

# 6

## SIZE RATIOS

The analysis of size ratios of coexisting species has been a major focus in evolutionary ecology for more than three decades. The idea that body size differences follow empirical “rules” and may reflect resource partitioning has arisen several times in the ecological literature. Dyar (1890) described a constant increment of hard part length for insects at each molt, so that body size ratios of successive molts within species would approximate 1.28 (Enders 1976). Both Huxley (1942) and Lack (1947) suggested that body size differences among predators should evolve to reduce the effects of competition for the same foods (Carothers 1986). Brown and Wilson (1956) described several examples of character displacement—species pairs whose body sizes differed more in sympatry than in allopatry. Empirical evidence for character displacement comes from studies of birds (Diamond et al. 1989), lizards (Losos 1990), fish (Schluter and McPhail 1992), mud snails (Fenchel 1975), and other taxa. A number of criteria are necessary to unequivocally establish character displacement in such natural experiments (Grant 1972a; Schluter and McPhail 1992).

Whereas character displacement analyses describe intraspecific variation in multiple assemblages, many size ratio tests compare coexisting species within a single assemblage. This practice can be traced to Hutchinson’s (1959) seminal paper “Homage to Santa Rosalia.” Hutchinson found that body size ratios of several pairs of sympatric bird and mammal species ranged from 1.1 to 1.4. The mean ratio, roughly 1.3, was “tentatively” interpreted as the amount of separation necessary to permit coexistence of species at the same trophic level.

This modest suggestion spawned a vast amount of ecological research. Ecologists measured size overlap in other assemblages, either confirming the “1.3 rule,” describing other axes of niche differentiation that “explained” size ratios less than 1.3, or invoking interference competition between species to explain size ratios greater than 1.3 (reviews in Roth 1981; Simberloff and Boecklen 1981; Simberloff 1983b). Both Hutchinson’s rule and Dyar’s constant were thought to reflect the same underlying pressure for divergence because of limited resources (Maiorana 1978).

Three patterns have been sought in reference to Hutchinson's rule (Simberloff 1983b): (1) minimum size ratios, below which species cannot coexist; (2) constant size ratios, in which the species in a community display an orderly spacing; (3) unusually large ratios in island assemblages, which are thought to be experiencing more severe competition than comparable mainland assemblages.

## ASSUMPTIONS

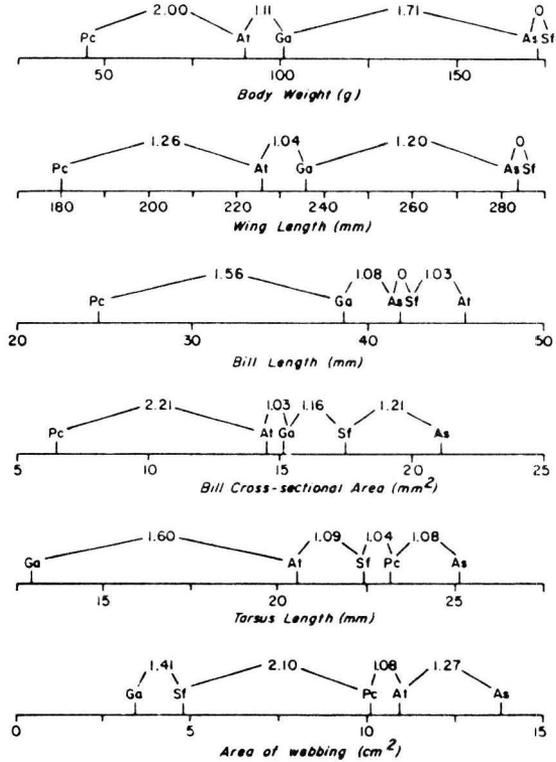
Simplistic analyses of these patterns depend on a number of assumptions:

1. *Morphology is linked to resource consumption, and the appropriate morphological features of the organisms have been measured.* Ratio analyses assume a linear and proportional mapping of morphology onto the resource axis. Although analyses of size ratios are frequently based on body size or trophic structures (bill depth, mandible length), it is by no means obvious what the appropriate morphological variables are that reflect resource use (Wilson 1975). For example, Carothers (1982) found that resource partitioning in a guild of Hawaiian honeycreepers depended on the morphology of the tongue rather than the size or shape of the bill. Moreover, character displacement may occur among distantly related taxa that do not have comparable body parts but nevertheless compete for limiting resources, such as Galápagos ground finches and carpenter bees that use flower nectar as a food resource (Schluter 1986a).

Nonetheless, analyses of body size ratios are usually restricted to small groups of closely related species. A frequently cited example is Ashmole's (1968) study of five species of tropical terns. However, different measurements of body size (body weight, tarsus length) and appendages (bill length, bill cross-sectional area) generated different size ratios and even different orderings of species (Figure 6.1). Hendrickson (1981) found that the choice of metric also affected the outcome of null model tests: birds of the Tres Marias Islands showed evidence of character displacement in wing length but not in bill length, which may be more directly related to resource use (Strong and Simberloff 1981). Because body size and shape are intimately related (Mosi-mann and James 1979), multivariate analyses of morphological displacement may be more appropriate than analyses of single characters such as bill size or body weight. We review these approaches later in this chapter.

2. *The underlying resource spectrum is symmetric.* This assumption underlies mathematical models of both character displacement (Slatkin 1980) and

Figure 6.1. Size ratios and species sequences depend on the particular morphological variable measured. Morphometric measurements of five species of sympatric tropical terns (Ashmole 1968). As = *Anous stolidus*; At = *Anous tenuirostris*; Ga = *Gygis alba*; Pc = *Procelsterna cerulea*; Sf = *Sterna fuscata*. From Wiens (1982). Reproduced with the permission of the Finnish Zoological and Botanical Publishing Board.



niche shift (MacArthur and Levins 1967), but it has rarely been verified in nature and may not be true (Schluter and Grant 1984). Symmetry of the resource spectrum affects the amount of divergence (Slatkin 1980) and hence the pattern of observed size ratios in a group of competitors.

3. *The environment is stable and the system has reached an ecological and evolutionary equilibrium.* If the environment is variable, divergence of competitors may not be pronounced, because specialization on a particular part of the resource spectrum will no longer be favored by selection (Gotelli and Bossert 1991) or promoted by extinction of overlapping species (Turelli 1978b). More importantly, if the system has not reached an equilibrium, then species composition will change through time, and the derived morphological “patterns” may be more apparent than real (Wiens 1981).

4. *Competition occurs only among adult organisms.* Size ratio analyses of adult animals ignore ontogenetic shifts in body size and resource use (Werner and Gilliam 1984), and the potential for different age classes to function as

ecological species (Polis 1984). The ontogenetic perspective was important in early analyses of Dyar's constant (Enders 1976) but has been neglected in most community studies. Interestingly, Hutchinson (1959) appreciated the importance of body size shifts during ontogeny. In addition to the "1.3 rule," he predicted that the larger of a pair of species of coexisting corixid beetles would breed earlier in the season to have enough time for growth to achieve its size advantage.

5. *Sexual dimorphism in body size is not important.* When species are sexually dimorphic in body size, differences between the sexes have been either averaged (e.g., Schoener 1984) or ignored, by restricting the analysis to a single sex (e.g., Hines 1982). Both approaches are unsatisfactory. Averaging body sizes may create an imaginary phenotype that does not exist in nature, whereas analyzing one sex ignores overlap and resource use by half the population. If both sexes are treated as distinct morphospecies (e.g., Dayan, Simberloff, et al. 1989, 1990), body size ratios are difficult to interpret, because overlap within a species may not be statistically or biologically equivalent to overlap between species (see Epilogue). Sexual dimorphism in body size is typically correlated with dimorphism in trophic appendages and feeding ecology (Shine 1989), and clearly needs to be considered in studies of resource partitioning. Unfortunately, there has been little theoretical research in this area (Slatkin 1984), and quantitative models for character displacement and sexual dimorphism in a suite of species are still needed (Dayan et al. 1990).

6. *Abundances of species are approximately equal.* The intensity of competition depends not only on the amount of overlap in resource use, but also on the densities of the two competing species. If one of the species is rare, it may overlap completely with its competitor but not contribute much to resource depletion. The relative abundance of competing species consequently is important in determining the degree of divergence of populations (Slatkin 1980). Some authors have restricted their analyses to "common" players (e.g., Bowers and Brown 1982) on the grounds that these "core" species (Hanski 1982a) are equilibrational and are more likely to be experiencing competition. At a regional level, Hanski (1982b) found some evidence that bumblebee proboscis lengths of core species were nonrandomly spaced, although this pattern was not confirmed by detailed analyses of local assemblages (Ranta 1982).

Given this list of restrictive and often unrealistic assumptions, why were size ratio analyses so popular and uncritically accepted in ecology? Wiens (1982) suggested several reasons. First, some communities seemed to conform to the

1.3 rule, and these examples reinforced the predictions of simple models of limiting similarity (MacArthur and Levins 1967). Second, size ratio studies were guided by an adaptationist view of organisms (Gould and Lewontin 1979)—traits such as bill length or body size represented adaptive “solutions” to ecological problems (limiting food resources). Third, it is far easier to measure morphological traits from museum collections than it is to conduct field studies to quantify food habits, dietary overlap, and ecological interactions (Wiens 1991a). Finally, size ratio analyses reflected a view that ecological systems were ordered and deterministic and that the underlying mechanisms would be revealed in simple rules that governed the assembly of communities.

## MODELS OF BODY SIZE DIVERGENCE

In parallel with empirical studies of character displacement, theoretical models also indicated the potential for divergence of competitors. Three basic mechanisms, two evolutionary and one ecological, could cause divergence of species in sympatry. First, evolutionary divergence of body sizes might be favored if it prevented hybridization (Bossert 1963). This mechanism has received little attention from animal ecologists, although it may be important in the divergence of morphology or phenology of closely related plant species (Levin 1971; see Chapter 5). Second, coevolutionary divergence might occur in sympatry through selection against intermediate phenotypes (Bulmer 1974). This evolutionary mechanism has been invoked for cases in which a species shows intraspecific variation in body size that is related to the presence or absence of competitors (Grant 1972a).

Finally, divergence may occur in the absence of evolutionary change through purely ecological mechanisms. Models of limiting similarity (MacArthur and Levins 1967) depict species resource utilization spectra as fixed, with no potential for evolutionary change. Species that overlap too much in resource use (and presumably in body size) will be driven to extinction; species that are widely separated along the resource axis survive, and the assemblage can subsequently be invaded by species with intermediate phenotypes. Through time, size assortment via colonization and extinction may lead to a community with a constant spacing of body sizes (Abrams 1986).

For both the evolutionary and the ecological models, divergence of competitors is by no means guaranteed, even in the face of persistent competition. Intrinsic factors, such as the amount of genetic variation underlying the trait (Slatkin 1980), and extrinsic factors, such as the variability of the environment

(Turelli 1978b), affect the degree of displacement. The terminology has been confused in this literature, as character displacement has been used to describe both the pattern and the mechanism of divergence. Strong et al. (1979) defined “community-wide character displacement” as any pattern of overdispersed body sizes, generated by either ecological or evolutionary mechanisms. In contrast, Case and Sidell (1983) developed null models to distinguish between “size-assortment” (ecological limits to similarity) and “size adjustment” (co-evolutionary character displacement). Rummel and Roughgarden (1983) also distinguished between invasion-structured and coevolution-structured competitive communities.

## **ALTERNATIVES AND ARTIFACTS**

In the early size ratio literature, competition was the only framework used to interpret patterns, and there was a distressing ignorance of alternative mechanisms. Yet, competitive interactions are only one of many forces that control body size. Selective responses to predation and the presence of enemy-free space may control the pattern of body sizes in an assemblage (Jeffries and Lawton 1984). Alternatively, body sizes of coexisting species may be phylogenetically constrained (Elgar and Harvey 1987) for reasons that have nothing to do with current or past biotic interactions. Finally, the ratios themselves may be mathematical artifacts that merely reflect the underlying distribution of body sizes in an assemblage (Tonkyn and Cole 1986; Eadie et al. 1987). For example, size ratios of tricycle wheels, iron skillet, and musical recorders also follow Hutchinson’s rule (Horn and May 1977), suggesting that the size ratios observed in natural communities may have no biological significance. Before ecological mechanisms are attributed to body size ratios, their mathematical, and hence nonbiological, properties need to be understood.

## **STATISTICAL PROPERTIES OF RATIOS**

The expected ratio in a large assemblage depends on the underlying distribution of body sizes and on the end points of possible body sizes (Tonkyn and Cole 1986). In ecological models of limiting similarity, this distribution represents the body sizes of existing phenotypes that could colonize a community. In evolutionary models of character displacement, this distribution represents the probability that a particular body size will evolve. Tonkyn and Cole (1986) assumed that the distribution of available body sizes was either uniform or

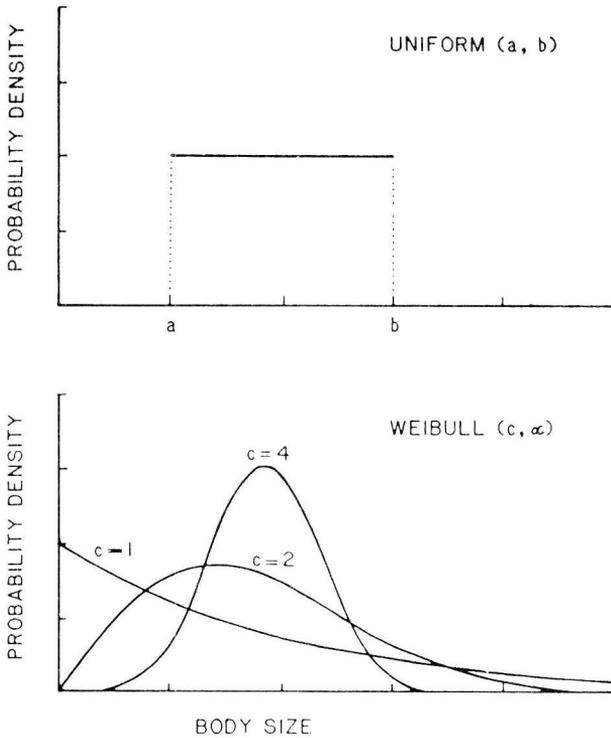


Figure 6.2. Hypothetical body size distributions ultimately determine the distribution of overlap ratios in null assemblages. From Tonkyn, D. W., and B. J. Cole. 1986. The statistical analysis of size ratios. *American Naturalist* 128:66–81. Copyright © 1986 by The University of Chicago. Reprinted by permission of the publisher.

unimodal, as fit by a Weibull distribution. The Weibull and the uniform captured the range of biologically realistic distributions, including many curves that could be characterized as log normal (Figure 6.2).

Regardless of the shape of the body size distribution, there were two general properties of size ratios for the theoretical assemblage. First, the most common size ratio for a pair of adjacent species was the minimum ratio of 1.0. The frequency of larger ratios in the assemblage decreased monotonically from this peak, and the shape of the curve was concave upward (Figure 6.3). Second, the more species in the assemblage, the smaller the expected body size ratio. This second property probably explains, in part, why ratios for island communities are often larger than those for corresponding mainland communities (Simberloff 1983b). By chance alone, we expect an island assemblage with few species to have larger size ratios

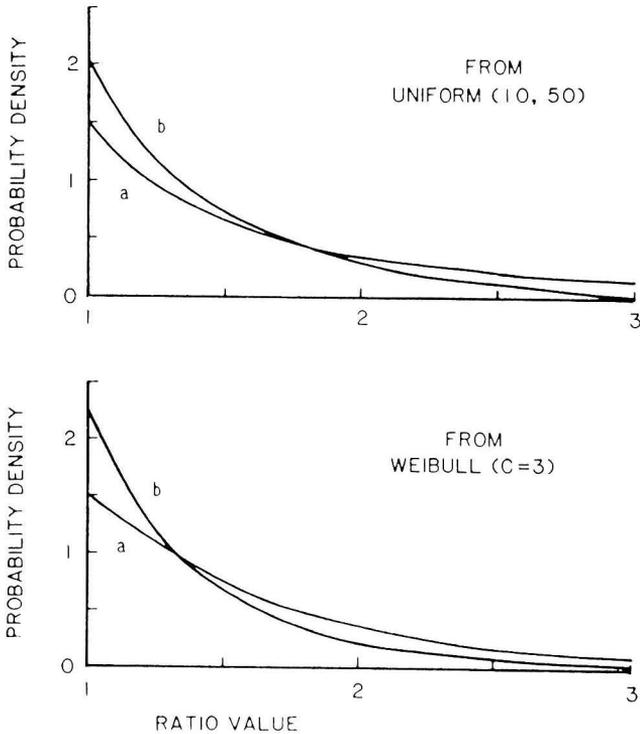


Figure 6.3. Expected ratio distributions for randomly assembled communities. The shape of these curves does not depend greatly on the body size distributions in Figure 6.2. a = two-species guilds; b = four-species guilds. From Tonkyn, D. W., and B. J. Cole. 1986. The statistical analysis of size ratios. *American Naturalist* 128:66–81. Copyright © 1986 by The University of Chicago. Reprinted by permission of the publisher.

than a corresponding mainland community, for the same reasons that we expect island species/genus ( $S/G$ ) ratios to be lower (see Chapter 1).

Because these predictions hold for any set of species that are drawn randomly, they constitute a simple null model for the distribution of ratios in a large assemblage. In contrast, a community that is competitively assembled should show a unimodal distribution of ratios, in which the mode represents the limit to similarity (Tonkyn and Cole 1986). Expected ratios would be the same for large communities as for small and perhaps would even increase in large communities because of diffuse competition (Pianka 1974).

Tonkyn and Cole (1986) examined two large data sets of size ratios, one compiled by Schoener (1965) for guilds of sympatric, congeneric bird species

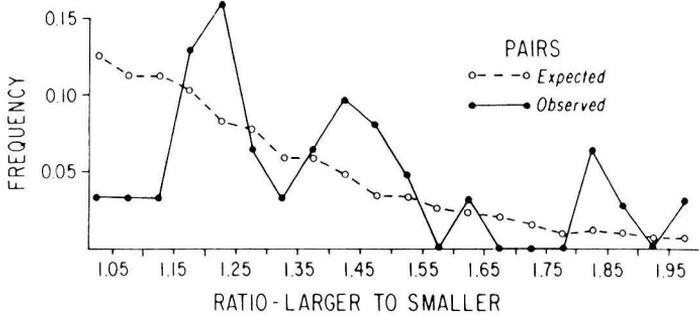


Figure 6.4. Expected and observed body size ratios for pairs of sympatric bird-eating hawks. Note the lack of very small size ratios in the observed assemblages compared to the null model. See also Figure 6.3. From Schoener, T. W. Size differences among sympatric, bird-eating hawks: a worldwide survey. In: *Ecological Communities: Conceptual Issues and the Evidence*. D. R. Strong, Jr., D. Simberloff, L. G. Abele, and A. B. Thistle (eds). Copyright © 1984 by Princeton University Press. Reprinted by permission of Princeton University Press.

(410 ratios) and one by T. Tomasi (unpublished) for guilds of sympatric, insectivorous bats (61 ratios). For both data sets, the distribution of observed ratios was monotonically decreasing and concave upward, in good agreement with the null hypothesis. For Schoener's (1965) data, mean ratios also decreased with increasing guild size, as predicted by the null model. These results contrast with the appearance of the histogram of ratios for sympatric bird-eating hawks, also compiled by Schoener (1984). The histogram of the hawk ratios, while not unimodal, was certainly not decreasing monotonically (Figure 6.4). Schoener's (1984) Monte Carlo procedures, described later, also confirmed that these ratios deviated significantly from null expectations.

## THE 1.3 RULE AS AN ARTIFACT

Roth (1981) adapted conventional statistical tests for evaluating whether ratios are significantly different from any expected value. Even without such tests, it is clear that some communities show no constancy in ratios or tendency toward a particular ratio (Wiens and Rotenberry 1981a). Nevertheless, the 1.3 rule has received such widespread attention that it is worth considering the hypothesis in some detail.

Eadie et al. (1987) offered the most satisfying general explanation for why ratios should tend toward this magic number. They showed that ratios of

approximately 1.3 are expected if the underlying distribution of body sizes is log normal, and if the variance about that distribution is relatively small. Both assumptions have empirical support. For 35 of 41 comparisons, body size data could not be distinguished from a log normal distribution (Van Valen 1973). The log normal distribution of body sizes could, itself, be a consequence of competition for food (Schoener 1965). However, the log normal characterizes so many biological (and nonbiological) systems (Koch 1966), that this seems unlikely.

The variance of body sizes in most log normal distributions is small, usually less than 1.0. Again, competition is not the most likely explanation. Instead, Eadie et al. (1987) pointed out that a log normal variance in body size of 1.0 implies a 400-fold range of body sizes on an arithmetic scale. Given that size ratios are usually calculated for small sets of closely related species, a small variance in body size is assured. If there is any tendency for body size ratios to conform to a value of 1.3, the analysis of Eadie et al. (1987) provides the most likely explanation, as well as accounting for 1.3 ratios in collections of inanimate objects (Horn and May 1977).

## NULL MODEL APPROACHES

Although the 1.3 rule has not stood up to detailed analysis, the hypothesis that body size differences are constrained by competition is still viable and deserves testing. Null models can be used to generate expected body size ratios in the absence of competitive interactions. Three different strategies have emerged for testing body size ratios. The first approach, pioneered by Simberloff and Boecklen (1981), adapts conventional statistical tests to examine patterns of regularity or unusual minima in size ratios. A single body size value for each species is used, and intraspecific variation in body size is not considered. The tests apply to a single assemblage of co-occurring species.

The second approach, pioneered by Strong et al. (1979), considers intraspecific variation in body size data among a set of communities, usually on islands. These data form the basis for a Monte Carlo simulation in which different species populations are sampled to generate null communities and size ratios that would be expected in the absence of competition. Intraspecific variation in body size among populations is preserved.

The third approach, pioneered by Schoener (1984) and by Hopf and Brown (1986), tests size spacing patterns in multiple communities containing the same guilds and usually some of the same species. These tests are more powerful than analyses of single communities, although they may be complicated by

variation in species number and the repeated occurrence of certain species in multiple assemblages (James H. Brown, personal communication).

All three approaches are controversial. The papers by Simberloff and Boecklen (1981) and by Strong et al. (1979), in particular, provoked a number of responses. After reviewing these exchanges, we consider null models of morphometrics that rely on multivariate analyses of size and shape of organisms. Finally, we review three case studies that were published in the wake of the original controversy. These studies incorporate additional data on geographic variation in morphology (Dayan, Tchernov, et al. 1989), experimental field manipulations (Juliano and Lawton 1990b), and field measurements of resource use (Schluter and Grant 1984) to provide insight into patterns of body size overlap.

## RATIO TESTS FOR SINGLE ASSEMBLAGES

Simberloff and Boecklen (1981) systematically tested literature claims of unusual ratio constancy or minima. Their tests arranged the logarithms of body masses of each species as points along a line, with the largest and smallest species in the assemblage representing the end points of the line. A hypothesis of constant size ratios on a linear scale means equal spacing on a logarithmic scale. For an assemblage of  $n + 1$  species, there are  $n$  line segments,  $n - 1$  interior points, and  $n(n - 1)/2$  ratios of line segments that can be formed. If size ratios are unusually constant, then the segments will be very similar in length, so that ratios of segments would be unusually small. The Barton and David (1956) test gives the probability that the ratio between two specified segments (small segment/large segment) is smaller than observed. Simberloff and Boecklen (1981) tested three segment ratios for each assemblage of more than three species. Examining more than one ratio ensured that the results were typical for the assemblage but introduced a problem of nonindependence. For assemblages with three species, there are only two segments, and their ratio defines the placement of the interior species.

The Barton and David (1956) test does not handle ties (which give line segments of length zero), although identical body sizes in an assemblage should be counted as evidence against the hypothesis of ratio constancy. For cases of ties, Simberloff and Boecklen (1981) substituted other ratios from the assemblage. The Poole and Rathcke (1979) test, described in Chapter 5, was more appropriate for data with many ties, although it was less powerful than the Barton and David (1956) test. The Barton and David (1956) test was also superior to the ratio tests of constancy proposed by Roth (1981), which require a minimum of six species for statistical power (Simberloff 1983b).

For claims of ratio minima, Simberloff and Boecklen (1981) adapted the Irwin (1955) test, which gives the probability that the minimum segment for the assemblage is smaller than observed. Both the Barton and David (1956) and the Irwin (1955) tests are appropriate for analyzing body size distributions in a single assemblage. For multiple assemblages, Simberloff and Boecklen (1981) tested the probability of obtaining an observed minimum segment for a set of sites (Pielou and Arnason 1966). All of these statistics are tests for evolutionary character displacement, because the null hypothesis is that any evolutionary arrangement of body sizes within the observed limits is possible.

Simberloff and Boecklen (1981) applied these tests to a variety of published studies claiming constant or minimum size ratios. They also devised statistical tests for related claims about ratios. For example, Schoener (1965) suggested that size differences must be more extreme when food is rare, so that ratios would increase, rather than remain constant for large-bodied species. Similarly, Oksanen et al. (1979) argued that ratios much larger than 1.3 for waterfowl assemblages were the result of interference competition between species (but see Nudds et al. 1981). Simberloff and Boecklen (1981) tailored several null models to investigate these and other ratio claims.

Simberloff and Boecklen (1981) examined three segment ratios for each assemblage and defined a “significant effect” as  $p < 0.05$  for more than one ratio. For 21 literature claims of ratio constancy, four were sustained at the 0.05 level and 11 were sustained at a  $p$  value of 0.30. By chance, one would have expected to find one significant claim at the 0.05 level and six at the 0.30 level. A few assemblages were nonrandom in the opposite direction: too many small ratios to be explained by chance. Of 18 claims of large minimum size ratios, only one was sustained at the 0.05 level and 13 were sustained at a  $p$  value of 0.30. Simberloff and Boecklen (1981) concluded that “evidence presented to date that sizes are competitively determined is weak, and that in particular the ‘1.3 rule’ was probably always a red herring and has certainly outlived its usefulness to evolutionary ecologists.”

## CRITICISMS OF RATIO TESTS

Losos et al. (1989) disagreed with Simberloff and Boecklen’s (1981) interpretation of these results. They reanalyzed a subset of the studies that included data for more than one assemblage. Combining probabilities within these studies, Losos et al. (1989) found that 47% of ratio claims of constancy were supported at the  $p = 0.05$  level and more than 70% at the  $p = 0.30$  level. They concluded that there was substantial evidence for ratio constancy and unusual

size minima, although the patterns may not be strong in any single assemblage. Whether or not published studies represent a random subset of natural assemblages, the evidence for non-random size ratios is certainly much weaker than claimed in most of the original papers.

Other objections to the Barton and David (1956) test have been statistical. The most common complaint has been that the assumption of a (log) uniform body size distribution biases the test against rejecting the null hypothesis (Case et al. 1983; Colwell and Winkler 1984; Schoener 1984). However, when body sizes were drawn randomly from a more realistic log normal distribution, the Barton and David (1956) test was slightly liberal for small assemblages and slightly conservative for large assemblages; overall differences were trivial and the performance was consistent and unbiased (Boecklen and NeSmith 1985). The Irwin (1955) test also performed well with a log normal distribution (Losos et al. 1989). These results are consistent with Tonkyn and Cole's (1986) finding that the expected distribution of ratios is insensitive to the distribution of body sizes. For a log normal distribution of body sizes, Sinclair et al. (1985a,b) derived test statistics for constant ratios and large minimum ratios. Given that most ratio tests are conducted on small assemblages of similar species (Eadie et al. 1987), the log uniform probably performs just as well as the log normal distribution and is unlikely to affect the outcome of the test.

A more serious problem is that ratio tests may not be powerful enough to detect character displacement when it is happening. Losos et al. (1989) explored the power of the Irwin (1955) test by randomly assembling hypothetical communities with minimum threshold ratios. For communities of three species, the Irwin (1955) test almost never rejected the null hypothesis, even when the threshold ratio was large ( $>1.3$ ). For communities of six species, the results were better, and the probability of falsely accepting the null hypothesis was never greater than 0.25, at least for small variances. The power of the Irwin (1955) test increased with large threshold ratios, large numbers of species, and small variances in body size (Losos et al. 1989). For multiple assemblages, the Barton and David (1956) test may also be less powerful than Monte Carlo simulations that are tailored to particular data sets (Schoener 1984).

Still another objection to the Barton and David (1956) test is that it relies on particular ratios for an assemblage that are arbitrary and not independent of one another, rather than summarizing dispersion of body sizes with a single "community-wide" parameter (Hopf and Brown 1986; Pleasants 1990). This criticism does not apply to the Irwin (1955) test, because the minimum ratio is a property of the entire assemblage. On the other hand, aggregate statistics do not always detect character displacement when present (Arita 1993). Although a single index is statistically desirable, it still may be appropriate to examine

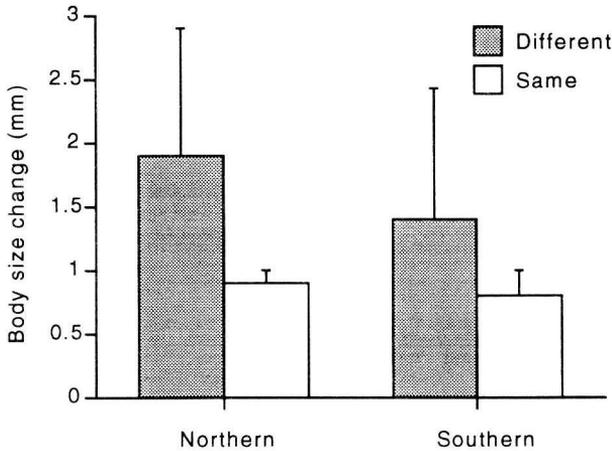


Figure 6.5. Phylogenetic evidence for character displacement in *Anolis* lizards of the Northern and Southern Lesser Antilles. Different = taxon occupied an island with a different number of species than the island occupied by its most immediate ancestor. Same = taxon occupied an island with the same number of species as the island occupied by its most immediate ancestor. Note that evolutionary change in body size is greatest when species number is different on islands with ancestral versus descendent taxa. Data from Losos (1990).

particular ratios within an assemblage when searching for ratio constancy. More recent analyses by Pleasants (1994) and Williams (1995) suggest that community-wide statistics, such as the variance of distances between body sizes (Poole and Rathcke 1979; see Chapter 5), are indeed most powerful for detecting displacement.

Finally, Tonkyn and Cole (1986) objected that ratio tests used the largest and smallest species in the assemblage to set the end points of possible body sizes. The effect of this procedure, if any, will be to bias the test toward rejecting the null hypothesis. If the actual range of possible body sizes is much greater than observed, then a pattern that appears evenly spaced by using species end points may be random or even clumped using true evolutionary end points (Simberloff and Boecklen 1981).

Using two of the species from a small assemblage to set the end points loses information, and it would be preferable to set the end points by some other criterion. An important issue in the analysis of body size patterns is to understand the evolutionary and ecological “boundaries” within which character displacement can occur. We think this is an important area of research that deserves additional attention.

Phylogeny is probably the key to understanding limits of body size evolution, and an exemplary analysis by Losos (1990) illustrates the power of incorporating phylogenetic information into null models. Islands of the Lesser Antilles support one or two species of *Anolis* lizard. On two-species islands, sympatric populations usually show substantial size differences, whereas allopatric populations on single-species islands are usually of intermediate size. A phylogenetic reconstruction (Huey and Bennett 1987) suggested that substantial evolutionary change in body size was rare and may have only occurred once within the *Anolis* clade. Nevertheless, size change was significantly greater when a descendant occurred on a two-species island than on a one-species island, supporting a model of evolutionary size adjustment (Figure 6.5). Size assortment was also revealed by a null model in which communities were randomly assembled by drawing populations from independent clades. This is essentially the same as the Strong et al. (1979) procedure, but it retains historical constraints and does not treat all island populations as equiprobable and independent of one another. Losos's (1990) approach is considerably more sophisticated and powerful than most null model tests of character displacement. However, like all comparative studies (Harvey and Pagel 1991), its application will be limited to assemblages for which reliable phylogenies exist.

## MONTE CARLO TESTS

Strong et al. (1979) used Monte Carlo simulations to investigate claims of character displacement for three island avifaunas: (1) the Tres Marias Islands of western Mexico, (2) the Channel Islands of southern California, and (3) the Galápagos Islands. They compiled morphological and distributional data for the islands and for adjacent mainland areas. Mainland source pool species were limited to those that might occur in the range of habitats present in each archipelago.

For the first two data sets, Strong et al. (1979) created 100 assemblages of species from each taxonomic subfamily present on the mainland by drawing randomly the same number of species that were present in the archipelago. However, ratios for individual islands were not calculated. Instead, they averaged ratios for each morphological character in the null assemblage and treated this as the expected value in the absence of character displacement. Next, they counted the number of times the observed ratios in the archipelago exceeded the expected ratios in the null assemblages. If bill and wing length were influenced by character displacement, observed ratios should frequently have exceeded this expectation. By a binomial test, an excess of large ratios was not

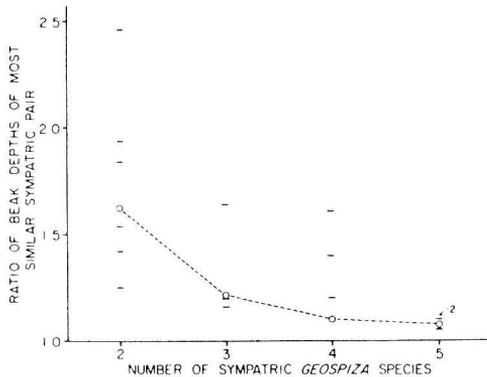


Figure 6.6. Observed (bars) and expected (open circles) minimum bill size ratios for *Geospiza* finches on the Galápagos Islands. Expected ratios were calculated by randomly sampling species and then populations. Note the lack of ratios consistently above or below the expectation. From Strong et al. (1979), with permission.

present in either assemblage, and Strong et al. (1979) concluded “a general trend of character displacement cannot be inferred from these traits.”

For the Galápagos finches, Strong et al. (1979) used different procedures. They used Lack’s (1947) published data on bill depth and bill length of species of *Geospiza* (ground finches) and of *Camarhynchus* (tree finches) and *Certhidea* (warbler finches) to test for character displacement. Because the Galápagos archipelago is 1,000 km offshore, there is no obvious mainland source pool for comparison with this isolated fauna. In addition, individual islands have morphologically distinct finch populations.

Strong et al. (1979) asked whether the particular combination of morphologically distinct populations on each island exhibited character displacement relative to random combinations of species and populations from throughout the archipelago. For an  $n$ -species island, they first chose  $n$  species randomly and equiprobably. Then, for each species, they randomly chose a subpopulation from the archipelago. Thus, for each island, both species identity and subpopulations were randomized, but species number was held constant. Compared to 100 random draws, there were never significantly more ratios above than below the null expectation (Figure 6.6).

Strong et al (1979) also searched for character displacement in bill shape. For each species population, they plotted bill shape (bill depth/length) as a function of bill size (length) and found that island populations for each species formed a well-defined polygon (Figure 6.7). In these simulations, Strong et al. (1979) retained species identities but randomly chose the particular island

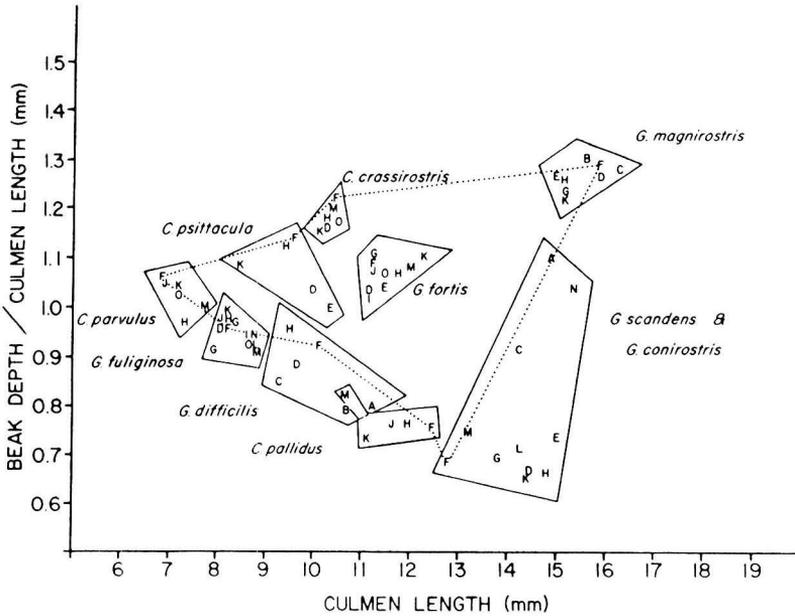


Figure 6.7. Bill shape as a function of bill size for populations of *Geospiza* finches on the Galápagos Islands. Solid polygons enclose populations of the same species. The dotted line connects coexisting species. For a pattern of overdispersion, the Euclidian distance separating these species would be unusually large. By this test, most communities were randomly spaced, or showed slight convergence. From Strong et al. (1979), with permission.

population from the archipelago list. Next, they calculated the average Euclidian distance between species of a given island. Character displacement would be expressed as unusually large Euclidian distances compared to those in randomly assembled communities.

These analyses again yielded no evidence of character displacement, although one of the islands (San Cristóbal) showed evidence of character convergence: observed Euclidian distances between species were unusually small. Analyses of the rank sum of size and shape variables also suggested a slight tendency toward character convergence, although the patterns were weak and variable from one island to the next. Strong et al. (1979) concluded, "Our approach emphasizes apparent randomness in actual communities, and that species often persist together independently of their morphological characteristics. We suggest that apparent randomness would account for a substantial proportion of variation in many real ecological communities."

## CRITICISMS OF MONTE CARLO TESTS

The study by Strong et al. (1979) was important for three reasons. First, it introduced a novel method of analysis for character displacement studies (see also Gatz 1979). Second, the results indicated little evidence of character displacement in the Galápagos finches, one of the classic textbook examples. Finally, Strong et al. (1979) argued for the “logical primacy” of null hypotheses as a method for analyzing community patterns in the absence of experimental tests (see Chapter 1).

Several authors raised biological and statistical objections to the procedures and conclusions of Strong et al. (1979). Grant and Abbott (1980) leveled five major criticisms. First, they objected to analyses at the family level (Tres Marias and Channel Islands avifaunas) because these taxonomic groupings were not equivalent to ecological guilds. Second, they argued that source pools for the Tres Marias and Channel Islands should have been based on species lists of mainland areas of approximately equal size, rather than on species lists from an entire country (see Epilogue). Third, Grant and Abbott (1980) claimed that the null model assumption of equiprobable species dispersal was unrealistic. Fourth, they argued that it was somewhat circular to use the observed data to generate null distributions that were then compared to the observed data. Finally, they disagreed with the statistical tests of Strong et al. (1979), in part because the binomial did not measure the magnitude of the differences between observed and expected ratios, and because the null model simulations tacitly assumed a model of “size assortment” (Case and Sidell 1983), rather than “size adjustment” (i.e., evolutionary character displacement).

Hendrickson (1981) corrected some errors in the source pool and island lists of Strong et al. (1979). He noted that the mean was a biased measure of central tendency, and that it was more appropriate to compare observed ratios with the median of the ratios for simulated communities. Using the same null model as Strong et al. (1979), his reanalysis revealed significant patterns consistent with the character displacement hypothesis for three variables: wing length of the Tres Marias birds and bill depth and length of Galápagos *Geospiza*. Strong and Simberloff (1981) responded that two or three significant results out of a battery of more than 15 statistical tests was hardly overwhelming evidence, and that means and medians of ratios for simulated communities were quite similar. They also suggested that if character displacement were important for the Tres Marias birds, its effects should have been manifest in some measure of bill morphology rather than in wing length.

Case and Sidell (1983) examined the performance of the Strong et al. (1979) procedure for a hypothetical community that was structured by competition.

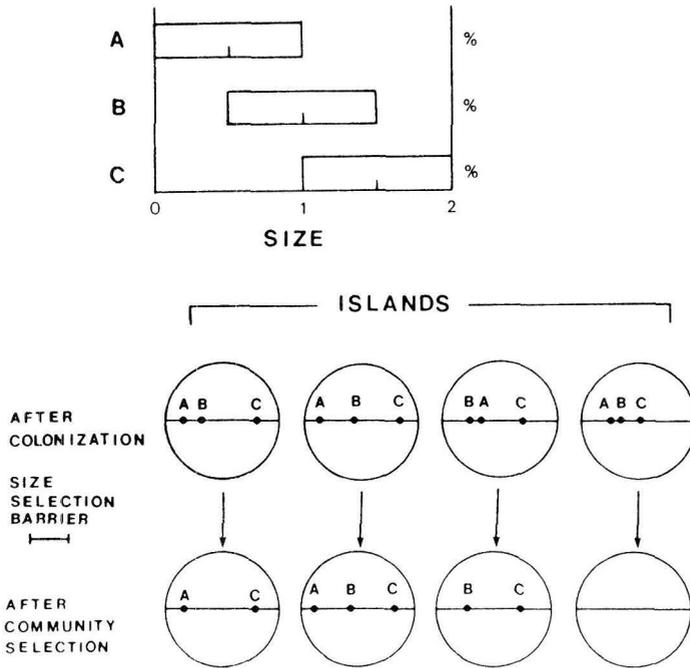


Figure 6.8. Community-wide selection following island colonization. The top figure gives body size distributions of three hypothetical colonizing species. Body sizes are drawn randomly from each distribution to seed the islands. If pairs of species are closer together than the size selection barrier, one of the pair becomes extinct. The resulting community has been randomly assembled with a limit to similarity. From Case and Sidell (1983), with permission.

They started with a three-species mainland community in which body size distributions were uniform and overlapped somewhat between species (median body size ranks:  $A < B < C$ ). Next, a series of islands was colonized from this source pool by drawing a colonist randomly from each of the three size distributions. These three-species assemblages were then subjected to a size difference barrier,  $s$ . If any pair of species was closer in size than  $s$ , the interior of the two went extinct, and the assemblage collapsed to two species. This scenario of “size-assortment” corresponds to nonevolutionary models of limiting similarity (MacArthur and Levins 1967), in which communities are sorted through colonization and extinction (Figure 6.8).

If  $s$  was zero, then all three species coexisted, and the observed size distribution matched that in the source pool. As  $s$  increased, the system was dominated by two-species islands. The AC combination occurred more frequently than AB

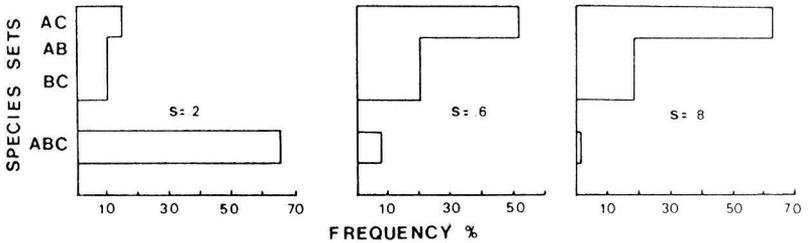


Figure 6.9. Expected frequency of different species combinations as a function of the size selection barrier ( $s$ ). As the limit to similarity increases, species pair AC comes to dominate the archipelago, because species A and C are most dissimilar in size in the initial source pool. From Case and Sidell (1983), with permission.

or BC, because species A and C were most dissimilar in size in the original mainland pool (Figure 6.9). Finally, the observed island assemblage (which was competitively structured) was randomized and tested with the Strong et al. (1979) procedure.

Average size differences between species in the observed assemblage were always greater than in the null assemblage, which is consistent with a pattern of size assortment. However, the expectation from the null model closely tracked the observed assemblage; the two values were usually within one standard deviation of each other. Thus, the Strong et al. (1979) test may lack power, because the null community too closely reflected the properties of the observed assemblage. Results were similar for models in which colonizing species were sampled equiprobably versus proportionally to their frequency of occurrence, in which the body size distributions were Gaussian versus uniform, and in which species number was low versus high.

The match between the observed and expected size differences seemed to reflect the fact that different species pairs behaved differently in Case and Sidell's (1983) model. Size assortment caused a divergence in species pairs AB and BC because these species could not coexist if they were too similar in size. However, species pair AC actually showed a convergence with size assortment—the observed size difference was smaller than expected. This unexpected result was due to interactions with species B. If species A and C were too dissimilar in size, then species B persisted. The only cases in which A and C coexisted but B did not were those in which the body size differences, by chance, were less than in the source pool fauna. Because the Strong et al. (1979) test relied on average size differences (or ratios) between similar (AB or BC) and dissimilar (AC) species pairs, these patterns would be obscured, and the null model expectation would resemble the observed community.

Using a more complex evolutionary model, Colwell and Winkler (1984) also had trouble detecting significant patterns of size displacement with randomization tests. They used models of stochastic phylogenies (Raup et al. 1973) with phenotypic characters (bill length and depth) that also evolved stochastically at each speciation event. Size barriers to coexistence on islands were established, and communities were randomized according to typical null model protocols. Euclidian distances in morphological space were calculated for all assemblages, before and after competition, and before and after randomization. Three potential biases in null model tests emerged from a comparison of these hypothetical assemblages:

1. *The Narcissus effect.* Sampling from a postcompetition pool underestimated the role of competition, because its effect was already incorporated in the source pool.
2. *The Icarus effect.* Correlations between vagility and morphology sometimes obscured the effects of competition in morphological comparisons of mainland and island biotas.
3. *The J. P. Morgan effect.* The weaker the taxonomic constraint on sampling, the harder it became to detect competition.

As in Case and Sidell's (1983) analyses, the direction of the null model comparison for observed and expected distances was correct (for first nearest neighbors in morphological space): observed communities were morphologically overdispersed compared to randomized assemblages. However, the power of the randomization was weak, compared to this evolutionary model of community assembly.

In spite of these results, Monte Carlo methods may still be valid if the patterns are strong. For example, the average size ratios for coexisting beetle species in the genus *Pterostichus* were larger in undisturbed habitats than ratios in randomly assembled communities (Brandl and Topp 1985), whereas a similar analysis yielded random or small ratios for tiger beetle assemblages (Pearson and Juliano 1991). Both studies showed that the results depended greatly on the particular habitat under consideration. We suspect that habitat affinities and source pool construction are likely to be at least as important as null model structure in determining the results of ratio tests (see Epilogue).

## TESTS FOR SIZE ADJUSTMENT AND SIZE ASSORTMENT

Case and Sidell (1983) proposed two tests for community-wide character displacement, based on an analysis of size differences of particular species

pairs in an archipelago. Their test for size assortment ranked all possible pairs, triplets, quintuplets, etc., of species by the average difference in body size within each combination. The median of this ranking was the null expectation for the size difference of a randomly chosen species combination. Next, the number of species combinations above and below this median was tallied. If size assortment is important, there should be an excess of species combinations with large body size differences and a deficit of species combinations with small body size differences.

To test for size adjustment, Case and Sidell (1983) created random assemblages following the Strong et al. (1979) procedure, using either weighted or unweighted colonization probabilities. After randomization, they calculated average (or minimum) size ratios for the observed and randomized communities. The difference between the observed ( $O$ ) and expected ( $E$ ) size ratio for each assemblage was measured as a standardized deviate  $((O - E)/E)$ . Next, the assemblages were ranked in order from those with the largest average body size differences to those with the smallest. Finally, the standardized deviate for each assemblage was plotted against its rank.

If the assemblages in an archipelago had been structured by evolutionary size adjustment, species combinations with dissimilar body sizes (low ranking) would show little or no divergence, whereas those combinations with similar body sizes (high ranking) would show substantial divergence. Therefore, size adjustment should generate a positive relationship between the standardized deviate for each community and its rank of body size difference.

Case and Sidell (1983) applied these tests in yet another reanalysis of the Galápagos finches, after first correcting some errors in the Strong et al. (1979) and Hendrickson (1981) data sets. For both *Geospiza* and *Camarhynchus*, Case and Sidell (1983) found a significant pattern of size assortment but no evidence of size adjustment. Size assortment was also evident in four feeding guilds (frugivores, gleaning insectivores, flycatchers, and nectarivores) of the West Indian avifauna (Case et al. 1983). Null expectations were generated both by randomization of the archipelago populations and by sampling from a mainland source pool of Colombian birds in these same guilds.

Case (1983b) also applied these tests to assemblages of *Cnemidophorus* lizards in three regions of North American desert, and found evidence for size assortment of coexisting *Cnemidophorus* in the Trans-Pecos region. However, Case's (1983b) analysis excluded sites that supported the common *C. exsanguis*-*C. gularis* species pair. When these sites are included in the analysis, there is no pattern of size assortment (Schall 1993).

A serious shortcoming of the Case and Sidell (1983) tests is the assumption of equiprobable dispersal by species in the absence of competitive effects. Two

species of dissimilar body size might not be competing with each other, but if they were both widely distributed, they would co-occur frequently, and this pattern would yield a significant result. Case and Sidell's (1983) test for size assortment further assumes that all islands are equally inhabitable for all species. The test for size adjustment does not make this assumption, because it holds observed species number constant in the simulations. Although Case and Sidell (1983) cautioned against weighting colonization probabilities by observed frequencies of occurrence, we think it is prudent to carry out weighted and unweighted analyses. If both give similar results, then the patterns are robust to assumptions about occurrence frequency. On the other hand, differing results would suggest that patterns of character displacement can be accounted for by different frequencies of occurrence of species on islands rather than by competitive structuring (see also Wilson 1995).

## RATIO TESTS FOR MULTIPLE ASSEMBLAGES

Whereas ratio tests for single assemblages may be weak in statistical power, multiple assemblages may reveal consistent ratio patterns. Schoener (1984) examined size ratios of sympatric bird-eating hawks, using a worldwide list of 47 species as his source pool. For sympatric assemblages of  $n$  species, he enumerated all possible  $n-1$  adjacent size ratios. Corresponding ratios in the source pool formed the null distribution. As in Tonkyn and Cole's (1986) analyses, these null distributions were concave upward and decreased from a maximum of very small size ratios (1.00–1.05). Schoener (1984) used a non-parametric Kolmogorov-Smirnov (K-S) test to compare the observed distribution of ratios with the source pool distribution. Most sets of ratios differed significantly by this test, with too many large ratios present in the sympatric groups compared to the source pool (Figure 6.4).

We note three points about Schoener's (1984) test. The first is that the K-S test does not indicate how the observed and expected distributions deviate from one another. For the hawk data, the absence or rarity of very small ratios was the most striking feature of the data. Other peaks in the observed distribution are more difficult to interpret. Second, the K-S test may be inappropriate for comparing observed and expected ratios because the two distributions are not independent of one another. Schoener (1984) controlled for this problem by removing species from the source pool that were actually present in a particular assemblage. Results were generally comparable with the original analysis, although sample sizes were seriously depleted. Schoener (1984) also obtained comparable results when the distributions were weighted by the amount of

geographic overlap between particular pairs of species. Finally, Schoener's (1984) test will be powerful only for large data sets. If small numbers of species in ecological guilds are tested, it will always be difficult to detect significant patterns.

Eldridge and Johnson (1988) used Schoener's (1984) method to analyze the distribution of size differences in mixed-species flocks of sandpipers. Size ratios in observed flocks did not differ from random subsets of species that use the North Dakota migration corridor. However, the assemblage of North Dakota species had too many intermediate-sized ratios ( $1.2 \leq r < 1.3$ ) compared to Western Hemisphere and global source pools. With respect to bill size, mixed species flocks were a random subset of a non-random source pool. Eldridge and Johnson (1988) speculated that past, but not present, competition may be important in the size structure of these assemblages.

Hopf and Brown (1986) presented an intriguing ratio test in which the ratio line segments were rescaled to a frequency of 1.0, projected from the edge of a unit simplex, and compared with the placement of random points on a "bull's-eye" target. Communities that were characterized by "even" distributions (i.e., constant size ratios) tended to fall disproportionately toward the center of the target. Hopf and Brown (1986) showed that for some assemblages, this bull's-eye test was more powerful than a modified version of the Simberloff and Boecklen (1981) protocol, although Arita's (1993) simulations suggested this will not always be true.

Two cautions should be noted for the bull's-eye test. First, the test is equivalent to using Simpson's diversity index to assess evenness of the distribution (Hopf and Brown 1986:1141), so it is vulnerable to the sample-size dependence that plagues most diversity indices (see Chapter 2). These problems are especially severe at small sample sizes, where systematic and stochastic errors of up to 20% are possible (Hopf and Brown 1986:1143). Second, the test was not designed to assess the significance of any particular assemblage, but rather to standardize the degree of evenness in body size distributions of repeated assemblages and reveal whether there was an overall trend toward evenness.

It is a difficult problem to determine how "independent" multiple assemblages really are. Only if each site contains a different set of species can we be confident that these represent distinct "natural experiments." But if some species occur repeatedly in different assemblages, the interpretation becomes more difficult. On the one hand, the recurrence of the same species combinations may, itself, be an important sign of nonrandom structuring of a community, as discussed in Chapter 7. On the other hand, certain species combinations may be the product of independent dispersal or habitat affinities that do not reflect interspecific interactions. In this case, a comparison of multiple assem-

blages inflates the degrees of freedom and may lead to an inappropriate rejection of the null hypothesis. The only possible solutions to this problem are to delete redundant species and repeat the analyses (e.g., Schoener 1984) or to use phylogenetic information to decide what constitutes independent species assemblages (e.g., Losos 1990).

## MULTIVARIATE ANALYSES

Although much of the size ratio literature is based on simple univariate measures of body size, these may be misleading. Tests of morphological patterns frequently yield different results depending on which particular character is chosen for analysis (Hendrickson 1981; Wiens 1982). At the same time, morphological characters are highly correlated with one another (Mosiman and James 1979, Bookstein 1984). Allometric constraints lead to predictable patterns of shape that are correlated with change in body size (Cherry et al. 1982). Finally, if niche segregation occurs along several different axes, it may be necessary to measure more than one morphological variable to adequately characterize niche segregation (Pianka et al. 1979).

For all these reasons, multivariate analyses of morphological pattern may give more insight into community structure than univariate tests. Karr and James (1975) pioneered the technique of using principal components analysis (and other multivariate methods) to study a set of correlated morphological variables. The principal components are a set of orthogonal composite variables that account for the maximum amount of variation in the original data (Hotelling 1933). Because much of the variation in morphology may be explained by the first few principal component axes, the dimensionality of the original data set can be reduced (Gauch 1982).

Unfortunately, multivariate analysis has been greatly abused by ecologists (James and McCulloch 1990). The common practice of representing species as points in multivariate space, drawing polygons (or amoebas) around groups of species, and interpreting the results often amounts to ecological palmistry. Ad hoc “explanations” often are based on the original untransformed variables, so that the multivariate transformation offers no more insight than the original variables did. Because multivariate techniques maximize the amount of variation explained by a transformed axis (principal components analysis), or maximize the separation between groups (discriminant analysis), apparent patterns in multivariate space may arise from data sets created with a random number generator (Karr and Martin 1981). Consequently, it is important to compare the patterns in multivariate analysis against a properly formulated null model.

Even apparently clear-cut patterns of convergence or divergence (e.g., Fuentes 1976) may arise from neutral models that do not include these processes (Crowder 1980; but see Fuentes 1980).

Two types of community structure have been inferred from multivariate analyses of morphology. (1) Overdispersion of morphology, which is measured within a community and is assumed to reflect competition for food resources (Ricklefs and Travis 1980). It is the multivariate analog of size ratio analyses. (2) Convergence of morphology, which is measured between communities and is assumed to reflect evolutionary convergence of unrelated species that live in similar environments (Cody 1974). Convergence can be measured for a single community (the opposite of overdispersion), but it usually involves broad morphological comparisons of communities with different evolutionary histories (Wiens 1991b). Null models have been used frequently for tests of overdispersion, but have been underutilized in studies of convergence. In the following sections, we review null model tests of these hypotheses.

### Overdispersion of Morphology

Gatz (1979) analyzed 56 morphological characters measured for co-occurring stream fishes of the Piedmont of North Carolina. Points representing each species were placed in a nine-dimensional factor space that was calculated from the original morphological variables. Null assemblages were constructed by choosing random points along each factor axis and then projecting them into the morphological space. For both real and simulated assemblages, Gatz (1979) calculated the Euclidian distances between all possible pairs of species and measured the number of times the observed distributions overlapped with the simulated. Compared to the null model, there were too many species pairs with unusually small or unusually large Euclidian distances.

Gatz's (1979) procedure is the multivariate analog of Lawlor's (1980b) randomization algorithm RA1 for resource utilization analyses (see discussion in Chapters 4 and 5). In RA1, resource utilizations for each species are replaced by a uniform [0,1] random number. The resulting overlaps always have a high mean and low variance (Pianka et al. 1979; Winemiller and Pianka 1990). Gatz (1979) concluded that Euclidian distances within his assemblage were nonrandom, but the comparisons were based on a rather extreme null model that did not preserve natural covariation among morphological traits.

Although Gatz (1979) attributed the nonrandomness to interspecific competition, he also explored the alternative hypothesis that phylogenetic correlates were responsible for the patterns. He analyzed Euclidian distances between sympatrically occurring members of a single family or genus and obtained

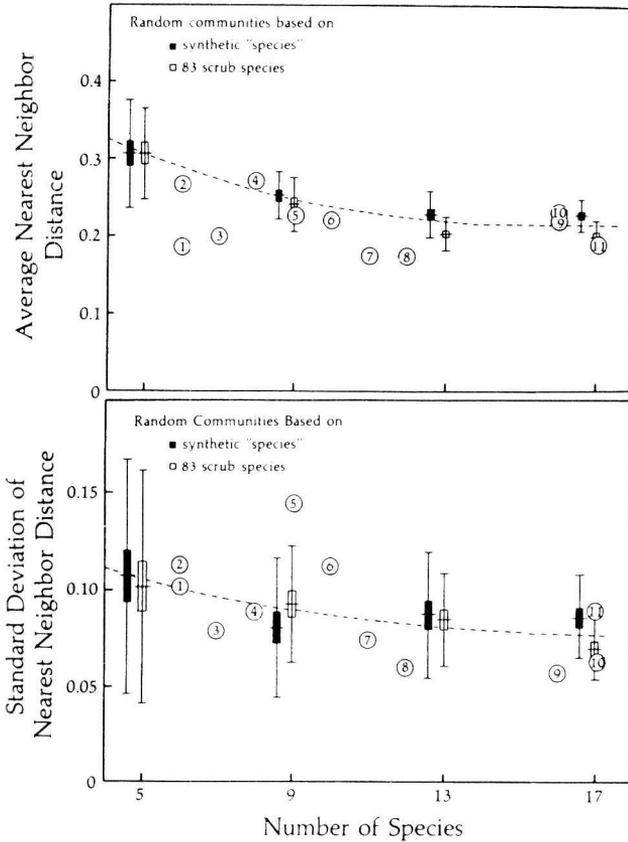


Figure 6.10. Morphological mean and standard deviation of nearest-neighbor distances for 11 shrubsteppe bird communities (Cody 1974). Each number refers to a different observed assemblage. Null distributions were based on 20 random draws of observed source pool species or construction of hypothetical “synthetic species.” Means were slightly less than expected, indicating weak convergence, whereas standard deviations matched the null model predictions. From Ricklefs and Travis (1980), with permission of the American Ornithologists’ Union.

comparable results with analyses of the complete assemblage. Although the randomization algorithm (RA1) was primitive, Gatz’s (1979) study is noteworthy for its use of null models and consideration of phylogeny in the context of ecomorphology.

Ricklefs and Travis (1980) constructed null assemblages by either drawing species lists from a larger source pool and maintaining observed morphological features, or randomly generating “synthetic species” by substituting each factor

score with a random, normal deviate, similar to Gatz's (1979) protocol. Species packing was measured by the average nearest-neighbor distance in morphological space, and the evenness of species packing was measured by the standard deviation of nearest-neighbor distances.

Ricklefs and Travis (1980) applied this protocol to morphological data for 11 temperate zone avian communities (Cody 1974). Mean nearest-neighbor distances were somewhat less than expected, whereas standard deviations usually matched the predictions of the null model (Figure 6.10). However, there were differences in the morphological structure of large and small assemblages. Species were added nonrandomly near the edge of morphological space, mostly along novel morphological dimensions. For Neotropical avian communities, the results varied widely among localities. Species packing was unusually even, and nearest-neighbor distances were overdispersed for communities on small islands in the Lesser Antilles (Travis and Ricklefs 1983). These results were consistent with guild analyses of tropical hummingbirds, which were also overdispersed in morphology (Brown and Bowers 1985; Ranta 1986).

Ricklefs et al. (1981) used similar methods to analyze the morphological structure of lizard assemblages in desert habitats. Australian assemblages were more loosely packed in morphological space than North American or African assemblages. However, the means and standard deviations of nearest-neighbor distances did not differ significantly from those of null communities that were randomly assembled from continental source pools. For North American lizard communities, observed nearest-neighbor distances were always greater than predicted by the null model, but all of these assemblages contained the same four core species and may not represent independent samples. Lizard communities from a variety of temperate-zone habitats in North America showed no evidence for overdispersion in either morphology or habitat use (Scheibe 1987).

### **Concordance of Morphology**

Studies of concordance in morphology involve comparisons of two or more assemblages of unrelated species that evolve morphological similarity in similar environments (Cody 1973). Relatively little null model work has been done on this topic. Current tests use analysis of variance models to partition morphological variation into components of "habitat," which presumably reflect convergence due to similar environments, and components of "region," which presumably reflect unique historical or phylogenetic effects on morphology (Schluter 1986b; Schluter and Ricklefs 1993). Schluter (1990) recommended tests for a more specific hypothesis of species-for-species matching, in which

ecologically equivalent sets of species are found in similar habitats at different locations. The null hypothesis here is that differences in body size of species matched between two assemblages are no smaller than would be expected by chance. Schluter (1990) found that the left-hand tail of conventional statistical distributions (chi-squared, *F*-ratio) matched simulated distributions in which species body sizes in a community were assigned randomly within a given range. He used the test to show that body size distributions of rodents from the Great Basin and the Sonoran Desert were “too close” to be expected by chance (but see Lomolino 1993). However, the species-for-species matching test may be unsuitable if some species occur in both locations or if species richness differs between locations. Because concordance is not necessarily the same as matching or similarity (Wiens 1991b), a broader test for convergence may require detailed comparisons of the morphology of closely related species in different habitats and different regions (e.g., Niemi 1985).

## Morphology and Abundance

The studies described so far have searched for associations between morphology and coexistence. These analyses assume not only that morphology and resource use are intimately related (Miles et al. 1987), but that competitive effects are strong enough to cause the extinction (or exclusion) of species. Alternatively, there may be relationships between morphology and abundance of coexisting species that do not reflect competitive exclusion. For example, the morphology of coexisting gastropod species in a New York lake was random when compared to an unweighted null model, but significantly overdispersed when compared to a model in which colonization probabilities were proportional to abundance (Dillon 1981).

James and Boecklen (1984) examined the relationship between abundance and morphology for an assemblage of forest birds in Maryland that was censused in seven consecutive years. Few significant patterns emerged. For example, density correlations (and nearest-neighbor distances) between species were unrelated to size and shape differences (Figure 6.11). Common species had no greater influence on the abundance of their nearest morphological neighbors than did rare species, and the density and variability of populations was not different for species near the edge versus the center of morphological space. Results were comparable for analyses at the guild level, and James and Boecklen (1984) concluded that competitive relationships did not structure the abundance or morphology of species in this assemblage. Wiens and Rotenberry (1980) arrived at a similar conclusion from a more detailed long-term study of

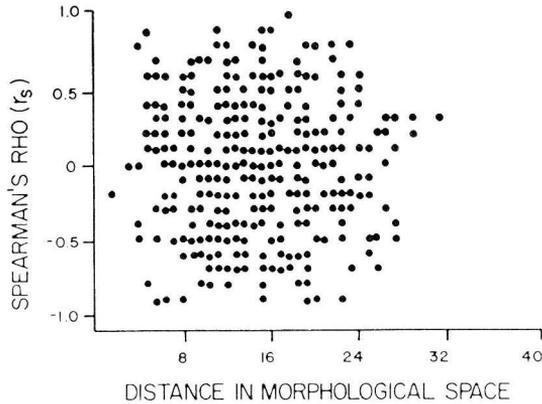


Figure 6.11. Pairwise density correlations and morphological distances for an avian assemblage in a Maryland woodland. Each point represents a different pair of species. If competition were organizing this community, morphologically similar species pairs should have shown negative density correlations. From James, F. C., and W. J. Boecklen. Interspecific morphological relationships and the densities of birds. In: *Ecological Communities: Conceptual Issues and the Evidence*. D. R. Strong, Jr., D. Simberloff, L. G. Abele, and A. B. Thistle (eds). Copyright © 1984 by Princeton University Press. Reprinted by permission of Princeton University Press.

shrubsteppe bird assemblages that included quantitative measures of habitat structure, resource availability, and dietary overlap.

In an assemblage of butterfly fishes, Findley and Findley (1985) also failed to find any relationship between morphology, diet, and abundance. These results, combined with the null model tests of Sale and Williams (1982), contradicted earlier claims of competitive structure in reef fish communities (Anderson et al. 1981). Although regional comparisons of community structure frequently are attributed to competitive effects, detailed studies of resource use, morphology, and abundance of local assemblages often do not support these explanations, even though this is the spatial scale at which strong competitive interactions ought to be most clearly expressed (Sale 1984).

### Morphology and Evolutionary Extinction

One difficulty with analyzing present-day assemblages is that ecological extinctions or evolutionary shifts in morphology are not observed but must be inferred indirectly. If we knew the temporal record of species extinctions, it

might be easier to evaluate the significance of morphological relationships among surviving species.

Fossil assemblages can give some insight into long-term patterns of extinction and morphology. For example, Foote (1991) tested whether species extinctions of fossil blastoids were random with respect to morphology. He measured eight morphological features of 85 species and characterized the position of each species in morphological space with a principal components analysis. Rarefaction and random sampling were then used to describe changes in taxonomic and morphological diversity in the fossil record. Generic richness of the blastoidea peaked in the Lower Carboniferous, but morphological diversity did not reach a peak until the Permian. Stochastic simulations of taxonomic and morphological evolution also yielded clades whose morphological diversity peaked later than taxonomic diversity. Because extinctions in the blastoid assemblage appeared to be random with respect to morphology, morphological diversity was maintained and even increased in the face of major taxonomic extinctions.

## Morphology and Ecological Extinction

In ecological time, can extinctions be predicted on the basis of morphology or other species attributes? From the conservation perspective, there is great interest in knowing whether extinctions occur randomly with respect to body size, habitat affinity, or trophic status (Karr 1982a,b; Diamond 1983; Pimm et al. 1988). However, extinctions are difficult to measure systematically, and we must often rely on natural experiments (and null models) to study the process. A controversial study of extinctions in the Hawaiian avifauna illustrates some of the difficulties.

Moulton and Pimm (1983, 1986, 1987) argued that extinction of introduced species in the Hawaiian Islands could be predicted on the basis of body size or morphology, and that these patterns were caused primarily by competition among introduced species. For example, introduced pairs of congeneric species that both survived on at least one island ( $n = 6$  pairs) differed more in bill length (22%) than congeneric pairs in which one of the pair went extinct (9%;  $n = 9$  pairs; Moulton 1985). The morphology of the surviving species also appeared non-random. For three of the six major Hawaiian islands, introduced species of forest passeriformes were overdispersed in morphological space compared to random draws of species from the set of all introduced forest passerines (Moulton and Pimm 1987). Introduced finches of Oahu (Moulton and Lockwood 1992) and introduced passeriformes of Tahiti (Lockwood et al. 1993) also exhibited morphological overdispersion (Figure 6.12).

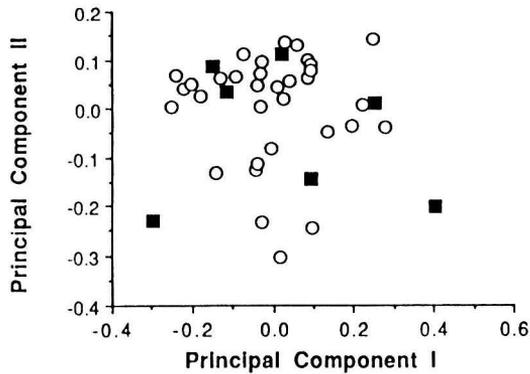


Figure 6.12. Morphological overdispersion of introduced bird species on the island of Tahiti. The surviving species (filled squares) were significantly overdispersed in comparison to random subsets of all introduced species. From Lockwood, J. L., M. P. Moulton, and S. K. Anderson. 1993. Morphological assortment and the assembly of communities of introduced passeriforms on oceanic islands: Tahiti versus Oahu. *American Naturalist* 141:398–408. Copyright © 1993 by The University of Chicago. Reprinted by permission of the publisher.

Simberloff and Boecklen (1991) challenged the Hawaiian results and argued that evidence for competition was not so clear-cut. They pointed out that most species introduced to the islands were either highly successful or consistent losers. Out of 41 species that were introduced on more than one island, 21 succeeded on all the islands to which they were introduced, 16 failed on all islands, and only four species showed mixed results. Apparently, each species was intrinsically successful or not at establishment, and successful introductions had little to do with the size or composition of the resident or introduced avifauna of the island. If this “all-or-none” hypothesis is true, then whatever patterns of overdispersion may have been present in the source pool would be retained in the subsets of introduced species on each island. Moulton (1993) countered that biases from phylogeny, errors in Simberloff and Boecklen’s (1991) data matrix, and differences in the number of islands per introduction refuted the all-or-none hypothesis and reaffirmed the role of competition.

The debate rests ultimately on the quality of the introduction and extinction data. In this case, the data are a heterogeneous collection of checklists (e.g., Henshaw 1902), surveys (e.g., Donagho 1965), secondary sources (Long 1981), and records from *Elepaio*, the monthly publication of the Hawaiian Audubon Society. This information required a number of arbitrary judgment calls to quantify introduction and extinction dates; even still, some species extinction dates could only be estimated to the nearest decade.

Consequently, the patterns described by Moulton and Pimm (1983, 1986, 1987) are extremely fragile and probably would not be replicated if another investigator extracted the data from the same sources. To cite just one example from Simberloff and Boecklen (1991), two species of cardinal (*Paroaria*) both survived on Hawaii, but Moulton (1985) did not include them in the analysis of congeneric species pairs because he felt it was likely they were allopatric. With such small sample sizes, this sort of judgment call will greatly affect any statistical analysis. We think there are so many problems of misidentification, errors in taxonomy and nomenclature, irregular and incomplete censuses, and sightings of vagrants, nonbreeders, or cage-released birds that it is impossible to evaluate the ecological significance of the results.

This is not the first time that avian ecologists have been plagued by unreliable census data. Controversies over avian extinctions in the California Channel Islands (Diamond 1969; Lynch and Johnson 1974; Jones and Diamond 1976), in remnant forest patches (Whitcomb et al. 1977; McCoy 1982), and on Barro Colorado Island (Willis 1974; Karr 1982a,b) highlight the dangers of inferring extinctions from historical records or short-term censuses. Extinction data for island birds of Great Britain probably are more reliable, and Pimm et al. (1988) concluded that large-bodied species were most vulnerable to extinction. However, these conclusions hinged upon the precise way that extinctions are measured (Haila and Hanski 1993). The disappearance of species may have depended more on island attributes such as area or isolation than on intrinsic properties of species (Tracy and George 1992; but see Diamond and Pimm 1993; Tracy and George 1993).

This is not to say that it is impossible to measure avian extinctions accurately, but the task requires intensive, systematic surveys using reliable census techniques (e.g., Haila and Järvinen 1981). The lesson is obvious: no matter how sophisticated the null model, the results are meaningless if the underlying data cannot be trusted.

## EMPIRICAL TESTS

The size ratio controversies have been good for community ecology. Not only have recent studies of size overlap used more rigorous statistical analyses, but they have incorporated data on geographic variation in morphology, patterns of resource use, and even experimental manipulations to understand the mechanisms behind the patterns. These approaches are illustrated in the following examples.

## Ecological Character Displacement in the Red Fox

Dayan, Tchernov, et al. (1989) provided persuasive biogeographic evidence suggesting character displacement between the red fox (*Vulpes vulpes*) and Ruppell's sand fox (*Vulpes ruppelli*). These species are sympatric in the Saharo-Arabian region, where they exhibit unusually constant size ratios (1.18–1.21) in lower carnassial length, as measured by the Barton and David (1956) test applied to the sexes separately. Both species are sympatric with Blanford's fox (*Vulpes cana*), in Israel where there was also unusual constancy in size ratios of tooth dimensions for sites with all three species present (Figure 6.13).

The red fox is allopatric with Ruppell's and Blanford's foxes throughout most of the Palearctic. In regions of allopatry, tooth size is strongly correlated with latitude and temperature, with large-toothed populations in the north (Bergmann's rule). But in regions of sympatry, the slope of the relationship is significantly more shallow. Thus, size ratios of sympatric populations are unusually constant, despite strong clinal variation in tooth size. This pattern meets one of the criteria set forth in Grant (1972a) for geographic tests of character displacement. Although resource availability is not known for the foxes, the pattern of constant size ratios in the face of latitudinal variation in tooth dimensions is compelling evidence for ecological character displacement.

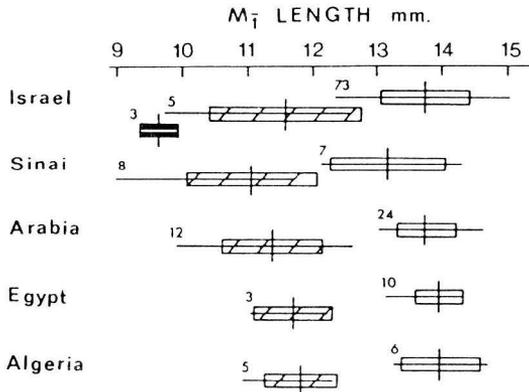


Figure 6.13. Lower carnassial lengths of three foxes in the Saharo-Arabian region. Vertical lines represent means, horizontal lines represent ranges, bars represent two standard deviations. Empty bars = red fox (*Vulpes vulpes*); shaded bars = Ruppell's sand fox (*Vulpes ruppelli*); blackened bar = Blanford's fox (*Vulpes cana*). Size ratios for these assemblages varied between 1.18 and 1.21, and were unusually constant by the Barton and David (1956) test. From Dayan, Tchernov, et al. (1989), with permission.

This study is one of several that have revealed evidence for character displacement in assemblages of mammalian carnivores. Other examples include canids of Israel (Dayan et al. 1990, 1992) and North American mustelids (Dayan, Simberloff, et al. 1989; but see Harvey and Ralls 1985). Patterns were much weaker for sympatric tropical cats (Kiltie 1984, 1988).

### Competition and Morphology of Co-occurring Dytiscid Beetles

Juliano and Lawton (1990a) tested for widely and regularly distributed body forms in assemblages of co-occurring dytiscid water beetles. The morphology of each species was represented by canonical discriminant analysis of four morphometric variables. The average nearest-neighbor distance and the variance of nearest-neighbor distances were interpreted as measures of species dispersion and regularity of niche packing.

Random assemblages were constructed by sampling from the total source pool (Strong et al. 1979) and by creating synthetic species in morphological space (Ricklefs and Travis 1980). Null assemblages were constructed for the entire species list, for the abundant species, and for species in the dominant genus *Hydroporus*. The results depended on site characteristics. For seven small sites that were acidic and lacked fish, real and null assemblages were not significantly different. For two large sites that were well buffered and had fish, morphological distances between species were larger and more regular than expected by chance.

Juliano and Lawton (1990b) next tested experimentally whether these patterns were the result of interspecific competition. The null model analyses led to the prediction of competition at large, well-buffered sites, but not at small, acidic sites. At one of the small, acidic ditches, the feeding rate of dytiscid beetles was unaffected by manipulation of adult density and was unrelated to the presence of competing species, regardless of their body size. However, there was evidence of competition for food and of cannibalism in the larval stages. In a large, well-buffered canal, the feeding rate declined with increasing density, but the effect was unrelated to body size of competing species, and adults did not appear to be food-limited at natural densities.

The results suggest that nonrandom body size spacing was not necessarily related to interspecific competition. The difference in body size patterns may instead have been related to the presence of predators, although a link between prey body morphology and predator avoidance has not been established in this system. The results also suggest that competition in the larval stages was considerably more important than interactions among adults in determining community composition.

Few other studies have used this double-barreled approach of null models combined with experimental manipulations. An interesting contrast to the dytiscid beetle system is the desert rodent assemblages of North America, for which null models and field experiments have yielded consistent results. Large-scale field manipulations have demonstrated competition for limited seed supplies (Brown et al. 1979; Brown and Munger 1985), and null model simulations revealed that common species of similar body size (mass ratios  $< 1.5$ ) co-occurred less frequently and overlapped less in geographic range than expected by chance (Bowers and Brown 1982). Body size distributions of local assemblages were also more uniform than in regional or continental source pools (Brown and Nicoletto 1991). Community ecology needs more of these studies that combine experimental manipulations with null model analyses.

### **Bill Sizes of Galápagos Finches**

Bill sizes of Galápagos ground finches (*Geospiza*) are one of the most well-known and controversial examples of character displacement (Lack 1947; Bowman 1961; Grant 1972a; Strong et al. 1979; Grant and Abbott 1980). Bill sizes for each species vary dramatically between islands, and the differences seem to correspond to the presence or absence of other finch species. These patterns may be explained by competition among species (Lack 1947) or by differences in food supply among islands (Bowman 1961). The initial null models of Strong et al. (1979) revealed little evidence for character displacement, whereas the reanalysis of Hendrickson (1981) and the null models of Case and Sidell (1983) suggested nonrandomness in morphology of coexisting species. Simberloff (1983b) reassessed the early null model studies and concluded that a substantial majority of tests revealed displacement in *Geospiza* bill morphology that was consistent with interspecific competition. All of these statistical analyses were based on lists of coexisting species on islands and measurements of average bill and body sizes for finch populations on each island.

Schluter and Grant (1984) resolved the controversy with new data on seed biomass and hardness, bill depth, and finch biomass, based on many years of field work. These data were incorporated into a series of null models that attempted to explain the co-occurrence and morphology of *Geospiza*. Schluter and Grant (1984) first estimated the expected population density on each island as a function of bill size for a hypothetical solitary species of granivorous finch (Figure 6.14).

These functions correspond to resource availability curves in models of character displacement and limiting similarity. In contrast to the simplifying assumptions of theory (MacArthur and Levins 1967; Slatkin 1980), these

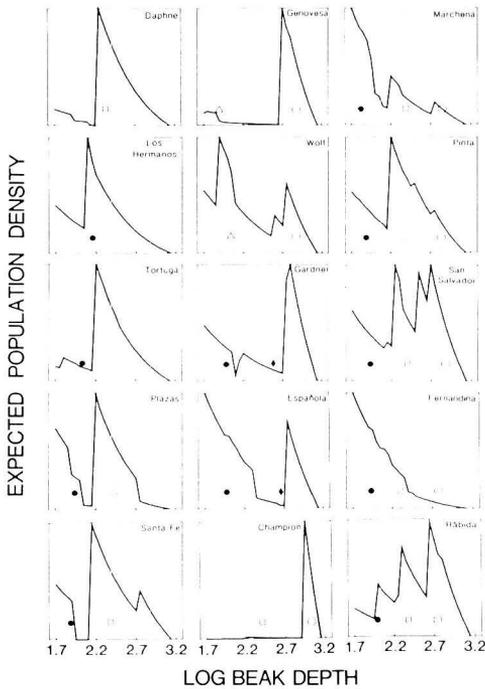


Figure 6.14. Expected population density of a solitary granivorous finch species on 15 Galápagos Islands. Points represent observed bill depths of species present on each island ● = *Geospiza fuliginosa*; □ = *G. fortis*; ○ = *G. magnirostris*; Δ = *G. difficilis*; ◆ = *G. conirostris*. Note the polymodal peaks of expected density and the wide spacing of species from one another. Such spacing was unlikely for null models that did not incorporate species interactions. From Schluter, D., and P. R. Grant. 1984. Determinants of morphological patterns in communities of Darwin's finches. *American Naturalist* 123:175–196. Copyright © 1984 by The University of Chicago. Reprinted by permission of the publisher.

empirically derived curves were not uniform or Gaussian in shape, but complexly polymodal. The modes represented bill sizes that might be expected for a species in the absence of competition. The curves formed the basis for five null models that were tested against the observed data. These null models predicted the minimum difference in bill size for communities that were assembled according to the following rules:

*Model 1. Random Assembly/Evolution.* Food supply set the range of permissible bill sizes, and all bill sizes were equiprobable within this range. This model is similar to the Barton and David (1956) test, although the

range was set by food supply rather than by the largest and smallest species in the assemblage.

*Model 2. Partly Directed Assembly.* All bill sizes were possible, but the probability of persistence on an island was proportional to available food supply. Thus, for a given island, the most likely bill sizes were those that corresponded to peaks in available food supply. Differences in morphology of finch populations between islands would reflect differences in resource availability, which was essentially Bowman's (1961) explanation.

*Model 3. Directed Evolution.* Only bill sizes corresponding to peaks in resource availability were permissible. This model assumed natural selection optimized the mean phenotype and that each species evolved independently. Bill size differences for a pair of species might be zero if they were randomly assigned to the same resource peak. This would always happen for islands that had fewer peaks than species.

*Model 4. Directed Assembly with Competitive Exclusion.* Whereas the first three models assumed no species interaction, this model assumed that the presence of other species could lead to competitive exclusion. Species were allowed to colonize an island randomly, but they would persist only if their bill sizes allowed them access to at least some exclusive food resources. This is a model of limiting similarity, in which colonization and extinction dynamics determine the combinations and sizes of species that can coexist, given the constraints of the available resource spectrum.

*Model 5. Directed Coevolution Under Interspecific Competition.* This model assumed that natural selection adjusted the mean phenotype of a species to maximize its population size, conditional on the morphology of other species present on the island. Coevolution between birds and seed plants was not considered. For this model, Schluter and Grant (1984) enumerated expected population densities of all possible species combinations on an island, and then found the particular combination of phenotypes that maximized population sizes. Presumably, coevolution would adjust bill sizes until this optimum was reached. For islands with multiple resource peaks, multiple solutions were possible.

For each model, Schluter and Grant (1984) computed the probability of randomly obtaining the minimum observed size difference on each island. Models 1–3 included no competitive effects and could not account for observed bill size differences. Across all islands, observed bill size differences were too large to be accounted for by these null models. In contrast, there was

a good fit between observed and expected bill sizes for models 4 and 5, although they were not assessed statistically in the same way as models 1–3. Schluter and Grant (1984) suggested that model 4 (size assortment) might be preferable to model 5 (size adjustment) because morphology on islands was predictable, but species combinations were not. This interpretation is consistent with Case and Sidell's (1983) analyses. On the other hand, the proximity of *Geospiza fortis* and *G. fuliginosa* to theoretical resource peaks was different in sympatry and allopatry, suggesting evolutionary size adjustment, at least for this species pair (Schluter et al. 1985). Given all the controversy surrounding the analysis of size ratios, the Galápagos *Geospiza* may represent one of the few unequivocal examples of character displacement.

## RECOMMENDATIONS

For tests of single assemblages, we recommend the Barton and David (1956) and Irwin (1955) tests as simple (but conservative) measures of ratio constancy and minima. For multiple assemblages with intraspecific variation, we recommend the Monte Carlo algorithm of Strong et al. (1979), but with the modifications suggested by Hendrickson (1981) and Grant and Abbott (1980). The weighted versions of Case and Sidell's (1983) tests may also be useful for distinguishing between patterns of size adjustment and size assortment. Schluter's (1990) test should be used for patterns of species-for-species matching, and Ricklefs and Travis's (1980) procedures should be used in the analysis of multivariate data. The simulation version of Hopf and Brown's (1986) bull's-eye test is appropriate for examining size ratios in multiple assemblages. All of these tests can be greatly strengthened by incorporating independent measures of resource availability (Schluter and Grant 1984) and phylogeny (Losos 1990).

