, gives the same result). Taking a different approach, Webb and Gaston instead compared each Asy to a theoretical expectation assuming a uniform distribution of range sizes and counted the number of observed values that exceed their expectation to determine overall significance. Although their approach could conceivably be modified to account for nonuniform range size distributions, the first approach has the clear advantage of avoiding assumptions about how range sizes are distributed.

In addition to skewness, Webb and Gaston expressed concern over the large number of ancestor-descendant pairs in Jablonski's data in which both species have very small ranges. They argued that the large number of points near the origin may spuriously influence even nonparametric analyses. While Webb and Gaston are correct in stating that there are a large number of species pairs in which both members have small ranges, this pattern is more appropriately interpreted as a consequence of range size heritability than as a statistical artifact. Or, to put it another way, one would not expect to see so many species pairs where both have small ranges unless range size were heritable. This can be seen by constructing a two-by-two contingency table, classifying ancestor and descendants by whether they have small (<500 km) or large (≥ 500 km) ranges (table 1). Table 1 shows the observed counts in each cell, followed by the expected counts if range size were independent in ancestors and descendants. A G -test of independence (Sokal and Rohlf 1995) confirms that there is significant overrepresentation of small-small and large-large range pairs ($G = 22.6$, $P < 10^{-5}$), as would be expected if range size were heritable. This significant non-independence is not sensitive to the exact cutoff between

Table 1: Two-by-two contingency table showing range sizes of 95 ancestor-descendant pairs of Cretaceous gastropod species from Jablonski (1987)

Ancestor range size	Descendant range size	
	Small (<500 km)	Large (≥ 500 km)
Small (<500 km)	37 (25.7)	10 (21.3)
Large (≥ 500 km)	15 (26.3)	33 (21.7)

Note: For each cell of the table, observed counts are followed in parentheses by the expected counts assuming independence of ancestral and descendant ranges. A G -test of independence finds a significant excess of small-small and large-large range pairs ($P < 10^{-5}$).

small and large ranges; any value between 10 and 3,800 km produces similar results.

It is also important to note that significant range size heritability does not result solely from an excess of ancestor-descendant pairs both with small ranges. Consider species pairs for which the ancestor has a large range (table 1). Although less than 50% of all species in this data set have large (>500 km) ranges, 69% (33/48) of the species originating from a large-range-size ancestor also have large ranges. This finding is unlikely unless large-range-size species preferentially give rise to large-range descendants ($P = .013$, binomial test of the probability of yielding 33 or more large-ranged descendants from 48 large-ranged ancestors given an equal chance of producing small- and large-ranged species). Moreover, even if all comparisons that include a species with a small (<500 km) range are excluded, there is still a significant rank correlation among the species pairs that remain ($r_s = 0.38$, $n = 34$, $P = .014$ based on 10,000 random permutations). Consistent with these results, the residuals of the regression of descendant range rank on ancestral range rank (see fig. 4) do not become larger in absolute magnitude with increasing ancestral range size, and range size asymmetry (as measured by *Asy*) is uncorrelated with the range size of the ancestral species ($r_s = -0.09$, $P = .39$; a positive correlation is expected if heritability is restricted to small-range-size species). All of these results strongly confirm that range size is significantly heritable in these Cretaceous gastropods and that this heritability holds across all range sizes.

The final concern of Webb and Gaston was that outlying observations may spuriously inflate correlation coefficients. Although often important in parametric correlations, outliers tend to have little effect on rank correlations. Figure 4 shows the Cretaceous gastropod data, with geographic ranges plotted as their rank. Visual inspection does not reveal any obvious outliers. A linear regression of descendant range rank on ancestral range rank finds three residuals more than two standard deviations from the regression line; these points are marked as filled squares in figure 4. Because these points are in the upper left and lower right of this plot, they tend to decrease rather than inflate the correlation between ancestral and descendant range ranks. These ranked data do not satisfy the assumptions of least-squares regression, and thus this approach to identify outliers is best considered heuristic. Nevertheless, if these outlying points have any effect on this analysis, it is to underestimate, rather than overestimate, the species-level heritability of geographic range size.

Discussion

As for the question of how best to assess geographic range size heritability, we do not advocate any particular ap-

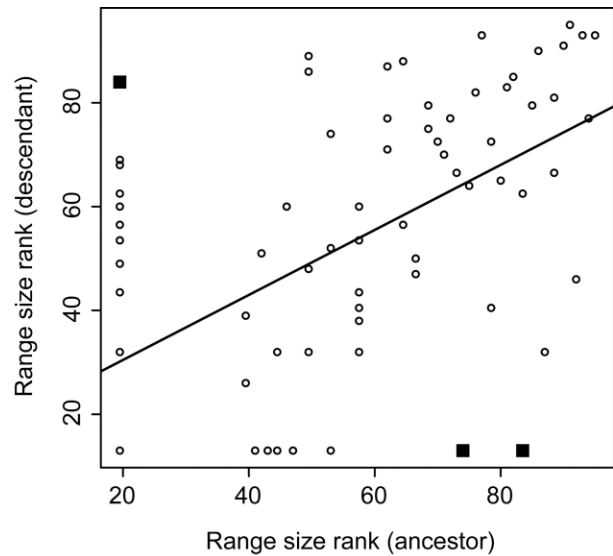


Figure 4: Ancestral versus descendant (or sister species) range size of Cretaceous gastropods plotted as their ranks. The line shows the linear (least-squares) regression of descendant range size rank on ancestral range size rank (switching the independent and dependent variables produces the same result). Potential outlying observations (those located more than two standard deviations from the regression line) are shown as filled squares. Note that fewer than $n = 95$ points appear on this plot because some points plot directly on top of one another.

proach beyond a strong preference for nonparametric statistics and randomization techniques to avoid problematic assumptions about the shapes of range size distributions. Although one could use *Asy* or other dissimilarity metrics to construct a test of range size heritability, it should be shown first that any new test is unbiased and reasonably powerful relative to more established methods. Based on the analyses presented here, we have not found any compelling evidence that supports the contention that the current arsenal of statistical approaches (e.g., nonparametric statistics, randomization/permutation tests, contingency tables), widely used in disciplines ranging from biological to social sciences, should yield systematically misleading conclusions when properly applied to the problem of range size heritability.

In this comment, we have considered the situation in which one has pairs of species that are closely related (sister species or ancestor-descendant pairs), but relationships among the pairs are not necessarily known (as is true for Jablonski 1987 and Webb and Gaston 2003). If, however, a well-supported phylogenetic hypothesis is available for the taxa under study, it may be preferable to instead assess heritability using any of several appropriate comparative methods (e.g., Cheverud et al. 1985; Freckleton et al. 2002; Housworth et al. 2004).

More studies are needed before the generality and macroevolutionary importance of range size heritability can be assessed definitively. For both Cretaceous gastropods and living birds, closely related species have geographic ranges that are more similar than would be expected by chance. On the other hand, the correlation is much stronger in Cretaceous gastropods than in living birds, with North American mammal and bird species (Brown 1995) and disjunct herbaceous (but not woody) plant genera (Ricklefs and Latham 1992; Qian and Ricklefs 2004) falling in between. If these few examples are any indication, then we can expect the level of geographic range heritability to vary among groups, just as body-mass heritability varies among species and populations (see, e.g., Falconer and Mackay 1996). Many interesting biological questions remain, including: How does the strength of range size heritability vary across different taxa? Do the levels of range size heritability change over the evolutionary history of a group (e.g., are the patterns in late Cenozoic gastropods different from those in the Cretaceous)? What are the biological underpinnings of differences in pattern or strength of range size heritability? Does the “snapshot” of geographic ranges inevitably used for extant taxa yield different results from paleontological data where maximum realized geographic range can be analyzed? Because narrow-ranging species at any instant in time could be in an early expansion stage, at a stable maximum range, or in decline, such snapshots may be analogous to a traditional heritability study that fails to standardize by ontogenetic age (see Jablonski 1987 and Gaston 1998 for discussion). Answering these questions will require many more case studies than are currently available, and new analyses should be careful to employ statistical methods that are insensitive to the skewed distributions typical of geographic range size data.

Conclusion

The method proposed by Webb and Gaston (2003) to assess the heritability of geographic range size is flawed in that it assumes that geographic range sizes are uniformly distributed. For the strongly right-skewed distributions that characterize virtually all empirical geographic range size data sets, Webb and Gaston’s methodology will often find that range sizes are more dissimilar than one would expect by chance, even when range sizes of close relatives are significantly correlated. Our reanalysis of Jablonski’s (1987) Cretaceous gastropod data set shows a strong and robust signal of significant species-level heritability in geographic range size. Results of randomizations presented by Webb and Gaston for 103 avian sister species pairs

suggest that a similar, albeit weaker, relationship also holds for birds.

Acknowledgments

We thank M. Foote, R. Lande, and T. J. Webb for their comments on an earlier version of this manuscript. This work was partially supported by grants from the National Science Foundation.

Literature Cited

- Brown, J. H. 1995. *Macroecology*. University of Chicago Press, Chicago.
- Cheverud, J. M., M. M. Dow, and W. Leutenegger. 1985. The quantitative assessment of phylogenetic constraints in comparative analyses: sexual dimorphism in body weight among primates. *Evolution* 39:1335–1351.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to quantitative genetics*. 4th ed. Addison Wesley Longman, Essex.
- Freckleton, R. P., P. H. Harvey, and M. Pagel. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *American Naturalist* 160:712–726.
- Gaston, K. J. 1998. Species-range size distributions: products of speciation, extinction and transformation. *Philosophical Transactions of the Royal Society of London B* 353:219–230.
- Gould, S. J. 2002. *The structure of evolutionary theory*. Harvard University Press, Cambridge, MA.
- Grantham, T. A. 1995. Hierarchical approaches to macroevolution: recent work on species selection and the “effect hypothesis.” *Annual Review of Ecology and Systematics* 26:301–322.
- Housworth, E. A., E. P. Martins, and M. Lynch. 2004. The phylogenetic mixed model. *American Naturalist* 163:84–96.
- Jablonski, D. 1986. Larval ecology and macroevolution in marine invertebrates. *Bulletin of Marine Science* 39:565–587.
- . 1987. Heritability at the species level: analysis of geographic ranges of Cretaceous mollusks. *Science* 238:360–363.
- . 2000. Micro- and macroevolution: scale and hierarchy in evolutionary biology and paleobiology. *Paleobiology* 26(suppl. to no. 4):15–52.
- Manly, B. F. J. 1997. *Randomization, bootstrap and Monte Carlo methods in biology*. Chapman & Hall, London.
- Okasha, S. 2003. The concept of group heritability. *Biology and Philosophy* 18:445–461.
- Qian, H., and R. E. Ricklefs. 2004. Geographical distribution and ecological conservatism of disjunct genera of vascular plants in eastern Asia and eastern North America. *Journal of Ecology* 92: 253–265.
- Ricklefs, R. E., and R. E. Latham. 1992. Intercontinental correlation of geographical ranges suggests stasis in ecological traits of relict genera of temperate perennial herbs. *American Naturalist* 139: 1305–1321.
- Sokal, R. R., and F. J. Rohlf. 1995. *Biometry*. 3rd ed. W. H. Freeman, New York.
- Webb, T. J., and K. J. Gaston. 2003. On the heritability of geographic range sizes. *American Naturalist* 161:553–566.
- Willis, J. C. 1922. *Age and area*. Cambridge University Press, Cambridge.